Sparse Autoencoder Features of 8-K Filings in a Beta-Neutral Equity Book

QUANTT
McGill–FIAM Asset Management Hackathon 2026, final round

October 2026

Abstract. A US firm must disclose a material event on Form 8-K within four business days. We ask whether the representation a language model forms while reading these filings predicts the cross-section of returns once a mandate's constraints and trading costs are imposed. Following Chen et al. (2025), each filing is passed through Gemma-2-9B-IT and the activations of the Gemma Scope sparse autoencoder at layer 20 are recorded. A ridge regression, re-estimated annually, maps firm-month activations to next-month return ranks, and the firms with the lowest forecasts form the short leg. The long leg ranks large firms on a characteristic score built with the shrinkage estimator of Kozak, Nagel, and Santosh (2020), augmented with card-transaction growth and supply-chain momentum. Weights are the projection of a beta-neutral target onto the constraint set at 8% ex-ante volatility. From January 2021 to August 2026 the portfolio earns, net of costs, an information ratio of 0.82 against T-bill + 4% with a realised market beta of 0.13.

1 Data and design

We use the FIAM US stock panel and its 8-K corpus, keyed by permno and dated by filing date. Signals are formed at each month-end t from information dated t or earlier. The portfolio is formed at the close of $t{+}1$ and its first return runs from that close to the next. Models are re-estimated each January on an expanding window that starts in February 2015. For test year Y, estimation ends in December $Y{-}3$ and hyperparameters are chosen on $Y{-}2$ and $Y{-}1$. We report a validation period, 2019 to 2020, and a test period from January 2021 to August 2026 (68 months, 1,421 trading days). Short candidates are the 2,000 stocks with the highest combined rank of market capitalisation and dollar volume, as in Frazzini, Israel, and Moskowitz (2018), with a market capitalisation of at least US$500m at formation. Long candidates are the 500 largest stocks with a characteristic score.

2 Signals

Sparse embeddings of 8-K text. Each filing is split into item sections. In the notation of Chen et al. (2025), let $r_\tau^{(l)} \in \mathbb{R}^{3584}$ be the residual stream of Gemma-2-9B-IT (Gemma Team, 2024) at layer $l = 20$ for token τ. The Gemma Scope autoencoder (Lieberum et al., 2024) maps it to the sparse code

$$z_\tau^{(l)} = f\big(W_e r_\tau^{(l)} + b_e\big) \in \mathbb{R}^{131{,}072}, \tag{1}$$

where f is the JumpReLU activation of Rajamanoharan et al. (2024). Let $\bar{z}_s$ be the mean of $z_\tau^{(l)}$ over the tokens of section s and $S_{i,t}$ the sections filed by firm i in month t. The firm-month embedding is

$$x_{i,t} = \log\Big(1 + \tfrac{1}{|S_{i,t}|}\textstyle\sum_{s\in S_{i,t}} \bar{z}_{s,\mathcal{F}}\Big) \in \mathbb{R}^{4096}, \tag{2}$$

where $\mathcal{F}$ holds the 4,096 features with the largest variance of $\log(1+z)$ among those active in at least 0.5% of pre-2019 text. Each feature is demeaned within month and divided by its pre-2019 standard deviation, giving $\tilde{x}_{i,t}$.

Return forecast. As in Chen, Kelly, and Xiu (2022), we regress the centred cross-sectional rank of next-month excess return, $y_{i,t+1} = \mathrm{rank}_t(r^e_{i,t+1})/N_t - \frac{1}{2}$, on $\tilde{x}_{i,t}$:

$$\hat{\beta}(\alpha) = \big(\tilde{X}^\top\tilde{X} + \alpha\bar{\lambda}I\big)^{-1}\tilde{X}^\top y, \tag{3}$$

where $\bar{\lambda}$ is the mean eigenvalue of $\tilde{X}^\top\tilde{X}$ and $\alpha \in \{10^{-2},\dots,10^{3}\}$. Forecasts $p_{i,t} = \tilde{x}_{i,t}^\top\hat{\beta}$ are averaged over a firm's recent filings with weight $2^{-k/h}$ on a filing k months old, $h \in \{1,3,6\}$. The pair (α, h) and the choice between narrative items alone and all sections maximise the mean monthly rank correlation with y over the validation years. The short score $s_{i,t}$ is the cross-sectional z-score of the forecast, ranked within GICS sector as in Asness, Porter, and Stevens (2000).

Short exclusions. Among stocks with a below-median forecast, we exclude any in the top 30% of the cross-section on one of three measures: analyst revision breadth, (up − down)/estimates for the current fiscal year (Chan, Jegadeesh, and Lakonishok, 1996); the change in breadth of ownership, measured by the number of 13F holders (Chen, Hong, and Stein, 2002); and the insider net purchase ratio, $(P-S)/(P+S)$ in shares over six months (Lakonishok and Lee, 2001).

Long score. Following Asness, Frazzini, and Pedersen (2019), each signal is converted to ranks and standardised, $z_k = (u_k - \bar{u}_k)/\sigma(u_k) = \sqrt{12}\,(u_k - \frac{1}{2})$, and the long score is

$$L_{i,t} = \tfrac{1}{2}\big(z^{\mathrm{SC}}_{i,t} + z^{\mathrm{card}}_{i,t} + z^{\mathrm{cust}}_{i,t} + z^{\mathrm{supp}}_{i,t}\big), \tag{4}$$

with a missing signal set to zero. The characteristic score z^{SC}, ranked within GICS sector, uses the dual-penalty estimator of Kozak, Nagel, and Santosh (2020). With F the returns of top-minus-bottom decile portfolios on 147 characteristics, and μ and Σ their sample mean and covariance, the SDF coefficients are

$$\begin{aligned}\hat{b} = \arg\min_b\ &(\mu - \Sigma b)^\top\Sigma^{-1}(\mu - \Sigma b)\\ &+ \gamma_2\, b^\top b + \gamma_1 \textstyle\sum_k |b_k|,\end{aligned} \tag{5}$$

with (γ_1, γ_2) chosen on the validation years, and a stock's score is $c_{i,t}^\top\hat{b}$, where $c_{i,t}$ holds its centred characteristic ranks, set to zero outside the top and bottom deciles. Card-transaction growth is $g_{i,t} = \ln\sum_{m=1}^{3} TX_{i,t-m} -$

$\ln \sum_{m=13}^{15} TX_{i,t-m}$, from Carbon Arc credit-card transaction counts, motivated by the evidence in Agarwal, Qian, and Zou (2021) that disaggregated consumer spending predicts returns. Customer momentum follows Menzly and Ozbas (2010), $\mathrm{CM}_{i,t} = \sum_{j \neq J(i)} \omega^{c}_{J(i)j}\, r^{e}_{j,t}$, where $r^{e}_{j,t}$ is the value-weighted excess return of industry j and $\omega^{c}_{J(i)j}$ is the share of industry $J(i)$'s output sold to industry j in the 2012 BEA use table. Supplier momentum uses input shares.

3 Portfolio construction

Selection. Let $\hat{\lambda}$ be the walk-forward slope of monthly returns on the score for each leg. A new position of size $v = \mathrm{NAV}/125$ in a stock with average daily dollar volume V_i, filled at 1% of V_i per day, takes $D_i = v/(0.01 V_i)$ days; the average fraction of a month of $H = 21$ days for which it is held is $\phi_i = 1 - D_i/(2H)$ for $D_i \le H$ and $H/(2D_i)$ otherwise. Each leg holds the 125 stocks with the highest net expected value

$$\mathrm{EV}_{i,t} = \hat{\lambda}\, |z_{i,t}|\, \phi_i - \kappa_i - c^{b}_{i}, \tag{6}$$

where κ_i is the expected price impact of entry and exit and c^{b}_{i} is one month of borrow fee for a short. Incumbents have $\phi_i = 1$ and $\kappa_i = 0$ and are retained while they rank in the top 625, a buy/hold band in the spirit of Novy-Marx and Velikov (2016).

Target weights. Longs are weighted in proportion to $\sqrt{\mathrm{ME}_i}$ and shorts in proportion to rank. The legs are scaled so that $\beta^{\top} w^{T} = 0$, with β the 60-month beta supplied in the panel.

Projection. The traded target is

$$\begin{aligned} w^{\star} = \arg\min_{w} \; & \|w - w^{T}\|_2^2 \\ \text{s.t.} \;\; & w^{\top} \hat{\Sigma} w \le (0.08)^2, \quad \beta^{\top} w = 0, \\ & \|w\|_1 \le 2, \quad |\mathbf{1}^{\top} w| \le 0.5, \\ & |w_i| \le 0.05, \quad w_i\, w^{T}_{i} \ge 0 \quad \forall i, \\ & \left| \textstyle\sum_{i \in g} w_i \right| \le 0.20 \quad \forall g \in \mathcal{G}, \end{aligned} \tag{7}$$

where $\mathcal{G}$ is the set of GICS sectors and $\hat{\Sigma}$ is 252 times the constant-correlation shrinkage estimator of Ledoit and Wolf (2004) applied to the previous 252 daily returns. The problem is a convex quadratically constrained program, solved with CVXPY (Diamond and Boyd, 2016) and Clarabel (Goulart and Chen, 2024).

Trading and costs. Orders are executed daily at no more than 1% of each stock's 126-day average dollar volume; drift between formations is not traded. The net exposure and sector limits are re-imposed at every close. Trading costs are IBKR Pro Tiered commissions plus regulatory fees and the price impact of Jensen et al. (2026), $0.1\,\tau^2/V_i$ for a trade of τ dollars. Borrow fees follow a curve fitted to the IBKR short-availability list.

4 Evaluation

With daily benchmark $b_d = \mathrm{TB3MS}_d/25200 + 0.04/252$, the information ratio is $\sqrt{252}$ times the mean of $r_d - b_d$ divided by its standard deviation. Monthly statistics use compounded monthly returns and $\sqrt{12}$. The Sharpe ratio and the Sortino ratio (Sortino and Price, 1994) use returns in excess of the T-bill, the latter with downside deviation measured below zero.

Table 1: Results net of all costs unless stated. IR is the information ratio against T-bill + 4%.

	Valid. 2019–20	Test 2021–26
IR, daily / monthly	1.33 / 1.43	0.82 / 0.83
IR before costs, daily		1.00
IR, flat borrow fee, daily		0.63
IR to / from 27 Jun 2024, daily		1.02 / 0.52
Sharpe ratio, daily	1.74	1.31
Sortino ratio, daily		1.98
Volatility p.a.	9.7%	8.2%
Beta to S&P 500, daily / monthly	0.20 / 0.09	0.13 / 0.14
Maximum drawdown	−9.1%	−5.9%
Ending value of US$1bn		US$2.16bn
FF5 + UMD alpha p.a. (t), daily		5.2% (2.11)
One-way turnover p.a.	4.9×	4.5×

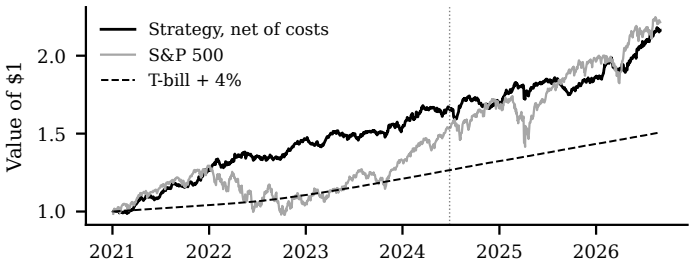


Figure 1: Value of $1, test period, daily. The dotted line marks 27 June 2024.

Beta is the OLS slope on the S&P 500. A separate program that reads only the holdings file verifies every mandate rule at each of the 1,421 closes.

5 Results

The average excess return of the portfolio after formation rises from 0.07% on the first day ($t = 1.06$) to 3.29% after 63 trading days ($t = 6.97$). Predictability accrues over weeks, hence monthly formation. In the Arcana US risk model, US$538m of the US$1,271m cumulative profit is attributed to residual returns. The short leg contributes US$382m of this residual although its total profit is −US$143m. Its stocks rise with the market but by less than their factor exposures imply.

6 Limitations

First, filings published before the release of Gemma-2 on 27 June 2024 may be part of its training data. The information ratio is 1.02 before that date and 0.52 after it, and we regard the later figure as the more reliable estimate. Second, Carbon Arc does not publish vintages of the card data, so the backtest uses the current vintage lagged one month, and revisions introduce look-ahead bias. Third, the realised beta of 0.13 arises from execution. Longs fill within a day and shorts within a week or more, so the book is net long between formations.

References

Agarwal, Sumit, Wenlan Qian, and Xin Zou, 2021, Disaggregated sales and stock returns, *Management Science* 67, 7167–7183.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Asness, Clifford S., R. Burt Porter, and Ross L. Stevens, 2000, Predicting stock returns using industry-relative firm characteristics, Working paper, AQR Capital Management.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1713.

Chen, Hui, Antoine Didisheim, Mohammad Pourmohammadi, Luciano Somoza, and Hanqing Tian, 2025, A financial brain scan of the LLM, arXiv preprint arXiv:2508.21285.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Chen, Yifei, Bryan T. Kelly, and Dacheng Xiu, 2022, Expected returns and large language models, Working paper, University of Chicago.

Diamond, Steven, and Stephen Boyd, 2016, CVXPY: A Python-embedded modeling language for convex optimization, *Journal of Machine Learning Research* 17, 1–5.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2018, Trading costs, Available at SSRN 3229719.

Gemma Team, 2024, Gemma 2: Improving open language models at a practical size, arXiv preprint arXiv:2408.00118.

Goulart, Paul J., and Yuwen Chen, 2024, Clarabel: An interior-point solver for conic programs with quadratic objectives, arXiv preprint arXiv:2405.12762.

Jensen, Theis I., Bryan T. Kelly, Semyon Malamud, and Lasse Heje Pedersen, 2026, Machine learning and the implementable efficient frontier, *Review of Financial Studies*, forthcoming.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lakonishok, Josef, and Inmoo Lee, 2001, Are insider trades informative?, *Review of Financial Studies* 14, 79–111.

Ledoit, Olivier, and Michael Wolf, 2004, Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management* 30, 110–119.

Lieberum, Tom, et al., 2024, Gemma Scope: Open sparse autoencoders everywhere all at once on Gemma 2, arXiv preprint arXiv:2408.05147.

Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance* 65, 1555–1580.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Rajamanoharan, Senthooran, et al., 2024, Jumping ahead: Improving reconstruction fidelity with JumpReLU sparse autoencoders, arXiv preprint arXiv:2407.14435.

Sortino, Frank A., and Lee N. Price, 1994, Performance measurement in a downside risk framework, *Journal of Investing* 3, 59–64.