Forecasting Swap Rate Spread from Real Economic Indicators

Name:	Thuy Anh Tran
Matriculation number:	31738A
Course:	Bank Loan Pricing

TABLE OF CONTENTS

1. Data Preparation_ Features

❸ Indicator name
❷ Release date
❶ Indicator value
Reference report date

	H	I	J	K	L	M
1		1		6.036363636	2	
2		GRIFPBUS Index			ITBCI Index	
3		IFO Business Climate			Manufacturing Confidence	
4						
5		2004-12-31	#N/A N/A	20041217	2004-12-31	102.7
6		2005-01-31	92.2	20050126	2005-01-31	102
7		2005-02-28	91.9	20050223	2005-02-28	101.1
8		2005-03-31	90.1	20050323	2005-03-31	100
9		2005-04-30	90	20050425	2005-04-30	97.1
10		2005-05-31	89.3	20050525	2005-05-31	97.2
11		2005-06-30	89.4	20050627	2005-06-30	96.9
12		2005-07-31	91.1	20050726	2005-07-31	100.4
13		2005-08-31	91.7	20050825	2005-08-31	103.7
14		2005-09-30	92.4	20050927	2005-09-30	104.3
15		2005-10-31	94.5	20051025	2005-10-31	105.1
16		2005-11-30	94.9	20051124	2005-11-30	105.8
17		2005-12-31	95	20051216	2005-12-31	106.9
18		2006-01-31	96.5	20060125	2006-01-31	108.7
19		2006-02-28	97.7	20060223	2006-02-28	109.2
20		2006-03-31	98.1	20060328	2006-03-31	110.3
21		2006-04-30	98	20060425	2006-04-30	112.1
22		2006-05-31	98.8	20060524	2006-05-31	110.8
23		2006-06-30	99.1	20060627	2006-06-30	113.9
24		2006-07-31	98.6	20060726	2006-07-31	111.8

- Data starts from column I, row 2.
- Index indicator and extract release date and corresponding indicator value

```
release_date_col_idx = 3 * idx + 2
value_col_idx = 3 * idx + 1
name_col_idx = 3 * idx
```

- Get indicator name from row 1 in **name_col_idx**

```
name = data_columns.iloc[0, name_col_idx]
```

1. Data Preparation_ Features

```
new_dataframe_adjusted
```

	GRIFPBUS Index - release date	GRIFPBUS Index	ITBCI Index - release date	ITBCI Index	MPMIFRMA Index - release date	MPMIFRMA Index	GRZEWI Index - release date	GRZEWI Index	GRZECURR Index - release date	GRZECURR Index	...	ETSLTOTL Index - release date	ETSLTOTL Index	ADP CHNG Index - release date	ADP CHNG Index
0	2004-12-17	NaN	2004-12-28	102.7	2005-01-03	NaN	2004-12-07	14.4	2004-12-07	-64.2	...	NaN	6.89	NaN	43.5
1	2005-01-26	92.2	2005-01-27	102.0	2005-02-01	NaN	2005-01-11	26.9	2005-01-11	-61.2	...	2005-02-25	7.10	NaN	106.4
2	2005-02-23	91.9	2005-02-23	101.1	2005-03-01	NaN	2005-02-15	35.9	2005-02-15	-58.7	...	2005-03-23	6.88	NaN	145.7
3	2005-03-23	90.1	2005-03-24	100.0	2005-04-01	NaN	2005-03-15	36.3	2005-03-15	-66.0	...	2005-04-25	6.96	NaN	190.9
4	2005-04-25	90.0	2005-04-28	97.1	2005-05-02	NaN	2005-04-19	20.1	2005-04-19	-73.0	...	2005-05-24	7.12	NaN	219.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
198	2021-06-24	101.9	2021-07-28	115.0	2021-07-01	59.0	2021-06-08	79.8	2021-06-08	-9.1	...	2021-07-22	5.87	2021-06-30	741.4
199	2021-07-26	100.8	2021-08-27	113.2	2021-08-02	58.0	2021-07-06	63.3	2021-07-06	21.9	...	2021-08-23	6.00	2021-08-04	321.5
200	2021-08-25	99.6	2021-09-24	113.0	2021-09-01	57.5	2021-08-10	40.4	2021-08-10	29.3	...	2021-09-22	5.88	2021-09-01	339.7
201	2021-09-24	98.8	NaN	NaN	2021-10-01	55.0	2021-09-07	26.5	2021-09-07	31.9	...	2021-10-21	NaN	2021-10-06	568.1
202	NaN	NaN	NaN	NaN	NaN	NaN	2021-10-12	22.3	2021-10-12	21.6	...	NaN	NaN	2021-11-03	NaN

203 rows × 62 columns

Indicator release date

Indicator value

1. Data Preparation_ Features

Statistical Summary for Each Indicator:

	Indicator	Start Date	End Date	Average Value	Variance	Missing Value Percentage	Missing Date Percentage
0	GRIFPBUS Index	2004-12-17	2021-09-24	97.057711	3.063265e+03	0.985222	0.492611
1	ITBCI Index	2004-12-28	2021-09-24	100.358706	7.575364e+03	0.985222	0.985222
2	MPMIFRMA Index	2005-01-03	2021-10-01	51.361111	2.748130e+03	82.266010	0.492611
3	GRZEWI Index	2004-12-07	2021-10-12	14.035468	1.113799e+05	0.000000	0.000000
4	GRZECURR Index	2004-12-07	2021-10-12	18.067980	3.054423e+05	0.000000	0.000000
5	GRIORTMM Index	2005-02-18	2021-12-06	0.251741	1.696331e+03	0.985222	0.000000
6	MPMIITMA Index	2005-01-03	2021-11-01	51.102778	4.134142e+03	82.266010	0.492611
7	MPMIDEMA Index	2005-01-03	2021-10-01	51.244444	7.807568e+03	82.266010	0.492611
8	MPMIEZMA Index	2005-01-03	2021-10-01	51.505556	4.775025e+03	82.266010	0.492611
9	MPMIESMA Index	2014-04-01	2021-11-01	50.786111	3.237780e+03	82.266010	54.679803
10	MPMINLMA Index	2014-04-01	2021-11-01	54.650000	5.219343e+03	82.266010	54.679803
11	ITPSSA Index	2004-12-21	2021-09-24	101.587065	7.552293e+03	0.985222	0.985222
12	CZRSYOY Index	2005-02-16	2021-12-06	3.666667	3.471273e+03	0.985222	0.000000
13	BEBCI Index	2004-12-22	2021-09-24	-4.831188	6.869788e+03	0.492611	0.492611
14	IERSVMOM Index	2005-02-18	2021-10-28	0.585750	6.899035e+03	60.591133	0.492611
15	SWRSAMM Index	2005-01-27	2021-10-28	0.285871	1.641511e+02	0.985222	0.492611
16	GRIFPEX Index	2004-12-17	2021-09-24	97.590050	2.683200e+03	0.985222	0.492611
17	IERSVYOY Index	2005-02-18	2021-10-28	3.208625	1.733359e+04	60.591133	0.492611
18	SWETSURV Index	2007-06-26	2021-09-29	101.038119	1.126072e+04	0.492611	15.270936
19	NAPMPMI Index	2005-01-03	2021-11-01	53.395050	2.294614e+03	0.492611	0.000000
20	CONSSENT Index	2004-12-23	2021-10-15	82.885714	1.401762e+04	0.000000	0.000000
21	CONCCONF Index	2004-12-28	2021-09-28	90.483465	7.437960e+04	0.492611	0.492611
22	RSTAMOM Index	2005-01-13	2021-11-16	0.346535	5.163396e+02	0.492611	0.492611
23	NHSLTOT Index	2005-01-31	2021-10-26	612.373134	7.279843e+06	0.985222	0.492611
24	MPMIUSMA Index	2012-06-01	2021-10-01	53.747222	3.212085e+03	82.266010	44.334975
25	NHSPSTOT Index	2005-01-19	2021-10-19	1149.029851	1.834655e+07	0.985222	0.492611
26	ETSLTOTL Index	2005-02-25	2021-10-21	5.211493	7.368608e+01	0.985222	0.985222
27	ADP CHNG Index	2006-08-02	2021-11-03	63.980198	2.083150e+08	0.492611	9.359606
28	LEI CHNG Index	2005-01-20	2021-10-21	0.088557	1.136518e+02	0.985222	0.492611
29	EMPRGBCI Index	2004-12-15	2021-10-15	7.573399	2.243763e+04	0.000000	0.000000
30	CHPMINDX Index	2004-12-30	2021-09-30	55.832673	6.238291e+03	0.492611	0.492611

Filter features having missing value percentage and missing date percentage less than 5%

```
valid_indicators = summary_df[(summary_df['Missing Value Percentage'] < 5) & (summary_df['Missing Date Percentage'] < 5) ]['Indicator']
```

→ **20 indicators left**

1. Data Preparation_ Features

- Fill in missing dates:
  - If the current month ≠ previous month + 1
  → Adjust current date = previous month + 1

- Fill in missing values:
  - If it's the first value in the series → $D_0 = D_1$
  - If it's not the first value in the series → $D_t = D_{t-1}$ if D_{t-1} is not NA.

DataFrame after filling NA values:

	GRIFPBUS Index - release date	GRIFPBUS Index	ITBCI Index - release date	ITBCI Index	GRZEWI Index - release date	GRZEWI Index	GRZECURR Index - release date	GRZECURR Index	GRIORTMM Index - release date	GRIORTMM Index	...	NHSPSTOT Index - release date	NHSPSTOT Index	ETSLTOTL Index - release date	E
0	2004-12	92.2	2004-12	102.7	2004-12	14.4	2004-12	-64.2	2005-02	6.4	...	2005-01	2042.0	2005-02	
1	2005-01	92.2	2005-01	102.0	2005-01	26.9	2005-01	-61.2	2005-03	-2.6	...	2005-02	2144.0	2005-03	
2	2005-02	91.9	2005-02	101.1	2005-02	35.9	2005-02	-58.7	2005-04	-1.8	...	2005-03	2207.0	2005-04	
3	2005-03	90.1	2005-03	100.0	2005-03	36.3	2005-03	-66.0	2005-05	2.0	...	2005-04	1864.0	2005-05	
4	2005-04	90.0	2005-04	97.1	2005-04	20.1	2005-04	-73.0	2005-06	-0.8	...	2005-05	2061.0	2005-06	

1. Data Preparation_ Target

ORIGINAL

	Date	USDSB3L10Y	USDSB3L5Y	USDSB3L2Y
0	2024-01-09	3.6020	3.5379	3.7170
1	2024-01-08	3.7210	3.6890	4.0092
2	2024-01-07	3.9424	3.9801	4.4429
3	2024-01-06	4.1525	4.2959	4.8460
4	2024-01-05	4.4600	4.6050	5.0990
...	...	...	...	...
166	2010-01-11	2.9660	1.7540	0.7560
167	2010-01-10	2.6800	1.4180	0.4940
168	2010-01-09	2.5610	1.4980	0.5950
169	2010-01-08	2.4810	1.6030	0.6490
170	2010-01-07	2.8860	1.7750	0.7200

171 rows × 4 columns

TRANSFORM

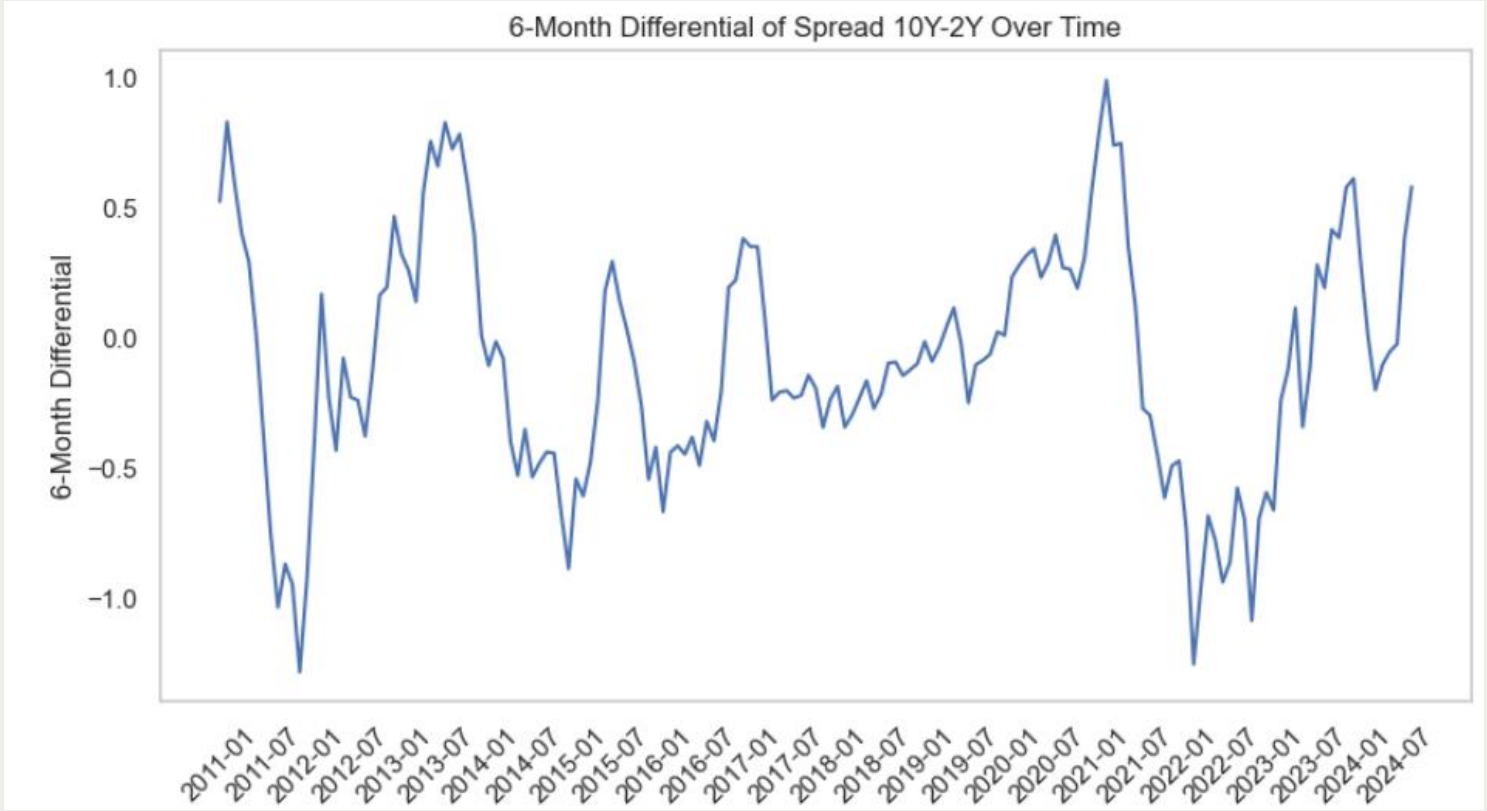
```
swap_df['Date'] = pd.to_datetime(swap_df['Date'], format='%Y-%m-%d', errors='coerce')

# Extract only the Year and Month as a separate column
swap_df['Year-Month'] = swap_df['Date'].dt.strftime('%Y-%d')

# Set the 'Year-Month' column as the index and drop the original column
swap_df.set_index('Year-Month', inplace=True)

# Sort the dataframe by Date in ascending order
swap_df = swap_df.sort_values(by='Date')

# Calculate the spread: S(t) = USDSB3L10Y - USDSB3L2Y
swap_df['S(t)'] = swap_df['USDSB3L10Y'] - swap_df['USDSB3L2Y']

# Calculate the 6-month differential: y_6m_diff = S(t) - S(t-6)
swap_df['y_6m_diff'] = swap_df['S(t)'].diff(periods=6)

# Drop rows with NaN in 'Date' or NaN in 'y_6m_diff'
swap_df = swap_df.dropna(subset=['Date', 'y_6m_diff'])

# Only keep the column y_6m_diff (Year-Month is the index)
swap_df = swap_df[['y_6m_diff']]
```

RESULT

Year-Month	y_6m_diff
2011-01	0.5237
2011-02	0.8290
2011-03	0.6000
2011-04	0.4020
2011-05	0.2940
...	...
2024-05	-0.1010
2024-06	-0.0515
2024-07	-0.0218
2024-08	0.3788
2024-09	0.5802

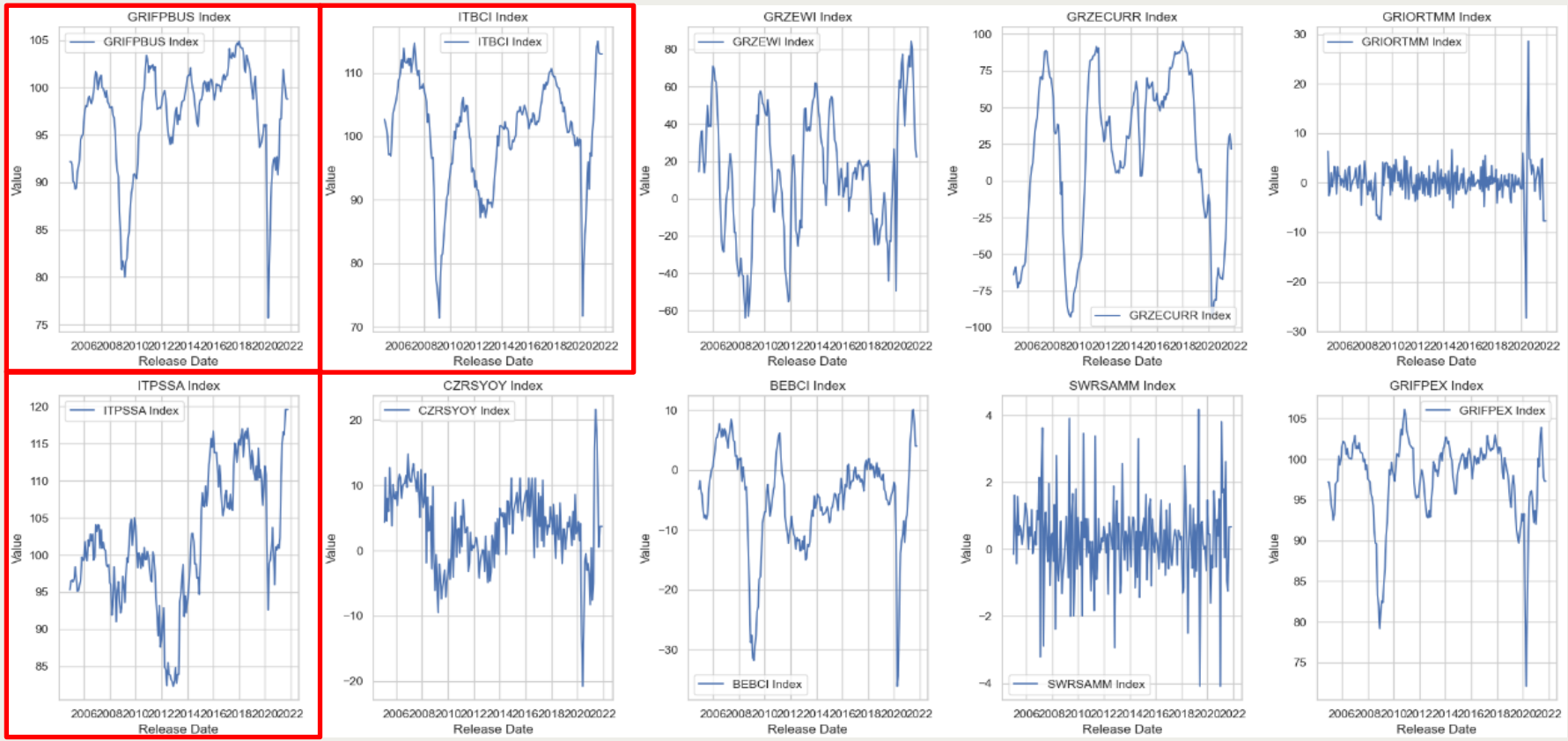
165 rows × 1 columns

2. Exploratory Data Analysis_ Target



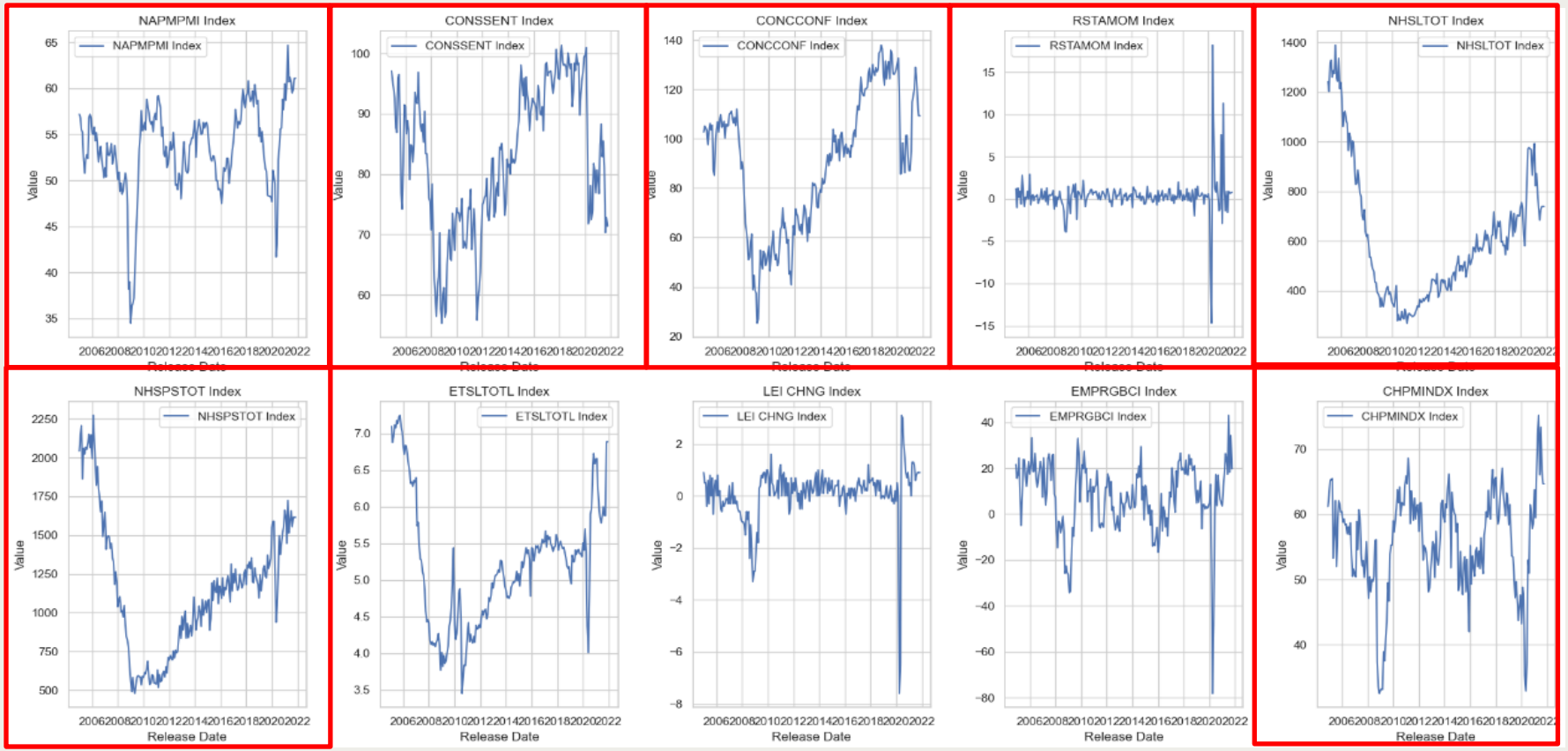
2. Exploratory Data Analysis_ Features' trends

- Non-stationary behavior, on different scale, extreme spikes



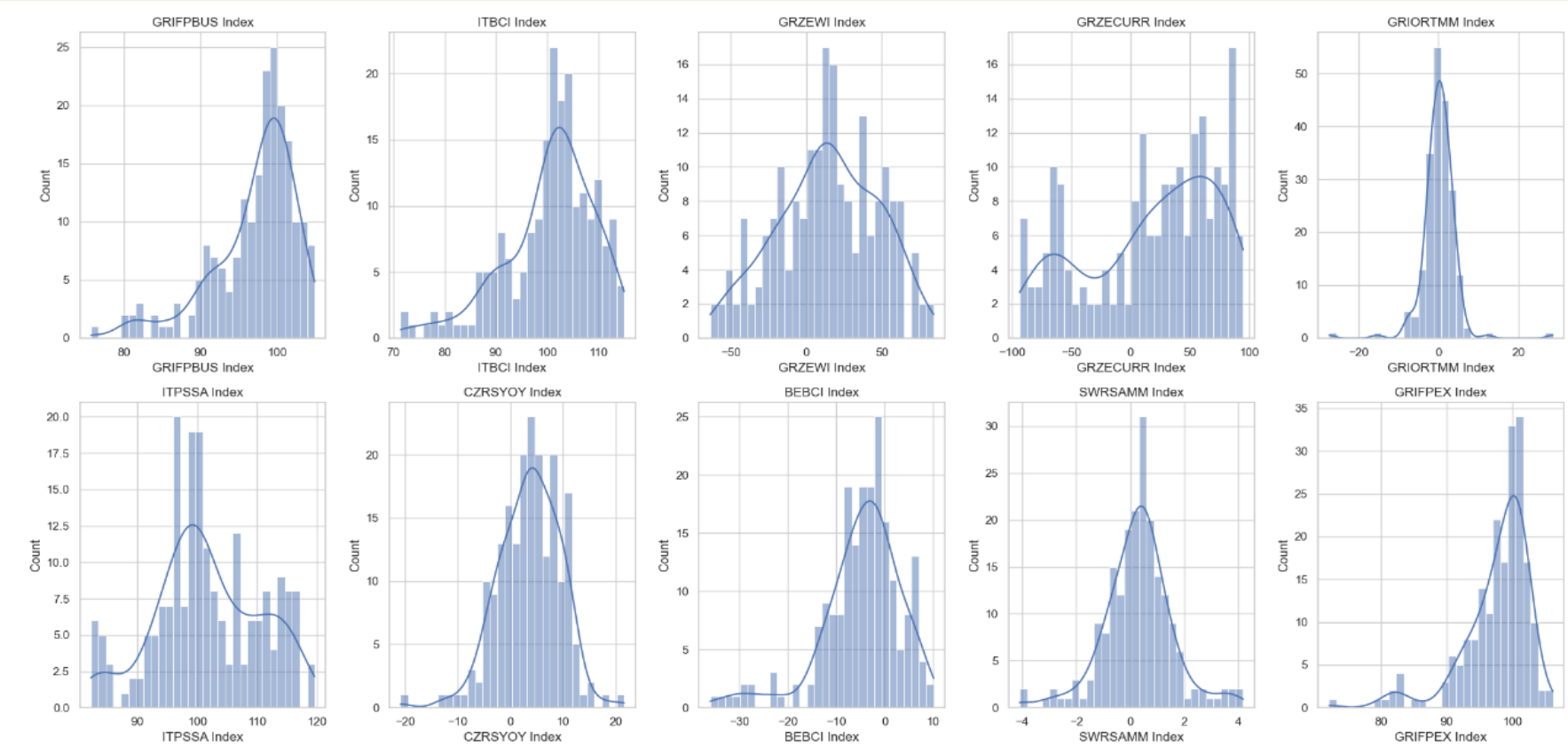
2. Exploratory Data Analysis_ Features' trends

- Non-stationary behavior, highly fluctuate, extreme spikes



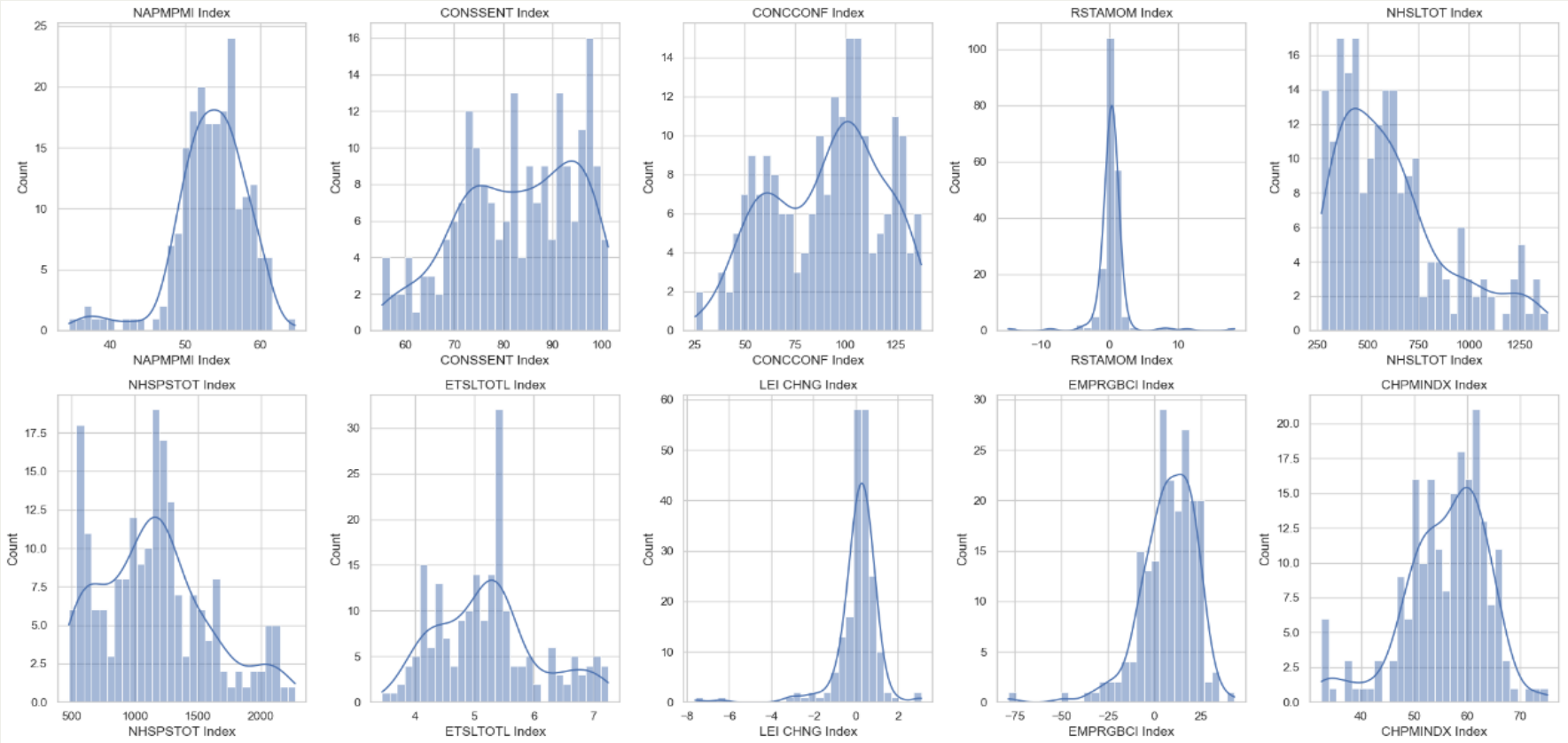
2. Exploratory Data Analysis _ Distributions

- Many of these distributions are not normally distributed. They exhibit various shapes, including right-skewed, left-skewed, multi-modal, and highly irregular patterns.

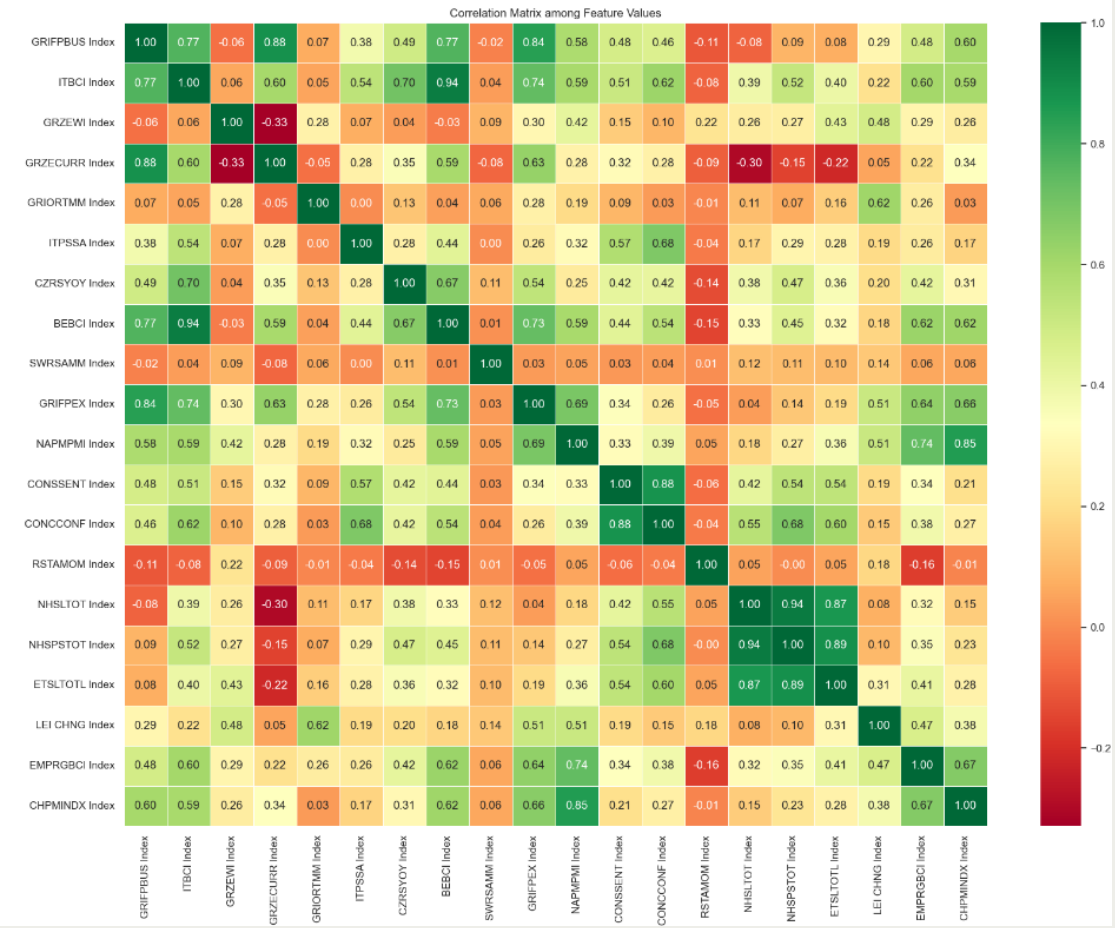


2. Exploratory Data Analysis_ Distributions

- Many of these distributions are not normally distributed. They exhibit various shapes, including right-skewed, left-skewed, multi-modal, and highly irregular patterns.



2. Exploratory Data Analysis_ Correlation matrix



Many feature have high correlation to each others

→ We need to do feature selection to choose those that contribute the most to explain the target.

2. Exploratory Data Analysis_ Transformation

#	Indicator	Details of indicator	Transformation
1	GRIFPBUS Index	IFO Business Climate	Log
2	ITBCI Index	Manufacturing Confidence	Log
3	GRZEWI Index	ZEW Survey Expectations	Difference
4	GRZECURR Index	ZEW Survey Current Situation	Difference
5	GRIORTMM Index	Factory Orders MoM	Difference
6	ITPSSA Index	Consumer Confidence Index	Log
7	CZRSYOY Index	Retail Sales YoY	Difference
8	BEBCI Index	Business Confidence	Difference
9	SWRSAMM Index	Retail Sales MoM	Difference
10	GRIFPEX Index	IFO Expectations	Log

#	Indicator	Details of indicator	Transformation
11	NAPMPMI Index	ISM Manufacturing	Log
12	CONSSENT Index	U. of Mich. Sentiment	Difference
13	CONCCONF Index	Conf. Board Consumer Confidence	Difference
14	RSTAMOM Index	Retail Sales Advance MoM	Difference
15	NHSLTOT Index	New Home Sales	Log
16	NHSPSTOT Index	Housing Starts	Log
17	ETSLTOTL Index	Existing Home Sales	Log
18	LEI CHNG Index	Leading Index	Difference
19	EMPRGBCI Index	Empire Manufacturing	Difference
20	CHPMINDX Index	MNI Chicago PMI	Log

Merge Features vs Target

Target

Start: 2011- Jan
End: 2024 - Sep

Final Dataframe

Start: 2011- Jan
End: 2021 - Oct

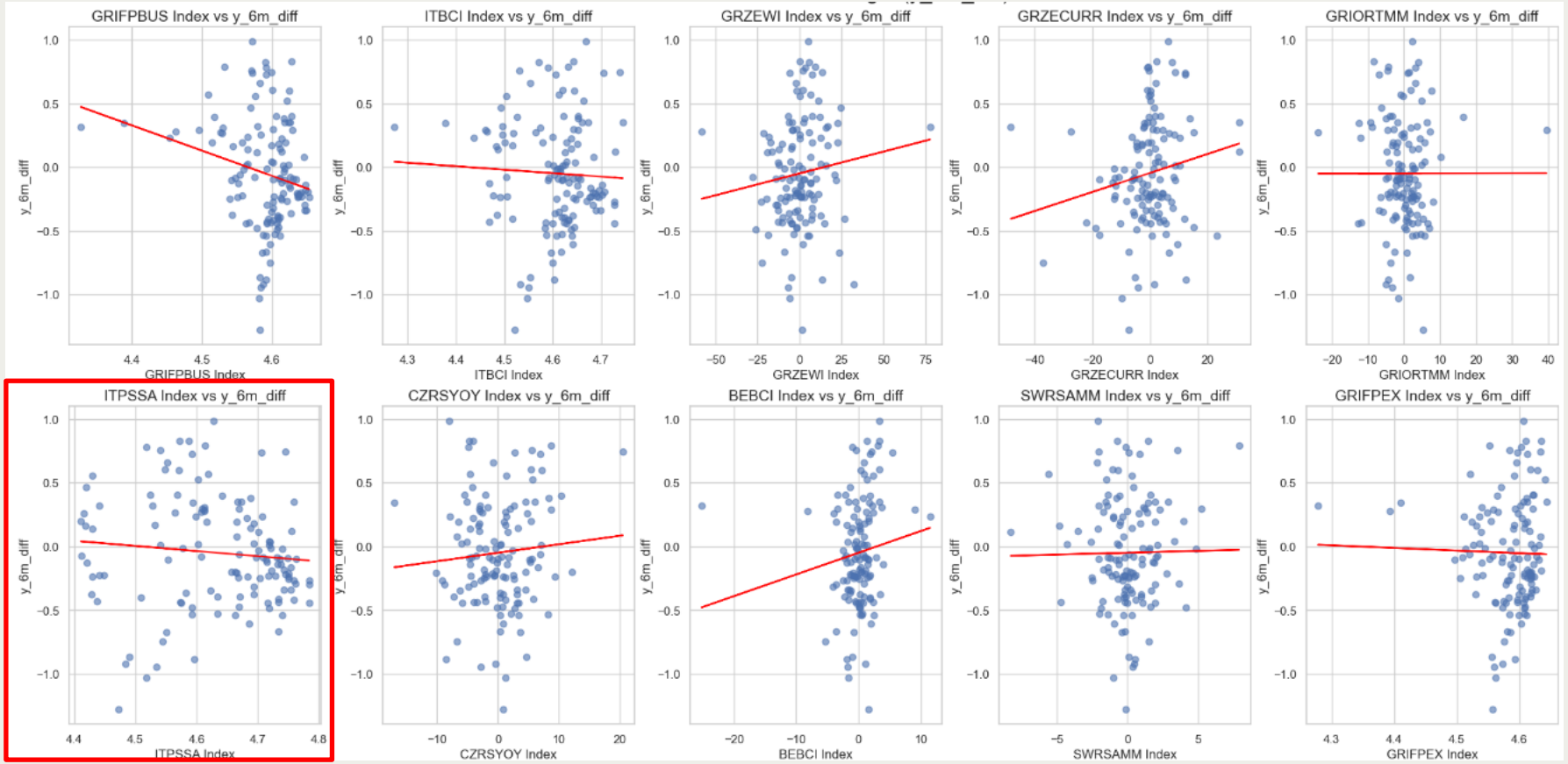
Features

Start: 2004 - Dec
End: 2021 - Oct

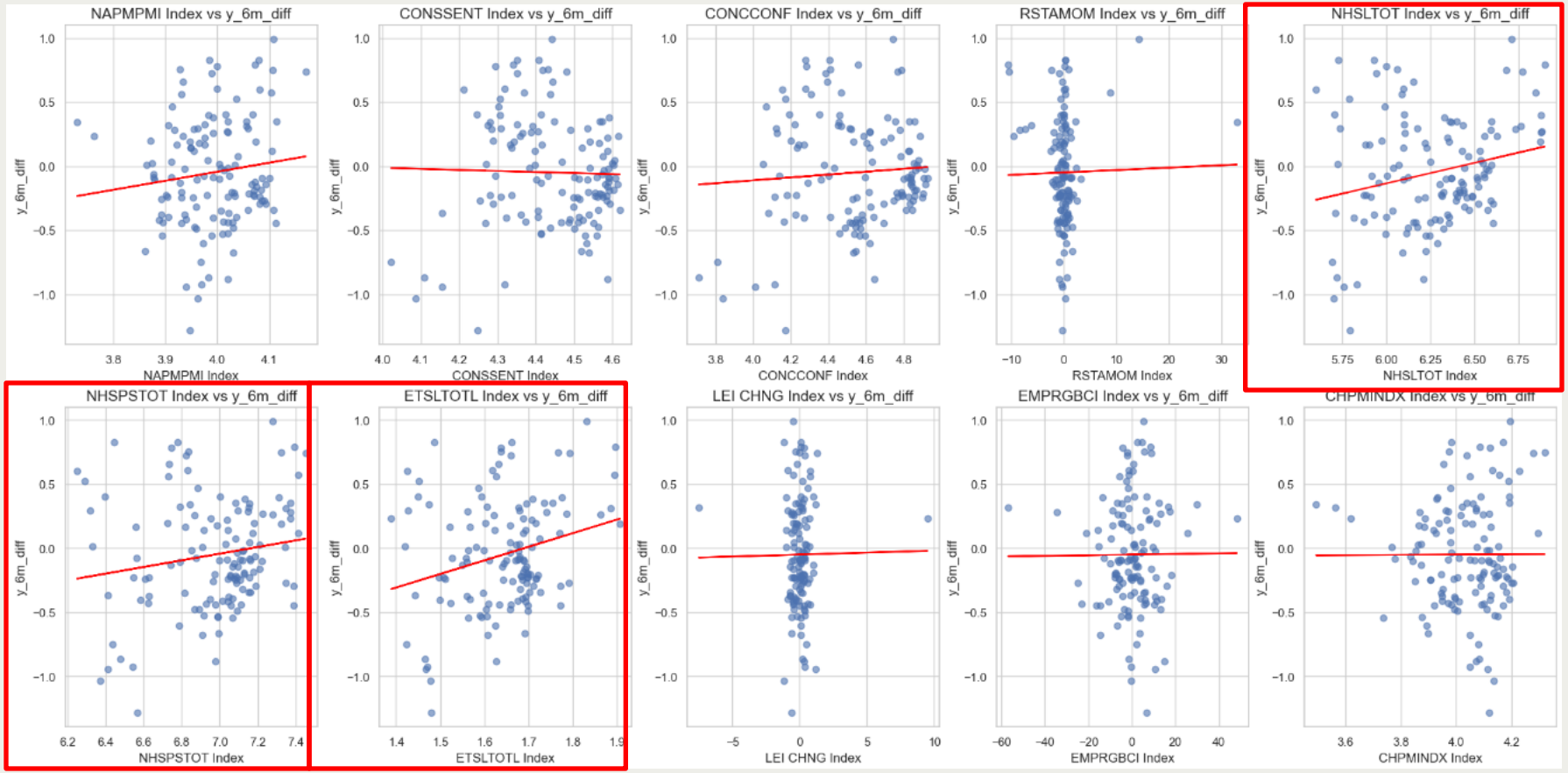
Cleaned DataFrame (with NaN rows dropped):

Year-Month	y_6m_diff	GRIFPBUS Index	ITBCI Index	GRZEWI Index	GRZECURR Index	GRIORTMM Index	ITPSSA Index	CZRSYOY Index	BEBCI Index	SWRSAMM Index	...	NAPMPMI Index	CONSSENT Index	CONCCONF Index	
Start date → 2011-01	0.5237	4.621044	4.664382	11.1	0.2	4.9	4.603168	5.7	1.4	-1.47	...	4.036009	4.306764	4.171151	
2011-02	0.8290	4.627910	4.643429	0.3	2.4	-8.5	4.587006	-4.7	1.3	1.47	...	4.079231	4.350278	4.276944	
2011-03	0.6000	4.624973	4.644391	-1.6	0.2	7.7	4.571613	7.2	0.4	-0.69	...	4.080922	4.212128	4.156067	
2011-04	0.4020	4.627910	4.653008	-6.5	1.7	-3.3	4.568506	-2.0	-3.4	-1.37	...	4.067316	4.245634	4.189958	
2011-05	0.2940	4.628887	4.652054	-4.5	4.4	-4.5	4.609162	-3.1	-3.3	5.21	...	4.058717	4.308111	4.122932	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2021-06	0.3515	4.623992	4.744932	-4.6	31.0	-1.8	4.758749	6.0	3.3	2.87	...	4.114147	4.448516	4.859037	
2021-07	0.1200	4.613138	4.729156	-16.5	31.0	-4.8	4.755313	-3.1	0.3	-3.53	...	4.104295	4.396915	4.829113	
2021-08	-0.2696	4.601162	4.727388	-22.9	7.4	8.0	4.784153	-9.2	-2.5	-0.34	...	4.085976	4.252772	4.746670	
2021-09	-0.2956	4.593098	4.727388	-13.9	2.6	0.3	4.784153	-8.8	-3.6	1.92	...	4.092677	4.287716	4.694096	
End date → 2021-10	-0.4446	4.593098	4.727388	-4.2	-10.3	-12.6	4.784153	3.2	0.0	0.00	...	4.112512	4.268298	4.694096	

2. Exploratory Data Analysis_ Scatter plots



2. Exploratory Data Analysis _ Scatter plots

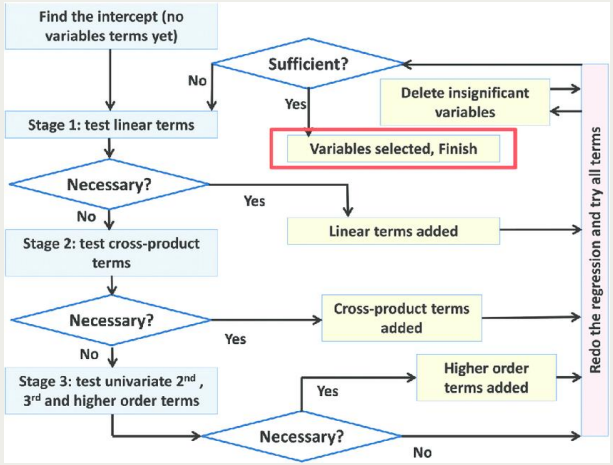


Linear Model Selection

Stepwise Regression

Hybrid approach: Forward selection + Backward elimination
Variables are added to the model sequentially, in analogy to forward selection. However, after adding each new variable, the method may also remove any variables that no longer provide an improvement in the model fit.

Sample workflow of Forward Stepwise regression



LASSO Regularization

A technique that regularizes the coefficient estimates, shrinks the coefficients towards zero. The lasso regression coefficient estimates $\widehat{\beta^L}$ are the values that minimize:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\underbrace{\sum_{j=1}^{p}|\beta_j|}_{\text{Shrinkage penalty}}$$

$\lambda >= 0$: tuning parameter

$|\beta|$: l1 norm

λ balances the tradeoff between bias and variance in the resulting coefficients.
As λ increases, the bias increases, and the variance decreases, leading to a simpler model with fewer parameters, avoid overfitting.

3. Model Selection _ Stepwise Regression Result

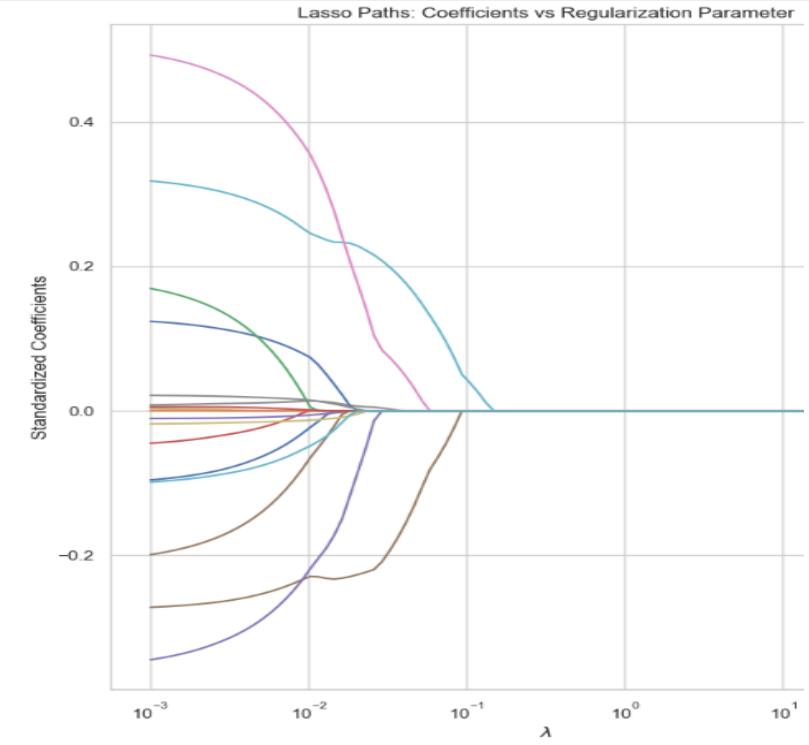
```
# Split the data into training and test sets
test_set = merged_df_cleaned.iloc[-36:]  # Get the last 36 rows for the test set
training_set = merged_df_cleaned.iloc[:-36]  # Get all rows except the last 36 for the training set
```

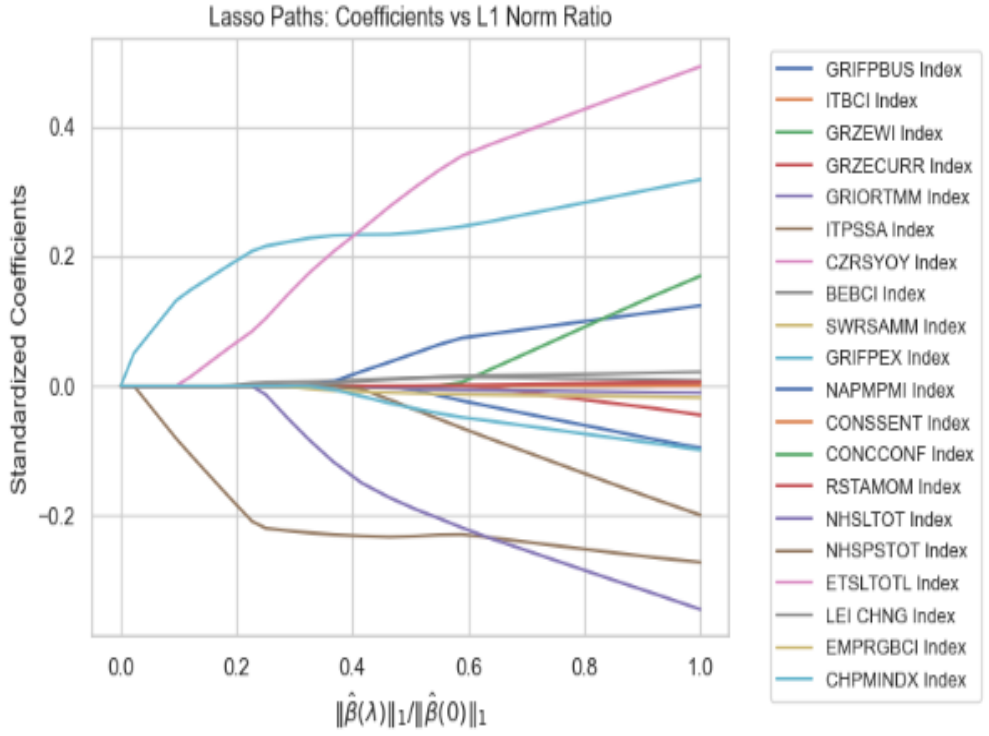
```
Add  GRIFPEX Index with p-value 0.0013064251441023346
Add  ITPSSA Index with p-value 1.608354550819898e-05
Add  ETSLTOTL Index with p-value 0.0005136148322723488
Add  NHSLTOT Index with p-value 0.0009678868493614711
                            OLS Regression Results
==============================================================================
Dep. Variable:              y_6m_diff   R-squared:                       0.438
Model:                            OLS   Adj. R-squared:                  0.413
Method:                 Least Squares   F-statistic:                     17.34
Date:                Sat, 23 Nov 2024   Prob (F-statistic):           1.49e-10
Time:                        19:22:00   Log-Likelihood:                -29.037
No. Observations:                  94   AIC:                             68.07
Df Residuals:                      89   BIC:                             80.79
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -39.0205      6.921     -5.638      0.000     -52.773     -25.268
GRIFPEX Index  10.9400      1.700      6.436      0.000       7.563      14.317
ITPSSA Index   -2.3379      0.536     -4.358      0.000      -3.404      -1.272
ETSLTOTL Index  5.0598      1.040      4.866      0.000       2.993       7.126
NHSLTOT Index  -1.4351      0.420     -3.413      0.001      -2.270      -0.600
==============================================================================
Omnibus:                        3.084   Durbin-Watson:                   0.690
Prob(Omnibus):                  0.214   Jarque-Bera (JB):                1.830
Skew:                          -0.002   Prob(JB):                        0.401
Kurtosis:                       2.317   Cond. No.                     1.87e+03
==============================================================================
```

```
Mean Squared Error (MSE) on Training Set: 0.1086
```

3. Model Selection _ LASSO Result

```
# Split the data into training and test sets
test_set = merged_df_cleaned.iloc[-36:]  # Get the last 36 rows for the test set
training_set = merged_df_cleaned.iloc[:-36]  # Get all rows except the last 36 for the training set
```





3. Model Selection _ LASSO Result

```
Lasso with alpha = 0.01
Selected Features: ['GRIFPBUS Index' 'GRIORTMM Index' 'ITPSSA Index' 'BEBCI Index'
 'GRIFPEX Index' 'NAPMPMI Index' 'CONCCONF Index' 'RSTAMOM Index'
 'NHSLTOT Index' 'NHSPSTOT Index' 'ETSLTOTL Index' 'LEI CHNG Index'
 'EMPRGBCI Index' 'CHPMINDX Index']
Mean Squared Error (MSE): 1.5579900639097444
R-squared: -14.036064611712659
Coefficients for Selected Features:
GRIFPBUS Index: -0.0246
GRIORTMM Index: -0.0060
ITPSSA Index: -0.2298
BEBCI Index: 0.0137
GRIFPEX Index: 0.2468
NAPMPMI Index: 0.0750
CONCCONF Index: 0.0090
RSTAMOM Index: 0.0012
NHSLTOT Index: -0.2221
NHSPSTOT Index: -0.0682
ETSLTOTL Index: 0.3583
LEI CHNG Index: 0.0149
EMPRGBCI Index: -0.0132
CHPMINDX Index: -0.0495

Lasso with alpha = 0.04
Selected Features: ['ITPSSA Index' 'GRIFPEX Index' 'ETSLTOTL Index']
Mean Squared Error (MSE): 0.795090189168343
R-squared: -6.67336566092928
Coefficients for Selected Features:
ITPSSA Index: -0.1607
GRIFPEX Index: 0.1799
ETSLTOTL Index: 0.0516

Comparison of Lasso Models:
Alpha = 0.01: MSE = 1.5579900639097444
Alpha = 0.04: MSE = 0.795090189168343
```

I choose model with alpha =0.04 with lower MSE, then refit chosen features in OLS

OLS Model Summary:

OLS Regression Results

Dep. Variable:	y_6m_diff	R-squared:	0.364
Model:	OLS	Adj. R-squared:	0.343
Method:	Least Squares	F-statistic:	17.20
Date:	Sun, 24 Nov 2024	Prob (F-statistic):	6.50e-09
Time:	13:18:39	Log-Likelihood:	-34.819
No. Observations:	94	AIC:	77.64
Df Residuals:	90	BIC:	87.81
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-40.8876	7.296	-5.604	0.000	-55.383	-26.392
ITPSSA Index	-3.1087	0.515	-6.041	0.000	-4.131	-2.086
GRIFPEX Index	11.2830	1.794	6.288	0.000	7.718	14.848
ETSLTOTL Index	1.9902	0.552	3.604	0.001	0.893	3.087

Omnibus:	3.117	Durbin-Watson:	0.527
Prob(Omnibus):	0.210	Jarque-Bera (JB):	2.105
Skew:	0.168	Prob(JB):	0.349
Kurtosis:	2.348	Cond. No.	1.38e+03

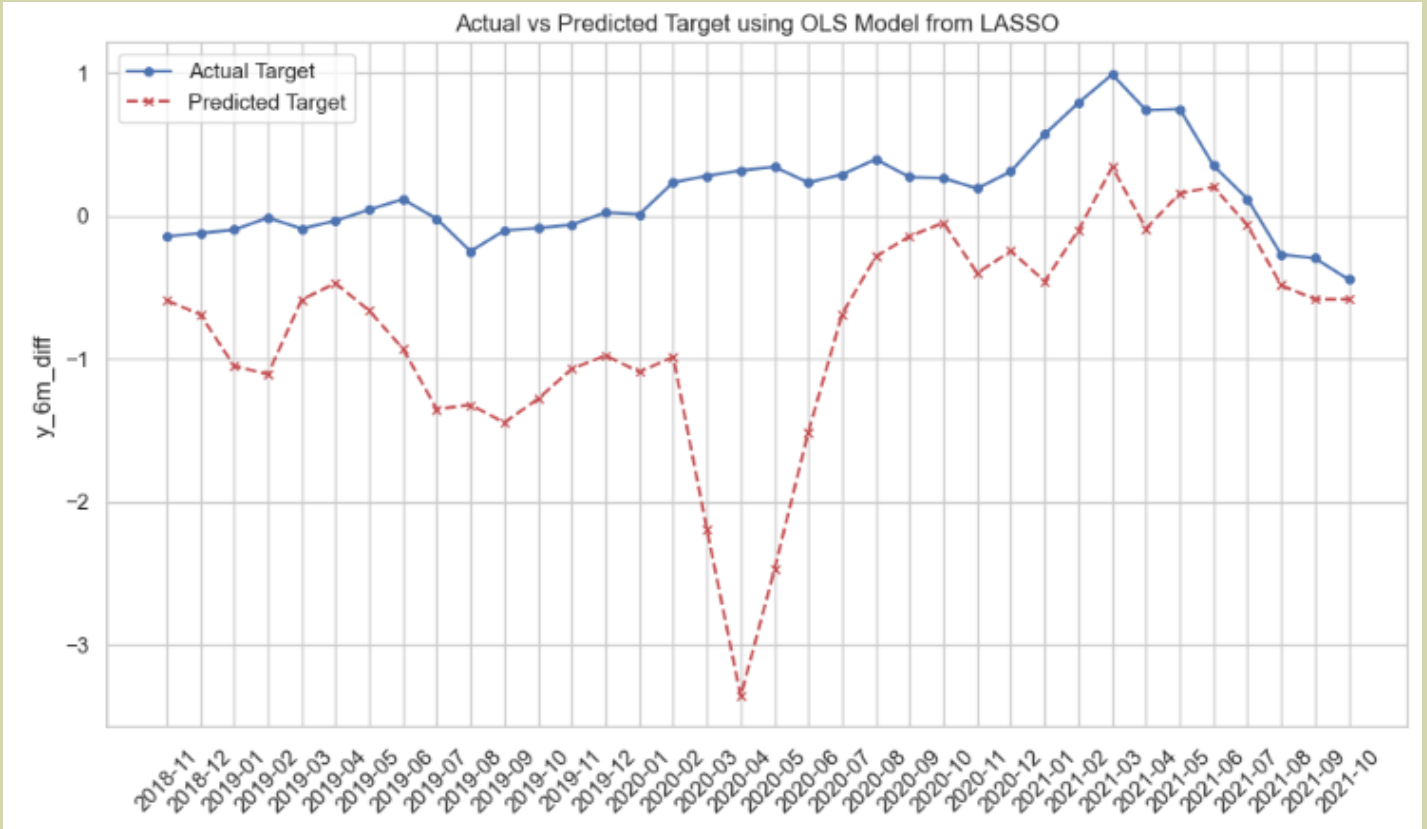
Mean Squared Error (MSE) on Train Set using OLS: 0.1228

4. Final Model

Compare results on both train set and test set of model from two techniques.

	Stepwise Regression	LASSO
Selected features	GRIFPEX Index (coef. = 10.94) ITPSSA Index (coef. = -2.34) ETSLTOTL Index (coef. = 5.06) NHSLTOT Index (coef. = -1.44)	GRIFPEX Index (coef. = 11.28) ITPSSA Index (coef. = -3.11) ETSLTOTL Index (coef. = 1.99)
MSE on train set	0.1086	0.1228
MSE on test set	2.2798	1.4429

4. Final Model_ Compare actual vs forecasted target



THANKS!