Asset Pricing with Attention Guided Deep Learning

Philippe Chatigny, University of Sherbrooke
Ruslan Goyenko, McGill University & visiting Yale School of Management
Chengyu Zhang, McGill University

May 9 2022

INQUIRE UK
Institute for Quantitative Investment Research

FIRM Labs - Financial Innovations and Risk Management Labs

What do we do?

- Optimal, bottom-up, portfolio construction: identify the best stochastic discount factor (SDF)
- *Admissible from a real-time investment perspective* SDF
- Our focus is on after- rebalancing, excluding difficult to arbitrage micro- and small-cap stocks, and after short-selling constraints, performance - after controlling for the limits to arbitrage economic conditions

Why Machine Learning

- More information (fundamental and alternative) is now-days available to investors which is a rich complement to traditionally used data, and permits to make better asset allocation decisions
- Machine Learning enables to use all these data
- ML allows these data to effect portfolio weights allocation in linear and *non-linear* ways

Literature

- ML approaches to portfolio construction has been shown to provide significant improvements compared to the traditional techniques(incomplete list of popular models):
- deep-to-shallow neural network approach of Gu, Kelly, and Xiu (2020, RFS), Chen, Pelger, and Zhu (2020, WP) - deep learning - a generative adversarial network (GAN), (IPCA) per Kelly, Pruitt, and Su (2019 JFE), the conditional *autoencoder* by Gu, Kelly, and Xiu (2021, JE), Risk Premium (RP), RP-PCA, Lettau and Pelger (2020 RFS), asset pricing trees (Bryzgalova et al 2020, WP), Alpha-portfolio, RL-reinforcement learning based, Cong te al (2021, WP).

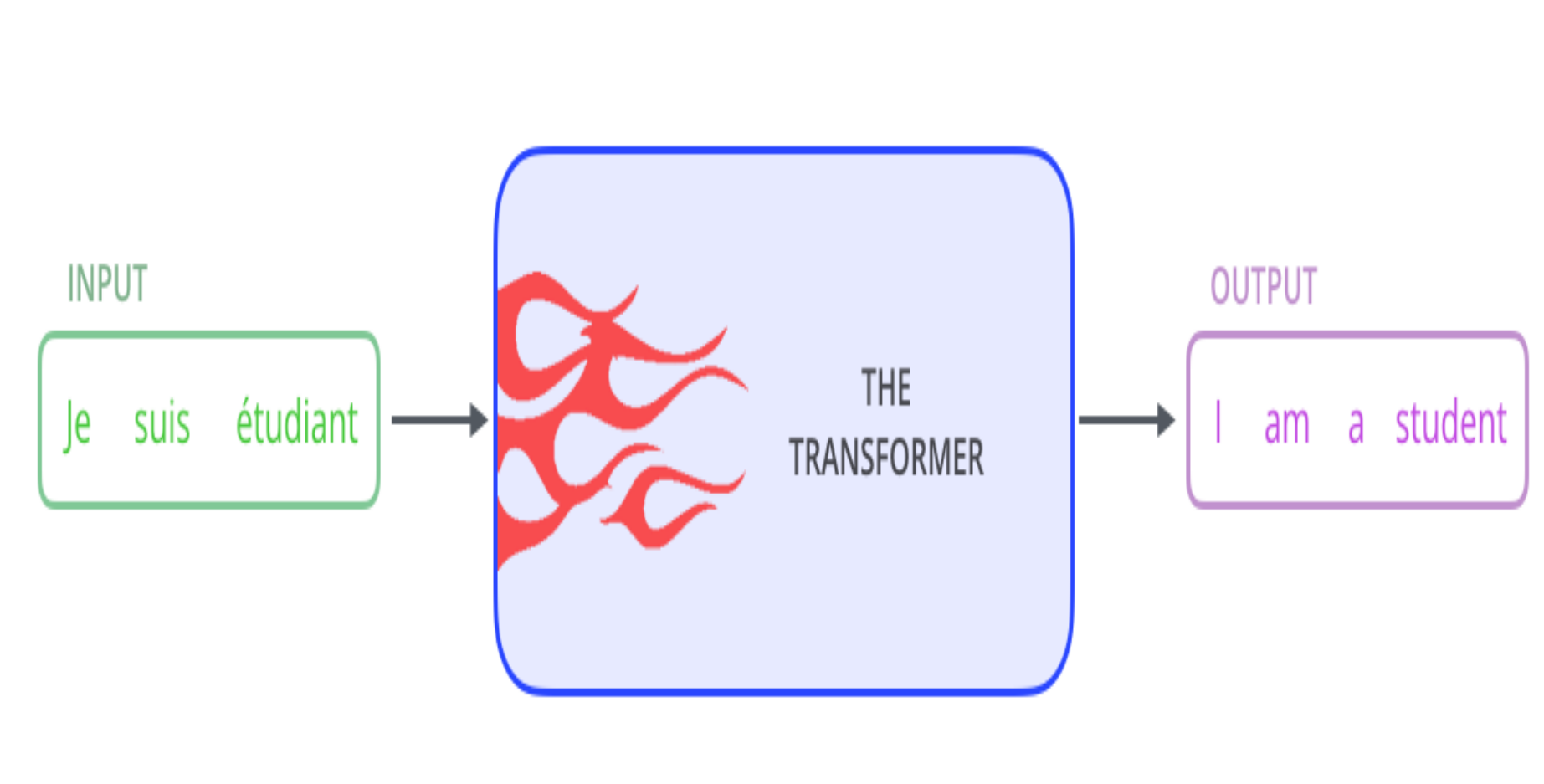
Challenge of this literature

- Avramov, Cheng and Metzker (2022, forthcoming) - *investments based on deep learning signals extract profitability from difficult-to-arbitrage stocks and during high limits-to-arbitrage market states*
- Abnormal performance of most ML models disappears *after excluding micro-cap stocks, firms that do not have credit ratings, after excluding financially distressed firms*
- High monthly turnover/weights in the long-short portfolio positions indicate that *the pricing kernel might be inadmissible from a real-time investment perspective*

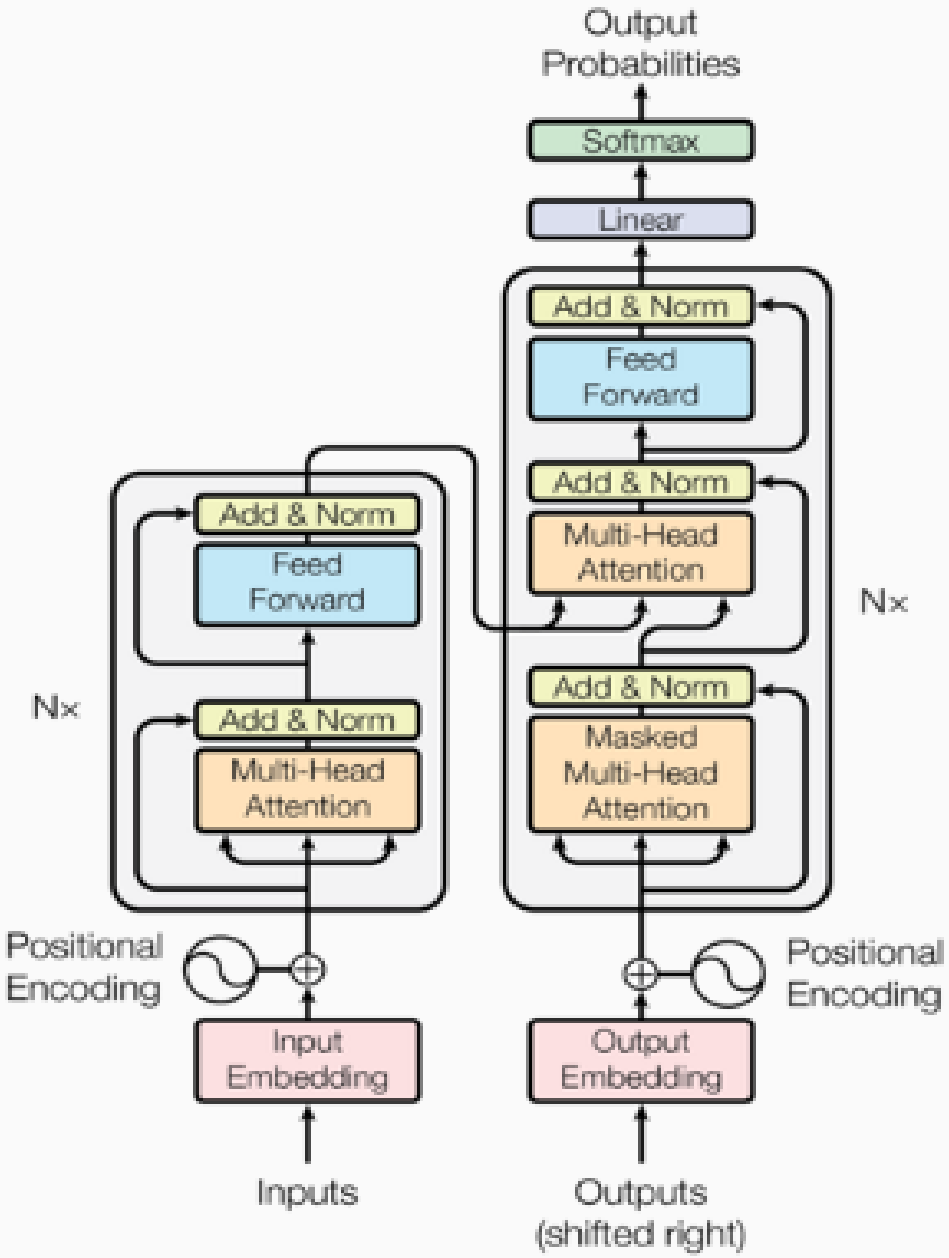
Our Contribution

- Methodology: A new deep learning architecture which incorporates "asset attention" layers
- Our SDF survives Avramov et al (2021) criticism
- *Machine Attention* (commonly used by Transformers - speech recognition and natural language processing) is critical for financial data applications
- using big data, multiple firm characteristics and macro-economic factors, and Deep Learning allows to make significantly (economically and statistically) better asset allocation decisions

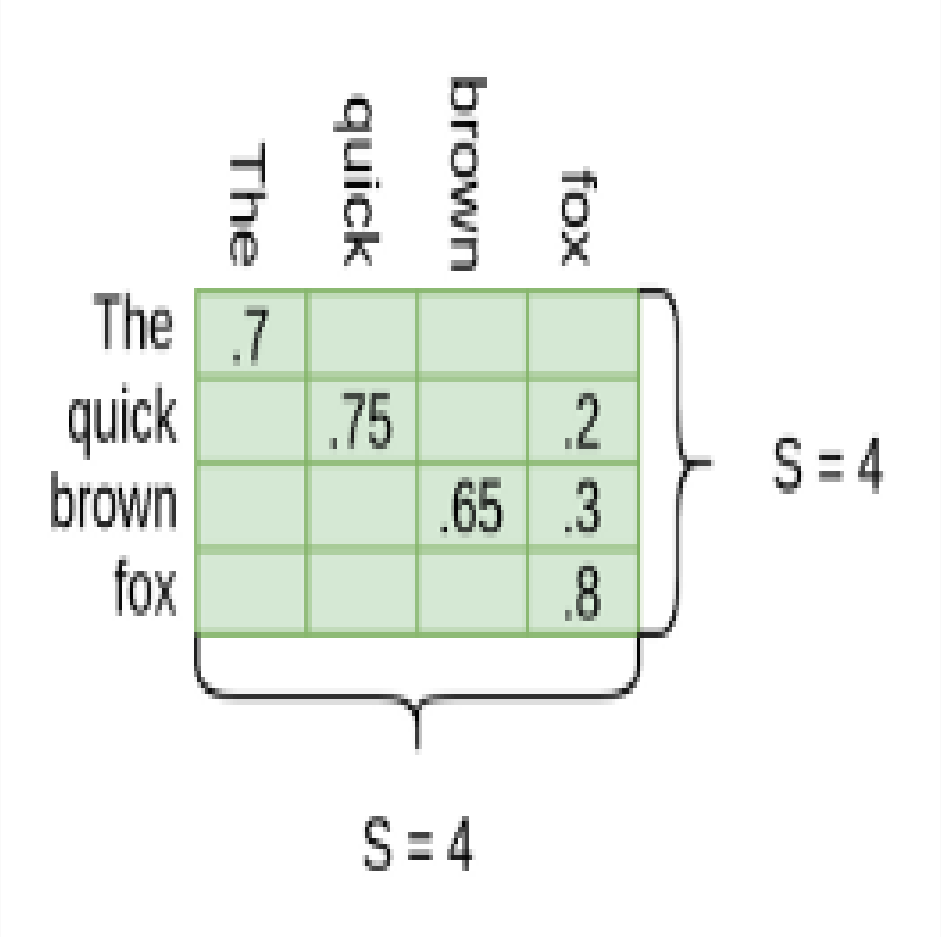
Transformers



Transformer Architecture



Embedding via Attention Layer



Stock Characteristics (Green et al (2017)): Features Part 1

No.	Acronym	Description	No.	Acronym	Description
1	absacc	Absolute Accruals	25	currat	Current ratio
2	acc	Working Capital accruals	26	depr	Deprecation/PP&E
3	aeavol	Abnormal earnings announcement volume	27	divi	Dividend initiation
4	age	# years since first Compustat coverage	28	divo	Dividend omission
5	agr	Asset growth	29	dolvol	Dollar trading volume
6	baspread	Bid-ask spread	30	dy	Dividend to price
7	beta	Beta	31	ear	Earnings announcement return
8	betasq	Beta squared	32	egr	Growth in common shareholder equity
9	bm	Book-to-market	33	ep	Earnings to price
10	bm_ia	Industry-adjusted book to market	34	gma	Gross profitability
11	cash	Cash holdings	35	grcapx	Growth in capital expenditures
12	cashdebt	Cash flow to debt	36	grltnoa	Growth in long-term net operating assets
13	cashpr	Cash productivity	37	herf	Industry sales concentration
14	cfp	Cash flow to price ratio	38	hire	Employee growth rate
15	cfp_ia	Industry-adjusted cash flow to price ratio	39	idiovol	Idiosyncratic return volatility
16	chatoia	Industry-adjusted change in asset turnover	40	ill	Amihud Illiquidity
17	chcsho	Change in shares outstanding	41	indmom	Industry momentum
18	chempia	Industry-adjusted change in employees	42	invest	Capital expenditures and inventory
19	chinv	Change in inventory	43	lev	Leverage
20	chmom	Change in 6-month momentum	44	lgr	Growth in long-term debt
21	chpmia	Industry-adjusted change in profit margin	45	maxret	Maximum daily return
22	chtx	Change in tax expense	46	mom12m	12-month momentum
23	cinvest	Corporate investment	47	mom1m	1-month momentum
24	convind	Convertible debt indicator			

Stock Characteristics (Green et al (2017)): Features Part 2

No.	Acronym	Description	No.	Acronym	Description
48	mom36m	36-month momentum	72	realestate	Real estate holdings
49	mom6m	6-month momentum	73	retvol	Return volatility
50	ms	Mohanram financial statement score	74	roaq	Return on assets
51	mve	Size	75	roavol	Earnings volatility
52	mve_ia	Industry-adjusted size	76	roeq	Return on equity
53	nincr	Number of earnings increases	77	roic	Return on invested capital
54	operprof	Operating profitability	78	rsup	Revenue surprise
55	orgcap	Organizational capital	79	salecash	Sales to cash
56	pchcapx_ia	Industry adjusted percentage change in capital expenditures	80	saleinv	Sales to inventory
57	pchcurrat	Percentage change in current ratio	81	salerec	Sales to receivables
58	pchdepr	Percentage change in depreciation	82	secured	Secured debt
59	pchgm_pchsale	Percentage change in gross margin less percentage change in sales	83	securedind	Secured debt indicator
60	pchquick	Percentage change in quick ratio	84	sgr	Sales growth
61	pchsale_pchinvt	Percentage change in sales less percentage change in inventory	85	sin	Sin stocks
62	pchsale_pchrect	Percentage change in sales less percentage change in A/R	86	sp	Sales to price
63	pchsale_pchxsga	Percentage change in sales less percentage chagne in SG&A	87	std_dolvol	Volatility of liquidity (dollar trading volume)
64	pchsaleinv	Percentage change in sales-to-inventory	88	std_turn	Volatility of liquidity (share turnover)
65	pctacc	Percent accruals	89	stdacc	Accrual volatility
66	pricedelay	Price delay	90	stdcf	Cash flow volatility
67	ps	Piotroski financial statements score	91	tang	Debt capacity/firm tangibility
68	quick	Quick ratio	92	tb	Tax income to book income
69	rd	R&D increase	93	turn	Share turnover
70	rd_mve	R&D to market capiitalization	94	zerotrade	Zero trading days
71	rd_sale	R&D to sales			

Macro-economic variables

No.	Acronym	Description	No.	Acronym	Description
1	dp	Dividend Price Ratio	10	dfy	Default Yield Spread
2	dy	Dividend Yield	11	dfr	Default Return Spread
3	ep	Earnings Price Ratio	12	infl	Consumer Price Index
4	svar	Stock Variance	13	spvw	S&P 500 Index Returns
5	bm	Book-to-Market Ratio	14	vix	VIX Index
6	ntis	Net Equity Expansion	15	amihud	Aggregate Amihud Illiquidity
7	tbl	Treasury-bill Rates	16	tedrate	TED Rate
8	ltr	Long Term Rate of Returns	17	bab	BAB Returns
9	tms	Term Spread	18	cs_baspread	Aggregate Bid-Ask Spread

The Model I

The no-arbitrage assumption implies the existence of stochastic discount factor, SDF, M_{t+1}, which for any excess return, $R^e_{t+1,i}$, satisfies the equation:

$$\mathbb{E}[M_{t+1}R^e_{i,t+1}] = 0 \Leftrightarrow \mathbb{E}_t[R^e_{i,t+1}] = \overbrace{\frac{\mathrm{Cov}_t(R^e_{i,t+1}, M_{t+1})}{\mathrm{Var}_t(M_{t+1})}}^{\beta_{t,i}} \cdot \overbrace{\frac{\mathrm{Var}_t(M_{t+1})}{\mathbb{E}_t[M_{t+1}]}}^{\lambda_t} = \beta_{i,t}\lambda_t$$

this maps to a factor model $R^e_{i,t+1} = \alpha_{i,t} + \beta'_{i,t}F_{t+1} + \epsilon_{i,t+1}$, where $\alpha_{i,t} = 0$ for all i in t.

The Model II

Following Hansen and Jagannathan (1991), the SDF can be formulated as

$$M_{t+1} = 1 - \sum_{i=1}^{N_t} \omega_{t,i} R^e_{i,t+1} = 1 - \boldsymbol{\omega}_t^\top \boldsymbol{R}^e_{t+1}$$

The tangency, optimal portfolio is then defined as $F_{t+1} = \omega_t^\top R^e_{t+1}$ and we denote this factor as the traded SDF.

Conditional SDF Loss Function

For each individual asset i, and time t we consider the following functional form for the weight,

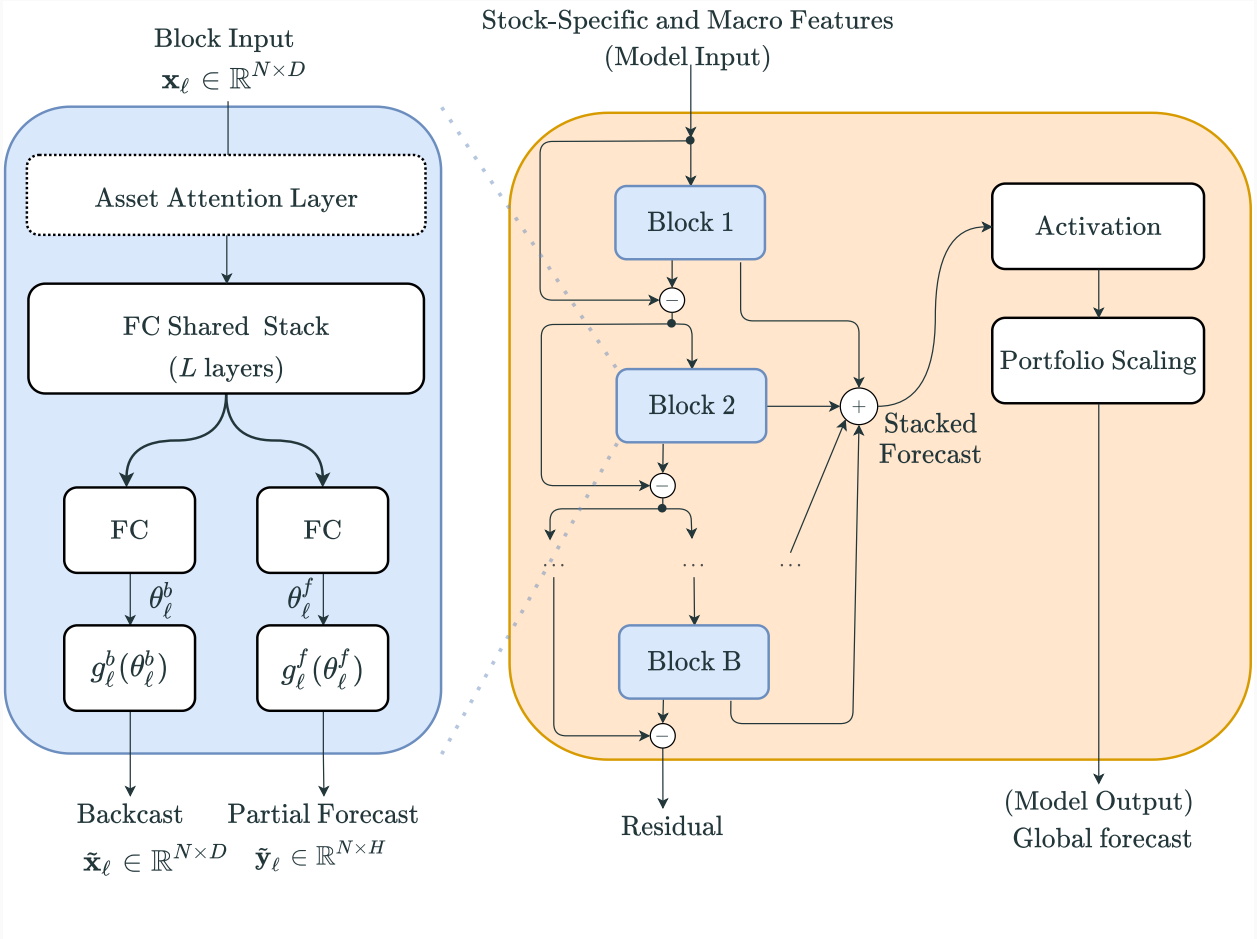
$$\omega_{t,i} = \omega\left(I_t, I_{t,i}\right)$$

where I_t is macro-economic conditioning variables that are not asset specific, and $I_{t,i}$ are firm characteristics which are asset specific (94 feature from Green et al (2017))

The general form of our optimization function is

$$\omega^*_{t,i} = \text{argmin}_{\omega_{t,i}} \frac{1}{N} \sum_{n=1}^{N} \left(\overbrace{\left(1 - \sum_{i=1}^{N_t} \omega_{t,i}\left(I_t, I_{t,i}\right) R^e_{t+1,i}\right)}^{M_{t+1}} R^e_{t+1,n} \right)^2$$

Model Architecture



Empirical Design

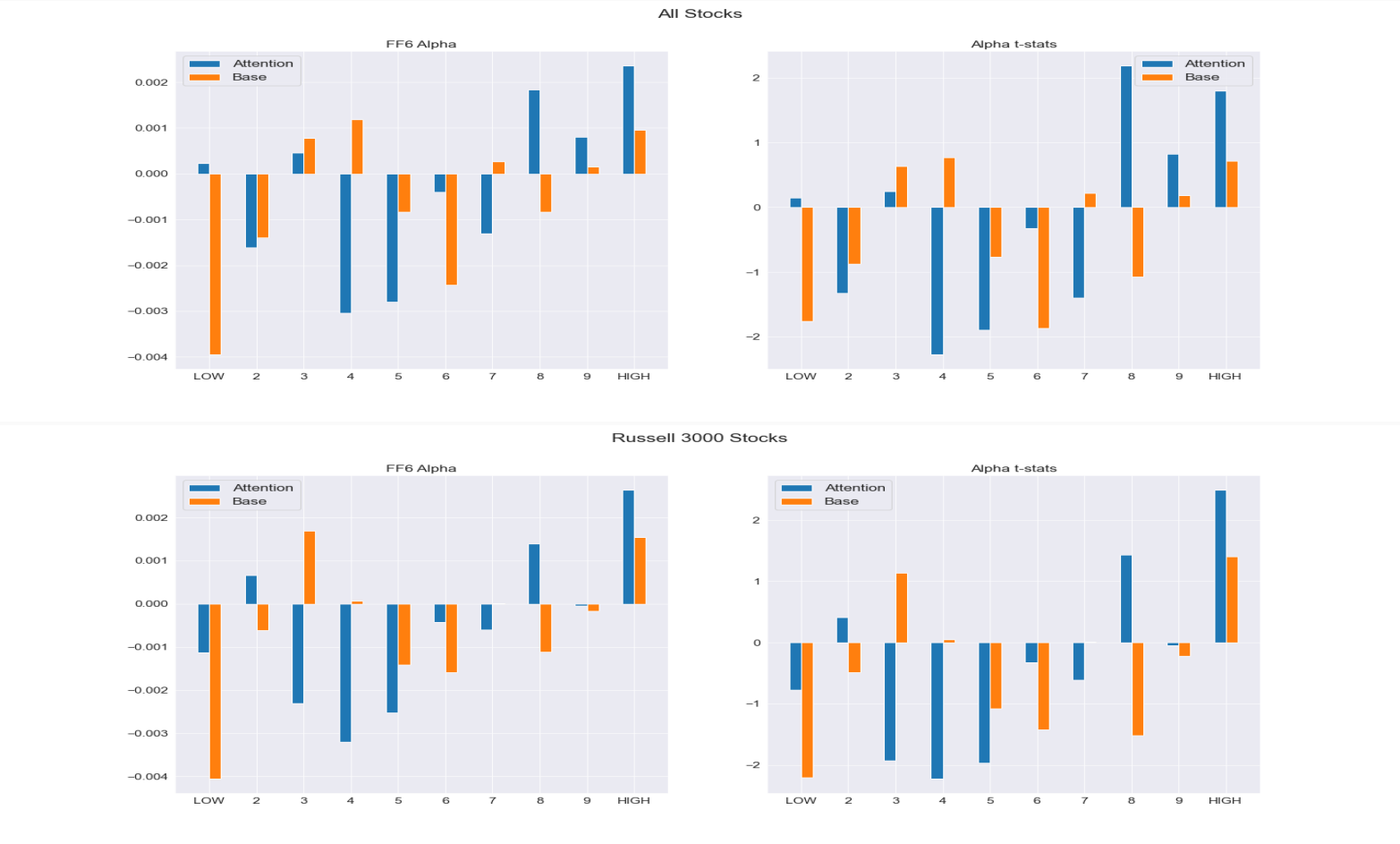
- Our sample period is from 1990-01 to 2020-12 - defined by the availability of CBOE VIX index
- We consider all universe of US publicly listed common stocks as well as the cross-section of 3000 components of Russell 3000 index
- We use the sample from 1990-01 to 2004-12, 180 months, as the first training sample. Our first out-of-sample, *OOS* prediction for SDF weights is therefore for January, 2005.
- We then roll the training sample by 1 month, keeping its 180-month length fixed, and our next *OOS* prediction is for February, 2005, and so on.
- Overall, our *OOS* period is from 2005-01 to 2020-12 for a total of 192 months.

OOS performance

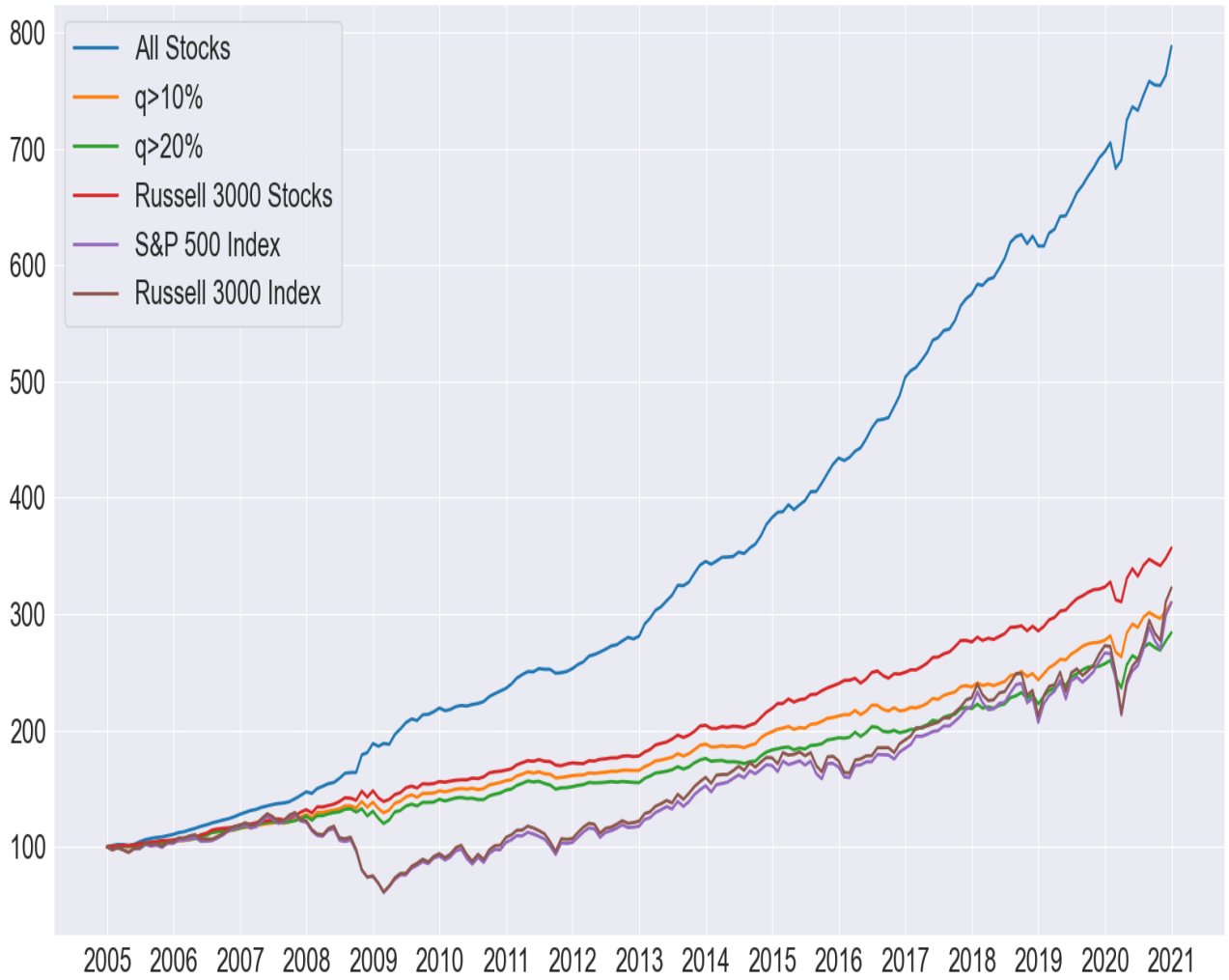
	All Stocks		q>10%		q>20%		Russell 3000 Stocks	
	Attention	**Base**	**Attention**	**Base**	**Attention**	**Base**	**Attention**	**Base**
Return	0.0099	0.0106	0.0050	0.0057	0.0046	0.0058	0.0057	0.0062
Std.Dev.	0.0121	0.0145	0.0138	0.0193	0.0157	0.0237	0.0133	0.0180
Alpha	0.0088	0.0095	0.0031	0.0026	0.0022	0.0017	0.0042	0.0039
Alpha t-stat	8.060	7.657	3.694	2.967	2.616	2.058	4.936	4.145
Sharpe	2.834	2.537	1.255	1.016	1.006	0.842	1.486	1.196
InfRatio	2.639	2.389	0.910	0.682	0.625	0.468	1.214	0.933
MaxDD	0.033	0.060	0.073	0.147	0.103	0.243	0.065	0.131
Max 1M Loss	-0.032	-0.030	-0.051	-0.083	-0.058	-0.107	-0.048	-0.083

OOS SDF Performance

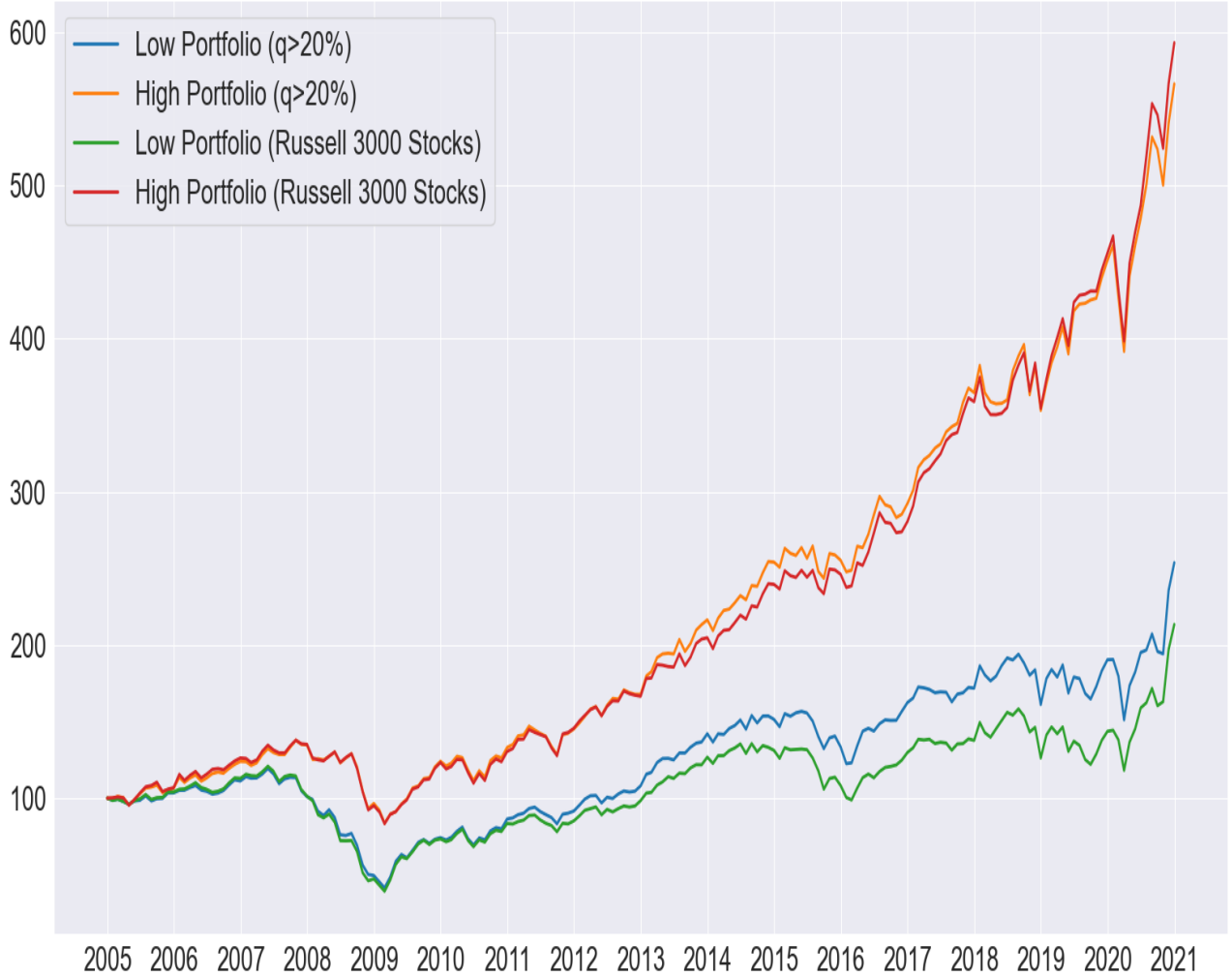
Portfolio Deciles' Performance, Attention vs Base, All Stocks & Russell 3000 stocks



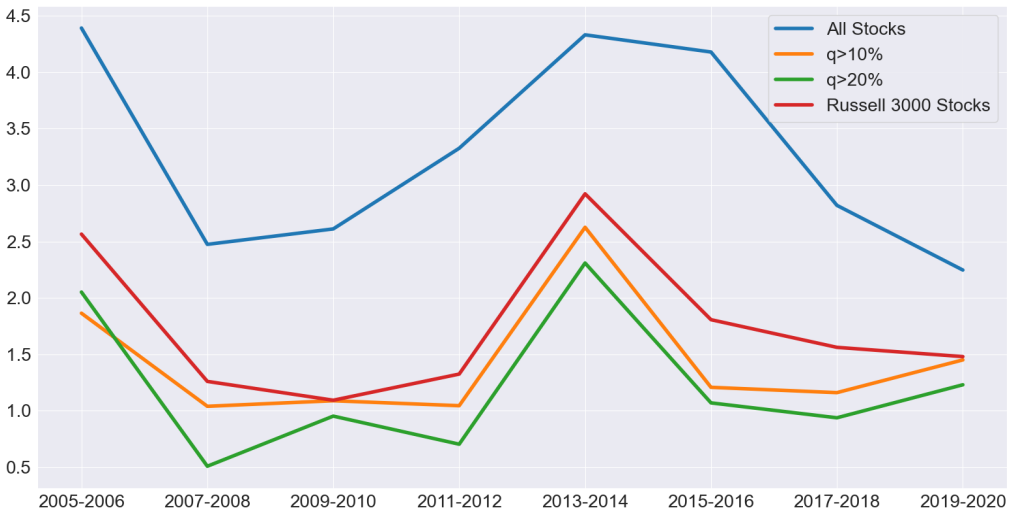
Attention Deep Learning Portfolio Performance



Attention Deep Learning Portfolio: Top and Bottom 20% sub-portfolios Performance



Bi-annual Sharpe Ratios



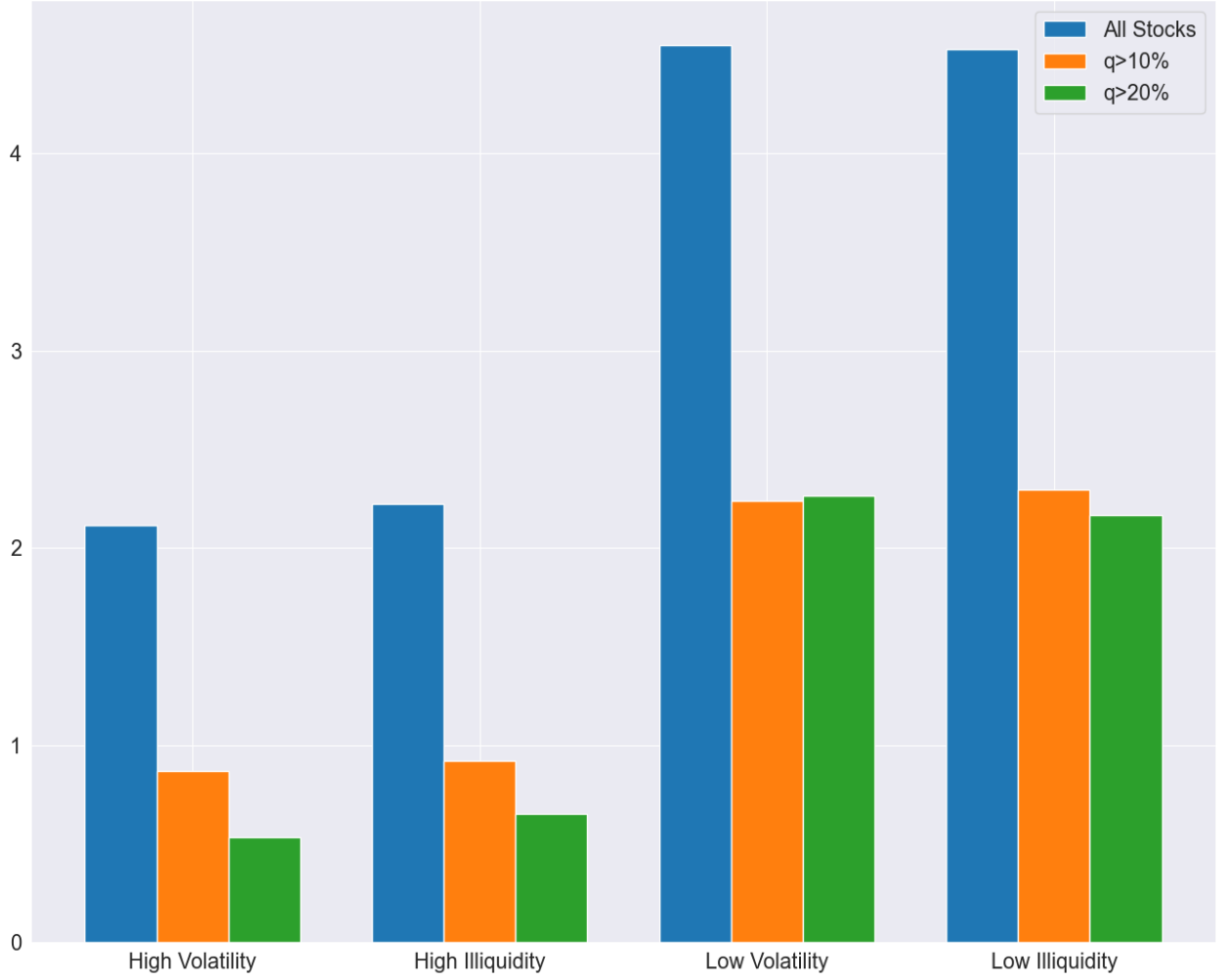
Trading Costs and Turnover: Trading Costs Analysis for *OOS* SDF Performance, Attention Model

	All Stocks	q>10%	q>20%	Russell 3000 Stocks
Turnover All	0.348	0.360	0.352	0.360
Turnover Long	0.227	0.221	0.220	0.223
Turnover Short	0.121	0.139	0.132	0.137
Eff. Spreads All	0.0115	0.0033	0.0025	0.0043
Eff. Spread Long	0.0111	0.0032	0.0024	0.0040
Eff. Spread Short	0.0120	0.0035	0.0026	0.0046
Avg Size Long ($mln)	7,070.83	10,981.45	13,132.47	9,756.96
Avg Size Short ($mln)	3,345.91	5,472.13	7,103.02	4,633.49

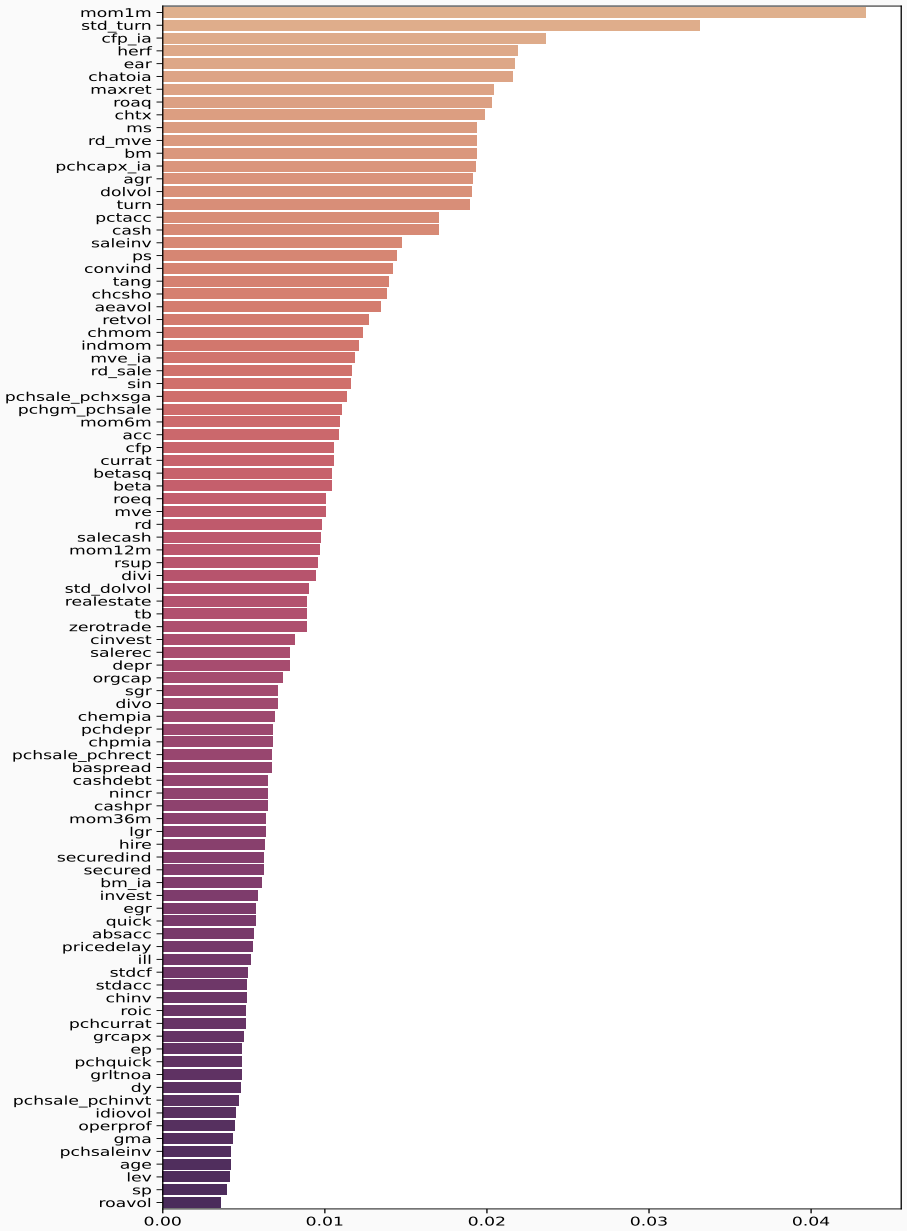
Why is Turnover Low?

- DeMiguel et al (2020, RFS) - important to consider *jointly* multiple firm characteristics for portfolio construction to implicitly account for the whole portfolio's trading costs
- Why? - the trades required to rebalance in one underlying stock based on different characteristics often cancel out
- Consequently - "paying *Attention*" to multiple stock characteristics reduces stocks' turnover on the whole portfolio level

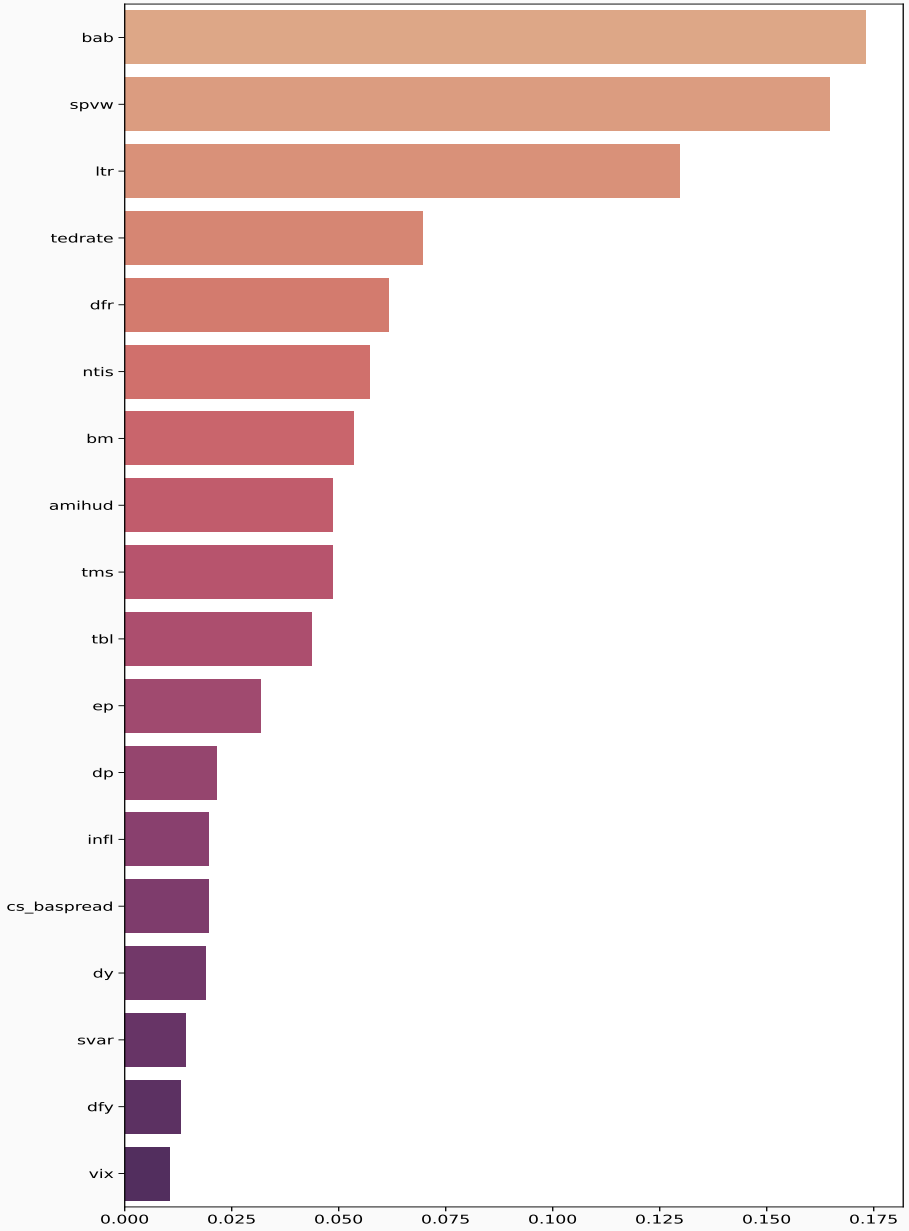
SDF Performance in High vs Low Market Volatility/Illiquidity regimes: annualized *OOS* Sharpe Ratios



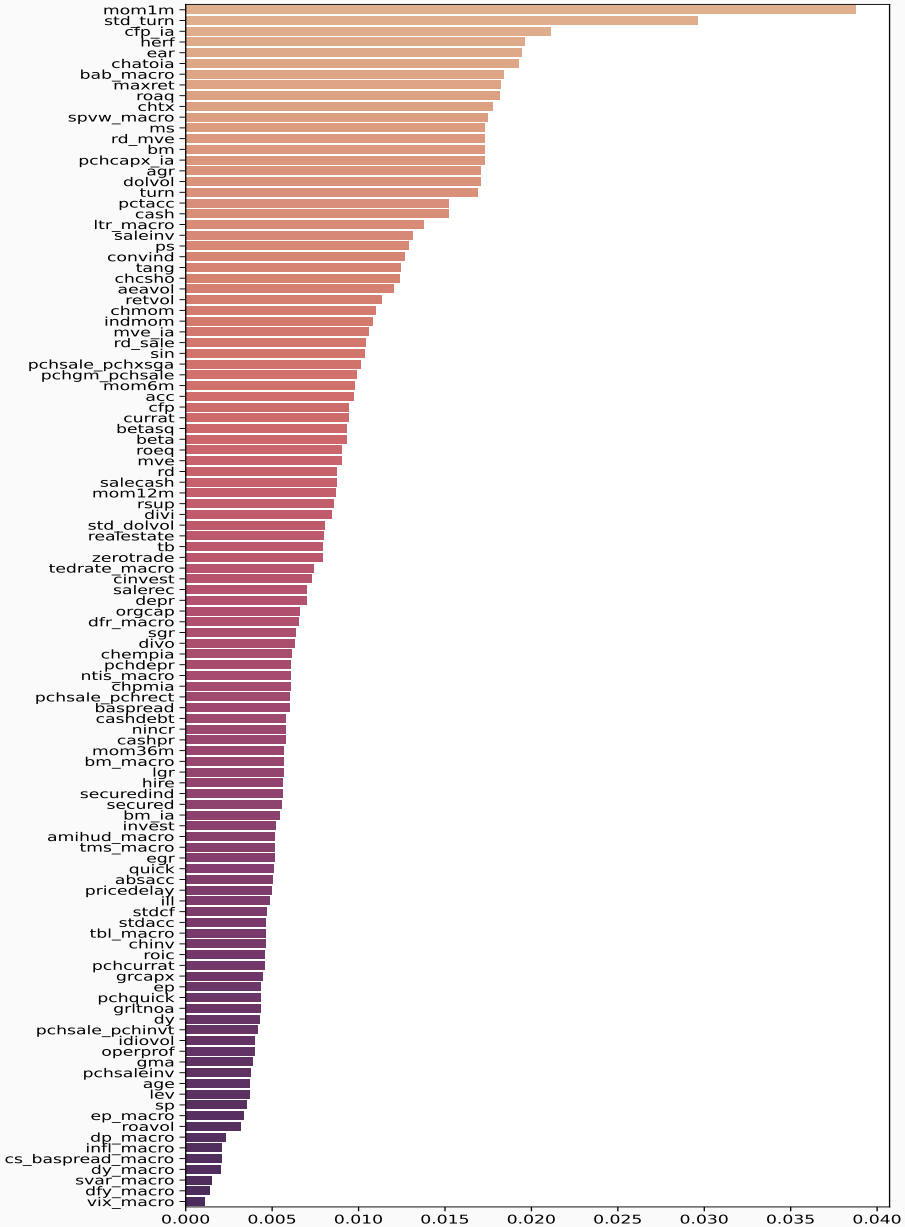
Variable Importance: stock characteristics



Variable Importance: Macro-economic variables



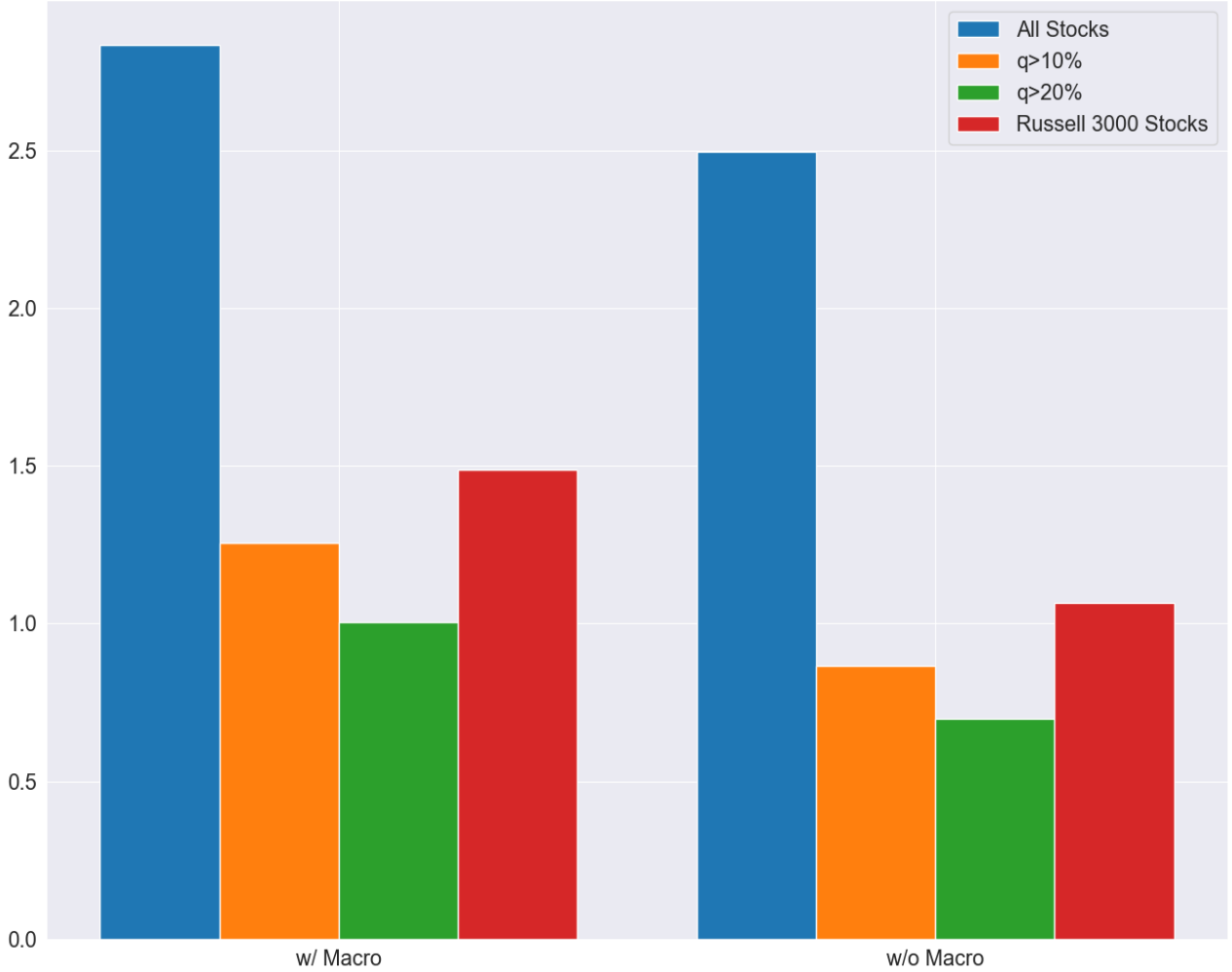
Variable Importance: Macro- & Stock specific variables



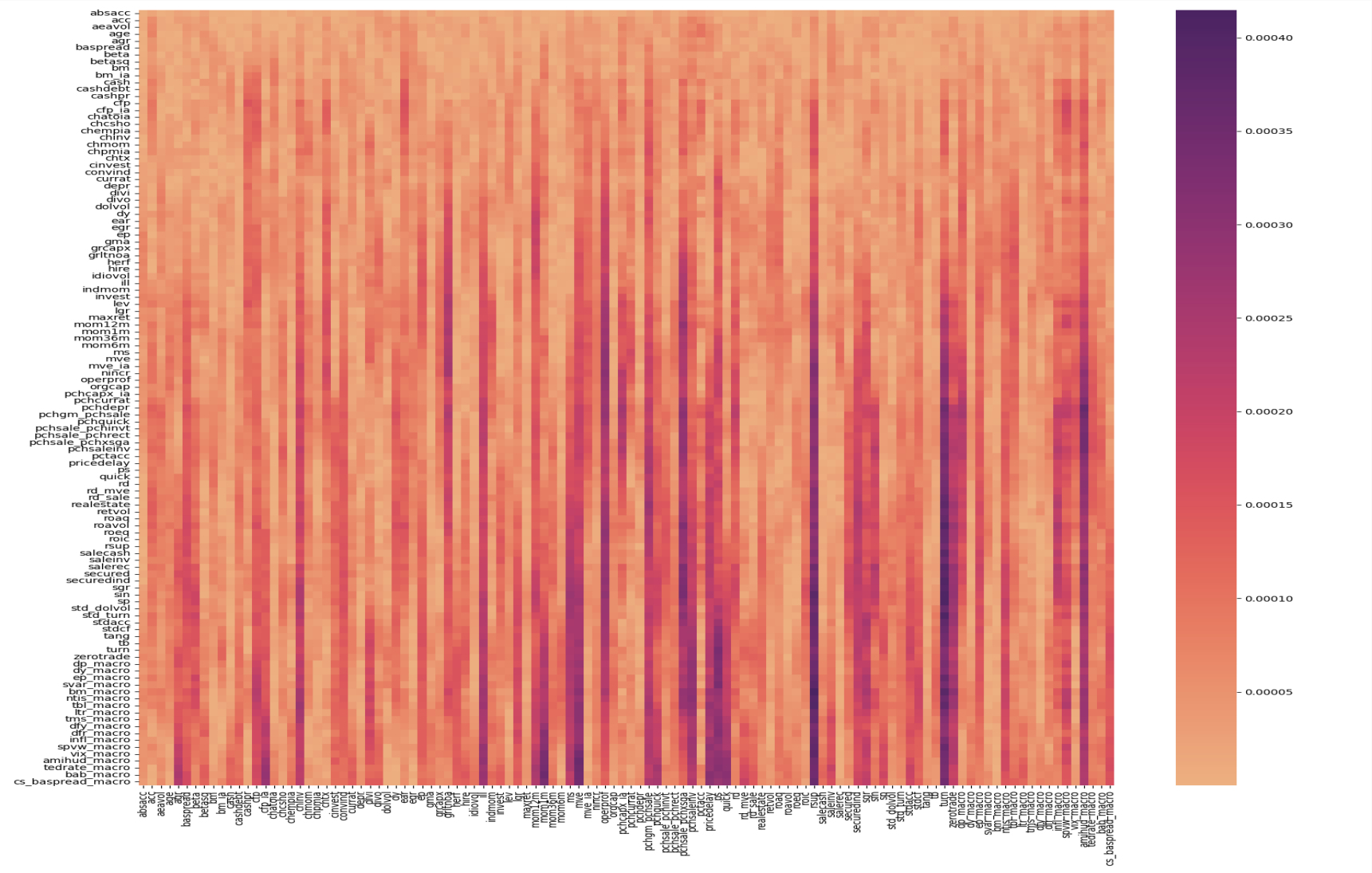
Variable Importance: Summary

- besides price trends and liquidity as in Gu et al (2020)
- other important variables are related to firms' fundamentals: earnings announcement returns (ear), industry-adjusted cash flow to price ratio (*cfp_ai*), Piotrosky financial statements score (ps), industry adjusted percentage change in capex (*pchcapx_ai*), and return on assets (roaq).
- These variables are then followed by return volatility and standard deviation of turnover, and then followed again by fundamentals: industry sales concentration (herf), change in tax expenses (chtx), and percent accruals (pctacc), i.e. firm fundamental and accounting balance sheet variables.
- Macro: Betting against Beta, BAB (Frazzini and Pedersen, 2014, JFE), the market itself, S&P500 index, Long Term rates, ltr.

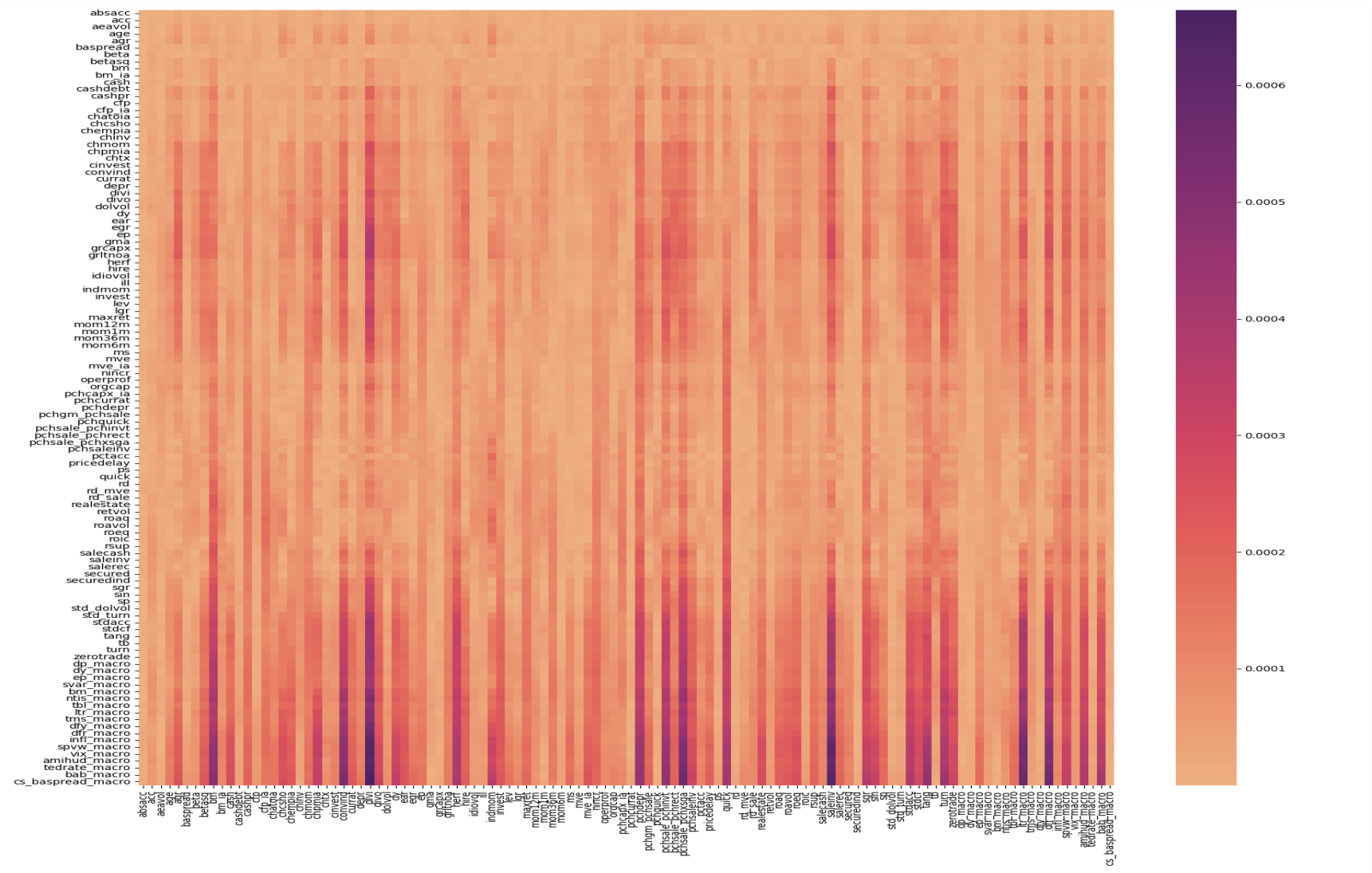
Importance of Macro- variables: OOS Sharpe ratios with and without macro



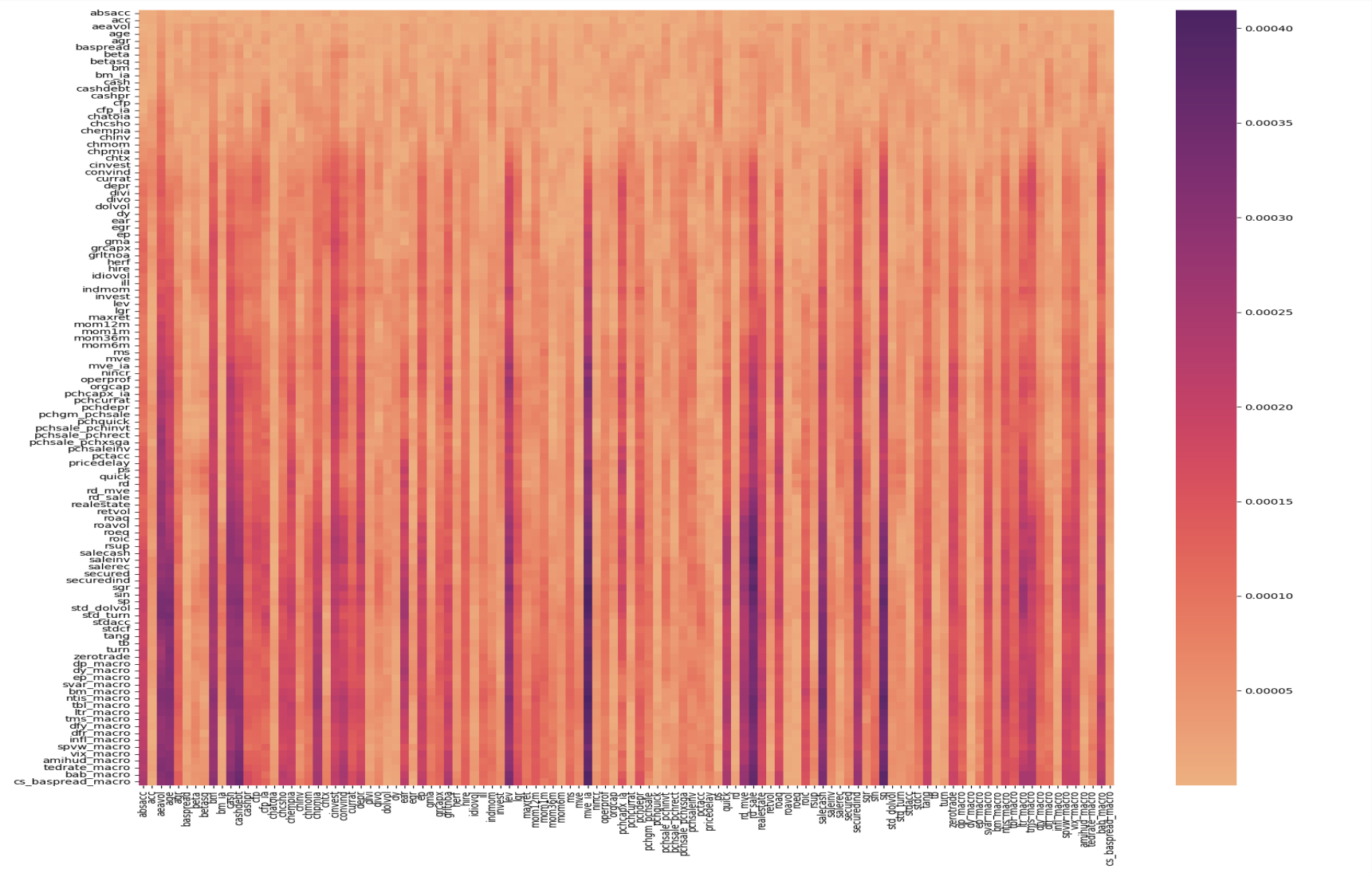
Attention for 12/2020 - Block 1: size, price delay, mom1ret, ps and quick (firm financial & liquidity scores), earning ann. returns, turnover



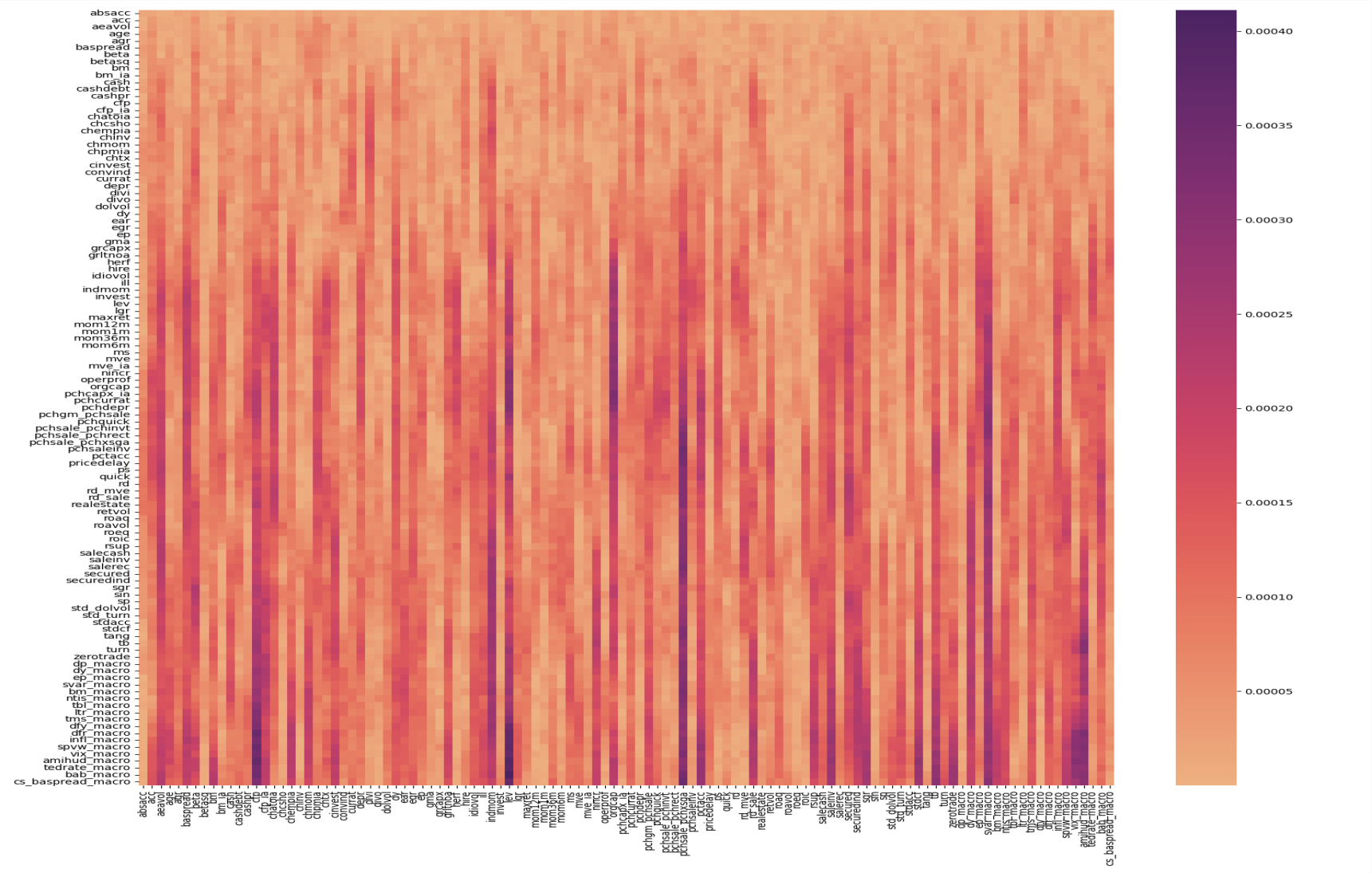
Attention for 12/2020 - Block 2: bm, dividend, percentage change in sales, macro: ltr, dfr



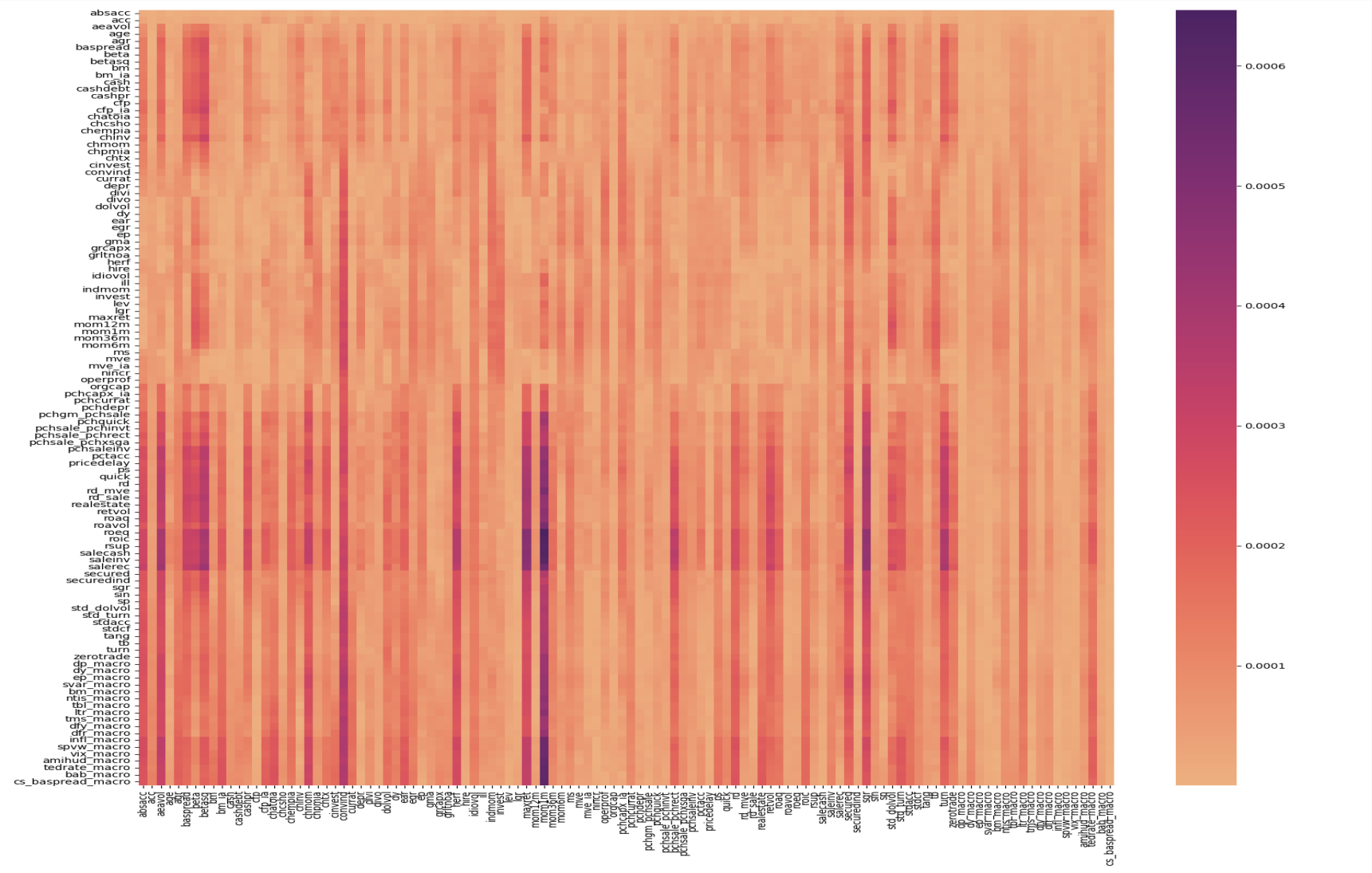
Attention for 12/2020 - Block 3: age, abn earning announce. volume, cash, leverage, ind-adjusted size, R&D to sales



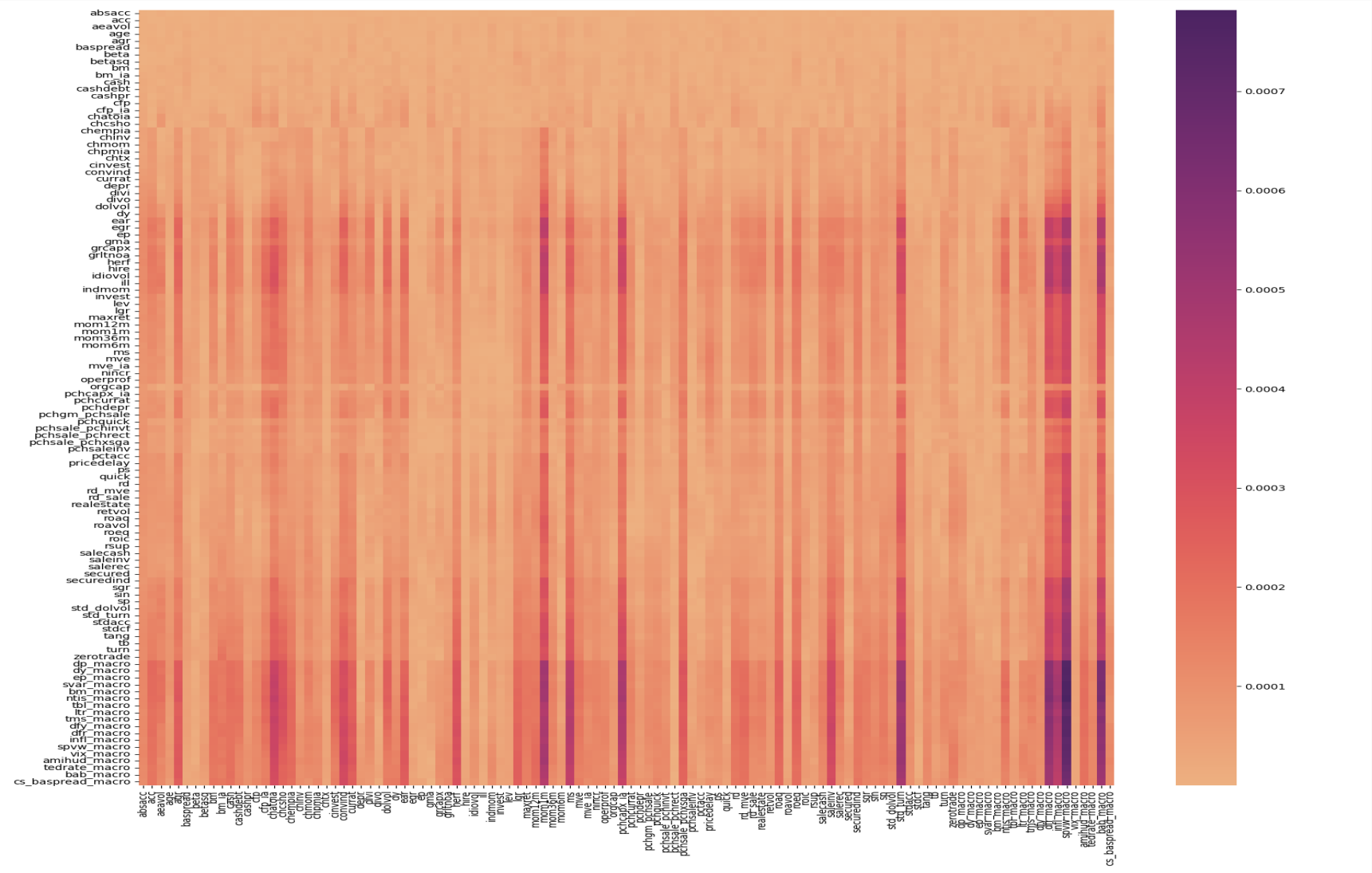
Attention for 12/2020 - Block 4: cash flow to price ratio, leverage, volatility, macro-variables: VIX, market volatility, market-wide illiquidity



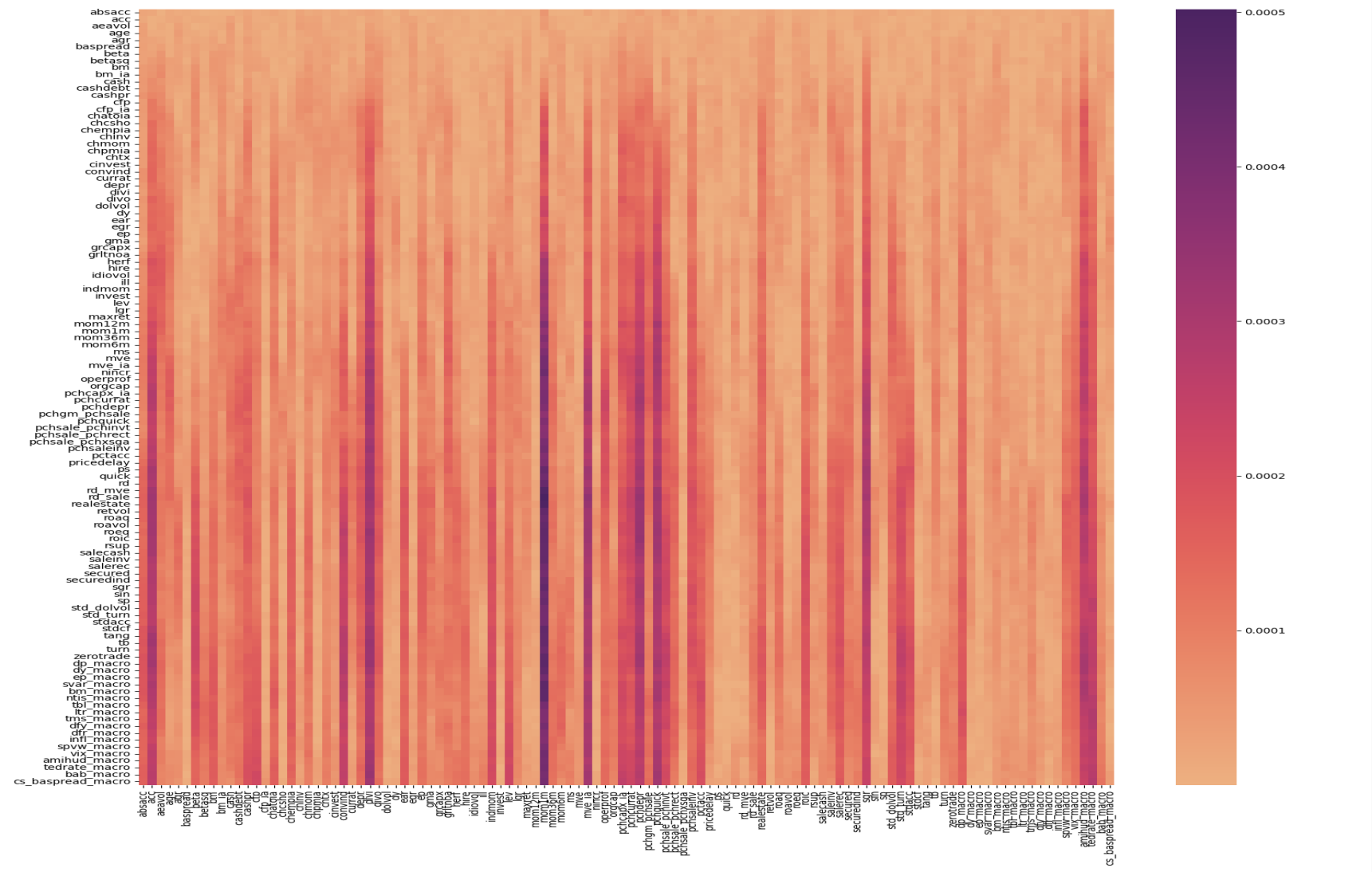
Attention for 12/2020 - Block 5: past returns interactions with sales



Attention for 09/2008 - Block 1: past return, financial statement score, and mostly macro: default rates and spreads, S&P500 ret, inflation

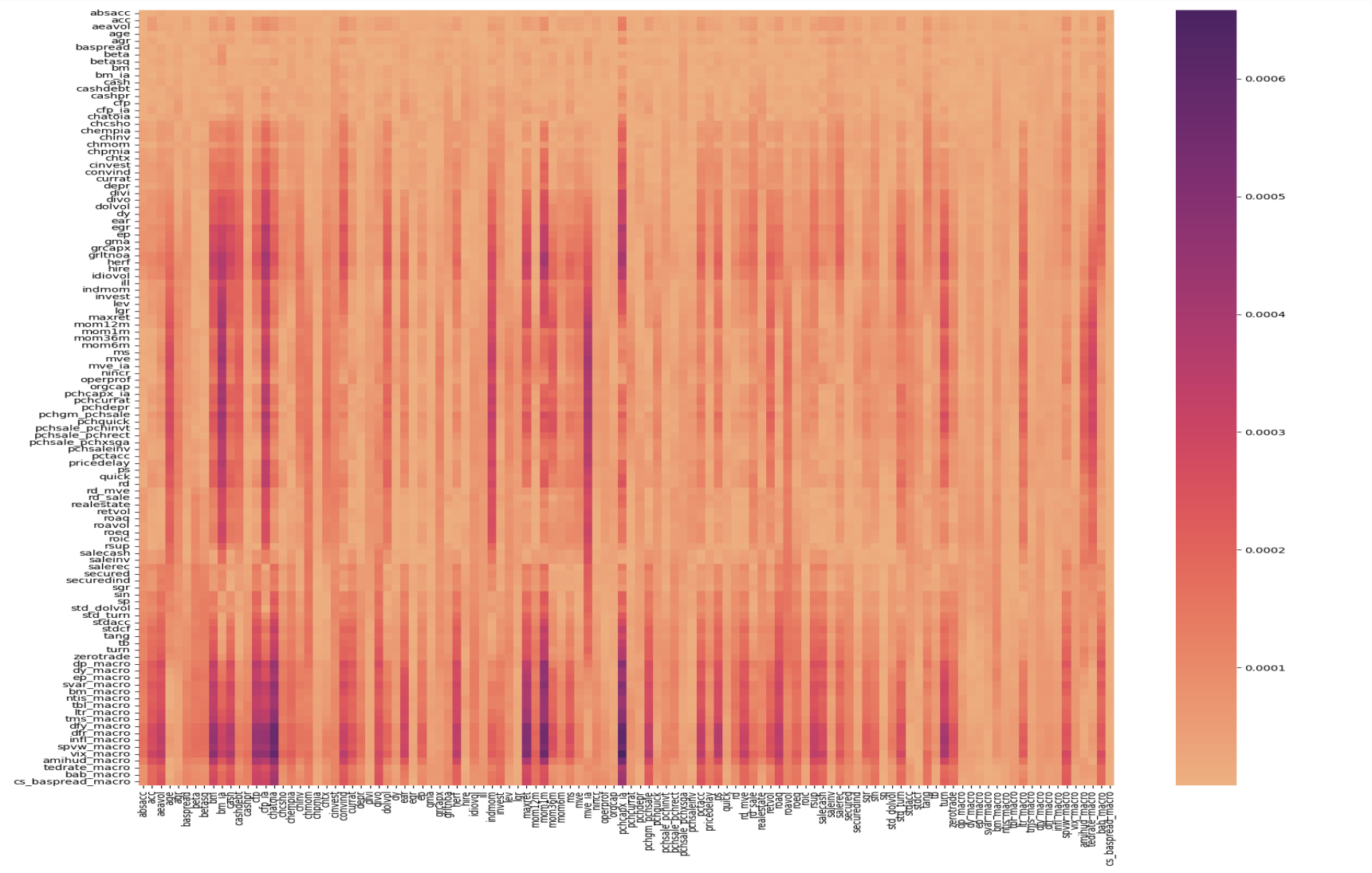


Attention for 09/2008 - Block 2: past returns

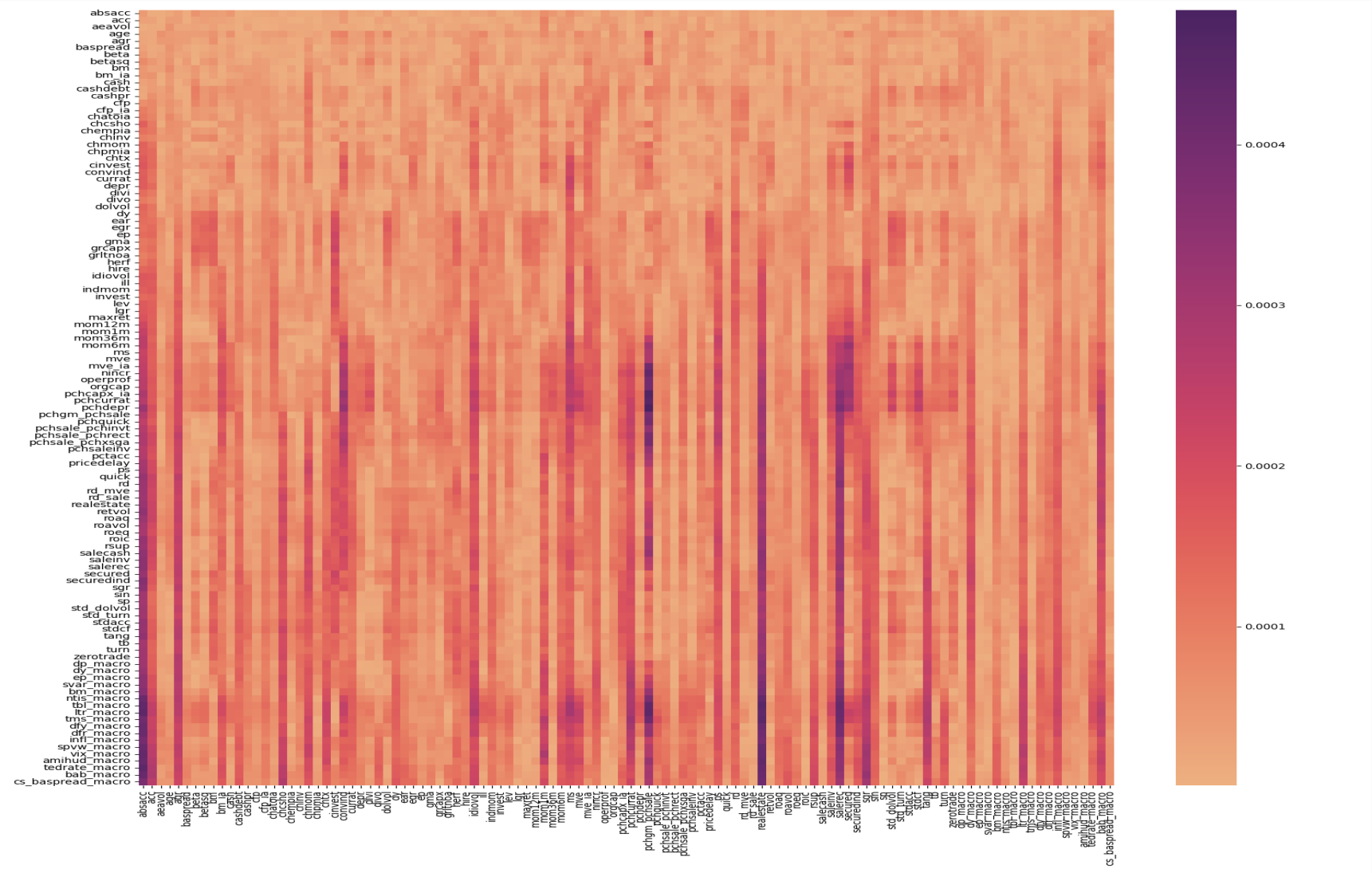


Attention for 09/2008 - Block 3: cash flow variables and changes

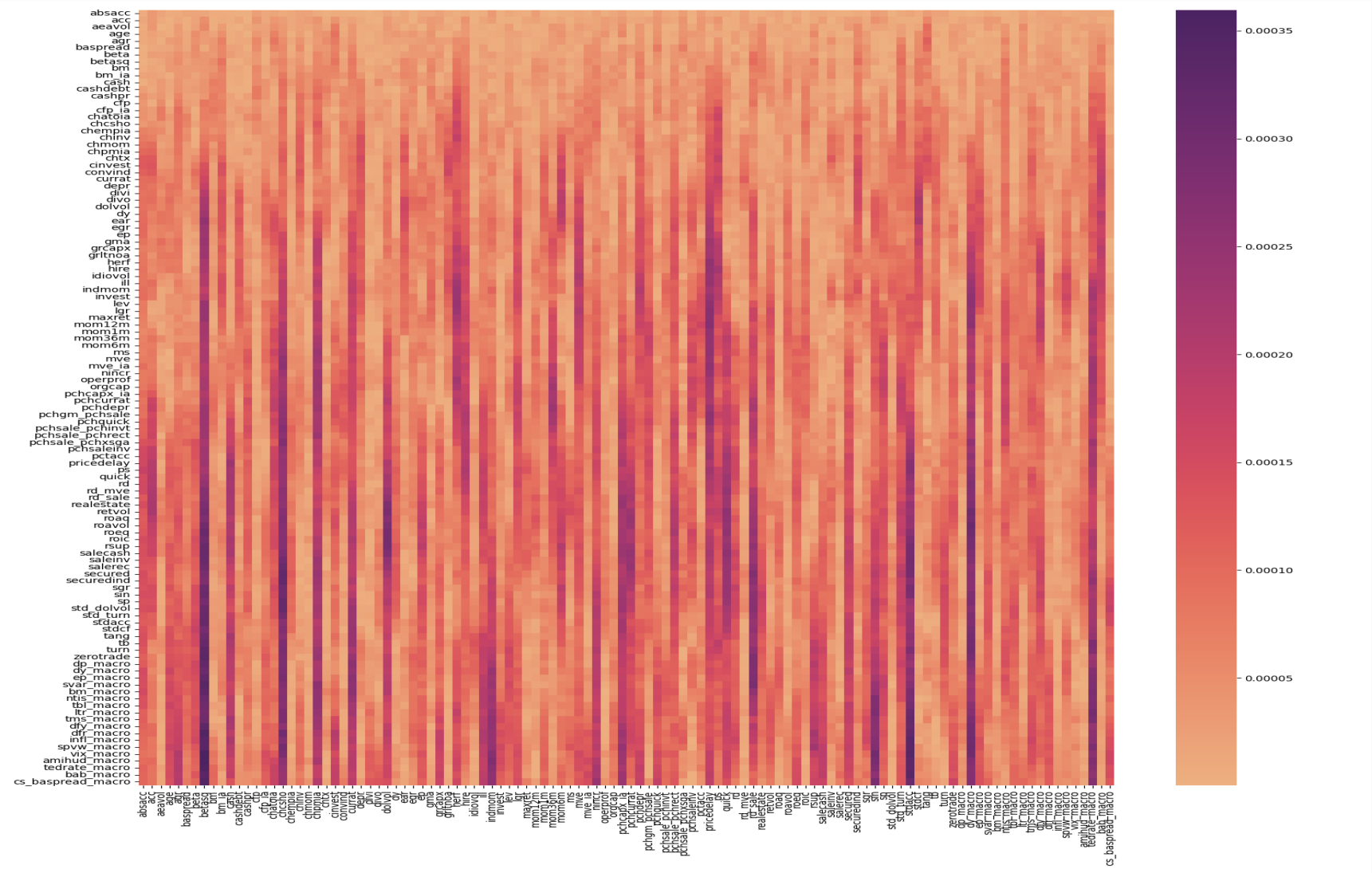
capex



Attention for 09/2008 - Block 4: abs accruals, real estate holdings, sales to receivables



Attention for 09/2008 - Block 5: betasq (measure of uncertainty), accrual volatility, macro: TED spread



Conclusions

- ML approaches to SDF (or optimal portfolio constructions) with big data have a great potential
- Deep Learning without Attention falls under umbrella of Avramov et al (2021) limits-to-arbitrage criticism
- Attention guided deep learning provides an empirical support to DeMiguel et al. (2020), who argue in favor of considering multiple firm characteristics *jointly* for portfolio construction to implicitly account for portfolio's rebalancing frequency.
- *Attention* is critical to reduce the noise and sparsity in financial data