Zeroing in on the Expected Returns of Anomalies

Andrew Y. Chen, Mihail Velikov

Journal of Finalcial and Quantitative Analysis, 2024

解读人：赵伟皓

武汉大学金融系

2025.12.14

Research Question

Whether stock anomalies generate achievable expected return under transaction costs, post publication decay, and modern market structure?

- Is the return recorded in historical literature truly achievable?
- **Not really**, this study constructed over 200 of the most commonly studied anomaly factors, **and their expected future returns are close to zero.**

Motivation I

1. **Number of anomalies is rapidly increasing, but feasibility is questioned**
    - Anomalies' returns recorded in literature far exceed the real achievable returns.
    - Lack of a unified standard for testing anomalies' achievable returns
2. **Importance of examining the net returns of anomalies**
    - Existence of anomalies is direct challenge to EMH;
        - If efficient market, anomalies will quickly fade after publication;
        - Does the anomaly reflect a risk premium or simply sample noise?
    - Related to cross-sectional asset pricing theory literature.
        - Testing anomalies for persistent cross-sectional predictive ability.
3. **Investors depend on academic anomalies as a profitable source of strategy**

Motivation II

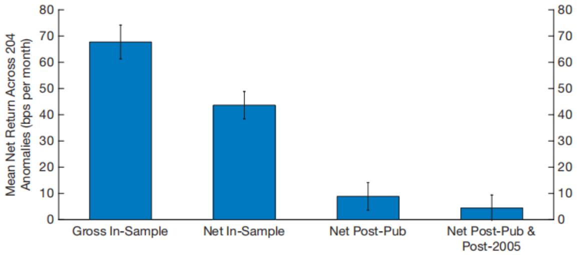
1. **The literature largely ignores trading costs**
    - For those used LF spreads, the transaction costs are biased. (DeMiguel et al., 2020).
    - LF data leads to an overestimation bias in modern market(Post 2005).
2. **Too early historical data is not representative of the future**
    - Stale data(1920s) may bias upward expected returns; (Detzel et al., 2021)
    - Data-mining bias: recent returns are much smaller; (McLean et al., 2016);
    - Revolution in information and trading technologies in 2000s. (Chordia et al., 2014)
3. **The actual expected returns of post pub and post 2005 have not been systematically evaluated**

How does this paper 'zero in'?

FIGURE 1

Anomaly Mean Long–Short Returns

The error bars in Figure 1 show 2 standard errors.



- These results omit additional trading costs such as short-sale costs.
- Short-sale costs may wipe out remaining profits. (Drechsler et al., 2016)

Average anomaly is unprofitable, how about the strongest anomalies?

Contribution

1. **Literature related to Transaction Costs' Calculation**

   Prior: Both early and recently literature exclusively used LF spreads.(Novy-Marx et al., 2016; Freyberger et al., 2020)

   Extend: HF spread measurement leads to two additional results:

   - LF effective spreads are upward biased compared to HF spreads.
   - Averaging LF spreads provide a more accurate estimate of HF spreads.

2. **Literature on Evaluating Anomalies' Expected Returns**

   Prior: Using samples that are inconsistent with current market structure.

   - Lack of systematic analysis of modern market in Post-2005;
   - Lack of systematic analysis of Post-publication decay.

   Extend: Calculation of rets inculding all conditions for 204 anomalies.

Hypothesis

H1: **Historical anomalies' returns do not** represent achievable expected returns.

- **After publication**, the net returns in the **modern market** declined to **nearly zero**.

H2: **Combining multiple factors** helps **obtain a significant net alpha**.

- **No.** Net return after deducting costs is low and not significant.

H3: Cost mitigation measures may increase the achievable returns.

- After cost-mitigation, the net profit of Post-2005 is **still small** (<5bps/month).

Anomalies Data

- **Sample: 1926-2020**
  - Data compiled from the **Chen et al. (2022) open-source** project.
  - Primary dataset covers **204 cross-sectional anomalies** found in the literature.
- **Comprehensive Coverage:**
  - Dataset covers nearly all predictors from literature (Harvey et al., 2016; Hou et al., 2020; McLean et al., 2016; Green et al., 2017).
  - Comprehensive dataset is **crucial for estimating average expected return**.
- **Implication for Prior Work:**
  - The overall performance of full anomalies is shown to be **significantly weaker** than focusing only on 23 strongest, most prominent anomalies (Novy-Marx et al., 2016).

Transcation Cost Measurement

- **HF Effective Spread**
  - Data based on individual transactions and quotes(from TAQ and ISSM).
  - The price of each transaction and the midpoint of the buying and selling price at that time are recorded.
- **Combining HF and LF Data**
  - **HF Data (TAQ/ISSM):** Estimating **post-publication** trading costs (1983-2020).
  - **LF Data: Pre-publication**, cost mitigation and when HF data is unavailable.
- **Combination Forecasts for LF Spreads**
  - Used the **simple average** of four different LF spread proxies (Gibbs, HL, CHL, VoV).
  - Averaging multiple proxies improves accuracy by reducing estimation error.

Portfolio Implementations

For each anomaly
- (1). Implementation in the original paper;
- (2). Cost optimization w.t equal-weighting;
- (3). Cost optimization w.t value-weighting;

Cost-mitigations
- **(1). Low-cost universe;**
- **(2). Reduced rebalancing;**
- **(3). Buy/hold spreads (a.k.a. banding);**

Final cost-optimization implementation:

(i). Implementation with highest in-sample net return across all above;

(ii). Same as (i), but only allow for value-weighted implementations;

H1: Zeroing In on the Average Anomalies

Anomalies' historical returns are generally overestimated.

	a	b	c	$d \approx b \times c$	$e = a - d$
	Gross Return	Turnover (2-Sided)	Ave. Spread Paid	Return Reduction	Net Return
Panel A. Original Paper Implementations					
In-sample	68 (3)	39 (3)	206 (6)	74 (7)	−7 (6)
Post-publication	28 (4)	40 (3)	85 (4)	30 (3)	−1 (5)
Post-pub and post-2005	19 (2)	41 (4)	68 (2)	24 (2)	−5 (3)
Panel B. Cost-Mitigated Implementation Selected to Maximize In-Sample Net Return					
In-sample	61 (3)	16 (2)	137 (6)	17 (1)	44 (3)
Post-publication	16 (3)	17 (2)	51 (4)	7 (1)	9 (3)
Post-Pub and Post−2005	9 (2)	17 (2)	40 (3)	5 (1)	4 (2)
Panel C. Cost-Mitigated, Value-Weighted Implementations Only, Selected In-Sample					
In-sample	47 (3)	18 (2)	78 (3)	12 (1)	35 (3)
Post-publication	5 (3)	20 (2)	21 (2)	4 (1)	1 (3)
Post-Pub and Post-2005	1 (3)	20 (2)	15 (1)	3 (1)	−2 (3)

Zeroing in on the Strongest Anomalies

When selecting the 'Strongest' anomalies, the selected group may have **data mining bias**.

$$\overline{r_i} = \mu_i + \epsilon_i$$

When we filter based on the observation value $\overline{r_i}$(top 25%):

$$\mathbb{E}(\overline{r_i}|\overline{r_i} > \overline{r}_{75}) = \mathbb{E}(\mu_i|\overline{r_i} > \overline{r}_{75}) + \underbrace{\mathbb{E}(\epsilon_i|\overline{r_i} > \overline{r}_{75})}_{>0}$$

- This leads to an **overestimation of future returns**.
- Using out-of-sample testing to eliminate data mining bias:
  - Sorting anomalies based on net return using in sample data (1985-2005).
  - Calculating average net return of each group during post-pub, post-05.

H1: Zeroing In on the Strongest Anomalies

The strongest anomalies’ expected returns are only around 10 bps per month.

	Post-Pub Post-05 Net Return (bps Monthly)			
	Predictor Quartile			
In-Sample Predictor	1 (Worst)	2	3	4 (Best)
Panel A. Including Equal-Weighted Implementations				
Net return	7.5	−0.9	3.1	$\boxed{9.5}$
	(4.2)	(5.5)	(4.5)	(5.3)
Net sharpe	8.2	−0.0	0.5	10.5
	(5.6)	(5.2)	(4.1)	(4.6)
1/Turnover	8.7	6.8	3.7	0.2
	(4.7)	(6.5)	(4.3)	(3.9)
Panel B. Value-Weighted Implementations Only				
Net return	−0.5	0.7	−9.7	2.1
	(4.7)	(6.5)	(4.9)	(5.0)
Net sharpe	−1.3	1.9	−9.1	1.1
	(4.9)	(6.6)	(5.1)	(4.4)
1/Turnover	1.1	−4.9	1.4	−5.2
	(5.0)	(6.2)	(4.1)	(5.8)

H2: Performance of Long–Short Strategies I

	Gross	Turnover	Net
Panel A. Fama–MacBeth			
Equal-weighted decile sorts on fitted expected returns			
1985–2005	374 (35)	64	86 (35)
2006–2020	80 (35)	55	31 (35)
With cost mitigation selected to maximize net return in 1985–2005			
1985–2005	246 (34)	13	188 (34)
2006–2020	31 (35)	12	$\boxed{21\ (35)}$
Panel B. Average Rank			
Equal-weighted decile sorts on fitted expected returns			
1985–2005	276 (47)	28	149 (47)
2006–2020	4 (42)	26	−22 (42)
With cost mitigation selected to maximize net return in 1985–2005			
1985–2005	240 (46)	14	177 (46)
2006–2020	−27 (40)	15	$\boxed{-41\ (40)}$
Panel C. IPCA			
Equal-weighted decile sorts on fitted expected returns			
1985–2005	379 (39)	63	94 (39)
2006–2020	46 (36)	44	3 (36)
With cost mitigation selected to maximize net return in 1985–2005			
1985–2005	242 (37)	12	185 (37)
2006–2020	12 (34)	10	$\boxed{2\ (34)}$

Compared to Novy-Marx et al. (2016)

TABLE 6

Reconciliation with Novy-Marx and Velikov (2016)

In Table 6, we vary the anomaly selection, implementation, sample period, and direct cost measurement to reconcile with Novy-Marx and Velikov (NV) (2016). Returns are in bps per month. NV (2016) anomalies are 23 of the “best known, and strongest performing anomalies.” CZ (Forth) anomalies are 204 anomalies covering the majority of the literature. Column 3 limits the CZ anomalies to continuous ones, in order to apply decile sorts. Low turnover is below 10%, mid turnover is between 10% and 50%, and high turnover is above 50% (1-sided, monthly). Anomaly selection and sample period both contribute to the lower net returns found in our results compared to NV (2016).

	1	2	3	4	5
Anomaly Selection	NV (2016)	NV (2016)	CZ (Forth)	CZ (Forth)	CZ (Forth)
Implementation	VW NYSE Deciles			Original Paper	
Sample	Full	Post-Pub05	Full	Post-Pub05	Post-Pub05
Direct Costs	Gibbs	Gibbs	Gibbs	Gibbs	Combined
Panel A. Mean Gross Return (bps monthly)					
Low turnover	36	−20	20	11	11
Medium turnover	78	27	54	28	28
High turnover	96	45	58	30	30
All	68	15	34	19	19
Panel B. Mean Net Return (bps monthly)					
Low turnover	28	−25	12	1	4
Medium turnover	34	−1	12	−16	−3
High turnover	−35	−39	−60	−92	−46
All	14	−19	3	−17	−5

Extension

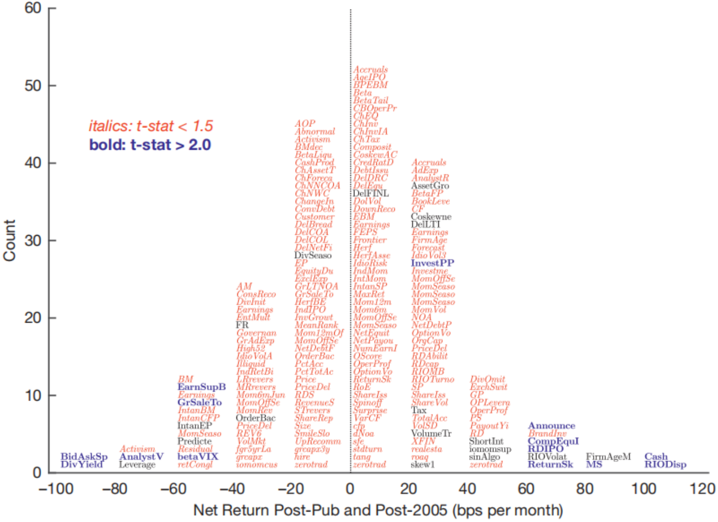
存在少数异象在 2005 年后净收益 80+ bps/月，t-stats 统计量大于 2.0。

- Cash, Mohanram G-score(MS), FirmAgeM, RIODisp;
- 这些异象 Net ret 若真实存在，盈利能力是否源于某些共同的特征或机制?
- 这种盈利能力的可持续性如何?

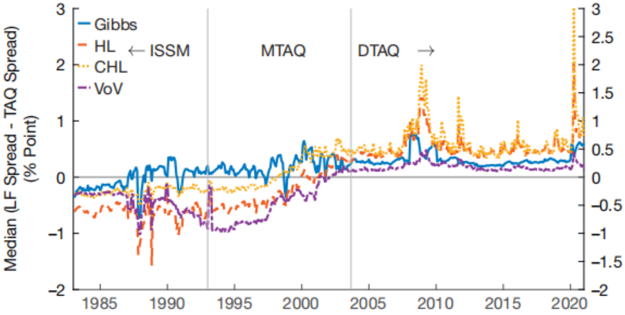
拓展思路

- 聚焦于构建异象的特征和构建方式，例如 Bowel et al. (2024) 发现异象收益主要发生在异象信号对应信息首次公开后的短期内。
- 这些“幸存”异象的盈利能力是否是状态依赖的? 例如，它们是否主要在市场恐慌期或特定宏观经济阶段（如衰退期）才产生收益?

Figure 7


italics: t-stat < 1.5
bold: t-stat > 2.0
Count
0
10
20
30
40
50
60
−100
−80
−60
−40
−20
0
20
40
60
80
100
120
Net Return Post-Pub and Post-2005 (bps per month)
BidAskSp
DivYield
Activism
AnalystV
Leverage
BM
EarnSupB
Earnings
GrSaleTo
IntanBM
IntanCFP
IntanEP
MomSeaso
Predicte
Residual
betaVIX
retCongl
AM
ConsReco
DivInit
Earnings
EntMult
FR
Governan
GrAdExp
High52
IdioVolA
Illiquid
IndRetBi
LRevers
MRevers
Mom6mJun
MomOffSe
MomRev
OrderBac
PriceDel
REV6
VolMkt
fgr5yrLa
grcapx
iomomcus
AOP
Abnormal
Activism
BMdec
BetaLiqu
CashProd
ChAssetT
ChForeca
ChNNCOA
ChNWC
ChangeIn
ConvDebt
Customer
DelBread
DelCOA
DelCOL
DelNetFi
DivSeaso
EP
EquityDu
ExclExp
GrLTNOA
GrSaleTo
HerfBE
IndIPO
InvGrowt
MeanRank
Mom12mOf
MomOffSe
NetDebtF
OrderBac
PctAcc
PctTotAc
Price
PriceDel
RDS
RevenueS
STrevers
ShareRep
Size
SmileSlo
UpRecomm
grcapx3y
hire
zerotrad
Accruals
AgeIPO
BPEBM
Beta
BetaTail
CBOperPr
ChEQ
ChInv
ChInvIA
ChTax
Composit
CoskewAC
CredRatD
DebtIssu
DelDRC
DelEqu
DelFINL
DolVol
DownReco
EBM
Earnings
FEPS
Frontier
Herf
HerfAsse
IdioRisk
IndMom
IntMom
IntanSP
MaxRet
Mom12m
Mom6m
MomOffSe
MomSeaso
NetEquit
NetPayou
NumEarnI
OScore
OperProf
OptionVo
ReturnSk
RoE
ShareIss
Spinoff
Surprise
VarCF
cfp
dNoa
sfe
stdturn
tang
zerotrad
Accruals
AdExp
AnalystR
AssetGro
BetaFP
BookLeve
CF
Coskewne
DelLTI
Earnings
FirmAge
Forecast
IdioVol3
InvestPP
Investme
MomOffSe
MomSeaso
MomSeaso
MomSeaso
MomVol
NOA
NetDebtP
OptionVo
OrgCap
PriceDel
RDAbilit
RDcap
RIOMB
RIOTurno
SP
ShareIss
ShareVol
Tax
TotalAcc
VolSD
VolumeTr
XFIN
realesta
roaq
skew1
DivOmit
ExchSwit
GP
OPLevera
OperProf
PS
PayoutYi
RD
ShortInt
iomomsup
sinAlgo
zerotrad
Announce
BrandInv
CompEquI
RDIPO
RIOVolat
ReturnSk
FirmAgeM
MS
Cash
RIODisp

低频数据导致价差偏差



- 现代交易时代（03 年后），LF 价差代理明显存在向上偏差（高估），约为 25-50 个基点。
- **现代异象检验需要使用高频数据进行准确的成本估算。**

高频有效买卖价差

这衡量的是单笔交易实际偏离“公允价格”（报价中点）的程度。

$$[\text{Effective Spread}] = 2 \times |log(P_k) - log(M_k)|$$

其中 P_k 是第 k 笔交易的价格，M_k 是匹配的合并最佳出价和报价（BBO）报价的中点。

- 其优势是精确测量“有效价差”；
- 本文关注的交易成本核心是有效价差，而非报价价差；
- 有效价差反映了交易者实际执行时面临的摩擦，通常由于“价格改善”而小于报价价差。
- 高频数据能直接、精确地计算出这个值，适用于本文的研究。

Appendix - 本文贡献 1 第二点

平均四个 LF 价差可以比使用单独 LF 价差更准确地估计 HF 价差。

Panel A. LF Spread Correlations (1926–2020)

	Gibbs	HL	CHL	VoV
Gibbs	1.00			
HL	0.63	1.00		
CHL	0.74	0.86	1.00	
VoV	0.74	0.53	0.73	1.00

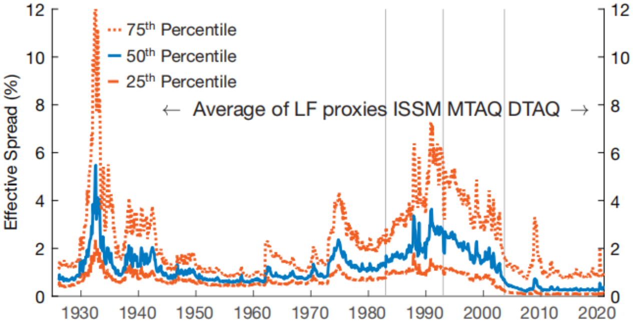
Panel B. Correlations with TAQ (1993–2020)

	TAQ	Gibbs	HL	CHL	VoV	LF_AVE
TAQ	1.00					
Gibbs	0.84	1.00				
HL	0.64	0.60	1.00			
CHL	0.79	0.72	0.85	1.00		
VoV	0.84	0.72	0.53	0.74	1.00	
LF_AVE	$\boxed{0.90}$	0.89	0.82	0.93	0.86	1.00

Panel C. Correlations with ISSM (1983–1992)

	ISSM	Gibbs	HL	CHL	VoV	LF_AVE
ISSM	1.00					
Gibbs	0.88	1.00				
HL	0.77	0.74	1.00			
CHL	0.83	0.78	0.88	1.00		
VoV	0.86	0.81	0.62	0.74	1.00	
LF_AVE	$\boxed{0.92}$	0.94	0.88	0.93	0.87	1.00

合并有效价差的演变过程



1. 1970 年代初，随着纳斯达克股票纳入 CRSP 指数体系，交易成本急剧攀升；
2. 1980 年代末，交易成本继续上涨，与其他研究一致（Abdi et al., 2017）；
3. 随着电子化交易提高了流动性，2000 年代的交易成本大幅下降。

交易缓解策略

1. **低成本股票池：** 不交易所有股票，而是只交易在每个规模十分组内交易成本最低的那一部分股票（例如，只交易成本最低的三分之一或一半的股票）。这直接降低了所交易股票的平均价差。
2. **降低调仓频率**：将原始策略的月度调仓改为每 3 个月、6 个月或 12 个月调仓一次。这直接大幅降低了换手率
3. **买入/持有区间**：也称“带状”交易。一个更精巧的降低换手率的方法。
   - 例：一个“20/40”规则意味着：当一只股票的异象信号强度进入前 20% 时买入，但只有当其信号跌出前 40% 时才卖出（做空端同理）。
   - 这创造了一个“缓冲带”，避免了信号在小范围内波动时就触发交易，从而大幅降低换手率。