Big Data in Finance

Gold Futures Return Prediction and Trading Strategy

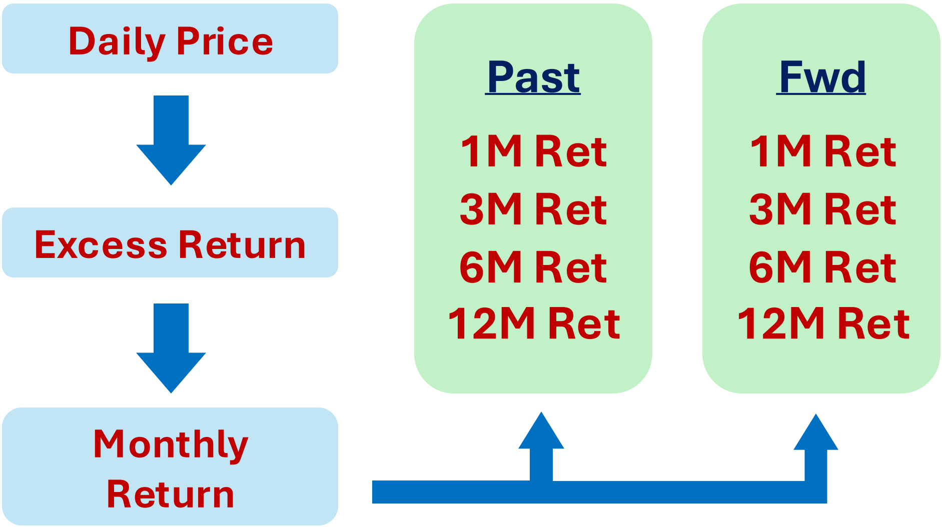
Team 3

Alfred Choi
Sicheng Li
Paul Janicot
Sonyi He

I. Data Preparation
How do we select the suitable data for prediction?

Gold Futures data (Bloomberg)
Choose the most liquid market with high open interest, which is current month +1 & +2.

Moskowitz, Ooi, and Pedersen (2012), Time Series Momentum
"we compute the daily excess return of the most ***liquid futures contract*** *(typically the* ***nearest or next nearest-to-delivery contract****), and* ***then compound the daily returns*** *to a cumulative return index from which we can compute returns at any horizon"*



Macroeconomic data (Bloomberg and Internet)
Gold futures is significantly influenced by many macro-economic factors, and considered as safe heaven asset.

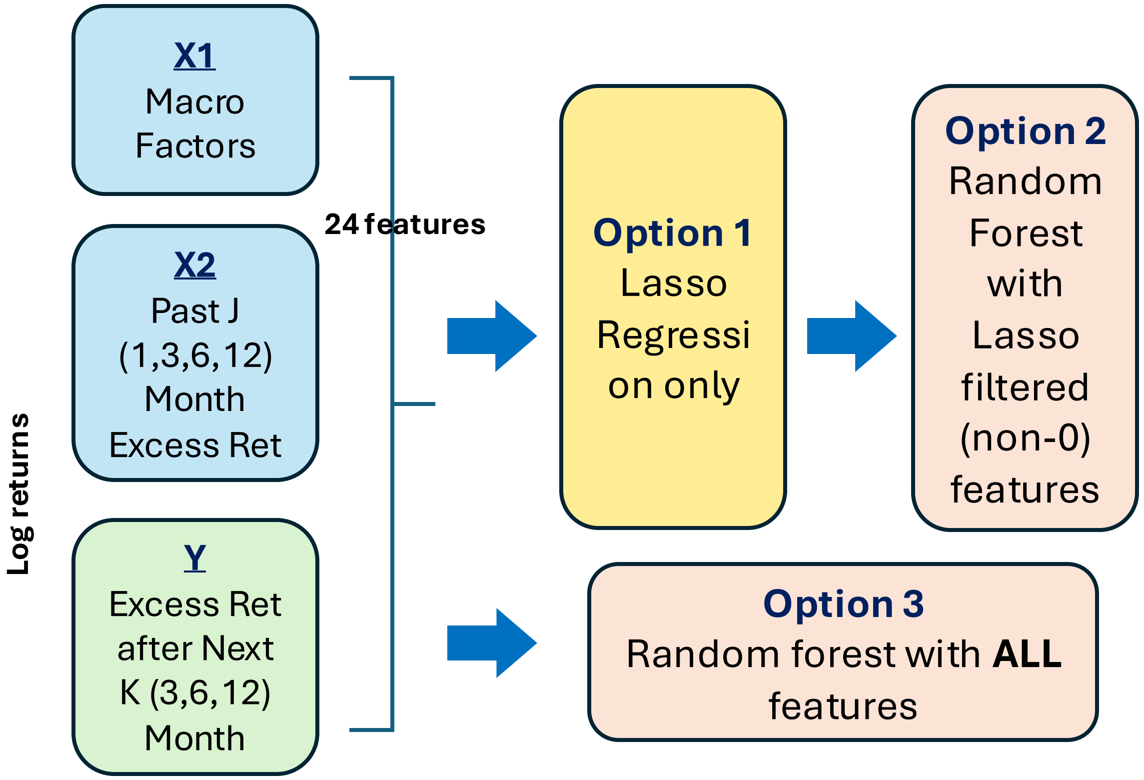
Categories	Data	Economic Justification
Currency	DXY (USD), JPY	Gold is priced in USD, JPY regarded as safe heaven
Yield	2Yr Yield, 10 Yr Yield, Fed Eff Rate	Affects excess return, increase opportunity cost of non-yield asset
Market Valuation	SPX P/E, SPX D/P	If market is overvalued, investors may invest in gold for diversification
Inflation	CPI, PCE, PPI	Gold serves as inflation hedge during high inflation and weak dollar
Econ Growth	GDP, Unemployment, Consumer Confidence	Historically, gold prices exhibited an inverse relationship with stock market and economic performance
Market	SPX, Nasdaq, VIX	
Gold Demand & Supply	ETF Ounce, Chinese premium disc, Indian premium disc	ETF market and local gold market reflect the demand of gold and price premium that affects price

II. Model Intuition

What models do we build?

We forecast **excess return** after N month holding period, with following ML models:

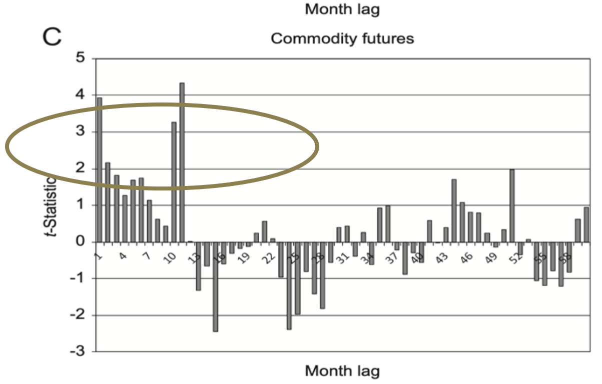
*We prefer excess return because it eliminates bias from risk-free return (compared to total return), and it gives clarity for **signal identification**, **direct comparison** and **absolute performance** (compared to sharpe ratio)*



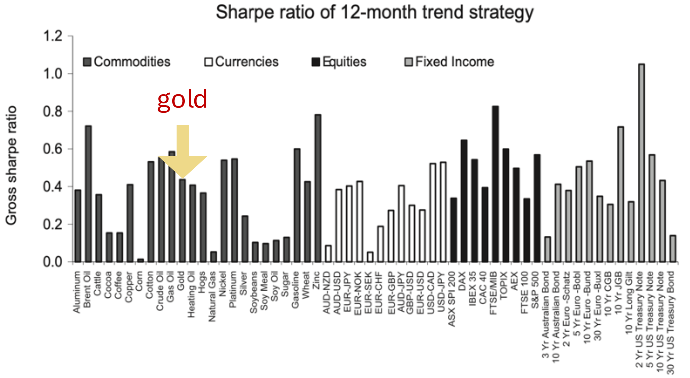
Inspired by academia **time series momentum** strategy:

- ***Moskowitz, Ooi and Pederson (2012)*** - right: Momentum works positively for month **lag 1 to 12** and significantly for gold, with sharpe ratio enhancement (12 month strategy)
- ***Jegadeesh/ Titman (1993)*** - left: For stocks market, long the best perform and short the worst perform bucket **in past J month for next K months** generates positive significant return.

		Panel A			
J	K =	3	6	9	12
3	Sell	0.0108	0.0091	0.0092	0.0087
		(2.16)	(1.87)	(1.92)	(1.87)
3	Buy	0.0140	0.0149	0.0152	.0156
		(3.57)	(3.78)	(3.83)	(3.89)
3	Buy-sell	0.0032	0.0058	0.0061	0.0069
		(1.10)	(2.29)	(2.69)	(3.53)
6	Sell	0.0087	0.0079	0.0072	0.0080
		(1.67)	(1.56)	(1.48)	(1.66)
6	Buy	0.0171	0.0174	0.0174	0.0166
		(4.28)	(4.33)	(4.31)	(4.13)
6	Buy-sell	0.0084	0.0095	0.0102	0.0086
		(2.44)	(3.07)	(3.76)	(3.36)
9	Sell	0.0077	0.0065	0.0071	0.0082
		(1.47)	(1.29)	(1.43)	(1.66)
9	Buy	0.0186	0.0186	0.0176	0.0164
		(4.56)	(4.53)	(4.30)	(4.03)
9	Buy-sell	0.0109	0.0121	0.0105	0.0082
		(3.03)	(3.78)	(3.47)	(2.89)
12	Sell	0.0060	0.0065	0.0075	0.0087
		(1.17)	(1.29)	(1.48)	(1.74)
12	Buy	0.0192	0.0179	0.0168	0.0155
		(4.63)	(4.36)	(4.10)	(3.81)
12	Buy-sell	0.0131	0.0114	0.0093	0.0068
		(3.74)	(3.40)	(2.95)	(2.25)



T.J. Moskowitz et al. / Journal of Financial Economics 104 (2012) 228–250



III. Data Cleansing and Model Construction

How do we cleanse data and train model appropriately?

Besides dataframe formatting, we have done the following data pre-processing…

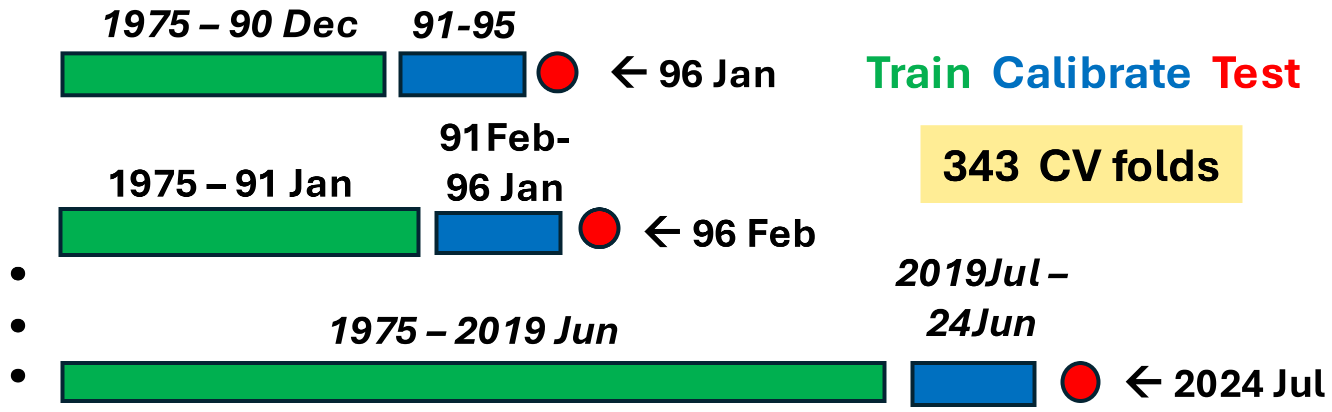
1. Shifting: To prevent data leakage, all features are shifted by one month. For macro features like CPI - released later than month-end, we apply an extra one-month shift to maintain data integrity. This ensures that **no future data, unavailable at the forecast time**, is used.

2. Standardization: Since LASSO regression requires all features on the **same comparable scale**, we **standardize all features** to prevent disproportionate shrinkage.

How do we perform in-sample training and out-of-sample testing?

1. Hyperparameter tuning: For Lasso regression model, we apply grid search to loop through the regularization parameter (λ) to find the best match that reduces error and eliminates some features.

2. In-Sample Train/ Out-of-Sample Test Split: We follow ***Goyal and Welch (2008)*** method, which "uses only the data **available up to the time at which the forecast** is made". For example:



IV. Model Evaluation

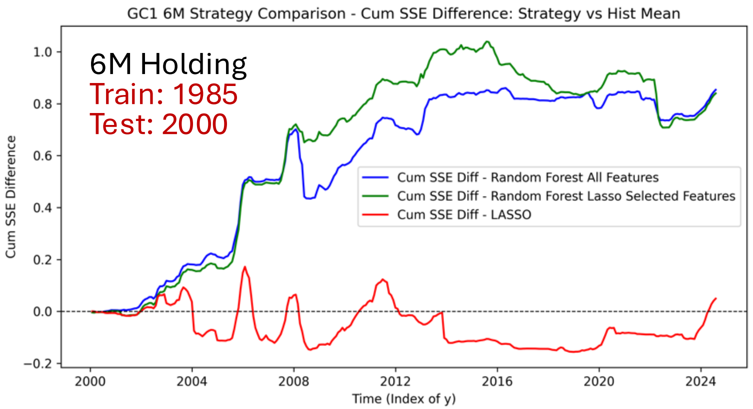
How well do the 3 models predict gold futures excess return?

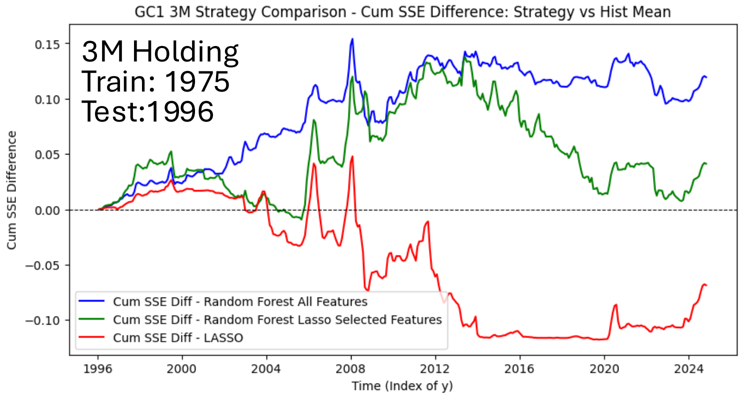
1. **Lasso performs poorly** in prediction (worst than historical mean), because it **cannot capture the non-linear relationships** of the features.

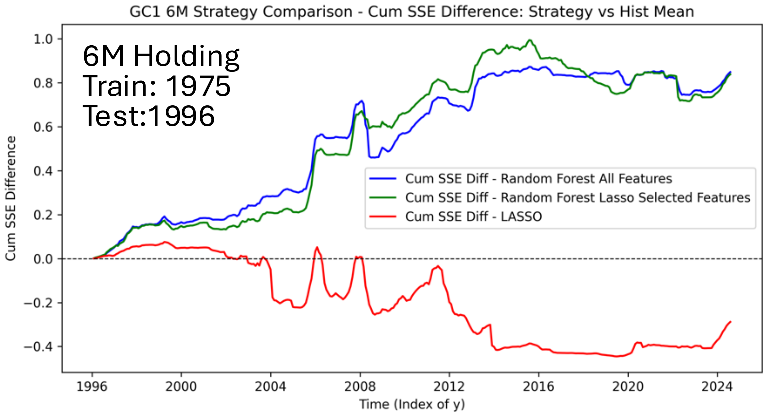
2. For cumulative SSE difference, **random forest performs better for longer-term holding** period returns (6 or 12 months), suggesting that **long-term trends are easier to forecast** than short-term fluctuations, which are often obscured by noise.

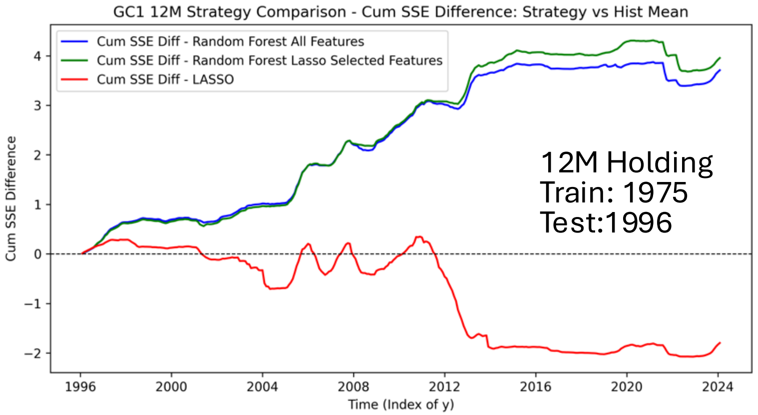
3. The **random forest with lasso-selected** features **fluctuates more than the full-feature model**, with a significant drop in cumulative SSE difference for 3M & 6M holding periods between 2016 and 2019, likely because the Lasso filter removed many features, adversely affecting its prediction performance.

4. Both random forest prediction models **struggle when the market moves sideways** rather than trending up or down.









Train start: 1975, Test start:1996

IV. Model Evaluation

Which features matter most in the LASSO filter and in the full-feature Random Forest model?

Lasso features that are used more than 33% of the time for either 3M, 6M, 12M

Features	3M	6M	12M
GC1_6M_Ret	0	343	337
GC1_12M_Ret	346	343	155
US Gov 10Yr Yield	85	272	299
US_Unemployment	141	261	337
GDP_Nominal_YOY	161	258	337
PPI	70	180	252
JPY Curncy	7	176	23
US Gov 2Yr Yield	229	174	83
GDP_Real_QoQ	118	174	337
US Real Interest Rate	71	139	311
ETF_Ounces	123	117	120
Fed Effective Rate	114	17	22
Consumer Confidence	36	67	251
Indian premium disc	67	41	133

Full-Feature Random Forest top 10 important features ranked (6M ,12M)

Rank	6M Features	Importance
1	PCE CYOY Index	0.167
2	GDP_Nominal_YOY	0.119
3	US_Unemployment	0.095
4	CPI YOY Index	0.091
5	GC1_12M_Ret	0.078
6	US Real Interest Rate	0.062
7	GDP_Real_QoQ	0.057
8	GC1_6M_Ret	0.055
9	PPI	0.054
10	ETF_Ounces	0.030

Rank	12M Features	Importance
1	PCE CYOY Index	0.200
2	GDP_Nominal_YOY	0.189
3	PPI	0.086
4	CPI YOY Index	0.086
5	US Real Interest Rate	0.084
6	GC1_12M_Ret	0.072
7	US_Unemployment	0.060
8	ETF_Ounces	0.041
9	GC1_6M_Ret	0.039
10	GDP_Real_QoQ	0.027

6-month and 12-month past returns, along with **macro** factors, are most influential for gold futures excess return prediction. **Stock market** movements and valuations **have little impact**.

V. Trading Strategy Setup

How can we build a trading strategy based on model predictions?

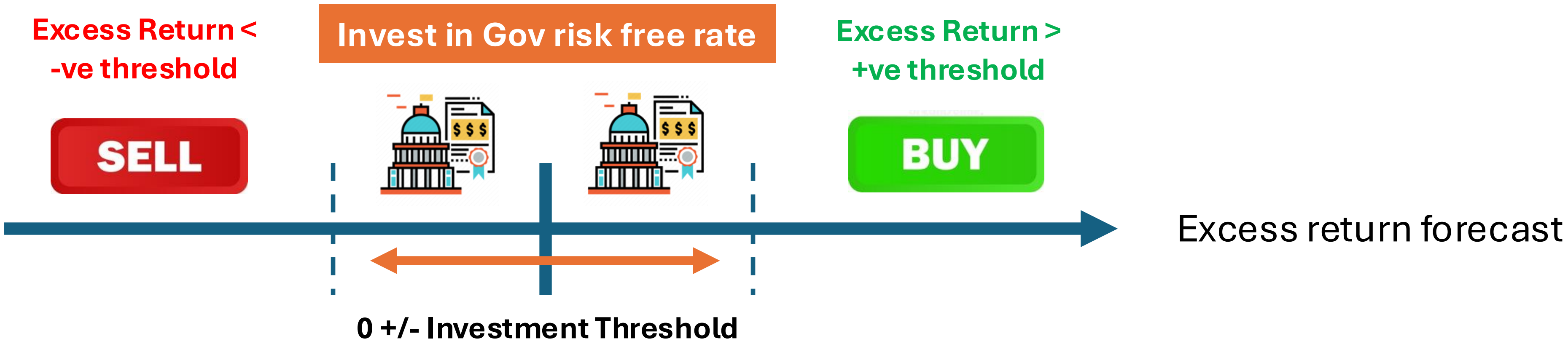
Back Testing Period Choice:
(i) 1996 Jan – 2024 Dec
(ii) 2000 Jan – 2024 Dec

Style:
(i) Long-Short
(ii) Long only

Transaction cost
(i) No cost
(ii) 10 bps/ trade (real-life)

Investment Threshold
(i) +/- 2% (can be tuned)

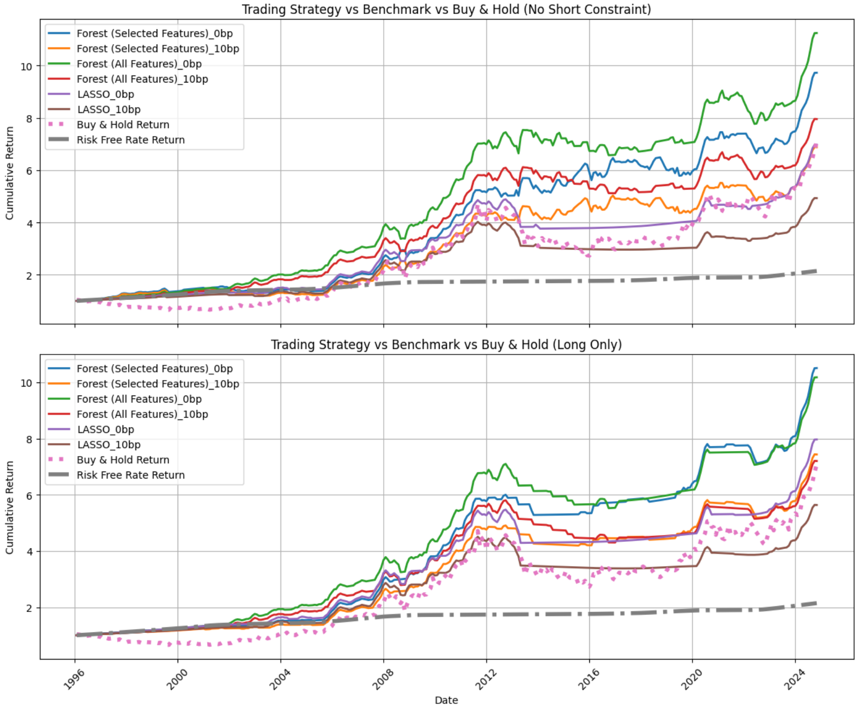
- Signal is **generated monthly** based on forecast.
- Capital is **allocated equally among holding periods**. Each month, only the portion indicated by the signal is invested, and the **realized holding period return** is **converted into a monthly return** for trading plot comparisons. *(e.g. 3% 6M return → 0.5% return for the invested month)*
- Signal is generated based on the following rule:



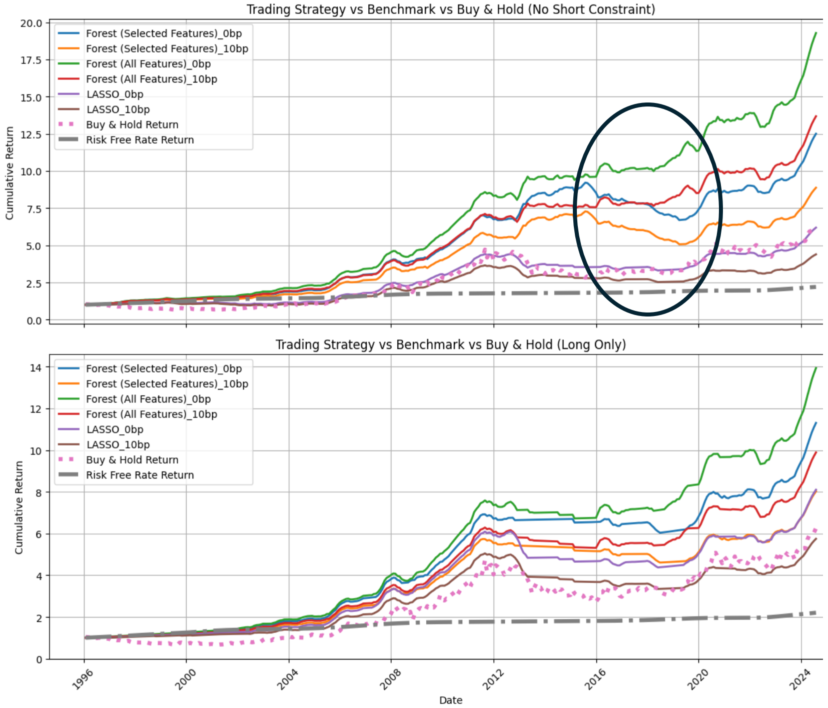
VI. Trading Strategy Evaluation

Does our strategy outperform buy-and-hold scenario?

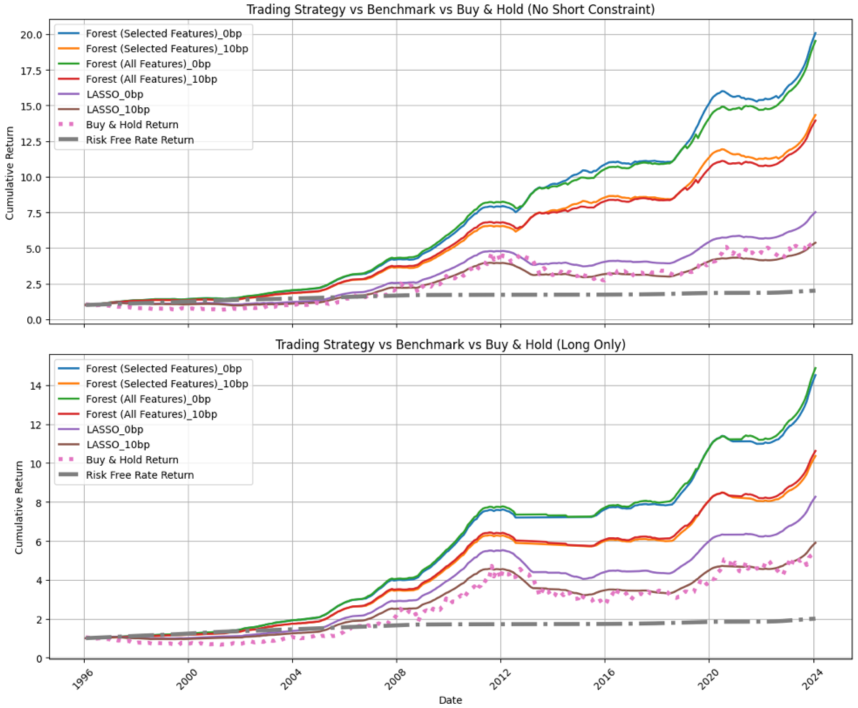
3M, Start test 1996



6M, Start test 1996



12M, Start test 1996



Lasso model does not provide good trading signal. Its performance after transaction cost is much worse than buy-and-hold strategy.

For 6M, **Full-feature random forest (green and red) predict better** and enhance investment return with **multiple risk-free and short sell actions** between 2016 and 2019.

Gold futures are **better suited for long-term forecasting and investment strategies**, as they deliver higher, smoother returns. It also **narrows the performance gap** between full-feature and selected-feature random forest models.

VI. Trading Strategy Evaluation
Does our strategy outperform buy-and-hold scenario?

6M Holding Period,
Train start: 1975, Test start:1996

Strategy	Transaction Cost	Final Value Per Dollar Invested	Annualized Return	6M Sharpe Ratio	Max Drawdown	Win Rate	R2
Buy & Hold	-	6.25	6.62%	0.24	-42.04%	n/a	n/a
Forest_All_F_LongShort	0bp	19.28	10.91%	1.57	-8.58%	73.18%	0.221
Forest_All_F_LongShort	10bp	13.68	9.58%	1.31	-9.12%	73.18%	0.221
Forest_All_F_LongOnly	0bp	13.93	9.65%	1.44	-11.35%	73.18%	0.221
Forest_All_F_LongOnly	10bp	9.89	8.35%	1.17	-15.49%	73.18%	0.221
Forest_Sel_F_LongShort	0bp	12.50	9.24%	1.20	-27.31%	69.68%	0.218
Forest_Sel_F_LongShort	10bp	8.87	7.94%	0.96	-30.44%	69.68%	0.218
Forest_Sel_F_LongOnly	0bp	11.30	8.85%	1.32	-12.76%	69.68%	0.218
Forest_Sel_F_LongOnly	10bp	8.02	7.56%	1.03	-19.67%	69.68%	0.218

6M Holding Period,
Train start: 1985, Test start: 2000

Better sharpe ratio and return

Strategy	Transaction Cost	Final Value Per Dollar Invested	Annualized Return	6M Sharpe Ratio	Max Drawdown	Win Rate	R2
Buy & Hold	-	8.38	9.03%	0.41	-42.04%	n/a	n/a
Forest_All_F_LongShort	0bp	13.39	11.13%	1.63	-8.58%	72.54%	0.154
Forest_All_F_LongShort	10bp	9.97	9.80%	1.39	-9.12%	72.54%	0.154
Forest_All_F_LongOnly	0bp	11.13	10.30%	1.57	-11.35%	72.54%	0.154
Forest_All_F_LongOnly	10bp	8.29	8.98%	1.31	-15.49%	72.54%	0.154
Forest_Sel_F_LongShort	0bp	10.91	10.21%	1.41	-21.20%	71.53%	0.150
Forest_Sel_F_LongShort	10bp	8.12	8.89%	1.17	-24.30%	71.53%	0.150
Forest_Sel_F_LongOnly	0bp	10.16	9.89%	1.47	-14.96%	71.53%	0.150
Forest_Sel_F_LongOnly	10bp	7.56	8.58%	1.22	-21.50%	71.53%	0.150

VII. Discussion and Next Step

For coding, please refer to "gold_futures_data_0223.ipynb", all dataframes and plot for the runs are saved in the folder name "generated_dataframe"

1. What is the commercial value for this study?

Investing in gold commodities provides **portfolio diversification**. Hedge fund managers may **speculate on commodities** to capitalize on directional trends. Machine learning systematic trading strategies offer commercial value to investors.

2. What economic interpretation can we learn from the features?

Focusing on longer holding periods (e.g., 6 and 12 months), past returns for these periods, as well as **macroeconomic factors** such as growth, unemployment, and inflation, are most influential. **Yield** also plays a role, as it **relates to the risk-free rate** and therefore **excess return** prediction.

As expected, **stock market factors are less significant** for gold returns, aligning with common financial findings.

3. Why does the trading strategy work?

Beyond model effectiveness, our focus is solely on **directional bets** rather than precisely predicting future returns, which is nearly impossible. We **establish a threshold** so that if the predicted excess return is close to 0%, **we opt for a risk-free investment** instead of taking a directional bet. This approach helps **prevent excessive trading or unconfident guesses**, **preserves capital for future growth**, and still generates some return—especially when the market is sideways or declining.

4. What can be further improved for the machine learning model and trading strategy?

We can further tune random forest hyperparameters and consider **switching from an expanding window**, which may overweight outdated data, to **a rolling window approach**. This is particularly relevant for older periods (e.g., 1970–1980) with volatile risk-free rates. Additionally, **optimizing the investment threshold** could enhance trading performance.