Preview Document: Monthly Country Industry Performance and Risk Forecast Model

Caleb Ong

Contents

1 Introduction

Global financial markets typically undergo various phases and cycles. This phenomenon is as a result of changing market and business environments.

The investment performance of market sectors and industries highly depend on the current business phase and current market conditions. Indications of market phases and investor sentiment are also reflected in macroeconomic indicators. Given these relationships, a sector risk and return forecasting model framework has been developed to provide monthly risk and return expectations of broad industries of multiple countries, by considering macroeconomic indicators and industry returns as the model inputs, and expected monthly returns as the model outputs.

- Model inputs:
  - Features: Macroeconomic indicators, industry returns
  - Target: Industry returns
- Model outputs:
  - Expected industry returns

2 Brief Summary

Industry forecasts are typically made at the beginning of the month, to form expectations on the industry return and risk characteristics for that month. Given that the release timings of macroeconomic data are usually asynchronous, the model has been structured to consider all the information set available at the time of the forecast; the model can be recalibrated during the month to incorporate the latest data, which could be released some time into the month, to update the industry return and risk expectations.

Time series of industry returns and macroeconomic indicators are extracted from data vendors. The data then undergoes cleaning, pre-processing, and feature engineering.

The data generally go through three passes:

- ARIMA
  - ARIMA model is used to fill in and forecast the expected value of the features, if the data had not been released yet, or is not available yet.
- Random Forest
  - Random Forest model is trained for each industry to identify and select the top indicators that have the most significant impact on industry returns.
- Xgboost
  - Xgboost model time series cross-validation is performed to obtain the model with the best parameters for model forecasting.

The finalized model is then tested on out-of-sample data to assess out-of-sample model performance for best model selection. Evaluation metrics are also gathered. Before making actual forecasts, the finalized model is recalibrated to now include out-of-sample data.

The models go through multiple runs to obtain simulated, empirical probability density estimates of return expectations.

3 Sample Outputs

This section showcases some examples of analytical outputs for the 11 GICs stock market sectors, as well as some investment style factors, for USA, for May 2023. 24 months of data are used for out-of-sample validation.

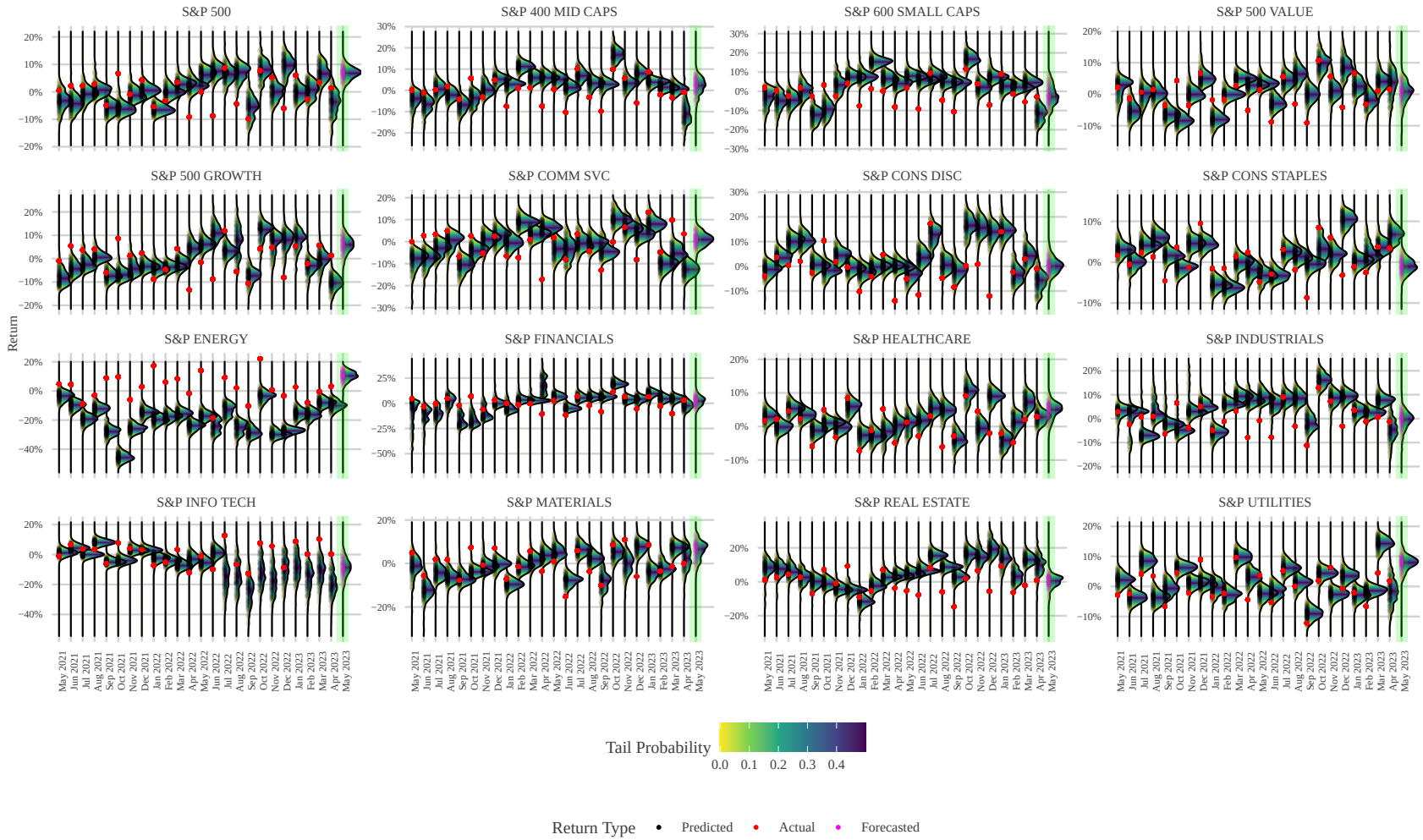

Figure 1: Overlay plot of out-of-sample predictive densities, overlaid with actual returns. The green shaded region highlights the estimated return densities, which were forecasted in the beginning of the month.

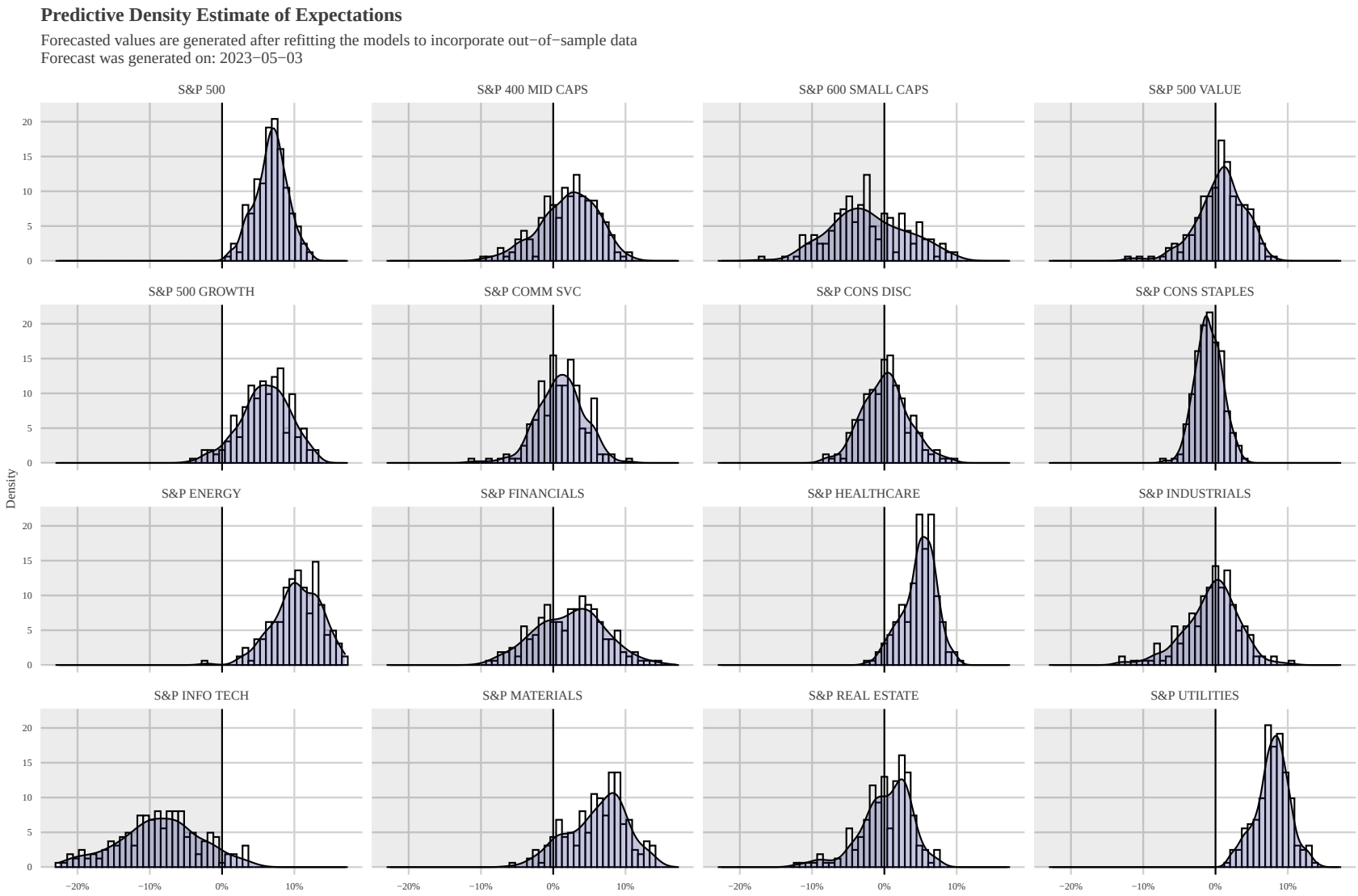

Figure 2: Estimated return densities for the month.

Table 1: Summary Statistics of Density Estimates

Industry	Average	Min	10% Quantile	25% Quantile	50% Quantile	75% Quantile	90% Quantile	Max
S&P 500	6.79%	0.918%	3.66%	5.26%	6.93%	8.18%	9.46%	12.3%
S&P 400 MID CAPS	2.11%	-9.68%	-3.52%	-0.433%	2.46%	5.06%	6.77%	10.7%
S&P 600 SMALL CAPS	-2.16%	-17%	-8.83%	-5.79%	-2.54%	1.86%	5.02%	9.37%
S&P 500 VALUE	0.71%	-12%	-3.43%	-1.13%	1.01%	2.91%	4.89%	8.11%
S&P 500 GROWTH	5.93%	-4.18%	1.41%	3.91%	6.13%	8.32%	10.1%	12.8%
S&P COMM SVC	1.02%	-11%	-2.81%	-1.08%	1.06%	3.08%	5.32%	10.4%
S&P CONS DISC	0.18%	-8.13%	-3.8%	-1.99%	0.154%	2.1%	4.33%	9.35%
S&P CONS STAPLES	-0.944%	-6.93%	-3.25%	-2.12%	-1.1%	0.366%	1.42%	4.05%
S&P ENERGY	10.4%	-2.08%	6.08%	8.49%	10.4%	13%	14.3%	17.1%
S&P FI-NANCIALS	2.37%	-9.2%	-3.84%	-1.09%	2.72%	5.41%	8.25%	14.7%
S&P HEALTH-CARE	4.85%	-2.34%	1.53%	3.57%	5.09%	6.53%	7.57%	10.1%
S&P INDUSTRI-ALS	-0.547%	-13%	-5.48%	-2.61%	-0.0765%	1.8%	3.79%	10.2%
S&P INFO TECH	-8.68%	-22.6%	-16.4%	-12.2%	-8.78%	-5.14%	-1.64%	3.48%
S&P MA-TERIALS	6.13%	-5.52%	0.464%	3.59%	6.74%	8.94%	10.6%	14.1%
S&P REAL ESTATE	0.227%	-12.3%	-4.36%	-1.65%	0.72%	2.78%	3.98%	7.36%
S&P UTILITIES	7.89%	1.75%	4.51%	6.66%	8.08%	9.38%	10.6%	14.1%

Table 2: Selected Out-of-sample Evaluation Metrics (using mean prediction for point estimate)

Industry	Proportion in Interquartile Range	Proportion between 5% to 95% Quantiles	Proportion in Predicted Range	RMSE	Number of Positive Predictions & Positive Returns	Number of Positive Predictions & Negative Returns	Number of Negative Predictions & Positive Returns	Number of Negative Predictions & Negative Returns	Proportion of Positive Predictions & Positive Returns	Proportion of Positive Predictions & Negative Returns	Proportion of Negative Predictions & Positive Returns	Proportion of Negative Predictions & Negative Returns	Accuracy	Precision	Sensitivity	Specificity
S&P 500	16.67%	37.5%	62.5%	0.0924	9	9	3	3	37.5%	37.5%	12.5%	12.5%	50%	50%	75%	25%
S&P 400 MID CAPS	20.83%	66.67%	79.17%	0.0707	8	6	6	4	33.33%	25%	25%	16.67%	50%	57.14%	57.14%	40%
S&P 600 SMALL CAPS	8.33%	41.67%	79.17%	0.0753	9	9	3	3	37.5%	37.5%	12.5%	12.5%	50%	50%	75%	25%
S&P 500 VALUE	12.5%	45.83%	70.83%	0.0921	5	5	8	6	20.83%	20.83%	33.33%	25%	45.83%	50%	38.46%	54.55%
S&P 500 GROWTH	29.17%	58.33%	70.83%	0.0554	12	5	1	6	50%	20.83%	4.17%	25%	75%	70.59%	92.31%	54.55%
S&P COMM SVC	25%	45.83%	62.5%	0.0946	10	5	1	8	41.67%	20.83%	4.17%	33.33%	75%	66.67%	90.91%	61.54%
S&P CONS DISC	16.67%	37.5%	66.67%	0.0482	9	7	3	5	37.5%	29.17%	12.5%	20.83%	58.33%	56.25%	75%	41.67%
S&P CONS STAPLES	0%	4.17%	16.67%	0.2400	0	0	15	9	0%	0%	62.5%	37.5%	37.5%	NaN%	0%	100%
S&P ENERGY	12.5%	45.83%	70.83%	0.1100	7	8	3	6	29.17%	33.33%	12.5%	25%	54.17%	46.67%	70%	42.86%
S&P FI-NANCIALS	20.83%	50%	75%	0.0468	10	5	3	6	41.67%	20.83%	12.5%	25%	66.67%	66.67%	76.92%	54.55%
S&P HEALTH-CARE	33.33%	45.83%	58.33%	0.0752	9	9	2	4	37.5%	37.5%	8.33%	16.67%	54.17%	50%	81.82%	30.77%
S&P INDUSTRI-ALS	25%	41.67%	70.83%	0.1250	5	2	9	8	20.83%	8.33%	37.5%	33.33%	54.17%	71.43%	35.71%	80%
S&P INFO TECH	29.17%	58.33%	79.17%	0.0679	4	5	7	8	16.67%	20.83%	29.17%	33.33%	50%	44.44%	36.36%	61.54%
S&P MA-TERIALS	8.33%	45.83%	58.33%	0.1070	10	9	2	3	41.67%	37.5%	8.33%	12.5%	54.17%	52.63%	83.33%	25%
S&P REAL ESTATE	8.33%	41.67%	62.5%	0.0948	5	5	7	7	20.83%	20.83%	29.17%	29.17%	50%	50%	41.67%	58.33%
S&P UTILITIES	20.83%	50%	62.5%	0.0491	8	4	4	8	33.33%	16.67%	16.67%	33.33%	66.67%	66.67%	66.67%	66.67%

Note:

For Accuracy, Precision, Sensitivity, and Specificity metrics, the binary categorical outcomes are defined as Positive Return, and Negative Return. A True Positive outcome is defined as an outcome with a Positive Prediction and a Positive Return

Table 3: Selected Out-of-sample Evaluation Metrics (using median prediction for point estimate)

Industry	Proportion in Interquartile Range	Proportion between 5% to 95% Quantiles	Proportion in Predicted Range	RMSE	Number of Positive Predictions & Positive Returns	Number of Positive Predictions & Negative Returns	Number of Negative Predictions & Positive Returns	Number of Negative Predictions & Negative Returns	Proportion of Positive Predictions & Positive Returns	Proportion of Positive Predictions & Negative Returns	Proportion of Negative Predictions & Positive Returns	Proportion of Negative Predictions & Negative Returns	Accuracy	Precision	Sensitivity	Specificity
S&P 500	16.67%	37.5%	62.5%	0.0918	9	9	3	3	37.5%	37.5%	12.5%	12.5%	50%	50%	75%	25%
S&P 400 MID CAPS	20.83%	66.67%	79.17%	0.0714	8	5	6	5	33.33%	20.83%	25%	20.83%	54.17%	61.54%	57.14%	50%
S&P 600 SMALL CAPS	8.33%	41.67%	79.17%	0.0753	9	9	3	3	37.5%	37.5%	12.5%	12.5%	50%	50%	75%	25%
S&P 500 VALUE	12.5%	45.83%	70.83%	0.0936	6	5	7	6	25%	20.83%	29.17%	25%	50%	54.55%	46.15%	54.55%
S&P 500 GROWTH	29.17%	58.33%	70.83%	0.0555	12	3	1	8	50%	12.5%	4.17%	33.33%	83.33%	80%	92.31%	72.73%
S&P COMM SVC	25%	45.83%	62.5%	0.0943	10	4	1	9	41.67%	16.67%	4.17%	37.5%	79.17%	71.43%	90.91%	69.23%
S&P CONS DISC	16.67%	37.5%	66.67%	0.0482	10	7	2	5	41.67%	29.17%	8.33%	20.83%	62.5%	58.82%	83.33%	41.67%
S&P CONS STAPLES	0%	4.17%	16.67%	0.2400	0	0	15	9	0%	0%	62.5%	37.5%	37.5%	NaN%	0%	100%
S&P ENERGY	12.5%	45.83%	70.83%	0.1100	6	8	4	6	25%	33.33%	16.67%	25%	50%	42.86%	60%	42.86%
S&P FI-NANCIALS	20.83%	50%	75%	0.0470	10	5	3	6	41.67%	20.83%	12.5%	25%	66.67%	66.67%	76.92%	54.55%
S&P HEALTH-CARE	33.33%	45.83%	58.33%	0.0756	9	9	2	4	37.5%	37.5%	8.33%	16.67%	54.17%	50%	81.82%	30.77%
S&P INDUSTRI-ALS	25%	41.67%	70.83%	0.1250	5	2	9	8	20.83%	8.33%	37.5%	33.33%	54.17%	71.43%	35.71%	80%
S&P INFO TECH	29.17%	58.33%	79.17%	0.0678	5	5	6	8	20.83%	20.83%	25%	33.33%	54.17%	50%	45.45%	61.54%
S&P MA-TERIALS	8.33%	45.83%	58.33%	0.1070	10	9	2	3	41.67%	37.5%	8.33%	12.5%	54.17%	52.63%	83.33%	25%
S&P REAL ESTATE	8.33%	41.67%	62.5%	0.0938	5	5	7	7	20.83%	20.83%	29.17%	29.17%	50%	50%	41.67%	58.33%
S&P UTILITIES	20.83%	50%	62.5%	0.0491	8	4	4	8	33.33%	16.67%	16.67%	33.33%	66.67%	66.67%	66.67%	66.67%

Note:

For Accuracy, Precision, Sensitivity, and Specificity metrics, the binary categorical outcomes are defined as Positive Return, and Negative Return. A True Positive outcome is defined as an outcome with a Positive Prediction and a Positive Return

4 Ending Notes

The evaluation metrics provide feedback on the degree of effectiveness for each model for each industry; there is certainly room for improvement in every step of the modelling process. Risk managers may be more concerned about the effectiveness of downside expectations (Specificity), while portfolio managers may be more concerned about the effectiveness of accurate predictions (Accuracy) and upside expectations (Sensitivity). The distribution shape can also serve as an indication of uncertainty of expectations.