
Generative AI and Asset Management

Jinfei Sheng, Zheng Sun, Baozhong Yang, and Alan Zhang

Discussion
Amit Goyal

23 May 2025
ABFER Annual Conference



Broad questions

1. Do hedge funds (HFs) use ChatGPT?
2. Does it help?

How to do it

Biggest challenge

How to measure HFs' reliance on ChatGPT?

Authors' solution

Check if output of ChatGPT correlates with Δ Holdings.

My first reaction

Clever idea. Seems reasonable.

Details

Notation: Fund i , Stock j , Quarter t .

$R_{Fundamental,i,t}^2$ from cross-sectional regression across stocks for each fund-quarter:

$$\Delta Holdings_{i,j,t} = \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}.$$

$R_{AI,i,t}^2$ from augmented cross-sectional regression:

$$\Delta Holdings_{i,j,t} = \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}.$$

Reliance on ChatGPT:

$$GenAIReliance_{i,t} = R_{AI,i,t}^2 - R_{Fundamental,i,t}^2.$$

It matters for performance

$$Return_{i,t} = \beta GenAIReliance_{i,t-1} + \gamma Controls_{i,t-1} + \alpha_t + e_{i,t}.$$

Dep. Var.	(1) <i>CAPM Alpha</i>	(2) <i>CAPM Alpha</i>	(3) <i>FF3 Alpha</i>	(4) <i>FF3 Alpha</i>	(5) <i>FF4 Alpha</i>	(6) <i>FF4 Alpha</i>
<i>GenAI Reliance</i>	1.533*** (3.14)	1.619*** (3.07)	1.784*** (3.59)	1.798*** (3.38)	1.879*** (3.54)	1.893*** (3.36)
<i>Size</i>	0.108 (1.59)	0.157** (2.10)	-0.026 (-0.40)	0.045 (0.61)	-0.020 (-0.32)	0.043 (0.59)
<i>Age</i>	-0.082*** (-5.02)	-0.023 (-1.15)	-0.011 (-0.69)	0.001 (0.07)	-0.033* (-1.87)	0.005 (0.24)
<i>Turnover</i>	0.514 (0.74)	0.917 (1.29)	1.604** (2.44)	1.865*** (2.76)	1.346** (2.05)	1.773*** (2.62)
<i>Risk</i>	3.216 (0.72)	6.011 (1.26)	8.007 (1.64)	8.789* (1.72)	9.550* (1.75)	10.055* (1.78)
<i>Past Return</i>	-0.122*** (-7.42)	-0.218*** (-6.61)	-0.012 (-0.83)	-0.129*** (-3.66)	-0.017 (-1.12)	-0.140*** (-4.11)
Observations	2,066	2,066	2,066	2,066	2,066	2,066
R-squared	0.059	0.099	0.014	0.053	0.015	0.057
Time FE	No	Yes	No	Yes	No	Yes

Corresponds to
1.6% per year.

It matters after ChatGPT introduction – DiD

$$Return_{i,t} = \beta_1 GenAIReliance_{i,t-1} \times PostGPT_t + \beta_2 GenAIReliance_{i,t-1} + \gamma Controls_{i,t-1} + \alpha_t + e_{i,t}.$$

Dep. Var.	(1) CAPM Alpha	(2)	(3) FF3 Alpha	(4)	(5) FF4 Alpha	(6)
<i>GenAI Reliance</i> × <i>Post GPT</i>	1.411** (2.39)	1.633*** (2.83)	2.629*** (4.81)	2.445*** (4.46)	2.725*** (4.69)	2.537*** (4.46)
<i>GenAI Reliance</i>	-0.536* (-1.77)	-0.396 (-1.27)	-0.673*** (-2.62)	-0.538** (-2.00)	-0.608** (-2.32)	-0.511* (-1.85)
<i>Size</i>	-0.118*** (-3.28)	-0.115*** (-3.23)	-0.112*** (-3.50)	-0.108*** (-3.32)	-0.112*** (-3.60)	-0.113*** (-3.52)
<i>Age</i>	0.002 (0.32)	0.004 (0.54)	0.002 (0.35)	0.002 (0.35)	-0.001 (-0.19)	0.002 (0.33)
<i>Turnover</i>	-0.444 (-1.35)	-0.447 (-1.37)	-0.633** (-2.19)	-0.558* (-1.86)	-0.658** (-2.28)	-0.547* (-1.82)
<i>Risk</i>	0.359 (0.19)	-2.541 (-0.79)	-5.556*** (-4.03)	-8.684*** (-3.43)	-5.878*** (-3.96)	-8.069*** (-2.95)
<i>Past Return</i>	-0.032*** (-5.25)	0.048*** (3.02)	-0.018*** (-3.44)	-0.016 (-1.19)	-0.017*** (-3.01)	-0.018 (-1.31)
<i>Post GPT</i>	-1.168*** (-8.04)		-0.592*** (-4.20)		-0.556*** (-3.72)	
Observations	11,921	11,921	11,921	11,921	11,921	11,921
R-squared	0.009	0.096	0.011	0.066	0.011	0.063
Time FE	No	Yes	No	Yes	No	Yes

AI is not proxying for ML

X below includes positive and negative sentiment using bag of words of Loughran and McDonald (2011)

$$\Delta Holdings_{i,j,t} = \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}$$

$$\Delta Holdings_{i,j,t} = \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}$$

$$GenAIReliance_{i,t} = R^2_{AI,i,t} - R^2_{Fundamental,i,t}$$

Dep. Var.	(1) <i>Return</i>	(2) <i>CAPM Alpha</i>	(3) <i>FF3 Alpha</i>	(4) <i>FF4 Alpha</i>
<i>GenAI Reliance_{LM}</i>	0.575 (0.93)	1.437** (2.48)	1.493*** (2.72)	1.528*** (2.68)
<i>Size</i>	0.064 (0.94)	0.132* (1.86)	0.023 (0.37)	0.015 (0.24)
<i>Age</i>	-0.003 (-0.16)	-0.023 (-1.24)	-0.001 (-0.08)	-0.002 (-0.12)
<i>Turnover</i>	0.762 (1.11)	1.105* (1.73)	2.191*** (3.71)	2.174*** (3.64)
<i>Risk</i>	24.453*** (4.74)	7.999* (1.87)	8.162* (1.93)	9.351** (1.99)
<i>Past Return</i>	-0.115*** (-3.59)	-0.188*** (-6.44)	-0.091*** (-3.11)	-0.093*** (-3.33)
Observations	1,977	1,974	1,974	1,974
R-squared	0.012	0.097	0.052	0.061
Time FE	Yes	Yes	Yes	Yes

I suggest move this table up from Internet Appendix Table IA.5.

Other tests

Non-HF do not show an improvement in performance.

There is heterogeneity across HFs.

HFs that have/hire AI-skilled workers do better.

Survey evidence.

My plan

Main comments.

Minor comments.

Philosophy.

Comments about GenAIReliance

1. Magnitude
2. Significance
3. Different measure
4. Different equation

Magnitude of GenAIReliance

$$\begin{aligned}\Delta Holdings_{i,j,t} &= \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t} \\ \Delta Holdings_{i,j,t} &= \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t} \\ GenAIReliance_{i,t} &= R^2_{AI,i,t} - R^2_{Fundamental,i,t}.\end{aligned}$$

How much is $R^2_{Fundamental,i,t}$?

- Kacperczyk and Seru (2007) report $\sim 20\%$ for MFs.
- I assume that it is 10-15% for HFs.

$GenAIReliance \approx 0.26\%$.

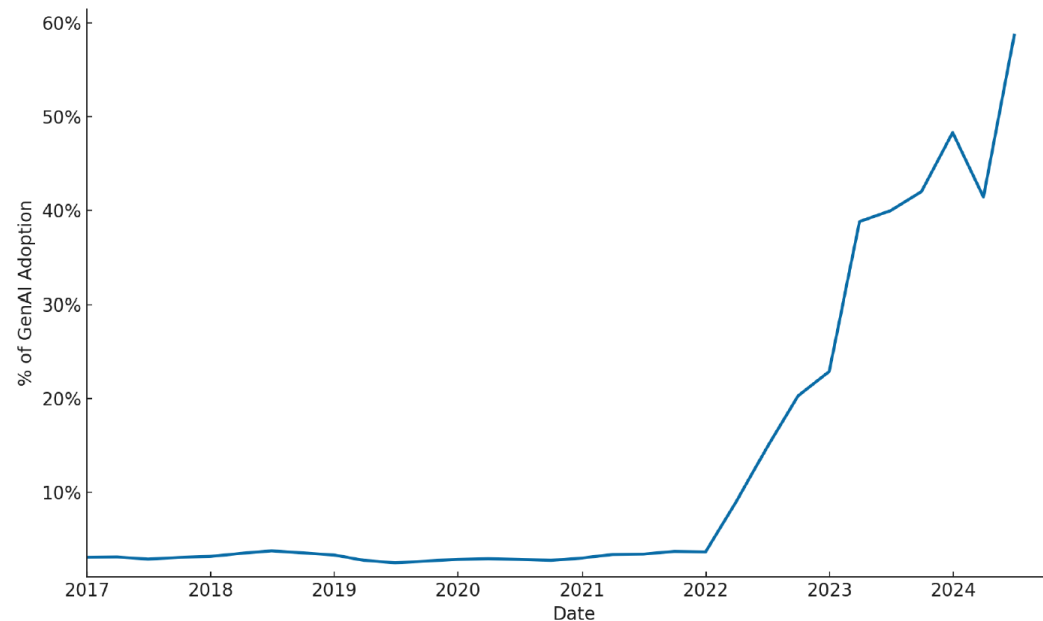
This does not sound like much. But it may be that small R^2 in (explaining) holdings can lead to large improvement in performance. Perhaps the authors can work out the math to show this.

Significance of GenAIReliance

Since we are measuring R^2 (and not \bar{R}^2), the R^2 of an augmented regression is necessarily higher than the one without. Meaning, by definition, $GenAIReliance_{i,t} > 0$, and for all i, t .

- This would mean that all HFs always use (have used) ChatGPT.

The authors are aware of this and conduct an F -test for the significance of difference in R^2 .



Significance of GenAIReliance ...

But then why not use the significance of difference in R^2 as the main explanatory variable? In other words, define an indicator variable as:

$$UseChatGPT = \begin{cases} 1 & \text{if } \Delta R^2 \text{ is significant} \\ 0 & \text{otherwise.} \end{cases}$$

Sometimes, high value of *GenAIReliance* may just be noise. Using *UseChatGPT* could even help make the results stronger.

Also currently, in the empirical specification, there is no HF with zero reliance of ChatGPT. That does not sound right. Using *UseChatGPT* would also solve that irregularity.

Different measure of GenAIReliance

Currently:

$$\begin{aligned}\Delta Holdings_{i,j,t} &= \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t} \\ \Delta Holdings_{i,j,t} &= \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t} \\ GenAIReliance_{i,t} &= R^2_{AI,i,t} - R^2_{Fundamental,i,t}.\end{aligned}$$

Why not just look at $\beta_{i,t}$?

- How big is it?
 - I assume that the Y and X variables are scaled, so β can be compared across time and funds. If not standardize left-hand and right-hand-side variables.
- How significant it is?

Different equation for GenAIReliance

Currently, the authors are correlating ChatGPT score with changes in holdings, without looking at direction.

$$\Delta Holdings_{i,j,t} = \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}.$$

ChatGPT does give an ordinal ranking. +2 is good while -2 is bad. Should we not also be looking at the sign of β ?

For example, it may be that in the current setup, high ΔR^2 comes from $\beta < 0$? That means that that HF did exactly the opposite of what the ChatGPT suggested. This is reliance on ChatGPT in a perverse sense.

I suggest that you modify this equation into positive and negative and consider only consistent sign slopes as evidence of reliance of ChatGPT.

Minor comments (1)

Timing of variables was confusing to me. Presumably both X_j and $ChatGPT_j$ are available to HFs at time t (not as in footnote 23).

$$\begin{aligned}\Delta Holdings_{i,j,t} &= \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t} \\ \Delta Holdings_{i,j,t} &= \beta_{i,t} ChatGPTScore_{j,t-1} + \gamma_{i,t} X_{j,t-1} + \epsilon_{i,j,t}.\end{aligned}$$

Why have a lag of an additional quarter in these equations?

Else, we are looking at how the holdings change in June, say, based on what the ChatGPT said in March (based on earnings calls in Jan-Mar).

Minor comments (2)

One should be able to do a more direct test of:

ChatGPT gives a high score to stock j .



Holdings of stock j by fund i increase.



This contributes to high return of fund i .

Currently, we only know that GenAIReliance by fund i leads to high return for fund i .

Minor comments (3)

It can happen that $GenAIReliance_{i,t} > 0$ but $GenAIReliance_{i,t+1} \approx 0$, or vice versa.

Meaning the HF i used ChatGPT in quarter t but stopped using it in quarter $t + 1$.

Obviously, this HF knows how to use ChatGPT even in quarter $t + 1$.

Why did it stop using it then?

Some investigation of 'switchers' could be illuminating.

Minor comments (4)

Why is reliance of ChatGPT is negative for performance before 2022?

Dep. Var.	(1) <i>CAPM Alpha</i>	(2)	(3) <i>FF3 Alpha</i>	(4)	(5) <i>FF4 Alpha</i>	(6)
<i>GenAI Reliance</i> × <i>Post GPT</i>	1.411** (2.39)	1.633*** (2.83)	2.629*** (4.81)	2.445*** (4.46)	2.725*** (4.69)	2.537*** (4.46)
<i>GenAI Reliance</i>	-0.536* (-1.77)	-0.396 (-1.27)	-0.673*** (-2.62)	-0.538** (-2.00)	-0.608** (-2.32)	-0.511* (-1.85)
<i>Size</i>	-0.118*** (-3.28)	-0.115*** (-3.23)	-0.112*** (-3.50)	-0.108*** (-3.32)	-0.112*** (-3.60)	-0.113*** (-3.52)
<i>Age</i>	0.002 (0.32)	0.004 (0.54)	0.002 (0.35)	0.002 (0.35)	-0.001 (-0.19)	0.002 (0.33)
<i>Turnover</i>	-0.444 (-1.35)	-0.447 (-1.37)	-0.633** (-2.19)	-0.558* (-1.86)	-0.658** (-2.28)	-0.547* (-1.82)
<i>Risk</i>	0.359 (0.19)	-2.541 (-0.79)	-5.556*** (-4.03)	-8.684*** (-3.43)	-5.878*** (-3.96)	-8.069*** (-2.95)
<i>Past Return</i>	-0.032*** (-5.25)	0.048*** (3.02)	-0.018*** (-3.44)	-0.016 (-1.19)	-0.017*** (-3.01)	-0.018 (-1.31)
<i>Post GPT</i>	-1.168*** (-8.04)		-0.592*** (-4.20)		-0.556*** (-3.72)	
Observations	11,921	11,921	11,921	11,921	11,921	11,921
R-squared	0.009	0.096	0.011	0.066	0.011	0.063
Time FE	No	Yes	No	Yes	No	Yes

Philosophical meandering (1)

ChatGPT is very good in predicting:

- corporate policies (Jha, Qian, Weber, and Yang 2025), and
- stock returns (Lopez-Lira and Tang 2024).

If today you ask ChatGPT, given the information it had in Dec-2017, what would it have predicted about returns on stocks in 2018, it would have done a very good job.

Philosophical meandering (2)

HFs (at least some of them) are pretty good at their job too.

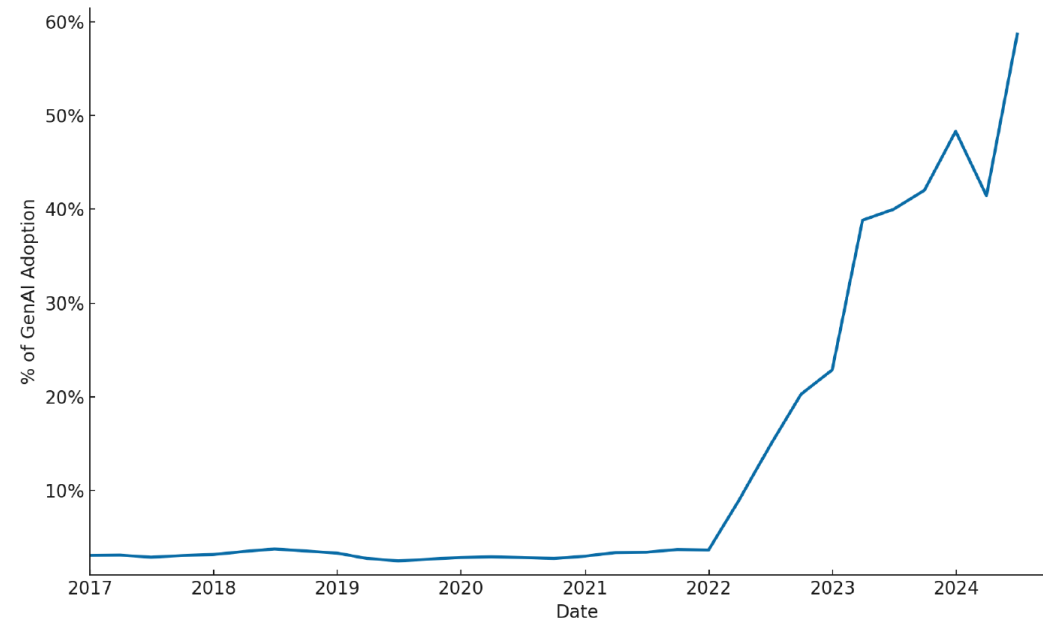
If you had asked them, in Dec-2017, about predicting returns on stocks in 2018, they would have done a decent job. And would have changed their holdings in the direction that they thought stock was going to go.

Philosophical meandering (3)

Putting (1) and (2) together, if, today, one correlates ChatGPT with Δ Holdings in Dec-2017, one should find a high correlation.

Even if, obviously, HFs could not have used ChatGPT in Dec-2017.

Why is it, then, that GenAIReliance is close to zero before 2022?



Philosophical meandering (4)

For the chart to make sense, it must be that, contrary to my assumptions, either (a) ChatGPT was bad; or (b) HFs were bad at their job before 2022.

Year	Long-short HF	S&P500	Difference	
2024	15%	23%	−9%	Average = −3%
2023	11%	24%	−13%	
2022	−6%	−19%	14%	
2021	8%	27%	−19%	Average = −10%
2020	8%	16%	−8%	
2019	12%	29%	−17%	
2018	−5%	−6%	2%	
2017	13%	19%	−6%	
2016	−3%	10%	−13%	

Final thoughts

I like the paper and believe the central result.

Main suggestions:

1. Refine the main independent variable.
2. Clarify the big picture mechanism.

I wish the authors best with the publication process, although I believe that they do not really need my wishes!