Predicting Anomalies

Boone Bowles, Adam Reed, Matt Ringgenberg, Jake Thornock

TEXAS A&M UNIVERSITY Mays Business School

UNC KENAN-FLAGLER BUSINESS SCHOOL

David Eccles School of Business THE UNIVERSITY OF UTAH

BYU MARRIOTT SCHOOL OF BUSINESS

Jackson Hole Finance Group Conference

January 2025

The Anomaly Zoo

Asset pricing anomalies:

- Variables that predict future abnormal returns.

There are now over 400 documented anomalies.

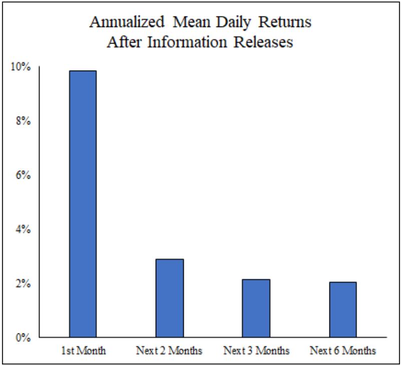
- McLean and Pontiff (2016) – 97
- Kakushadze and Serur (2018) – 151 (18 asset classes)
- Hou et al. (2020) – 452
- Chen and Zimmermann (2021) – 207
    - www.openassetpricing.com

The interpretation of these anomalies remains the source of much disagreement.

When do anomalies earn returns?

In “Anomaly Time” (2024), we study ***when*** anomaly returns occur, in event time.

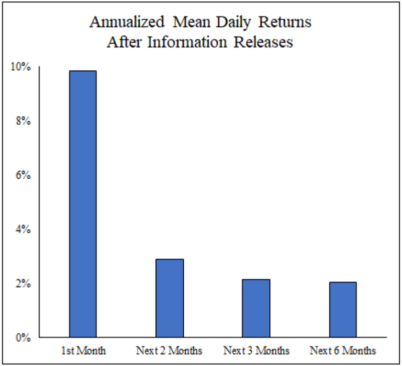
- Anomaly returns are earned immediately after information releases.



When do anomalies earn returns?

In "Anomaly Time" (2024), we study ***when*** anomaly returns occur, in event time.

- Anomaly returns are earned immediately after information releases.

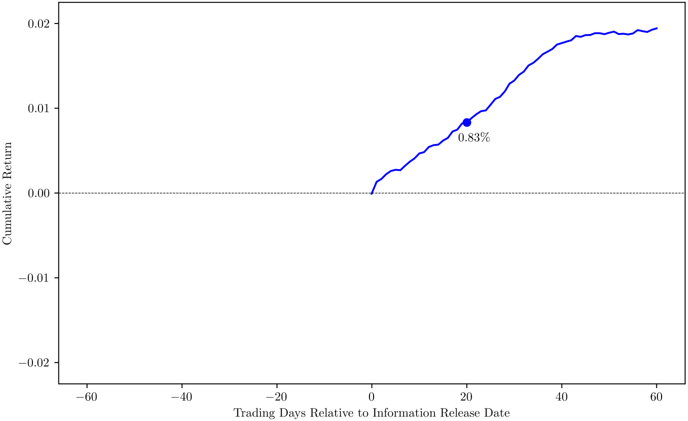


- But what happens ***before***?
- Can we trade early?
- Can we predict the predictors?
- What are the implications?

This paper

Returns exhibit patterns *prior* to information releases.

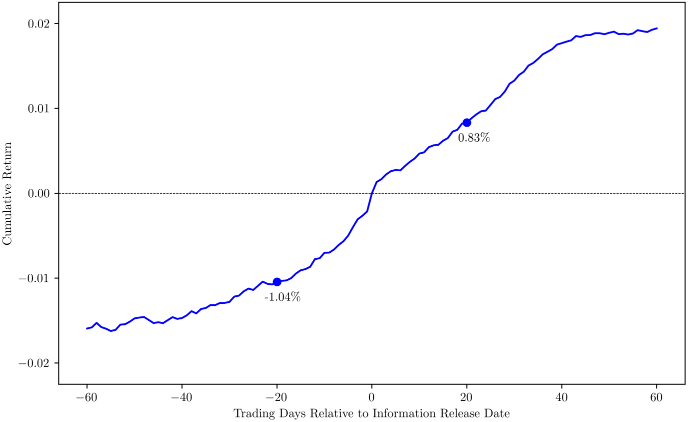
- So what? We can't trade prior to information releases, because we don't know which stocks should be in the portfolio...



This paper

Returns exhibit patterns *prior* to information releases.

- So what? We can't trade prior to information releases, because we don't know which stocks should be in the portfolio...



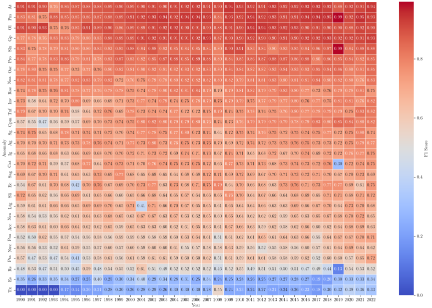
This paper

Returns exhibit patterns *prior* to information releases.

- So what? We can't trade prior to information releases, because we don't know which stocks should be in the portfolio...

Most anomaly signals are predictable!

- A very simple model reliably predicts which stocks are in the long and short portfolios one quarter before annual information releases.

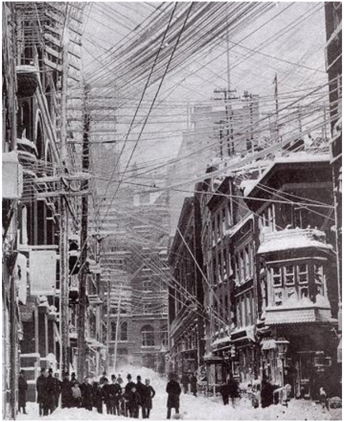


Several economic interpretations of anomalies

1. Anomalies never existed or were once real but disappeared.
   - McLean and Pontiff (2016).
     - Anomalies are "real"... but arbitrageurs eliminated them.
   - Harvey et al. (2016) & Hou et al. (2020).
     - Anomalies are not "real" ... but are spurious correlations found by data mining.
     - "...most claimed research findings in financial economics are likely false."
2. Anomalies exist and can be explained.
   - Cooper et al. (2008), Sloan (1996), Novy-Marx (2013), etc.
     - Authors document an anomaly, then argue for explanations based on over- or under-reaction, overconfidence, limited attention, or limits to arbitrage.
3. Bowles, Reed, Ringgenberg, and Thornock (2024).
   - Accounting-based anomalies are real but returns are concentrated in weeks right after information release.
   - Ivkovic and Zekhnini (2024) shows that this patterns starts after academic research first documents an anomaly.

The timing of information releases is important

Wall Street has always understood that prices respond to new information, so new information is valuable.



But because of data limitations, academics often ignore the moment information released.

The timing of information releases is important

Information releases create a race to trade on new information *before* everyone else.

NYT article 1982: *"vying for the same information is the highly competitive corps of professional researchers. They have their own telephones and are usually the first to reach for documents as they arrive from the S.E.C.'s filing room... files are often misplaced or even stolen."*


S.E.C. Data: Difficult Hunt

- Gao and Huang (2020) and Kothari, Zhang, and Zuo (2023).

The timing of information releases is important

Information releases create a race to trade on new information *before* everyone else.

NYT article 1982: *"vying for the same information is the highly competitive corps of professional researchers. They have their own telephones and are usually the first to reach for documents as they arrive from the S.E.C.'s filing room... files are often misplaced or even stolen."*


S.E.C. Data: Difficult Hunt

But this all begs the question: what happens *prior* to the information releases?

Anomalies and information release dates

Problem 1: Which anomalies? What information release dates?

Anomalies and information release dates

Problem 1: Which anomalies? What information release dates?

- We study anomalies that have clear, discrete, observable information release dates.
  - The *anomaly signal* is based on financial data released in 10-K, 10-Q, or earnings announcements.
  - E.g., Asset Growth, Asset Turnover, Profitability, etc.
  - Start with 97 anomalies from McLean and Pontiff (2016). Limit to "fundamental" or "event" anomalies (60 left).

Anomalies and information release dates

Problem 1: Which anomalies? What information release dates?

- We study anomalies that have clear, discrete, observable information release dates.
  - The *anomaly signal* is based on financial data released in 10-K, 10-Q, or earnings announcements.
  - E.g., Asset Growth, Asset Turnover, Profitability, etc.
  - Start with 97 anomalies from McLean and Pontiff (2016). Limit to "fundamental" or "event" anomalies (60 left).

Problem 2: 10-K dates or earnings announcement dates?

Anomalies and information release dates

Problem 1: Which anomalies? What information release dates?

- We study anomalies that have clear, discrete, observable information release dates.
  - The *anomaly signal* is based on financial data released in 10-K, 10-Q, or earnings announcements.
  - E.g., Asset Growth, Asset Turnover, Profitability, etc.
  - Start with 97 anomalies from McLean and Pontiff (2016). Limit to "fundamental" or "event" anomalies (60 left).

Problem 2: 10-K dates or earnings announcement dates?

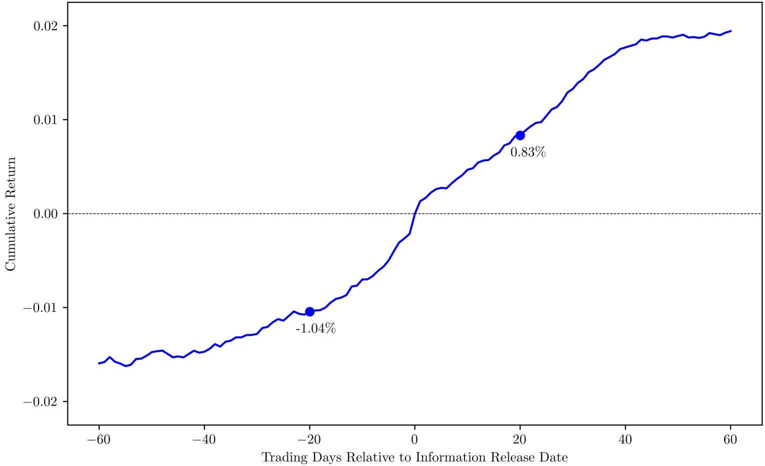
- We rely on the Compustat Snapshot database to identify when anomaly signals are first publicly released.
  - *The Snapshot database creates a historical investment environment by showing the information that was available at that time in history.*
  - Limiting ourselves to anomalies that can be replicated using Snapshot... **28** in our final sample **(1990-2023)**.

What we analyze

We focus on the months prior to annual information release dates and:

1. Analyze returns to a *Perfect Foresight* model trading one quarter early;
2. Examine the predictability of anomaly portfolios;
    - Can we approximate the Perfect Foresight model?
3. Analyze returns to the various prediction models.
    - Can our approximation make money?

Returns drift up PRIOR to information releases!



Not a random walk!

Returns drift up PRIOR to information releases!

	Compound Return (bps)	Basis Points Per Day	Annualized Return (%)
Return from trading ***after*** the information release date and holding for 3 months.	178	3.1	7.7
Additional return for for trading ***before*** the information release date:	Compound Return (bps)	Basis Points Per Day	Annualized Return (%)
1 trading day before	39	19.6	49.4
3 trading days before	53	13.2	33.2
1 week before	69	11.5	29.0
2 weeks before	89	8.1	20.4
1 month before	120	5.7	14.4
2 months before	161	3.9	9.9
3 months before	171	2.8	7.1

And effects get stronger as we approach the information release date.

Returns drift up PRIOR to information releases!

	Compound Return (bps)	Basis Points Per Day	Annualized Return (%)
Return from trading ***after*** the information release date and holding for 3 months.	178	3.1	7.7
Additional return for for trading ***before*** the information release date:	Compound Return (bps)	Basis Points Per Day	Annualized Return (%)
1 trading day before	39	19.6	49.4
3 trading days before	53	13.2	33.2
1 week before	69	11.5	29.0
2 weeks before	89	8.1	20.4
1 month before	120	5.7	14.4
2 months before	161	3.9	9.9
3 months before	171	2.8	7.1

But who cares? We don't have perfect foresight. Or a time machine.

Anomaly signals (and rankings) are highly correlated...

	Correlations with $Q4_t$					
	Signals			Rankings (1-10)		
Anomaly	$Q3_t$	$Q2_t$	$Q4_{t-1}$	$Q3_t$	$Q2_t$	$Q4_{t-1}$
Average	.41	.31	.05	.81	.71	.37
Accruals	.59	.60	-.48	.75	.62	.22
Asset Growth	.94	.99	.01	.84	.69	.26
Asset Turnover	.11	.02	.07	.97	.96	.92
Chg in Asset Turnover	.18	.01	-.62	.77	.67	.12
Chg in Profit Margin	.53	.48	-.64	.82	.70	-.03
Earnings Consistency	.38	.23	.81	.83	.73	.63
Earnings Surprise	.16	.10	-.07	.46	.30	-.37
Gross Profitability	.45	.19	.04	.97	.95	.88
Inventory Growth	.80	.14	.07	.70	.56	.10
Investments	.15	.18	.01	.88	.77	.31
Net Working Capital	.70	.71	-.50	.65	.51	-.16
Operating Leverage	.34	.22	.12	.97	.96	.91
Profit Margin	.67	.60	.29	.97	.95	.88
Profitability	.56	.38	.17	.93	.88	.72
Return on Equity	.00	.00	.00	.87	.79	.59
Revenue Surprise	.23	.03	-.00	.67	.43	-.23
Sales Growth	.90	.87	.85	.84	.79	.76
Tax	.00	.01	-.00	.56	.53	.57
Total External Financing	.65	.59	.58	.86	.73	.41

Anomaly signals (and rankings) are highly correlated...

	Correlations with $Q4_t$					
	Signals			Rankings (1-10)		
Anomaly	$Q3_t$	$Q2_t$	$Q4_{t-1}$	$Q3_t$	$Q2_t$	$Q4_{t-1}$
Average	.41	.31	.05	.81	.71	.37
Accruals	.59	.60	-.48	.75	.62	.22
Asset Growth	.94	.99	.01	.84	.69	.26
Asset Turnover	.11	.02	.07	.97	.96	.92
Chg in Asset Turnover	.18	.01	-.62	.77	.67	.12
Chg in Profit Margin	.53	.48	-.64	.82	.70	-.03
Earnings Consistency	.38	.23	.81	.83	.73	.63
Earnings Surprise	.16	.10	-.07	.46	.30	-.37
Gross Profitability	.45	.19	.04	.97	.95	.88
Inventory Growth	.80	.14	.07	.70	.56	.10
Investments	.15	.18	.01	.88	.77	.31
Net Working Capital	.70	.71	-.50	.65	.51	-.16
Operating Leverage	.34	.22	.12	.97	.96	.91
Profit Margin	.67	.60	.29	.97	.95	.88
Profitability	.56	.38	.17	.93	.88	.72
Return on Equity	.00	.00	.00	.87	.79	.59
Revenue Surprise	.23	.03	-.00	.67	.43	-.23
Sales Growth	.90	.87	.85	.84	.79	.76
Tax	.00	.01	-.00	.56	.53	.57
Total External Financing	.65	.59	.58	.86	.73	.41

...and anomaly portfolios are PERSISTENT

	Pct. of Q4 Portfolio from Prior Quarter		
	$Q3_t$	$Q2_t$	$Q4_{t-1}$
Anomaly	Remained	Remained	Remained
Average	66	56	32
Accruals	63	50	22
Asset Growth	72	59	23
Asset Turnover	85	82	60
Chg in Asset Turnover	67	59	26
Chg in Profit Margin	68	57	19
Earnings Consistency	57	43	42
Earnings Surprise	26	13	2
Gross Profitability	85	78	66
Inventory Growth	68	56	22
Investments	66	54	31
Net Working Capital	57	45	11
Operating Leverage	84	78	63
Profit Margin	87	81	64
Profitability	78	68	52
Return on Equity	75	65	44
Revenue Surprise	49	24	4
Sales Growth	57	50	45
Tax	35	30	25
Total External Financing	68	55	31

Maybe we can predict anomaly signals/portfolios?

To examine if annual anomaly signals ($x_{4,t}$) are predictable, we explore:

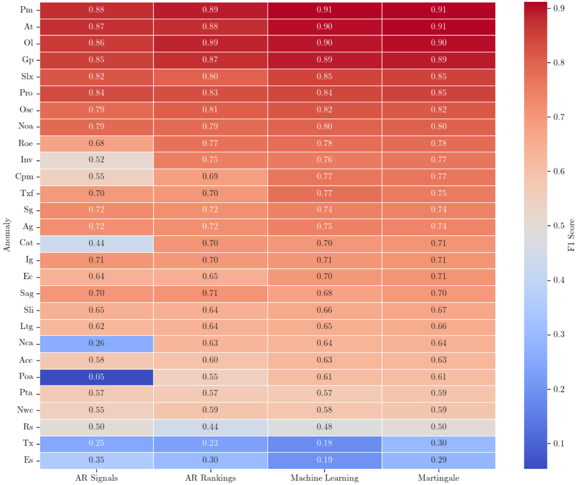
1. AR(1) model of the form: $x_{4,t} = \delta + \phi_1 x_{3,t} + \phi_2 x_{4,t-1} + w_{4,t}$
2. Machine learning model: LightGBM (gradient boosting machine)
3. Martingale model of the form: $E[x_{4,t}] = x_{3,t}$

We also examine whether it is better to predict anomaly signals, relative rankings (1-10), or portfolio assignment (long, short, or out).

- Answer: it is better to predict relative rankings and portfolio assignment.

We review *F1 Scores*: harmonic mean of precision (Type 1 errors) and recall (Type 2 errors).

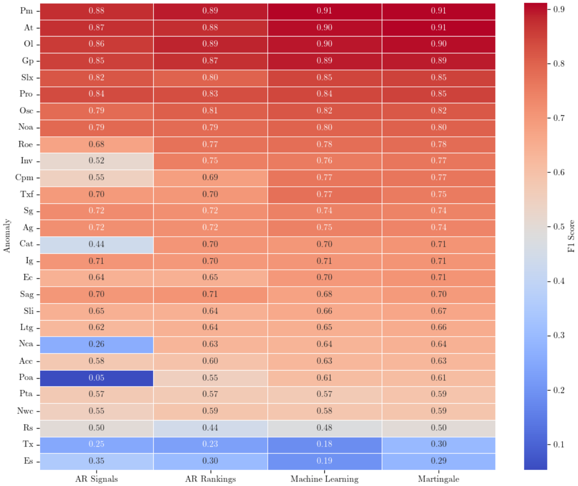
Finding 1: The models have very similar performance...



Anomaly	AR Signals	AR Rankings	Machine Learning	Martingale
Pm	0.88	0.89	0.91	0.91
At	0.87	0.88	0.90	0.91
Ol	0.86	0.89	0.90	0.90
Gp	0.85	0.87	0.89	0.89
Slx	0.82	0.80	0.85	0.85
Pro	0.84	0.83	0.84	0.85
Osc	0.79	0.81	0.82	0.82
Noa	0.79	0.79	0.80	0.80
Roe	0.68	0.77	0.78	0.78
Inv	0.52	0.75	0.76	0.77
Cpm	0.55	0.69	0.77	0.77
Txf	0.70	0.70	0.77	0.75
Sg	0.72	0.72	0.74	0.74
Ag	0.72	0.72	0.75	0.74
Cat	0.44	0.70	0.70	0.71
Ig	0.71	0.70	0.71	0.71
Ec	0.64	0.65	0.70	0.71
Sag	0.70	0.71	0.68	0.70
Sli	0.65	0.64	0.66	0.67
Ltg	0.62	0.64	0.65	0.66
Nca	0.26	0.63	0.64	0.64
Acc	0.58	0.60	0.63	0.63
Poa	0.05	0.55	0.61	0.61
Pta	0.57	0.57	0.57	0.59
Nwc	0.55	0.59	0.58	0.59
Rs	0.50	0.44	0.48	0.50
Tx	0.25	0.23	0.18	0.30
Es	0.35	0.30	0.19	0.29

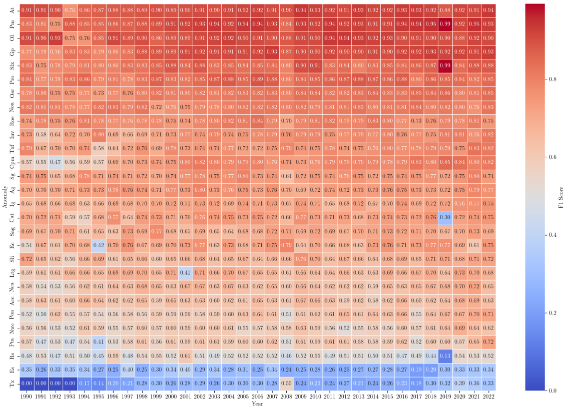
... but the AR(1) models are the weakest...

Finding 1: The models have very similar performance...



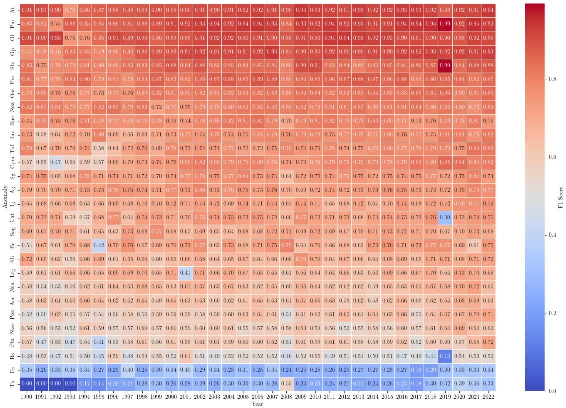
... while the Machine Learning model and the Martingale model are very similar.

Finding 2: Very little variation through time



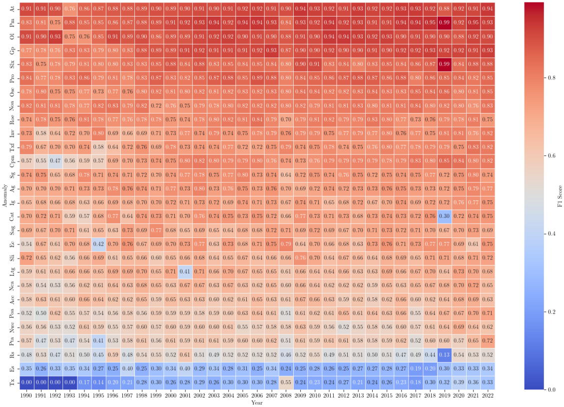
Not a lot of changes going from left (1990) to right (2023).

Finding 3: Good and Excellent F1 Scores



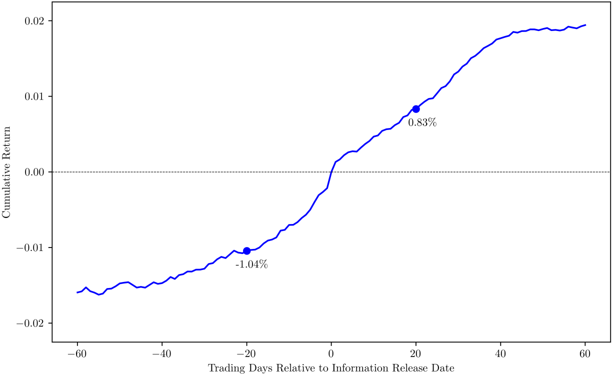
For most anomalies the models are pretty good.

Finding 3: Good and Excellent F1 Scores



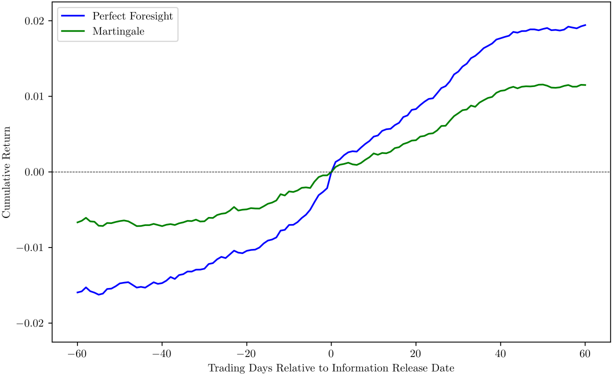
Now we'll use the Martingale model to predict annual anomaly portfolios and trade one quarter early.

Returns to predicting and trading early



Remember, high returns with perfect foresight...

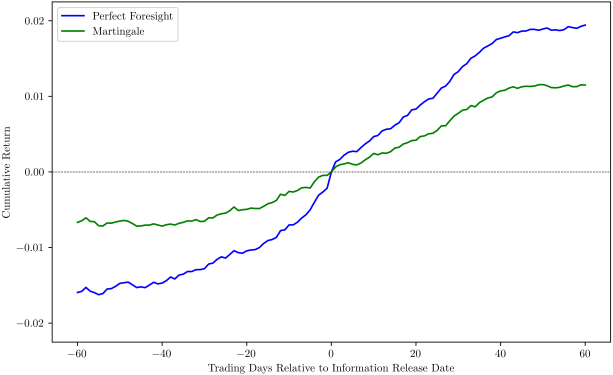
Returns to predicting and trading early



Remember, high returns with perfect foresight...and returns are still there using a prediction model.

Implication: anomalies may be more anomalous than previously realized.

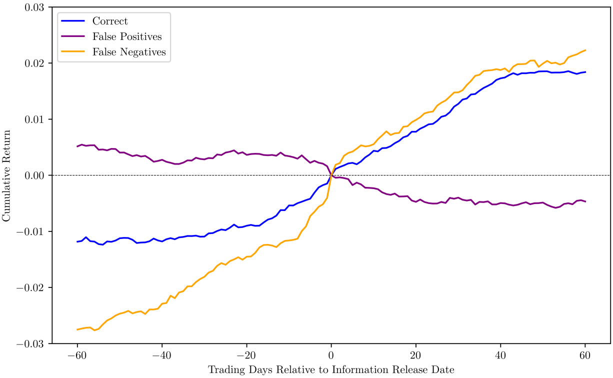
Returns to predicting and trading early



Remember, high returns with perfect foresight...and returns are still there using a prediction model.

But we've lost half of the return!?

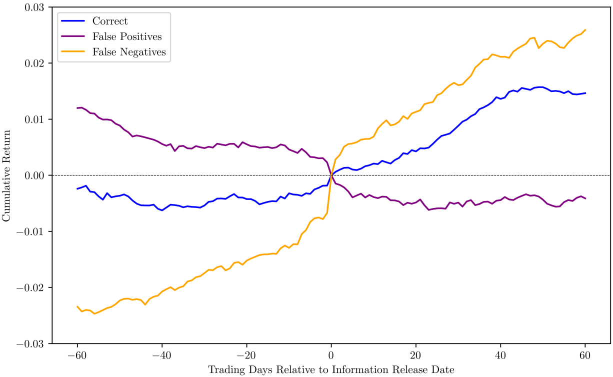
Mistakes are costly... especially Type 2 errors!



Type 1 errors (false positives) are sort of costly, but what really hurts is missing out on the Type 2 errors (false negatives).

Implication: Anomaly returns are closely tied to *surprising* information.

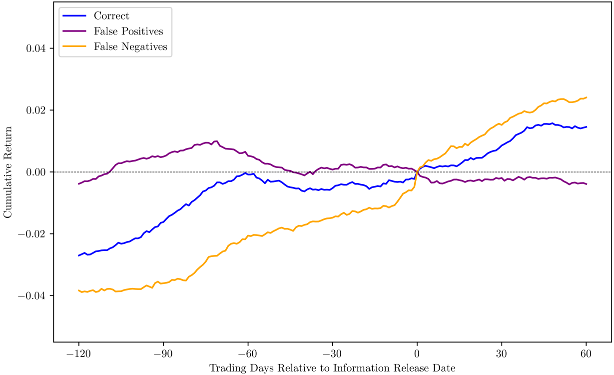
Mistakes are costly... especially Type 2 errors!



Type 1 errors (false positives) are sort of costly, but what really hurts is missing out on the Type 2 errors (false negatives). And mistakes have become more costly in recent years.

Implication: Arbitrageurs are correcting prices.

Returns to predicting and trading EARLIER



Mistakes aren't as costly if we back up and trade 2 quarters early (Greenwood and Sammon (2022)).

Implication: The starting gun for trading on new anomaly information is going off earlier and earlier.

Conclusion

Three main facts:

1. Anomalies earn predictable returns before information releases.
2. Many anomaly signals (and portfolios) can be predicted with a simple model.
3. Making predictions and trading early is no longer profitable since making mistakes is increasingly costly.

Implications:

1. Anomalies are real and may be more anomalous than previously realized.
2. Tight connection between anomaly returns and the release of ***new/surprising*** information.
3. Arbitrageurs are getting better (faster) at making predictions.
4. Important to understand *when* investors trade and *when* returns are earned because it impacts measurements and our economic interpretations.
   - Are commonly used factors using *stale* data?