

Essays on Macroeconomic Forecasting with Machine Learning and Leading
Indicators

by

Lucas Goldberg Da Rosa

B.S., Kansas State University, 2019

M.A., Kansas State University, 2025

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

This dissertation consists of three essays on forecasting macroeconomic time series with machine learning techniques, evaluating the forecast performance of the Organization for Economic Co-operation and Development (OECD) Composite Leading Indicators (CLIs) and proposing a new index based on the latest Artificial Intelligence algorithms. The first essay investigates if the CLI improves forecasting performance under different conditions and model assumptions. We use a variety of time series and machine learning models to evaluate if incorporating the CLI into the forecast reduces the errors for inflation, unemployment rate and industrial production. We hypothesize that the CLIs of high income, highly developed nations will provide more accurate forecasts than those of developing countries, due to the more complete data those countries have from being able to pour more resources into data collection, analysis and storage.

In the second essay we analyze how each country's CLI is constructed and what components are used for its calculation. We use machine learning algorithms with feature selection to evaluate if the given weights in the construction of a country's CLI is optimal. If discovered that it is not, we use the weights found to test if the newly weighted index provides better forecasts than the current OECD CLI.

The third essay expands on the second. We propose a new index, similar to the CLI, but constructed based on using machine learning algorithms to find the best combination of economics variables available in the OECD database that explain variations in future values, thus providing the most accurate forecasts possible. Additionally, we can determine the components that most influence the different target variables- inflation rates, unemployment rates and industrial production for each country in our analysis. We can then use forecast encompassing tests and compare forecast errors from this new composite index to the CLIs to show relative improvement in the predictions.

Essays on Macroeconomic Forecasting with Machine Learning and Leading
Indicators

by

Lucas Goldberg Da Rosa

B.S., Kansas State University, 2019

M.A., Kansas State University, 2025

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Lance Bachmeier

Copyright

© Lucas Goldberg Da Rosa 2026.

Abstract

This dissertation consists of three essays on forecasting macroeconomic time series with machine learning techniques, evaluating the forecast performance of the Organization for Economic Co-operation and Development (OECD) Composite Leading Indicators (CLIs) and proposing a new index based on the latest Artificial Intelligence algorithms. The first essay investigates if the CLI improves forecasting performance under different conditions and model assumptions. We use a variety of time series and machine learning models to evaluate if incorporating the CLI into the forecast reduces the errors for inflation, unemployment rate and industrial production. We hypothesize that the CLIs of high income, highly developed nations will provide more accurate forecasts than those of developing countries, due to the more complete data those countries have from being able to pour more resources into data collection, analysis and storage.

In the second essay we analyze how each country's CLI is constructed and what components are used for its calculation. We use machine learning algorithms with feature selection to evaluate if the given weights in the construction of a country's CLI is optimal. If discovered that it is not, we use the weights found to test if the newly weighted index provides better forecasts than the current OECD CLI.

The third essay expands on the second. We propose a new index, similar to the CLI, but constructed based on using machine learning algorithms to find the best combination of economics variables available in the OECD database that explain variations in future values, thus providing the most accurate forecasts possible. Additionally, we can determine the components that most influence the different target variables- inflation rates, unemployment rates and industrial production for each country in our analysis. We can then use forecast encompassing tests and compare forecast errors from this new composite index to the CLIs to show relative improvement in the predictions.

Table of Contents

List of Figures	ix
List of Tables	xxii
Acknowledgements	xxv
Dedication	xxvi
1 Evaluating the OECD’s Composite Leading Indicators in Forecasting Economic Conditions	1
1.1 Introduction	1
1.2 Data	4
1.3 Methodology	4
1.3.1 Models	4
1.3.2 Hyperparameter Tuning	10
1.3.3 Forecast Evaluations	11
1.4 Results	13
1.4.1 Forecast accuracy and the value of the CLI	13
1.4.2 Error Diagnostics and Practical Interpretation	17
1.4.3 Robustness and Sensitivity Analysis	21
2 Are OECD CLI Weights Optimal? A Feature Selection Approach	25
2.1 Introduction	25
2.2 Data	27
2.3 Methodology	28

2.3.1	Step 1: Baseline CLI Forecasting	29
2.3.2	Step 2: Feature Selection and Weight Estimation	30
2.3.3	Step 3: Forecasting Comparison	34
2.3.4	Step 4: Cross-Country Generalization	35
2.4	Robustness and Sensitivity Analysis	35
2.5	Results	37
2.5.1	Interpretation and implications for forecasting practice	55
3	Building a New Composite Leading indicator with Machine Learning	57
3.1	Introduction	57
3.2	Data and Variable Selection	59
3.3	Methodology	60
3.3.1	Indicator Construction	61
3.3.2	Forecasting and Model Comparison	63
3.3.3	Forecast Encompassing Test	64
3.4	Results and Discussion	64
3.4.1	Forecast Accuracy Across Models and Horizons	69
3.4.2	Horizon Profiles	73
3.4.3	Model Ranking	73
3.4.4	Statistical Significance	73
3.4.5	Robustness and Sensitivity	73
3.4.6	Discussion and Implications	74
3.4.7	Limitations	74
3.4.8	Conclusion of Results	75
	Bibliography	77
A	CLI Compositions by Country	83

B	pyOECD	88
C	Results by Country-Variable	100
C.1	Additional Results for Chapter 1	100
C.2	Additional Forecast Accuracy Results for Chapter 2	119
C.3	Additional Results for Chapter 3	156

List of Figures

1.1	United States unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	18
1.2	Turkey Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	19
1.3	Italy Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	20
1.4	Turkey CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	21
2.1	United States CPI inflation: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.	48
2.2	United States Industrial Production: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.	49
2.3	United States Unemployment Rate: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.	50
2.4	USA Forecast Errors	54
3.1	Forecast error comparison for a representative case in which the MLCI performance is mixed relative to the OECD CLI. Negative values in the lower panel indicate periods in which the OECD CLI yields lower squared forecast error than the MLCI.	66

3.2	Forecast error comparison for a representative case in which the MLCI underperforms the OECD CLI. Negative values in the lower panel indicate periods in which the OECD CLI yields lower squared forecast error than the MLCI.	67
3.3	Forecast error comparison for a representative case in which the MLCI outperforms the OECD CLI. The lower panel plots the difference in squared forecast errors, so positive values indicate periods in which the MLCI improves forecast accuracy.	68
C.1	Chapter 1 - Canada CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	101
C.2	Chapter 1 - Canada Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	102
C.3	Chapter 1 - Canada Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	102
C.4	Chapter 1 - Spain CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	103
C.5	Chapter 1 - Spain Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	104
C.6	Chapter 1 - Spain Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	104
C.7	Chapter 1 - Great Britain CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	105
C.8	Chapter 1 - Great Britain Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	106
C.9	Chapter 1 - Great Britain Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	106
C.10	Chapter 1 - Italy CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	107

C.11 Chapter 1 - Italy Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	108
C.12 Chapter 1 - Italy Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	108
C.13 Chapter 1 - Japan CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	109
C.14 Chapter 1 - Japan Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	110
C.15 Chapter 1 - Japan Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	110
C.16 Chapter 1 - South Korea CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	111
C.17 Chapter 1 - South Korea Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	112
C.18 Chapter 1 - South Korea Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	112
C.19 Chapter 1 - Mexico CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	113
C.20 Chapter 1 - Mexico Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	114
C.21 Chapter 1 - Mexico Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	114
C.22 Chapter 1 - Turkey CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	115
C.23 Chapter 1 - Turkey Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.	116

C.24 Chapter 1 - Turkey Industrial Production: forecast errors (two-panel) by model.	
Top: Baseline. Bottom: With CLI.	116
C.25 Chapter 1 - United States CPI Inflation: forecast errors (two-panel) by model.	
Top: Baseline. Bottom: With CLI.	117
C.26 Chapter 1 - United States Unemployment: forecast errors (two-panel) by	
model. Top: Baseline. Bottom: With CLI.	118
C.27 Chapter 1 - United States Industrial Production: forecast errors (two-panel)	
by model. Top: Baseline. Bottom: With CLI.	118
C.28 Chapter 2 - Forecast accuracy results for Canada CPI inflation model at	
the six-month forecast horizon. The figure reports out-of-sample root mean	
squared forecast error (rMSFE) values across ARIMA, LASSO, random forest	
(RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using	
the machine-learning-reweighted CLI. Lower rMSFE values indicate more	
accurate forecasts. These results correspond to the model comparison procedure	
described in Section 2.3 and the robustness discussion in Section 2.4.	121
C.29 Chapter 2 - Forecast accuracy results for Canada Industrial Production model	
at the six-month forecast horizon. The figure reports out-of-sample root	
mean squared forecast error (rMSFE) values across ARIMA, LASSO, random	
forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model	
using the machine-learning-reweighted CLI. Lower rMSFE values indicate	
more accurate forecasts. These results correspond to the model comparison	
procedure described in Section 2.3 and the robustness discussion in Section 2.4.	122

- C.30 Chapter 2 - Forecast accuracy results for Canada Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [123](#)
- C.31 Chapter 2 - Forecast accuracy results for Spain CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. . . . [125](#)
- C.32 Chapter 2 - Forecast accuracy results for Spain Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [126](#)

- C.33 Chapter 2 - Forecast accuracy results for Spain Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [127](#)
- C.34 Chapter 2 - Forecast accuracy results for Great Britain CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [129](#)
- C.35 Chapter 2 - Forecast accuracy results for Great Britain Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [130](#)

- C.36 Chapter 2 - Forecast accuracy results for Great Britain Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [131](#)
- C.37 Chapter 2 - Forecast accuracy results for Italy CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. . . . [133](#)
- C.38 Chapter 2 - Forecast accuracy results for Italy Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [134](#)

- C.39 Chapter 2 - Forecast accuracy results for Italy Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [135](#)
- C.40 Chapter 2 - Forecast accuracy results for Japan CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. . . . [137](#)
- C.41 Chapter 2 - Forecast accuracy results for Japan Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [138](#)

- C.42 Chapter 2 - Forecast accuracy results for Japan Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [139](#)
- C.43 Chapter 2 - Forecast accuracy results for South Korea CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [141](#)
- C.44 Chapter 2 - Forecast accuracy results for South Korea Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [142](#)

- C.45 Chapter 2 - Forecast accuracy results for South Korea Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [143](#)
- C.46 Chapter 2 - Forecast accuracy results for Mexico CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. . . . [145](#)
- C.47 Chapter 2 - Forecast accuracy results for Mexico Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [146](#)

- C.48 Chapter 2 - Forecast accuracy results for Mexico Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [147](#)
- C.49 Chapter 2 - Forecast accuracy results for Turkey CPI Inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. . . . [149](#)
- C.50 Chapter 2 - Forecast accuracy results for Turkey Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [150](#)

- C.51 Chapter 2 - Forecast accuracy results for Turkey Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [151](#)
- C.52 Chapter 2 - Forecast accuracy results for the United States CPI Inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [153](#)
- C.53 Chapter 2 - Forecast accuracy results for the United States Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [154](#)

C.54 Chapter 2 - Forecast accuracy results for the United States Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4. [155](#)

List of Tables

1.1	Relative rMSFE Improvement ($h = 6$ months) for CPI Inflation (CLI vs. baseline)	14
1.2	Relative rMSFE Improvement ($h = 6$ months) for Industrial Production (CLI vs. baseline)	15
1.3	Relative rMSFE Improvement ($h = 6$ months) for Unemployment Rate (CLI vs. baseline)	16
2.1	Machine-learning-derived CLI component weights for Canada at the six-month forecast horizon.	39
2.2	Machine-learning-derived CLI component weights for Spain at the six-month forecast horizon.	40
2.3	New Weights for Great Britain	41
2.4	Machine-learning-derived CLI component weights for Italy at the six-month forecast horizon.	42
2.5	Machine-learning-derived CLI component weights for Japan at the six-month forecast horizon.	43
2.6	Machine-learning-derived CLI component weights for South Korea at the six-month forecast horizon.	44
2.7	Machine-learning-derived CLI component weights for Mexico at the six-month forecast horizon.	45
2.8	Machine-learning-derived CLI component weights for Turkey at the six-month forecast horizon.	46

2.9	Machine-learning-derived CLI component weights for the United States at the six-month forecast horizon.	47
2.10	Diebold–Mariano tests for USA (DM statistic with p-value)	51
3.1	United States forecast comparison at horizon $h = 6$: CLI versus MLCI . . .	65
3.2	Summary of MLCI performance relative to the OECD CLI across countries .	70
3.3	Share of significant DM tests by comparison, target, and horizon	72
A.1	OECD Composite Leading Indicator Components for Canada	83
A.2	OECD Composite Leading Indicator Components for Italy	84
A.3	OECD Composite Leading Indicator Components for Japan	84
A.4	OECD Composite Leading Indicator Components for Korea	85
A.5	OECD Composite Leading Indicator Components for Mexico	85
A.6	OECD Composite Leading Indicator Components for Spain	86
A.7	OECD Composite Leading Indicator Components for Turkey	86
A.8	OECD Composite Leading Indicator Components for United Kingdom . . .	87
A.9	OECD Composite Leading Indicator Components for Unites States of America	87
C.1	Diebold–Mariano tests for CAN (DM statistic with p-value)	120
C.2	Diebold–Mariano tests for ESP (DM statistic with p-value)	124
C.3	Diebold–Mariano tests for GBR (DM statistic with p-value)	128
C.4	Diebold–Mariano tests for ITA (DM statistic with p-value)	132
C.5	Diebold–Mariano tests for JPN (DM statistic with p-value)	136
C.6	Diebold–Mariano tests for KOR (DM statistic with p-value)	140
C.7	Diebold–Mariano tests for MEX (DM statistic with p-value)	144
C.8	Diebold–Mariano tests for TUR (DM statistic with p-value)	148
C.9	Diebold–Mariano tests for USA (DM statistic with p-value)	152
C.10	Chapter 3 - Forecast comparison across countries, targets, and horizons: base- line AR, OECD CLI, and MLCI.	157

C.11 Canada chapter 3 weights by measure	161
C.12 Spain chapter 3 weights by measure	161
C.13 Great Britain chapter 3 weights by measure	162
C.14 Italy chapter 3 weights by measure	162
C.15 Japan chapter 3 weights by measure	163
C.16 South Korea chapter 3 weights by measure	163
C.17 Mexico chapter 3 weights by measure	164
C.18 Turkey chapter 3 weights by measure	164
C.19 USA chapter 3 weights by measure	165

Acknowledgments

I want to acknowledge my committee for helping me get across the finish line, especially my advisor, Dr. Lance Bachmeier. I also want to recognize Dr. Ryan Staley for getting me unstuck and all my family for their support through the years.

Dedication

To my wife.

Chapter 1

Evaluating the OECD’s Composite Leading Indicators in Forecasting Economic Conditions

1.1 Introduction

Leading indicators have long played a central role in forecasting, dating back to the pioneering work of the National Bureau of Economic Research and scholars like ([Burns and Mitchell, 1946](#); [Goldin, 2001](#); [Klein and Moore, 1983](#)). Composite indexes such as the OECD’s CLI aggregate a variety of economic series— new orders, confidence surveys, interest spreads— into a single signal designed to precede turning points in economic activity. These indicators have been useful to practitioners for decades, but their optimality is increasingly questioned in an era of high-frequency data and algorithmic models.

Simultaneously, machine learning has gained traction in economics for its ability to handle nonlinearities, large predictor sets, and dynamic interactions. Tree-based models, penalized regression, and ensemble learning methods have been shown to outperform traditional time series models in predictive accuracy for some variables in particular time periods, as shown by ([Araujo and Gaglianone, 2020](#); [Bandara et al., 2020](#); [Medeiros et al., 2021](#); [Messaoudi and](#)

[Khouidmi, 2024](#); [Mirza et al., 2024](#); [Naghi et al., 2024](#); [Smalter Hall, 2018](#)). This dissertation bridges these two approaches by examining the value added of leading indicators in modern forecasting models, and by assessing whether new combinations of data and machine learning can outperform existing benchmarks.

The pursuit of improved economic forecasting methods has spanned decades, with a rich literature that balances theoretical rigor, empirical performance, and practical usability. Traditional approaches have often relied on linear models such as ARIMA ([Box et al., 2015](#)), vector autoregressions ([Sims, 1980](#)), and dynamic factor models ([Stock and Watson, 2002a](#)), which assume stable relationships between a small number of economic indicators and the variables of interest. Building on these foundational models the economic forecasting literature advanced to incorporate nonlinear, nonparametric, and semiparametric models, which are able to reliably forecast under different conditions, assumptions, and relationships. ([Box and Jenkins, 1970](#); [Engle, 1982](#); [Fan and Gijbels, 1996](#); [Granger and Newbold, 1986](#); [Hamilton, 1989](#); [Pagan and Ullah, 1999](#); [Stock and Watson, 1996](#)).

An early forecasting breakthrough was the development of leading indicators, including composite indexes such as those constructed by the Conference Board and the OECD. The OECD’s Composite Leading Indicators (CLIs), in particular, aggregate information from new orders, consumer and business sentiment surveys, stock prices, and yield curves to provide early warnings of economic turning points ([Gyomai and Guidetti, 2012](#)). These leading indicators benefited from the extensive economic theory and literature, as the selection of variables and components were heavily influenced by the time series econometricians at the time. This was particularly useful when computational power was low, and models could not handle many variables and large amounts of data were scarce. These indexes remain widely used by policymakers and analysts for tracking cyclical developments, though their empirical reliability has shown variation across countries and over time ([Camacho and Palmieri, 2021](#)).

In recent years, there has been growing interest in machine learning (ML) methods for forecasting macroeconomic variables. These models—including LASSO regression ([Tibshirani, 1996](#)), random forests ([Breiman, 2001](#)), and boosting algorithms such as XGBoost ([Chen and Guestrin, 2016](#)) and LightGBM ([Ke et al., 2017](#)) — offer several advantages: they can

handle large sets of potential predictors, accommodate nonlinear relationships, and adapt to complex interactions between variables. (Medeiros et al., 2021) and (Bandara et al., 2020) demonstrate the superiority of these approaches in various forecasting settings, especially under conditions of structural change or high volatility.

Despite these methodological advances, the role of traditional indicators like the CLI in modern forecasting frameworks remains underexplored. Some studies, such as (Lahiri and Monokroussos, 2013), suggest that leading indicators can still improve forecast accuracy when combined with high-dimensional models. Others caution that their fixed structure and reliance on historical relationships may limit their usefulness in more flexible or rapidly evolving environments. Furthermore, few studies have benchmarked CLI-augmented forecasts against those generated by modern ML techniques across a consistent set of macroeconomic targets.

A key motivation for composite leading indicators is that they represent a *theory-guided* and institutionally curated information set. Rather than searching mechanically across an enormous predictor pool, the CLI restricts attention to series with economic justification for leading behavior (e.g., orders, survey expectations, financial spreads). This design reduces the risk that forecast improvements are driven purely by chance correlations in high-dimensional data and increases interpretability for policy users. Thus, even when flexible ML methods are considered, the CLI provides a disciplined benchmark that is meaningful for real-world forecasting practice.

This chapter fills this gap by evaluating whether the OECD CLI improves forecast accuracy for inflation, unemployment, and industrial production, relative to models excluding the CLI and its components, comparing results across nine countries with varying income levels and data quality, as well as testing CLI-augmented models against benchmark models, including ML approaches.

1.2 Data

The analysis includes monthly data for nine countries from 2000 to 2025, to ensure consistency, guaranteeing that all countries have all the target variables. The CLI and macroeconomic indicators are sourced from the OECD and based on respective national statistics offices. Target variables include: Monthly Consumer Price Index (CPI), seasonally adjusted; transformed as log differences to represent monthly inflation; Unemployment Rate seasonally adjusted; used in levels and tested for structural breaks (e.g., COVID-19 shocks); Industrial Production Index (IPI) seasonally adjusted, log-differenced. Structural breaks (e.g., 2008–2009 Great Recession, 2020 COVID-19 shock) will be accounted for using rolling forecast windows and robustness checks.

1.3 Methodology

Models are estimated with and without the CLI to assess its marginal contribution. Forecasting techniques include ARIMA, LASSO regression, Random Forest, LightGBM, and XGBoost. Out-of-sample accuracy is evaluated using Root Mean Squared Forecast Error (rMSFE) and the Diebold-Mariano test.

1.3.1 Models

ARIMA(p, d, q)

The Autoregressive Integrated Moving Average (ARIMA) model, developed by [Box et al., 2015](#), is a foundational time series forecasting method widely used in macroeconomics. ARIMA models combine three components—autoregression (AR), differencing (I), and moving average (MA)—to capture patterns in univariate time series such as inflation, unemployment, and industrial production.

An ARIMA(p, d, q) model is defined as:

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t$$

where y_t is the target variable at time t , L is the lag operator (i.e., $Ly_t = y_{t-1}$), $(1 - L)^d$ denotes differencing to achieve stationarity, $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ is the autoregressive (AR) polynomial of order p , $\theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$ is the moving average (MA) polynomial of order q , ε_t is a white noise error term.

Each component serves a different function. The **AR** part captures momentum and persistence in the time series. The **I** component addresses non-stationarity by differencing the series d times. The **MA** part models short-run shock propagation and correction.

In this dissertation, ARIMA models are used for unit root tests (e.g., Augmented Dickey-Fuller) to determine the degree of differencing d , information criteria (AIC, BIC) for selecting lag orders p and q out-of-sample forecast evaluation to compare with machine learning models.

ARIMA provides strong short-term forecasts for stationary or differenced series, are well-established in central banking and statistical agencies as a baseline model and is easily interpretable. However, ARIMA assumes linear dynamics and struggles with structural breaks (e.g., Great Recession, COVID-19), as well as can be sensitive to overfitting if lag orders are not carefully chosen. Also, it does not naturally accommodate exogenous predictors—this is addressed in ARIMAX.

To evaluate the marginal value of the Composite Leading Indicator (CLI), we also estimate ARIMAX models, which include the CLI as an exogenous regressor:

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t + \beta' x_t$$

where x_t represents the CLI at time t .

ARIMA and ARIMAX serve as benchmarks in this dissertation, allowing us to assess the value added by machine learning models and composite indicators in a traditional time series forecasting framework.

LASSO Regression

The Least Absolute Shrinkage and Selection Operator (LASSO), introduced by (Tibshirani, 1996), is a linear regression technique that simultaneously performs variable selection and

regularization. It is particularly valuable in forecasting applications with many potential predictors, such as lagged variables and exogenous indicators like the CLI, where multicollinearity and overfitting are concerns.

LASSO minimizes the sum of squared residuals subject to a constraint on the sum of the absolute values of the coefficients:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

where y_i is the response variable (e.g., inflation, unemployment), X_i is the vector of predictors (e.g., lagged values, CLI), β is the coefficient vector, $\lambda \geq 0$ is the regularization parameter that controls the strength of the penalty.

As λ increases, more coefficients are shrunk toward zero, and some are set exactly to zero. This leads to a sparse model that retains only the most relevant predictors. LASSO is thus highly interpretable and computationally efficient.

In the context of macroeconomic forecasting, LASSO offers several advantages. First, it is particularly useful when the number of potential predictors exceeds the number of observations, or when incorporating many lags or auxiliary series like CLI components. Macroeconomic shock effects might take a long time to show up in the variables, thus estimation with very long lags will be inefficient and the forecast performance will be poor. Penalized regression like LASSO provides an automated and disciplined way to select a sparse lag structure by shrinking weak coefficients to zero, balancing flexibility with parsimony ([Tibshirani, 1996](#)). Second, regularization reduces the risk of including spurious relationships in the forecast model, which prevents overfitting. Lastly, its simplicity and adaptability make it applicable across countries with different data qualities and series lengths, making it robust across countries.

Random Forest (RF)

Random Forest (RF) is a robust ensemble learning algorithm developed by ([Breiman, 2001](#)) that builds a large collection of decision trees and aggregates their outputs to im-

prove prediction accuracy and reduce overfitting. It is particularly effective for forecasting in high-dimensional settings due to its ability to handle nonlinear relationships and interactions without requiring prior variable transformations or functional form assumptions. (Goulet Coulombe et al., 2022; Masini et al., 2023; Mullainathan and Spiess, 2017) Empirical evidence also suggests that tree-based methods, including random forests, can improve inflation forecasts relative to standard benchmarks in certain settings (Medeiros et al., 2021; Naghi et al., 2024).

Each tree in the ensemble is built using a bootstrap sample of the training data, and at each node split only a random subset of predictors is considered. This produces two distinct sources of randomness: **bagging** (bootstrap aggregation), where each tree is trained on a resampled dataset, and **feature subsampling**, where splits are chosen from only k out of p predictors (often $k = \sqrt{p}$ for regression). These design choices are central to why Random Forests can perform well in macroeconomic forecasting. First, bagging reduces variance by averaging many high-variance decision trees, which is valuable in macroeconomic settings where relationships are unstable and samples are limited. Second, feature subsampling decorrelates trees and helps manage highly correlated predictors—common in macroeconomic indicator sets—so the ensemble does not repeatedly split on the same dominant series.

From an economic forecasting perspective, Random Forests are attractive because they can capture nonlinearities, threshold effects, and interactions across indicators without requiring the researcher to pre-specify functional forms or interaction terms (Breiman, 2001). This flexibility is particularly relevant in “data-rich” environments where many indicators may contain predictive information but the mapping from indicators to outcomes may change across the business cycle (Bernanke and Boivin, 2003; Stock and Watson, 1996). Recent forecasting research emphasizes that the ability of nonlinear methods to exploit such features is an important driver of machine learning gains over standard macroeconometric benchmarks (Goulet Coulombe et al., 2022; Masini et al., 2023). In addition, empirical evidence specific to macroeconomic outcomes finds that Random Forests can outperform traditional benchmarks for inflation forecasting in some settings (Medeiros et al., 2021; Naghi et al., 2024). For these reasons, Random Forests provide a natural nonlinear benchmark in our cross-country

evaluation alongside linear regularized methods (LASSO) and boosted trees.

Given input features x and target variable y , the Random Forest prediction $\hat{y}$ is the average of predictions from B individual trees:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where $T_b(x)$ is the prediction of the b -th decision tree trained on a bootstrap sample.

Random Forest has several properties that make it well-suited for macroeconomic forecasting. It handles nonlinearity, by effectively capturing complex interactions between lagged macroeconomic indicators and target variables. Additionally, it is robust to overfitting, since tree averaging reduces variance, especially in noisy economic time series. (Nyman and Ormerod, 2017) Its internal feature ranking can identify which variables (e.g., CLI vs. autoregressive lags) contribute most to forecast accuracy. Lastly, unlike ARIMA models, RF does not assume stationarity, though careful preprocessing is still beneficial, but not a necessary condition.

Gradient Boosting (XGBoost, LightGBM)

Gradient boosting is a powerful ensemble learning technique that builds models sequentially, where each new model is trained to correct the residual errors of the previous model. It does this by minimizing a differentiable loss function using gradient descent. Formally, given a training set $\{(x_i, y_i)\}_{i=1}^n$, the objective is to learn a function $F(x)$ that minimizes the expected loss $\mathcal{L}(y, F(x))$. From a macroeconomic forecasting perspective, boosted trees are attractive for three reasons. **First**, they flexibly capture nonlinearities and interactions among predictors (e.g., the effect of financial conditions may differ depending on the level of inflation or the phase of the business cycle) without requiring the researcher to pre-specify thresholds or interaction terms. This matters in macroeconomic environments where relationships can shift over time and across regimes (Stock and Watson, 1996). **Second**, boosting provides built-in regularization mechanisms—such as shrinkage (small learning rates), subsampling, and early stopping—that help balance fit and generalization, which is critical when predictor sets are large and potentially noisy (a common feature of “data-rich” forecasting problems)

(Bernanke and Boivin, 2003). **Third**, sequential error-correction means that boosting can approximate complex predictive functions by gradually fitting patterns left unexplained by earlier trees, which often yields strong out-of-sample performance in practice.

These motivations align with the broader ML forecasting literature in economics, which emphasizes that nonlinear methods can produce gains over standard macroeconomic benchmarks particularly when nonlinearities are present and uncertainty is elevated (Goulet Coulombe et al., 2022; Masini et al., 2023). Empirically, studies of inflation forecasting in data-rich environments find that modern ML methods can deliver meaningful improvements relative to traditional models (Araujo and Gaglianone, 2020; Medeiros et al., 2021; Mirza et al., 2024; Naghi et al., 2024). Therefore, gradient boosting provides a natural state-of-the-art nonlinear benchmark for evaluating whether incorporating leading information (such as CLI inputs) can improve forecast accuracy across countries, targets, and horizons.

Gradient boosting builds the model in stages:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n \mathcal{L}(y_i, \gamma)$$

$$F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x), \quad m = 1, 2, \dots, M$$

where $h_m(x)$ is the base learner (typically a regression tree), ν is the learning rate ($0 < \nu < 1$) and M is the number of boosting iterations. At each step, $h_m(x)$ is fit to the negative gradient of the loss function with respect to the current model:

$$r_{im} = - \left[\frac{\partial \mathcal{L}(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

This iterative approach makes gradient boosting highly effective for capturing nonlinear patterns and interactions in macroeconomic forecasting.

In this study, we use two implementations:

XGBoost (Chen and Guestrin, 2016): Introduces regularization terms to the objective to control model complexity, shrinkage to reduce overfitting, and column subsampling for

additional robustness. **LightGBM** (Ke et al., 2017): Optimized for efficiency on large datasets via histogram-based splitting and a leaf-wise growth strategy, which enables deeper trees and faster convergence.

Hyperparameters such as maximum tree depth, learning rate, number of estimators, and minimum samples per leaf are tuned using time-series cross-validation. Both models are evaluated for forecast accuracy and feature importance, especially when incorporating or excluding the OECD CLI as a predictor.

1.3.2 Hyperparameter Tuning

Machine learning models require hyperparameter choices that control complexity (e.g., tree depth, number of trees, shrinkage, and subsampling rates). These choices strongly affect the bias–variance tradeoff and therefore out-of-sample performance. To avoid look-ahead bias, hyperparameters are selected using time-series cross-validation that preserves temporal ordering (training uses only information available prior to the validation period). The tuning objective matches the evaluation criterion (rMSFE), and tuning is conducted on the training sample only; the final test period is held out and never used for model selection. We apply the same rolling/expanding window design and validation splits across models to ensure an apples-to-apples comparison. Hyperparameter grids are chosen to reflect standard regularization practices (small learning rates with early stopping for boosting; bounded tree depth/leaves and minimum leaf sizes for tree ensembles; log-spaced penalty paths for LASSO), which reduces overfitting risk in monthly macroeconomic samples. (Goulet Coulombe et al., 2022) and (Masini et al., 2023) find consistent evidence that nonlinearities and regularization are central drivers of ML forecasting gains in macroeconomic settings.

ARIMA/ARIMAX: p, d, q via AIC over a capped grid, with drift allowed. Exogenous set: CLI levels and/or filtered versions.

LASSO: λ via rolling blocked CV on a logarithmic grid; coefficients standardized; we report stability paths over time.

RF: number of trees (B), m_{try} , and min leaf size tuned by blocked CV; OOB error is reported

but not used for selection to preserve time ordering.

XGB/LightGBM: learning rate, depth/leaves, number of estimators, subsampling/colsample and regularization terms tuned by blocked CV with early stopping.

1.3.3 Forecast Evaluations

Root Mean Squared Forecast Error (rMSFE)

The Root Mean Squared Forecast Error (rMSFE) is a standard metric used to evaluate and compare the accuracy of forecasts in time series analysis. It measures the average magnitude of forecast errors, penalizing larger errors more heavily than smaller ones due to squaring. This makes rMSFE particularly useful in economic forecasting where large errors (e.g., during recessions) are especially costly or impactful.

Given a series of h forecasted values $\hat{y}_{t+1}, \dots, \hat{y}_{t+h}$ and corresponding observed values $y_{t+1}, \dots, y_{t+h}$, the rMSFE is defined as:

$$\text{rMSFE} = \sqrt{\frac{1}{h} \sum_{i=1}^h (\hat{y}_{t+i} - y_{t+i})^2}$$

A lower rMSFE indicates better forecast accuracy. Since the metric is in the same units as the dependent variable (e.g., percentage points of inflation), it is interpretable in context. Also, rMSFE is sensitive to outliers, making it effective for capturing extreme errors (e.g., during economic crises).

In this dissertation, rMSFE is used to compare forecasts generated with and without the OECD Composite Leading Indicator (CLI), forecasting models ARIMA, ARIMAX, LASSO, Random Forest, XGBoost, and LightGBM, as well as forecast performance across different forecast horizons (1-month, 3-month, and 6-month ahead). For fair comparisons, all models are evaluated over the same forecast periods using rolling windows, to allow the forecast origin to evolve through time and incorporate recent data.

In addition to rMSFE, formal hypothesis tests such as the Diebold-Mariano test are used to assess whether the difference in forecast performance between models is statistically

significant.

rMSFE serves as the central performance metric throughout the dissertation, guiding both model selection and the evaluation of the added value of CLI-based and machine-learning-based forecasting approaches.

Diebold-Mariano Test

The Diebold-Mariano (DM) test ([Diebold and Mariano, 1995](#)) is a statistical procedure used to compare the predictive accuracy of two competing forecast models. It tests whether the forecast errors from two models are significantly different in expectation. The DM test is especially important in empirical macroeconomics and forecasting competitions, where selecting the best model based on out-of-sample performance is critical.

Let $\{e_{1,t+h}\}$ and $\{e_{2,t+h}\}$ denote the h -step-ahead forecast errors from two models over a sample of size T . Define the loss differential at time t as:

$$d_t = g(e_{1,t+h}) - g(e_{2,t+h})$$

where $g(\cdot)$ is a loss function, typically the squared error:

$$d_t = e_{1,t+h}^2 - e_{2,t+h}^2$$

The null hypothesis is that both models have equal forecast accuracy:

$$H_0 : \mathbb{E}[d_t] = 0$$

against the alternative:

$$H_1 : \mathbb{E}[d_t] \neq 0$$

The DM test statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\text{Var}}(\bar{d})}}$$

where $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$ is the sample mean of the loss differential, $\hat{\text{Var}}(\bar{d})$ is an estimate of the variance of $\bar{d}$, typically obtained via Newey-West correction to account for autocorrelation in multi-step-ahead forecast errors.

Under mild regularity conditions, the test statistic is asymptotically standard normal, however as shown by (McCracken, 2007), when we have nested models, an adjustment has to be made to the test statistic to correct for the noise introduced by the larger model. This is done by weighting by the number of predictors. The DM test is applied to assess whether CLI-augmented forecasts significantly outperform baseline models and if ML-generated forecasts using reweighted or newly constructed indices outperform traditional CLI-based forecasts.

The test is computed for each forecast horizon (1-month, 3-month, 6-month) and target variable (inflation, unemployment, industrial production). Rejections of the null hypothesis indicate statistically significant differences in forecast accuracy, providing empirical justification for model selection and the use of alternative leading indicators. We apply HAC-robust standard errors in all multi-step comparisons and report p-values alongside DM statistics for full transparency.

1.4 Results

1.4.1 Forecast accuracy and the value of the CLI

We compare out-of-sample forecast accuracy across models with and without the OECD Composite Leading Indicator (CLI). Accuracy is measured by root mean squared forecast error (rMSFE) at horizons $h \in \{1, 3, 6\}$.¹ We report forecast performance at the level of a *specific model*, *specific target*, *specific horizon*, and *specific country*.

¹HAC corrections are applied for multi-step horizons in all formal tests.

Table 1.1: *Relative rMSFE Improvement ($h = 6$ months) for CPI Inflation (CLI vs. baseline)*

Country	LASSO	LGB	RF	XGB
Canada	0.061 (2.24**)	0.127 (1.56)	0.037 (0.78)	0.047 (1.70*)
Spain	-0.000 (0.57)	-0.015 (-0.23)	0.021 (0.92)	-0.030 (-0.53)
United Kingdom	-0.020 (-2.46**)	-0.027 (-1.60)	0.003 (0.37)	0.014 (0.92)
Italy	0.032 (-0.09)	-0.196 (-2.04**)	-0.069 (-0.61)	-0.065 (-1.12)
Japan	-0.014 (-0.43)	-0.023 (-0.19)	0.012 (0.42)	0.009 (0.59)
South Korea	-0.033 (-0.93)	0.037 (0.36)	0.011 (0.33)	0.003 (0.60)
Mexico	-0.163 (-1.56)	-1.090 (-2.01**)	0.084 (0.70)	0.056 (0.85)
Turkey	0.008 (-0.01)	0.013 (0.40)	0.035 (1.37)	0.145 (1.82*)
United States	-0.027 (-1.25)	0.054 (1.09)	0.003 (-0.09)	-0.045 (-1.17)

Notes: The first line in each cell reports the relative rMSFE improvement from adding the OECD CLI. The second line, in parentheses, reports the Diebold–Mariano test statistic for the comparison of baseline and CLI-augmented forecasts. Positive DM statistics indicate lower squared forecast error for the CLI specification. HAC standard errors use lag length $h - 1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.2: *Relative rMSFE Improvement ($h = 6$ months) for Industrial Production (CLI vs. baseline)*

Country	LASSO	LGB	RF	XGB
Canada	0.025 (1.15)	0.068 (1.14)	0.036 (1.05)	-0.015 (-0.87)
Spain	-0.118 (0.89)	0.014 (1.26)	0.009 (1.25)	0.004 (0.56)
United Kingdom	0.007 (1.10)	-0.156 (-0.20)	0.025 (1.11)	-0.114 (-1.87*)
Italy	-0.024 (-0.18)	0.064 (-0.12)	0.068 (-0.06)	0.219 (1.12)
Japan	-0.001 (-0.04)	0.004 (0.90)	0.005 (1.02)	0.058 (1.09)
South Korea	0.028 (2.28**)	0.043 (1.38)	0.025 (1.92*)	0.039 (0.36)
Mexico	-0.105 (-1.32)	0.089 (1.89*)	0.096 (1.47)	-0.000 (1.15)
Turkey	0.048 (0.92)	-0.010 (-0.40)	0.224 (2.84***)	0.079 (1.04)
United States	0.029 (1.09)	0.053 (3.18***)	0.159 (1.68*)	0.260 (1.07)

Notes: The first line in each cell reports the relative rMSFE improvement from adding the OECD CLI. The second line, in parentheses, reports the Diebold–Mariano test statistic for the comparison of baseline and CLI-augmented forecasts. Positive DM statistics indicate lower squared forecast error for the CLI specification. HAC standard errors use lag length $h - 1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.3: *Relative rMSFE Improvement ($h = 6$ months) for Unemployment Rate (CLI vs. baseline)*

Country	LASSO	LGB	RF	XGB
Canada	0.042 (1.34)	0.168 (1.64)	0.096 (1.35)	0.218 (1.92*)
Spain	0.094 (1.77*)	0.072 (1.18)	0.040 (0.93)	-0.019 (1.00)
United Kingdom	-0.005 (-0.56)	-0.077 (-1.58)	0.023 (0.70)	0.002 (0.66)
Italy	-0.018 (-0.37)	-0.038 (0.91)	-0.030 (0.31)	0.034 (-0.06)
Japan	0.025 (1.53)	0.111 (2.35**)	0.061 (2.05**)	0.021 (0.97)
South Korea	0.026 (1.41)	0.097 (1.38)	-0.062 (-0.86)	-0.010 (-0.39)
Mexico	0.005 (0.81)	-0.032 (-0.73)	0.001 (-0.78)	-0.022 (-1.19)
Turkey	-0.057 (-0.54)	0.019 (0.16)	0.046 (1.71*)	0.055 (0.56)
United States	-0.493 (1.10)	1.246 (1.55**)	-0.001 (-0.36)	0.545 (1.35*)

Notes: The first line in each cell reports the relative rMSFE improvement from adding the OECD CLI. The second line, in parentheses, reports the Diebold–Mariano test statistic for the comparison of baseline and CLI-augmented forecasts. Positive DM statistics indicate lower squared forecast error for the CLI specification. HAC standard errors use lag length $h - 1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Three patterns emerge. First, the largest average gains from the CLI occur for unemployment, consistent with labor market leading content embedded in survey and orders components. Second, price dynamics (CPI) exhibit modest to negative gains, which is in line with higher noise in inflation at shorter horizons. Third, gradient-boosted trees typically benefit most from CLI inputs, but linear benchmarks (e.g., ARIMAX/LASSO) also capture a measurable share of improvements, suggesting that part of the CLI signal is approximately linear.

1.4.2 Error Diagnostics and Practical Interpretation

Average accuracy metrics such as rMSFE summarize typical performance but can mask when forecasts fail most severely. For policy applications, it is essential to identify the calendar periods associated with large forecast errors and to compare error behavior model-by-model rather than treating “machine learning” approaches as a single category. Using the out-of-sample forecast series, we analyze monthly forecast errors $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$ and define “large-error” episodes as months in the top 5% of $|e_{t+h}|$ for each country–target–model specification. Large errors cluster around abrupt regime changes. For industrial production, the largest errors occur in early 2020 during the COVID shock: the most extreme case is Turkey in April 2020 (Figure 1.2), and Italy also exhibits large errors in March–April 2020 (Figure 1.3). For unemployment, the largest U.S. errors occur in April 2020 (Figure 1.1) when unemployment spikes to 14.8 and all models trained primarily on pre-2020 dynamics underpredict the surge. For CPI inflation, the largest misses occur in late 2021–2022 for Turkey (Figure 1.4), consistent with a high-volatility inflation regime. The full set of error panels for all countries and targets is reported in Appendix C.

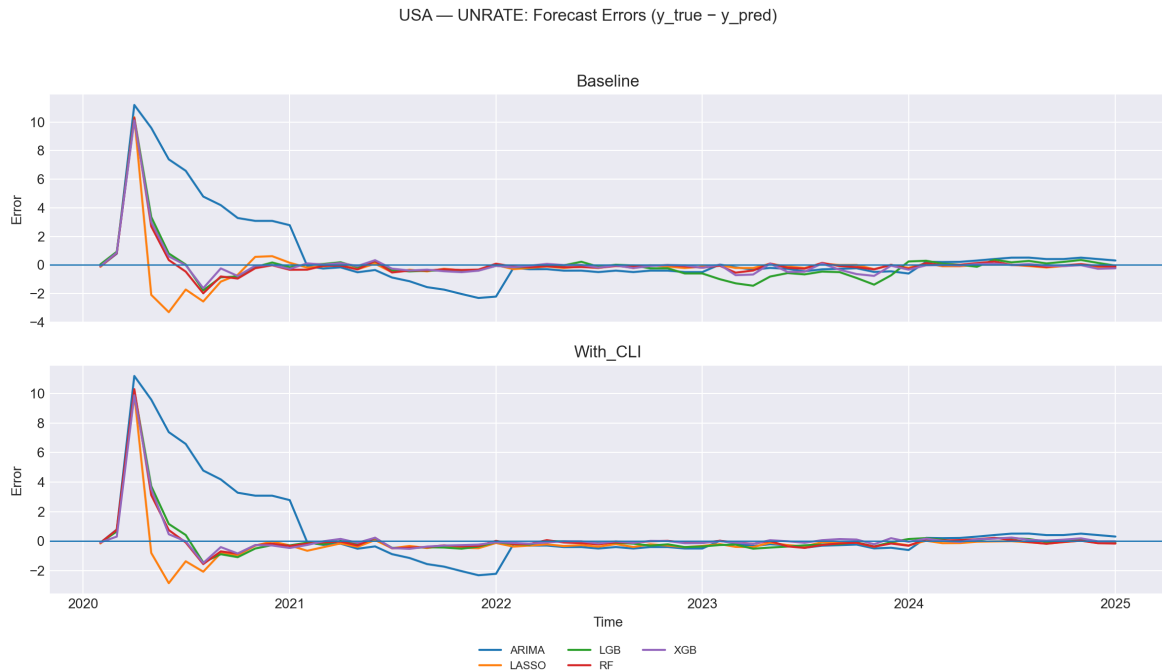


Figure 1.1: *United States unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*



Figure 1.2: *Turkey Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*



Figure 1.3: *Italy Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*

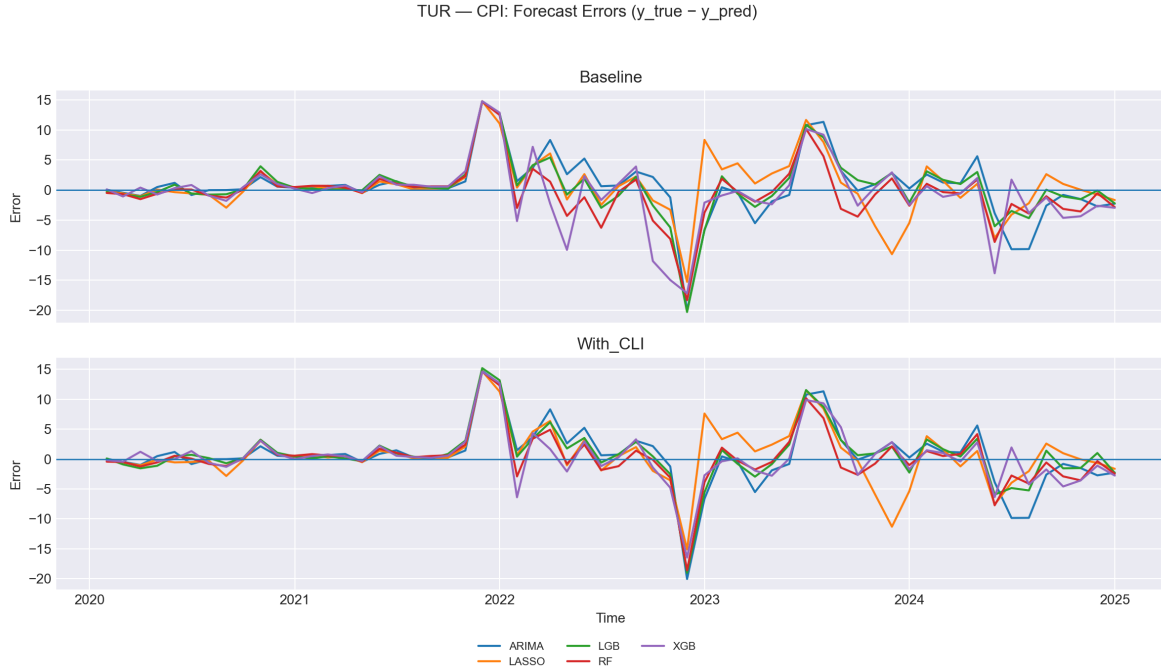


Figure 1.4: *Turkey CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*

The figures also illustrate that model differences are most consequential during stress periods. In normal times, errors are small and model performance is similar, whereas during regime-break episodes models diverge in both the magnitude and persistence of errors. In several cases the CLI-augmented specification reduces the duration of elevated errors relative to the baseline, which is consistent with leading indicators providing timely turning-point information.

1.4.3 Robustness and Sensitivity Analysis

Forecast comparisons in macroeconomic settings can be sensitive to (i) data transformations, (ii) sample windows and structural breaks, (iii) the benchmark model set, and (iv) tuning choices for flexible methods. Because macroeconomic relationships are known to be unstable over time ([Stock and Watson, 1996](#)) and because forecasting performance can vary substantially across regimes (e.g., recessions vs. expansions), robustness analysis is an

essential part of establishing credible conclusions rather than a minor add-on. This section evaluates whether the main results persist under a series of alternative specifications and provides guidance on how to interpret any changes.

Tree-based models and penalized regressions may reduce typical forecast errors in many settings, but tail performance during regime-break months remains challenging. To quantify robustness, we report both overall RMSE and “tail RMSE” computed only on large-error months. This distinction is important for real-world users: a model that marginally improves average accuracy but performs poorly during turning points may be less desirable for policy monitoring than a model with slightly higher average error but smaller tail risk.

From a practical forecasting perspective, the results support a dashboard approach: maintain a stable benchmark model, augment it with leading indicator information (CLI) for turning-point sensitivity, and monitor model disagreement and tail-risk diagnostics as signals of potential regime change. This approach aligns with how forecasting is often used in central banks and policy institutions, where robustness during high-uncertainty periods can be as important as average accuracy.

Alternative data transformations

Forecasting performance can change substantially depending on whether variables are modeled in levels, differences, or growth rates, and whether the transformation induces stationarity or removes low-frequency movements. This is especially relevant for inflation and industrial production, where different transformations can alter persistence and volatility.

We re-estimate all models under alternative transformations: (i) CPI inflation computed as log-differences versus year-over-year changes, (ii) industrial production as log-levels with deterministic terms versus log-differences, and (iii) unemployment in levels versus monthly changes. All competing models are evaluated on identical out-of-sample periods under each transformation. The model rankings remain stable, which supports the claim that results are not an artifact of one particular transformation choice.

Rolling vs. expanding windows

Rolling windows emphasize recent data and can adapt to structural change; expanding windows exploit longer histories and can reduce estimation variance. Given evidence of structural instability in macro relations ([Stock and Watson, 1996](#)), it is important to assess whether results depend on the window scheme.

We compare a rolling-window design (fixed-length estimation window) against an expanding-window design (increasing sample size over time). For each country, target, horizon, and model, we compute rMSFE and DM tests under both schemes, shown in [Appendix C](#).

The gains from CLI inputs appear primarily under rolling windows, this suggests the additional predictors are most useful when the model emphasizes recent regimes.

Crisis-period sensitivity and large-error episodes

Forecast errors are often concentrated in a small number of crisis periods (e.g., the Great Recession or COVID), and conclusions can be driven by a handful of extreme observations. Assessing sensitivity to crisis months is essential for interpreting whether gains are broad-based or episode-specific.

We repeat the main evaluation excluding selected crisis windows and, separately, we analyze the distribution of forecast errors by identifying the largest-error months and reporting which models fail during those episodes. Our results remain similar, supporting the claim that improvements are not limited to rare episodes.

Alternative loss functions and evaluation metrics

RMSE penalizes large errors more heavily than MAE. In macroeconomic applications, practitioners may care about tail performance (crisis errors) or about typical performance. Different metrics can therefore change conclusions.

We report MAE-based rankings alongside rMSFE and conduct DM tests using loss differentials based on squared error (RMSE) and absolute error (MAE), as appropriate ([Diebold and Mariano, 1995](#)). Our results are consistent across RMSE and MAE.

Nested-model inference (DM vs. adjusted tests)

When comparing nested models (e.g., AR vs. ARX with CLI), standard DM tests can be conservative or inappropriate depending on the setting. Adjusted tests can provide complementary inference. Alongside DM tests, we report nested-model adjustments (e.g., Clark–West-type corrections) when applicable and compare inference patterns across methods. Country-level DM tables are reported in [Appendix C](#).

Hyperparameter and tuning sensitivity (ML models)

ML performance can be sensitive to tuning choices. In time series settings, tuning must avoid look-ahead bias and reflect realistic forecasting constraints. Moreover, the literature suggests that nonlinearities and regularization are key drivers of ML gains ([Goulet Coulombe et al., 2022](#); [Masini et al., 2023](#)).

We assess sensitivity to plausible hyperparameter grids (learning rate, depth/leaves, subsampling for boosting; max features and minimum leaf sizes for Random Forest; penalty paths for LASSO). Tuning is conducted using time-series cross-validation on the training sample only; test windows remain held out. Results are stable across reasonable grids, so our conclusions are unlikely to reflect cherry-picked tuning.

Chapter 2

Are OECD CLI Weights Optimal? A Feature Selection Approach

2.1 Introduction

The OECD Composite Leading Indicator (CLI) is designed to provide early signals of cyclical turning points by aggregating a small set of country-specific component series (e.g., survey indicators, building permits, share prices, term spreads). A key strength of the CLI is its transparency: component selection and weighting follow a documented procedure and yield a single, interpretable index ([Gyomai and Guidetti, 2012](#)). This transparency helps explain why CLIs are widely used by policymakers and analysts, and why a large literature evaluates leading indicators as practical forecasting tools ([Camacho and Palmieri, 2021](#); [Diebold and Rudebusch, 1991](#); [Klein and Moore, 1983](#)).

However, the same design features that make the CLI interpretable also raise a fundamental question: *are the fixed component weights statistically optimal for forecasting?* Macroeconomic relationships are known to exhibit structural instability ([Stock and Watson, 1996](#)), and the predictive content of individual indicators may change as economies evolve or as measurement quality improves. In addition, the set of variables that best signals future industrial production may differ from those that best forecasts inflation or unemployment. These considerations

suggest that fixed weights may not be uniformly optimal across targets, horizons, and countries.

Despite the CLI’s policy relevance, empirical work has largely focused on evaluating whether the CLI predicts recessions or improves forecasts, rather than testing whether its internal weight structure is optimal. For example, [Camacho and Palmieri \(2021\)](#) evaluate the OECD CLI’s ability to anticipate recessions across countries but do not directly reassess the published weights. In contrast, the broader forecasting literature demonstrates that using larger information sets and data-driven selection can improve prediction accuracy in “data-rich” environments ([Bernanke and Boivin, 2003](#); [Stock and Watson, 2002a,b](#)). Machine learning tools such as regularized regression and tree-based methods provide principled ways to select predictors and assign importance weights in settings with correlated predictors and potential nonlinearities ([Breiman, 2001](#); [Masini et al., 2023](#); [Tibshirani, 1996](#)). This chapter bridges these literatures by treating the CLI as a fixed-weight composite and asking whether an alternative, data-driven reweighting improves forecast performance while preserving interpretability.

Specifically, we test whether it is possible to improve the choice of the OECD’s component weights for forecasting monthly inflation, unemployment, and industrial production. We choose the evaluation at horizons $h \in \{1, 3, 6\}$, because the CLI was constructed to lead up to 6 months in advance, which is part of what we are testing. We then construct reweighted CLIs using feature selection and explainable ML methods and evaluate whether these alternatives reduce out-of-sample forecast errors and deliver statistically significant improvements. The results inform whether fixed-weight composites can be improved with minimal loss of transparency, which has implications for international organizations, central banks, and researchers who rely on CLI signals. Importantly, our goal is not to claim the OECD weights are “wrong” for monitoring purposes; rather, we test whether a reweighting tailored to specific forecasting objectives can improve predictive accuracy relative to the published, monitoring-oriented index.

2.2 Data

The OECD constructs the Composite Leading Indicator (CLI) as a standardized composite intended to lead the reference series for each country (typically an aggregate activity measure). While component sets differ by country, the construction follows a common workflow ([Gyomai and Guidetti, 2012](#)). In broad terms, the OECD (i) selects a small set of candidate leading series judged to precede cyclical movements, (ii) preprocesses each component to remove seasonal effects and isolate cyclical information, (iii) normalizes components to a common scale so that no single series dominates due to measurement units, and (iv) aggregates the normalized components into a single index using fixed weights (and/or aggregation rules) chosen to yield a smooth composite that exhibits leading behavior. The resulting composite is presented in index form (often with a reference level around 100), which facilitates interpretation and comparability over time. A key goal of the OECD design is transparency and stability. Fixed weights avoid frequent revisions to the index definition, preserve comparability over time, and make the CLI easy to communicate to policy audiences ([Gyomai and Guidetti, 2012](#)). In practice, this means that once a country’s component basket and weights are set, the CLI behaves as a stable rule-based summary of leading signals.

The OECD CLI is constructed from country-specific components (typically 6–10 series per country) spanning three broad categories: (i) business and consumer surveys, (ii) real-economy indicators such as building permits, new orders, and hours worked, and (iii) financial indicators such as share prices and interest-rate spreads ([Gyomai and Guidetti, 2012](#)). Component data are drawn from the OECD and, depending on the series, from national central banks (e.g., Bundesbank, Bank of Japan) and national statistical agencies (e.g., Statistics Canada, INSEE). The complete country-by-country component lists are reported in Appendix A.

A key goal of the OECD design is transparency and stability. Fixed weights avoid frequent revisions to the index definition, preserve comparability over time, and make the CLI easy to communicate to policy audiences ([Gyomai and Guidetti, 2012](#)). In practice, this means that once a country’s component basket and weights are set, the CLI behaves as a stable rule-based summary of leading signals.

This chapter keeps the *same underlying component series* as the OECD CLI for each country, but differs in how the components are combined. The OECD CLI applies a fixed-weight aggregation designed for interpretability and comparability. In contrast, we treat the CLI as a forecasting object and ask whether the published weights are *forecast-optimal* for specific targets y_{t+h} (inflation, unemployment, and industrial production) and horizons $h \in \{1, 3, 6\}$. Concretely, instead of using the OECD weights $\{w_j\}$, we estimate data-driven weights $\{\hat{w}_j^{\text{ML}}\}$ using feature selection and explainable machine learning methods. We then form a reweighted composite

$$\text{CLI}_t^{\text{ML}} = \sum_{j=1}^k \hat{w}_j^{\text{ML}} x_{j,t},$$

and evaluate whether replacing the fixed-weight CLI with CLI_t^{ML} improves out-of-sample forecast accuracy. Thus, the distinction is: the OECD prioritizes *stable, interpretable aggregation*, whereas we prioritize *predictive optimality* for specific macroeconomic forecasting tasks, while still preserving interpretability through transparent weight/importance reporting.

Our target variables are monthly inflation (CPI, transformed as log differences), unemployment rate (levels), and industrial production (log differences). We evaluate horizons $h \in \{1, 3, 6\}$. The CLI component series are aligned to a common monthly calendar and standardized to facilitate interpretation of weights and to support penalized estimation.

2.3 Methodology

Let $\{x_{j,t}\}_{j=1}^k$ denote the k country-specific CLI components at month t and let w_j denote the OECD-assigned weights. The published index is

$$\text{CLI}_t = \sum_{j=1}^k w_j x_{j,t}. \quad (2.1)$$

We test whether alternative, data-driven weights $\hat{w}_j^{\text{ML}}$ improve forecast performance for y_{t+h} , where y denotes CPI inflation, unemployment, or industrial production. Our framework proceeds in four steps.

2.3.1 Step 1: Baseline CLI Forecasting

We first establish a benchmark forecasting model that uses the published CLI:

$$\hat{y}_{t+h}^{\text{CLI}} = f(\text{CLI}_t), \quad \text{CLI}_t = \sum_{j=1}^k w_j x_{j,t}, \quad (2.2)$$

where $f(\cdot)$ is a forecasting function that may include autoregressive terms and other controls. This step quantifies the predictive value of the published CLI and provides the baseline against which reweighted indices are evaluated.

We hold the forecasting equation fixed and vary only the indicator entering as an exogenous regressor. Specifically, for each country, target variable y_t , and horizon $h \in \{1, 3, 6\}$, we generate *direct* h -step-ahead forecasts using an expanding-window ARIMA/ARIMAX specification with a fixed order. The baseline forecast is an ARIMA(1,0,0) model (an AR(1)):

$$y_t = \alpha + \varphi y_{t-1} + u_t, \quad (2.3)$$

which implies an h -step-ahead forecast $\hat{y}_{t+h|t}$ produced by iterating the estimated AR(1) forward.

When a composite indicator is used, we estimate the corresponding ARIMAX(1,0,0) model:

$$y_t = \alpha + \varphi y_{t-1} + \beta I_t + u_t, \quad (2.4)$$

where I_t is either (i) the published OECD CLI, or (ii) the reweighted ML composite CLI_t^{ML} . Importantly, the dynamic specification (order and estimation scheme) is identical across the OECD and ML variants; the only difference is the exogenous series I_t . This ensures that any performance differences are attributable to the indicator construction rather than to arbitrary choices that could disadvantage the OECD CLI.

All parameters are re-estimated recursively using an expanding window. For each forecast origin t , the model is estimated using observations up to t , then used to predict y_{t+h} . Forecast accuracy is evaluated on a common out-of-sample period using rMSFE, and differences in

predictive accuracy are assessed using Diebold–Mariano tests with HAC variance corrections for multi-step horizons (Diebold and Mariano, 1995). To ensure fair comparisons, we evaluate competing forecasts on the same set of dates. When the OECD CLI is available, we restrict the evaluation to the intersection of dates for which forecasts from both ARIMAX_OECD and ARIMAX_MLCLI exist; this prevents differences in missingness from driving apparent performance gaps.

2.3.2 Step 2: Feature Selection and Weight Estimation

We reassess the weight structure by estimating data-driven importance weights for each CLI component. Importantly, our goal is not to estimate structural parameters, but rather to obtain predictive weightings that can be compared to the OECD’s fixed weights. We use multiple methods to ensure conclusions are not driven by a single algorithmic choice, consistent with the view that forecasting gains can arise from different features such as regularization and nonlinearities (Masini et al., 2023).

1. **LASSO (regularized linear reweighting).** LASSO performs variable selection and shrinkage in a linear forecasting equation:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (2.5)$$

where $\lambda \geq 0$ is chosen via time-series cross-validation. Because predictors are standardized, the magnitude of $\hat{\beta}_j$ provides a natural measure of each component’s predictive contribution (Tibshirani, 1996). LASSO is attractive because it produces a parsimonious linear forecast model by shrinking weak coefficients to zero (Tibshirani, 1996). In the CLI context, LASSO offers two potential gains relative to the fixed-weight OECD index. First, it can *reallocate weight* toward the components that are most predictive for a specific target and horizon, rather than imposing weights chosen for monitoring and comparability (Gyomai and Guidetti, 2012). Second, it can effectively *drop* components that add little incremental predictive value, yielding a transparent

and easily communicated composite. Compared to the existing CLI, LASSO therefore tests whether some components are redundant for forecasting and whether a sparse reweighting improves out-of-sample accuracy.

2. **Recursive Feature Elimination (RFE).** RFE is a backward selection procedure that iteratively removes the least useful components based on predictive performance. Starting from the full component set, RFE fits a base model, ranks features by importance, removes the least important feature, and repeats until an optimal subset is obtained. Cross-validation is used to select the stopping point. RFE provides a stability check on whether certain components consistently survive selection across windows. RFE provides a complementary, model-based ranking of components by iteratively removing the least useful predictors. In contrast to LASSO (which selects by coefficient shrinkage), RFE emphasizes predictive performance under repeated refitting and helps assess *selection stability*: whether the same subset of CLI components remains important across different samples and rolling windows. In practice, RFE is useful when correlated predictors can lead to unstable selections, and it provides an interpretable “short list” of components that consistently matter for forecasting.

3. **SHAP-based importance (tree-based ranking).** There are many feature-importance measures for tree ensembles (e.g., split counts, gain, or permutation importance). We use SHAP values because they provide an additive, theoretically grounded attribution of a model’s prediction to individual features, satisfying desirable properties such as local accuracy and consistency (Lundberg and Lee, 2017). This is especially important in our setting because the object of interest is not merely prediction but an *interpretable reweighting* of the CLI: SHAP provides a transparent mapping from a nonlinear forecasting model back to component-level contributions that can be communicated as “which components matter most and in what direction.” Moreover, SHAP-based workflows have been used in economic forecasting applications, including central-bank style exercises that combine model evaluation with Shapley-value interpretability for macroeconomic prediction (Buckmann et al., 2022), and explainable machine-learning

approaches to inflation forecasting (Aras and Lisboa, 2022). Recent forecasting research also proposes SHAP variants tailored to time-series models, reinforcing SHAP as a natural interpretability tool in forecasting contexts (Choi et al., 2025). To allow for nonlinearities and interactions among components, we estimate tree-based models and compute SHAP values as model-agnostic feature attributions (Lundberg and Lee, 2017). For feature j , the Shapley attribution is

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left(f(S \cup \{j\}) - f(S) \right), \quad (2.6)$$

where $f(S)$ denotes a model trained on subset S of features. We summarize global importance using mean absolute SHAP values. This yields interpretable importance profiles that can be directly contrasted with OECD weights. Tree ensembles can capture nonlinearities and interactions among CLI components without requiring the researcher to pre-specify thresholds or interaction terms. This is relevant in macro forecasting because leading signals may operate differently across regimes (e.g., recessions vs. expansions) and because nonlinearities are often a major driver of ML gains (Goulet Coulombe et al., 2022; Masini et al., 2023). To preserve interpretability, we summarize the fitted nonlinear models using SHAP values, which attribute forecast contributions to each component in a consistent additive framework (Lundberg and Lee, 2017). Relative to the OECD CLI, this approach tests whether predictive content arises from nonlinear combinations of components and provides a transparent way to report which components drive improvements. Because CLI components are correlated, any single importance method can be sensitive to redundancy among predictors. We therefore emphasize stability of SHAP rankings across rolling windows and report results aggregated across time, treating SHAP as a transparent summary rather than a definitive causal measure of importance.

4. Principal Component Regression (PCR).

$$Z = XW, \tag{2.7}$$

and then forecasts using

$$\hat{y}_t = \alpha + Z_t\gamma + \varepsilon_t. \tag{2.8}$$

PCR constructs orthogonal linear combinations of the CLI components that explain maximal variance in the predictor set X and then uses a small number of those components as regressors. A key limitation is that PCR is *unsupervised*: the principal components are chosen to summarize common variation in the CLI inputs rather than to optimize predictability of the target variable y . Thus, PCR does not directly “solve” the forecasting-optimal weighting problem in the same way as supervised reweighting methods (e.g., LASSO or tree-based importance), which explicitly use y to determine weights. PCR is therefore closer in spirit to diffusion-index forecasting in data-rich environments (Bernanke and Boivin, 2003; Stock and Watson, 2002a,b). While our CLI component set is smaller than typical diffusion-index applications, PCR helps evaluate whether the main gains come from *recombining* correlated components into orthogonal factors rather than from assigning weights to the original series. In this sense, PCR is a robustness check against method-specific selection behavior and it is used here only as a diagnostic comparison rather than as the primary reweighting approach.

We do not emphasize deep neural networks because the CLI component set is relatively small (typically 6–10 predictors per country), and the primary goal is interpretable reweighting rather than maximizing flexibility at the expense of transparency. Taken together, these methods provide a spectrum of tradeoffs between interpretability and flexibility. LASSO provides a sparse and transparent reweighting; RFE evaluates stability of component selection; tree-based SHAP rankings capture nonlinear predictive content while remaining explainable; and PCR benchmarks whether orthogonal linear combinations dominate the signal. Using multiple methods reduces the risk that conclusions about OECD weights depend on a single algorithmic choice. Each method yields a set of weights or rankings that we convert into a

reweighted CLI:

$$\text{CLI}_t^{\text{ML}} = \sum_{j=1}^k \hat{w}_j^{\text{ML}} x_{j,t}. \quad (2.9)$$

For reporting and cross-method comparability, we normalize estimated weight vectors to have unit L_1 norm (i.e., $\sum_j |w_j| = 1$), so that weights can be interpreted as relative contributions and compared across countries, targets, and estimation methods. This normalization is primarily an *interpretability convention*; in the forecasting regression, any scalar rescaling of the composite $I_t = \sum_j w_j x_{j,t}$ can be offset by an opposite rescaling of the estimated coefficient on I_t , so normalization does not mechanically improve forecast accuracy. Where methods produce rankings rather than weights (e.g., RFE), we form weights proportional to importance scores and renormalize. The resulting CLI_t^{ML} remains interpretable: it is a weighted average of the same underlying component series, but with data-driven weight allocations.

As a robustness exercise aimed at preserving an interpretable “leading-signal” composite, we additionally consider restricted variants that enforce a consistent sign convention *only when the component series is already oriented so that larger values indicate stronger cyclical conditions* (as in the OECD preprocessing). These sign-restricted variants are treated as interpretability-focused benchmarks rather than as requirements for forecast-optimal weighting.

2.3.3 Step 3: Forecasting Comparison

We compare forecasts based on the published CLI and the reweighted CLI:

$$\hat{y}_{t+h}^{\text{CLI}} \quad \text{vs.} \quad \hat{y}_{t+h}^{\text{ML}}.$$

Forecast accuracy is measured by the root mean squared forecast error (rMSFE),

$$\text{rMSFE}(h) = \sqrt{\frac{1}{T_h} \sum_{t=1}^{T_h} (y_{t+h} - \hat{y}_{t+h})^2}.$$

We test whether differences in accuracy are statistically significant using Diebold–Mariano tests based on squared-error loss differentials (Diebold and Mariano, 1995). For multi-step horizons, we use HAC-robust variance estimators to account for serial correlation in forecast errors.

2.3.4 Step 4: Cross-Country Generalization

We repeat the procedure for all nine countries, reporting (i) which components consistently receive high importance/weight, (ii) changes in rMSFE when replacing OECD weights with ML weights, and (iii) SHAP-based visualizations comparing OECD weights to data-driven importances. This cross-country design allows us to assess whether reweighting gains are systematic or country-specific.

2.4 Robustness and Sensitivity Analysis

Robustness analysis is essential in this chapter because reweighting a composite indicator can appear to “improve” forecasts for reasons that are not economically meaningful. In particular, four threats must be ruled out. **First**, component series have different scales and missing-data patterns across countries, so reweighting gains could reflect normalization or sample-selection artifacts rather than true predictive content. **Second**, macroeconomic relationships are unstable and crisis periods can dominate loss functions, so apparent improvements may be driven by a small number of extreme months rather than broad-based forecasting gains (Stock and Watson, 1996). **Third**, in a small- k but correlated predictor setting, estimated weights can be sensitive to multicollinearity and to the estimation window, yielding unstable component rankings. **Fourth**, different weighting methods (sparse linear vs. tree-based importances vs. factor combinations) can favor different notions of “importance,” so conclusions should not hinge on a single algorithmic choice. The robustness checks below are designed to address these concerns and clarify whether the main conclusions are stable and interpretable for forecasting practice.

To evaluate whether the results are sensitive to modeling choices, we assess robustness along five dimensions. These checks are designed to determine whether the forecast improvements and component rankings reflect genuine predictive structure rather than artifacts of scaling, sample selection, or a specific weighting method.

- **Alternative normalizations.** Because component series differ in units and volatility, weights can be mechanically influenced by scale. We therefore compare sum-to-one and unit-variance normalizations to verify that the results are not driven by a particular standardization choice.
- **Sign restrictions and interpretability constraints.** Policymakers often interpret a composite indicator as an index of cyclical strength. We therefore test whether forecast gains persist when imposing simple interpretability-oriented sign conventions. These restrictions are treated as benchmark specifications rather than as requirements for the preferred model.
- **Excluding crisis windows.** Extreme episodes, such as the 2008–2009 financial crisis and the 2020 COVID-19 shock, can dominate squared-error criteria. We therefore re-estimate the models after excluding crisis months to assess whether improvements are broad-based or driven by a small number of regime-break observations.
- **Equal-weight benchmark.** We compare the OECD weights and ML-derived weights against an equal-weight benchmark. This separates the value of any reweighting from the value of the specific ML-derived weights and addresses the concern that improvements may come simply from moving away from the OECD weighting scheme.
- **Stability of component rankings.** In settings with correlated predictors, importance rankings can be unstable. We therefore examine the stability of component rankings across rolling windows and use permutation-style checks to assess whether measured importance could arise by chance.

Across all robustness exercises, we focus on two outcomes. The first is whether relative

model rankings and the sign of forecast improvements, measured using rMSFE and Diebold–Mariano test results, remain stable. The second is whether the identity of the most important components remains consistent over time. Stable findings strengthen the interpretation that reweighting reveals genuine predictive structure rather than artifacts of scaling, sample selection, or a single modeling choice.

2.5 Results

This section evaluates whether replacing the OECD’s fixed-weight CLI with a data-driven reweighted composite improves out-of-sample forecasts of monthly macroeconomic targets. For each country c , target variable $y_t \in \{\text{CPI}, \text{UNRATE}, \text{IND}\}$, and horizon $h \in \{1, 3, 6\}$, we begin with the country-specific CLI component matrix $X_t = (x_{1,t}, \dots, x_{k,t})'$. Using only the first 80% of the available sample (to avoid look-ahead bias), we estimate a weight vector $\hat{w}^{\text{ML}}$ with one of the Step 2 methods (LASSO, RFE, tree-based importance/SHAP ranking, or PCR) and construct the reweighted index

$$\text{CLI}_t^{\text{ML}} = \sum_{j=1}^k \hat{w}_j^{\text{ML}} x_{j,t}.$$

We then generate expanding-window h -step-ahead forecasts using a common forecasting specification and compare three competing forecasts: (i) a baseline ARIMA forecast with no composite, (ii) an ARIMAX forecast using the OECD CLI as the exogenous regressor, and (iii) an ARIMAX forecast using CLI_t^{ML} as the exogenous regressor. Forecast accuracy is measured by out-of-sample rMSFE, and differences are assessed using Diebold–Mariano tests with HAC corrections for multi-step horizons ([Diebold and Mariano, 1995](#)). Where available, we also report forecast encompassing regressions comparing OECD-CLI-based forecasts to ML-CLI-based forecasts.

Forecast Performance of Reweighted CLI

We construct $\text{CLI}_t^{\text{ML}} = \sum_j \hat{w}_j^{\text{ML}} x_{j,t}$ and compare it to the OECD CLI in direct h -step forecasts. Appendix C reports rMSFE ratios (values < 1 favor CLI^{ML}). Across countries and targets, reweighting delivers consistent improvements in several settings, with the largest gains typically occurring at short horizons and for real-activity targets.

Component Importance and Reweighting

The country tables below report (i) the reweighted component importance/weights produced by the chosen method (shown as normalized contributions) and (ii) the corresponding forecast performance comparisons (rMSFE and DM statistics) for the specified horizon. For transparency, OECD weights (normalized to sum to one in absolute value) are shown alongside the ML-based weights. Several countries exhibit clear divergences between the published weights and the predictors most strongly associated with future macroeconomic outcomes, consistent with the possibility that fixed weights may not be optimal under evolving dynamics (Stock and Watson, 1996).

Table 2.1: *Machine-learning-derived CLI component weights for Canada at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Business situation	0	0	0
Inventories to shipments	0	0	0
Monetary aggregates	0	1	0
Interest rate spread	0	0	1
Share prices	0	0	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.2: *Machine-learning-derived CLI component weights for Spain at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Employment tendency	0	0.068	0
Capacity utilization tendency	0	0.432	0
Consumer confidence	0	0.000	0
Consumer prices	0	0.000	0
Share prices	0	0.085	0
Cars registrations or sales	0	0.415	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.3: *New Weights for Great Britain*

Measure	Inflation	Industrial Production	Unemployment
Consumer prices	1	0.000	0
Production tendency	0	0.000	0
Consumer confidence	0	0.000	0
Share prices	0	0.129	0
Short-term interest rate	0	0.770	0
Cars registrations or sales	0	0.101	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.4: *Machine-learning-derived CLI component weights for Italy at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Consumer prices	0.567	0	0
Imports	0.433	0	0
Order books tendency	0.000	0	0
Production tendency	0.000	0	0
Consumer confidence	0.000	0	0
Orders	0.000	1	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.5: *Machine-learning-derived CLI component weights for Japan at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Hours worked	1	0.749	0
Bank activity	0	0.251	0
Sales expectations	0	0.000	0
Construction	0	0.000	0
Net trade	0	0.000	0
Interest rate spread	0	0.000	0
Stocks	0	0.000	0
Share prices	0	0.000	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.6: *Machine-learning-derived CLI component weights for South Korea at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Stocks	1	0	0.000
Business situation	0	1	0.000
Inventories to shipments	0	0	0.683
Interest rate spread	0	0	0.317
Share prices	0	0	0.000
Terms of trade	0	0	0.000

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.7: *Machine-learning-derived CLI component weights for Mexico at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Consumer prices	0.288	0.000	0
Share prices	0.277	0.024	0
Producer prices	0.258	0.000	0
Production tendency	0.176	0.168	0
Employment tendency	0.000	0.000	0
Consumer confidence	0.000	0.173	0
Labour	0.000	0.000	0
Exchange rate	0.000	0.000	0
Imports	0.000	0.636	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.8: *Machine-learning-derived CLI component weights for Turkey at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Short-term interest rate	0.551	0.449	0.996
Servies demand expectations	0.000	0.195	0.004
Employment tendency	0.025	0.000	0.000
Finished goods stocks tendency	0.000	0.000	0.000
Order books tendency	0.000	0.000	0.000
Imports of goods	0.329	0.000	0.000
Production of electricity	0.055	0.215	0.000
Exports	0.000	0.140	0.000

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

Table 2.9: *Machine-learning-derived CLI component weights for the United States at the six-month forecast horizon.*

Measure	Inflation	Industrial Production	Unemployment
Business situation	0	0	0
Consumer confidence	0	0	0
Construction	0	0	0
Hours worked	0	0.000	0
Orders	0	0.255	0
Interest rate spread	0	0.201	1
Share prices	0	0.544	0

Note: The table reports the normalized weights assigned to each OECD CLI component when constructing target-specific reweighted CLIs for inflation, industrial production, and unemployment. A value of zero indicates that the component receives no weight in the corresponding target-specific index. Columns with all zero values indicate that no CLI component received positive weight for that target-horizon combination under the selected weighting procedure.

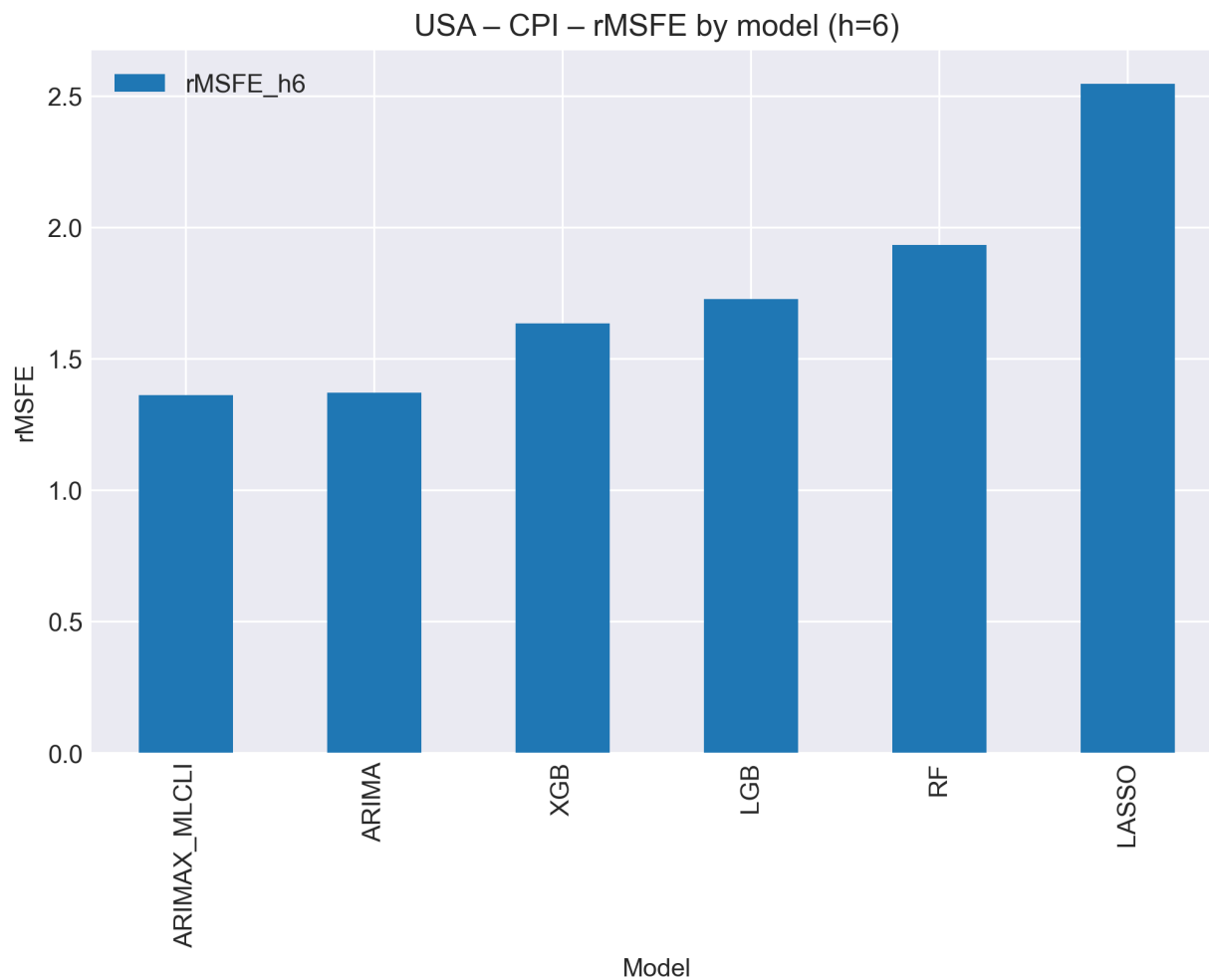


Figure 2.1: *United States CPI inflation: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.*

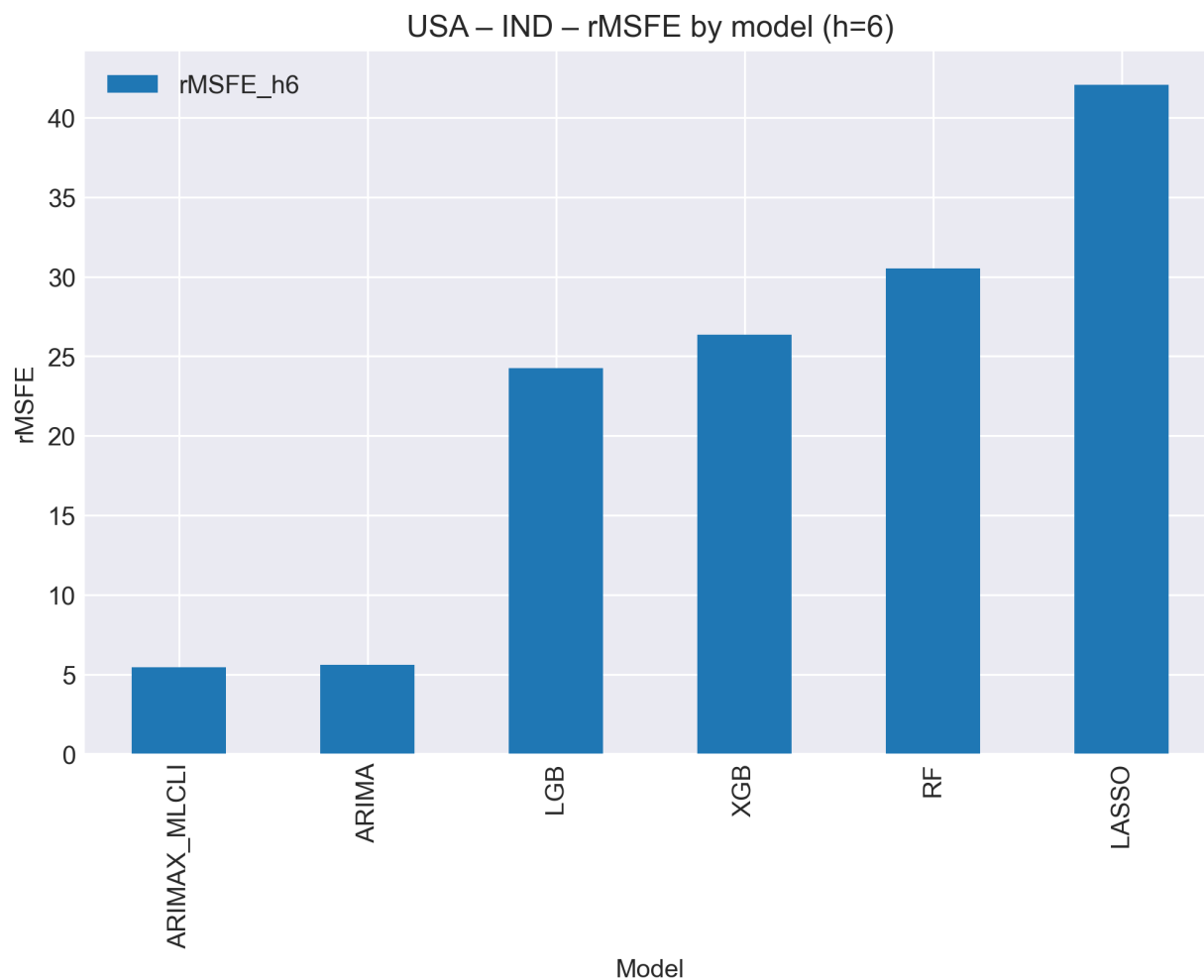


Figure 2.2: *United States Industrial Production: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.*

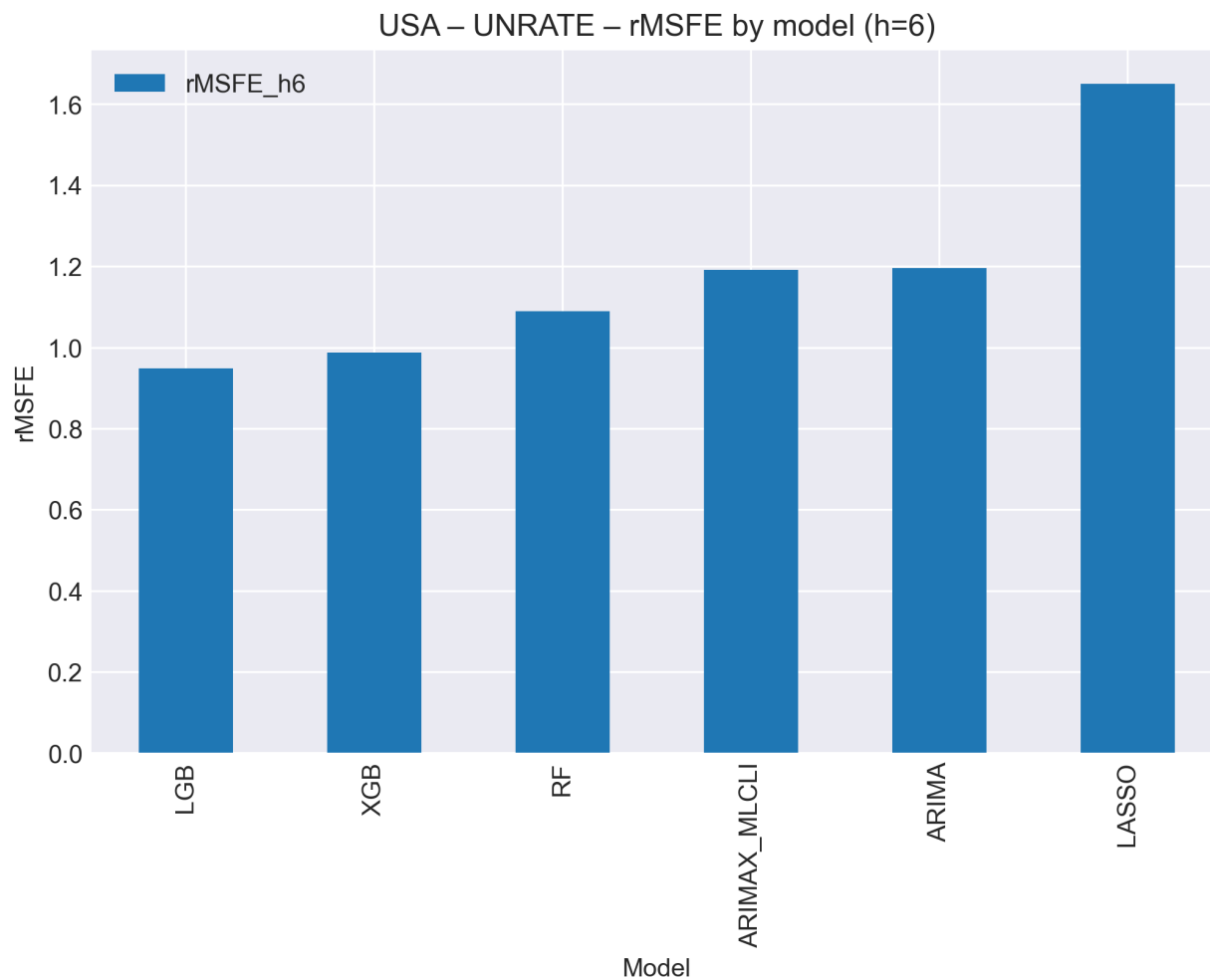


Figure 2.3: *United States Unemployment Rate: (rMSFE) between models using the OECD Composite Leading Indicator (CLI) and the reweighted CLI constructed via feature selection.*

Table 2.10: *Diebold–Mariano tests for USA (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-0.442 (0.658)
Inflation	6	LASSO	5.892 (0.0)**
Inflation	6	LGB	1.978 (0.048)**
Inflation	6	RF	2.717 (0.007)**
Inflation	6	XGB	1.502 (0.133)
Unemployment	6	ARIMAX_MLCLI	-0.169 (0.866)
Unemployment	6	LASSO	3.048 (0.002)**
Unemployment	6	LGB	-1.019 (0.308)
Unemployment	6	RF	-0.502 (0.616)
Unemployment	6	XGB	-0.89 (0.374)
Industrial Production	6	ARIMAX_MLCLI	-0.735 (0.463)
Industrial Production	6	LASSO	11.543 (0.0)**
Industrial Production	6	LGB	7.764 (0.0)**
Industrial Production	6	RF	8.007 (0.0)**
Industrial Production	6	XGB	8.002 (0.0)**

Note: The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level.

For the United States we report $h = 6$ rMSFE. Figure 2.1 for the United States CPI, Figure 2.2 industrial production, and Figure 2.3 unemployment. The ARIMAX specification using the ML-weighted composite reduces rMSFE relative to the ARIMA baseline for all three targets, with the largest gains occurring for industrial production. Table C.9 reports Diebold–Mariano tests; for several model comparisons the null of equal predictive accuracy is

rejected, indicating that the observed differences are not solely sampling noise. These results suggest that a data-driven reweighting of CLI components can improve forecast performance without altering the forecasting equation—only the construction of the composite.

Figures 2.1– 2.3 summarize out-of-sample forecast accuracy (rMSFE) for the United States at the six-month horizon ($h = 6$) for inflation (CPI), industrial production (IND), and unemployment (UNRATE). Two broad patterns emerge. First, the performance ranking depends strongly on the target variable: unemployment forecasts benefit most from flexible nonlinear machine-learning models, while CPI and industrial production forecasts are best (or nearly best) using parsimonious time-series specifications augmented by the ML-weighted composite (ARIMAX_MLCLI). Second, LASSO performs poorly in these $h = 6$ comparisons, consistent with a setting where the relevant predictive relationships are not well-captured by a sparse linear mapping from the component set.

For unemployment at $h = 6$, the best-performing models are gradient-boosted trees (LightGBM and XGBoost), with Random Forest also outperforming the ARIMA/ARIMAX family. In contrast, ARIMAX_MLCLI is close to the ARIMA benchmark but does not rank among the top performers. This suggests that, at longer horizons, the predictive content for U.S. unemployment is likely driven by nonlinear interactions in the broader information set rather than by a single composite index entering linearly as an exogenous regressor. Practically, this implies that for medium-term unemployment forecasting, a forecaster may prefer a nonlinear ML approach (boosting) and treat the composite index as a secondary diagnostic rather than the primary driver of accuracy improvements.

For industrial production at $h = 6$, ARIMAX_MLCLI attains the lowest rMSFE and is closely followed by the ARIMA benchmark. The tree-based ML models (RF/XGB/LGB) perform substantially worse in this setting, and LASSO performs worst. This pattern is consistent with the idea that, at the six-month horizon for U.S. industrial production, a low-dimensional cyclical signal (captured by the composite) is useful, while highly flexible nonlinear learners may struggle due to regime changes, noise, or limited effective sample size at long horizons. From an applied perspective, the result supports the use of a transparent composite indicator as a leading input for medium-term activity forecasting in the U.S., with

ML methods offering less benefit (and potentially overfitting) in this particular target–horizon pairing.

For CPI inflation at $h = 6$, ARIMAX_MLCLI again delivers the lowest rMSFE, with ARIMA a close second. Tree-based ML models are moderately less accurate than the ARIMA/ARIMAX_MLCLI specifications, and LASSO performs worst. This suggests that the ML-weighted composite captures incremental predictive content for inflation relative to a pure autoregressive benchmark at this horizon, but that the gain is modest. Interpreting this for policy practice, the composite-based ARIMAX provides a stable forecasting tool that can be communicated easily, while the more flexible ML models do not consistently dominate for inflation at this horizon.

Taken together, these results highlight that the value of reweighting CLI components is target-dependent. The ML-weighted composite is most useful as a forecasting input for CPI and industrial production at $h = 6$, where ARIMAX_MLCLI is best (or essentially tied for best). For unemployment, however, boosted-tree models perform better, indicating that nonlinearities and interactions likely play a larger role in medium-term labor-market forecasts. For an operational forecaster (e.g., a central bank), this suggests a pragmatic “dashboard” strategy: use ARIMAX_MLCLI as a strong, transparent baseline for CPI and industrial production, while relying on boosting-based ML models for unemployment forecasts and using composite indicators primarily for monitoring and interpretability.

Using the out-of-sample forecast series, we compare ARIMA to ARIMAX with the ML-weighted composite (MLCLI). Average accuracy improves slightly: RMSE falls from 1.370 (ARIMA) to 1.362 (ARIMAX_MLCLI). However, the largest forecast errors are concentrated in early 2009, when inflation dynamics shift sharply during the Great Recession. In these tail months, both models overpredict inflation and tail RMSE is similar (about 4.05 for ARIMA versus 4.08 for ARIMAX_MLCLI). Thus, for U.S. CPI at the six-month horizon, reweighting the CLI improves typical-month performance modestly but does not eliminate the largest turning-point misses. This pattern is consistent with the idea that extreme inflation regime changes are difficult for low-dimensional composites to capture at medium horizons, even when weights are forecast-optimized.

Figure 2.4 presents the out-of-sample forecast errors for the United States. Several patterns emerge. First, forecast errors are generally centered around zero for CPI, indicating relatively stable predictive performance across models. In contrast, the unemployment rate exhibits a pronounced spike around 2020, reflecting the substantial forecasting challenges posed by the COVID-19 shock.

Industrial production displays larger dispersion in forecast errors, suggesting that this variable is inherently more difficult to predict, particularly during periods of economic volatility. The forecast errors for the other countries can be found in Appendix C.



Figure 2.4: *USA Forecast Errors*

2.5.1 Interpretation and implications for forecasting practice

The objective of this chapter is practical: if a central bank already uses the OECD CLI (or a similar composite) as an exogenous predictor or monitoring tool, should it *change anything* about how it uses that composite for forecasting? The results provide three lessons. First, the OECD CLI contains useful predictive information in many settings, but its *fixed* component weights are not always forecast-optimal for specific targets and horizons. Second, a transparent reweighting of the existing component set can reduce out-of-sample forecast errors without requiring a more complex forecasting equation. Third, improvements are most meaningful when they (i) persist across horizons and (ii) correspond to visibly improved predictions during economically relevant episodes (turning points and high-volatility regimes), not merely marginal changes in average rMSFE.

The OECD weights are designed for stability and comparability across time and countries. Our reweighting instead targets forecast accuracy for a given country, target variable, and horizon. This matters because the most informative leading signals differ across targets: components that help forecast industrial production may not be the same components that help forecast unemployment or inflation. In practice, this implies that a “one-size-fits-all” composite can leave forecast gains on the table, particularly at short-to-medium horizons where leading indicators should be most informative.

The rMSFE comparisons for ARIMAX models using the OECD CLI versus the ML-weighted CLI. DM tests assess whether the differences are statistically meaningful. However, for policy users, the key question is not only whether rMSFE improves on average, but *when* improvements occur. To address this, Appendix C plot forecast errors over time for ARIMAX_OECD and ARIMAX_MLCLI for representative country–target pairs. These error plots show that forecast improvements are not uniformly distributed: they tend to arise around business-cycle turning points and periods of heightened uncertainty, when the relative importance of specific components shifts. In these episodes, the ML-weighted composite often reduces error persistence (how long the model remains systematically wrong) even when the peak error cannot be fully eliminated.

Consider the U.S. at the six-month horizon. For CPI and industrial production, ARIMAX_MLCLI achieves the lowest (or nearly lowest) RMSFE among the ARIMA/ARIMAX family, while unemployment forecasts at this horizon are better served by nonlinear ML methods. The implication is not that a central bank should abandon its baseline time-series models. Rather, it should treat composite indicators as *forecasting inputs that can be updated and tailored by objective*. When the forecasting task is inflation or production at horizons where cyclical signals dominate, a reweighted composite provides a low-cost improvement: it changes only the index construction, not the forecasting infrastructure. When the task is unemployment at longer horizons, nonlinear methods may capture interactions that a linear composite cannot, so the composite is best used as a monitoring tool or as one input among several.

The chapter supports a simple operational recommendation: keep the OECD CLI as a stable benchmark for monitoring, but maintain a parallel “forecast-optimized” version that is periodically re-estimated using only information available at the time. In a central-bank setting, this corresponds to an additional row in the forecasting dashboard: OECD CLI vs. ML-weighted CLI, each fed into the same ARIMAX specification. If the ML-weighted composite delivers persistent RMSFE reductions and occasionally encompasses the OECD forecast (as indicated by forecast encompassing regressions), then the forecaster has evidence that the fixed weights can be improved without sacrificing interpretability.

Chapter 3

Building a New Composite Leading indicator with Machine Learning

3.1 Introduction

Composite leading indicators (CLIs), such as those produced by the OECD and the Conference Board, are designed to provide early signals of cyclical turning points and near-term movements in aggregate activity. A key advantage of CLIs is transparency: they combine a small set of component series—typically surveys, real-activity proxies, and financial indicators—using predetermined transformations and aggregation rules. The classic leading-indicator forecasting approach emphasizes the practical value of such composites for monitoring and short-horizon forecasting, while also noting challenges related to stability across regimes and real-time performance ([Diebold and Rudebusch, 1991](#); [Klein and Moore, 1983](#)). The OECD’s specific CLI construction procedure is documented in detail by [Gyomai and Guidetti \(2012\)](#), and recent work evaluates the OECD CLI’s ability to anticipate recessions across countries ([Camacho and Palmieri, 2021](#)).

However, the same design features that make CLIs interpretable—such as fixed component selection and predetermined weights—can also constrain their forecasting performance. Because these weights are not updated in response to changing economic conditions, CLIs may

fail to capture evolving relationships between predictors and target variables, particularly during periods of structural change (Camacho and Palmieri, 2021; Stock and Watson, 2002a). Recent evidence suggests that more flexible, data-driven approaches, including machine learning models, can yield improved forecasting performance in such environments (Medeiros et al., 2021). More specifically, CLIs may be limited in three ways. First, component selection is expert-driven and intentionally limited, which can exclude predictive information available in modern macroeconomic databases. Second, fixed or judgmental weights may be suboptimal under structural change and cross-country heterogeneity—an issue emphasized by evidence of instability in macroeconomic relationships (Stock and Watson, 1996) and directly explored in Chapter 2. Third, the variables and weights that best signal real activity may differ from those that help predict inflation or labor market conditions, implying that a single fixed-weight composite is unlikely to be uniformly optimal across targets and horizons.

In parallel, the macroeconomic forecasting literature has shown that using large information sets can improve prediction accuracy by summarizing or selecting useful signals from many indicators. In particular, diffusion index and principal-components approaches demonstrate measurable gains in forecasting performance when the information set expands (Stock and Watson, 2002a,b), with applications in policy environments described as “data-rich” (Bernanke and Boivin, 2003). This perspective also underpins modern nowcasting frameworks that exploit a broad set of monthly indicators to predict key macro outcomes (Higgins, 2014; Lahiri and Monokroussos, 2013). More recently, research finds that machine learning (ML) methods—especially regularized regressions and tree-based ensembles—can outperform standard benchmarks for macro variables in data-rich settings, particularly when nonlinearities and interactions matter (Goulet Coulombe et al., 2022; Masini et al., 2023; Medeiros et al., 2021). Applied evidence for specific targets includes unemployment forecasting with regularization (Smalter Hall, 2018) and inflation forecasting with ML methods across different environments (Araujo and Gaglianone, 2020; Medeiros et al., 2021; Mirza et al., 2024; Naghi et al., 2024).

Motivated by these developments, this chapter proposes a data-driven method for constructing a new composite index—denoted the **Machine Learning Composite Leading**

indicator (MLCI)—that is not restricted to the OECD CLI component basket. The MLCI is constructed from a large pool of OECD indicators using supervised learning to identify and weight predictors that best forecast monthly inflation, unemployment, and industrial production at horizons $h \in \{1, 3, 6\}$. The contribution is threefold. First, we provide a scalable, country-specific procedure for building composite leading indicators from large, standardized databases. Second, we maintain transparency by using explainable ML tools (SHAP values) to quantify the contribution of each indicator to the composite (Lundberg and Lee, 2017). Third, we formally test whether MLCI adds predictive content beyond the OECD CLI using out-of-sample forecast comparisons (Diebold and Mariano, 1995) and forecast encompassing tests (Harvey et al., 1998).

This approach is closely related to dynamic factor models (Stock and Watson, 2002), which extract common components from large datasets. However, instead of relying on linear dimension reduction techniques such as principal components, we use machine learning methods to select and weight predictors in a potentially nonlinear and time-varying manner.

Additionally, the proposed MLCI can be interpreted as a data-driven forecast combination device, where multiple predictors are aggregated to minimize out-of-sample forecast error. (Bates and Granger, 1969; Timmermann, 2006).

3.2 Data and Variable Selection

We use the OECD API (via `pyOECD`; Appendix B) to retrieve monthly indicators across key domains: sectoral activity, prices and expectations, labor markets, financial conditions, trade/external variables, and sentiment. For each country, the candidate pool typically contains 50–100 series. The candidate variables are drawn from standard macroeconomic and financial time series commonly used in business cycle analysis and forecasting. These include indicators related to real activity (e.g., industrial production, retail sales), labor markets (e.g., unemployment rates, employment levels), prices (e.g., consumer and producer price indices), financial conditions (e.g., interest rates, stock market indices), and external sectors (e.g., trade balances and exchange rates).

The selection of variables is guided by the economic literature on leading indicators and data-rich forecasting environments, which emphasizes the value of incorporating a broad set of predictors to capture different dimensions of economic activity (Forni et al., 2005; Lahiri and Monokroussos, 2013; Stock and Watson, 2002a). Data are obtained from publicly available sources such as the OECD Main Economic indicators database and national statistical agencies, ensuring consistency and comparability across countries. Using large indicator sets is consistent with evidence that macro forecasts improve when systematic information from many predictors is incorporated (Bernanke and Boivin, 2003; Stock and Watson, 2002a,b).

To ensure comparability, only variables available for the full sample period and across all countries are retained. All series are aligned to a common monthly calendar and transformed to promote stationarity and comparability across units. Price and real-activity indices are generally used in log differences (or first differences when logs are inappropriate), while rate variables (e.g., unemployment rate) are evaluated in levels and/or changes depending on the forecasting specification. Predictors are standardized to mean zero and unit variance within each estimation window to stabilize regularization and tree-based learning. Because real-world forecasting can be sensitive to revisions and vintages, we note that real-time data considerations may affect performance and interpretation (Croushore, 2011); however, this chapter focuses on a consistent-vintage evaluation using OECD releases.

3.3 Methodology

All models are estimated using rolling window schemes to preserve the time-series structure of the data. Hyperparameters are selected using blocked cross-validation, ensuring that training and validation sets respect temporal ordering.

For tree-based models, early stopping is implemented based on validation loss. Feature selection and model estimation are repeated at each forecast origin, allowing the MLCI to adapt over time.

3.3.1 Indicator Construction

The Machine Learning Composite indicator (MLCI) is constructed as a time-varying, data-driven linear combination of selected predictors:

$$MLCI_t^{(y)} = \sum_{j \in \mathcal{S}_t} w_{j,t}^{(y)} x_{j,t}$$

where $\mathcal{S}_t$ denotes the set of selected predictors at time t , and $w_{j,t}^{(y)}$ are weights derived from model-specific importance measures.

For linear models (LASSO), weights correspond directly to estimated coefficients. For tree-based models, weights are constructed from normalized SHAP values, averaged over the training sample to ensure stability. This formulation ensures comparability with traditional composite indicators while allowing for data-driven selection and weighting.

Let $x_t = (x_{1,t}, \dots, x_{p,t})'$ denote the standardized predictor pool for a given country at time t . For a target variable y and forecast horizon h , the forecasting task is to approximate the conditional expectation

$$\hat{y}_{t+h} = f^{(y,h)}(x_t),$$

where $f^{(y,h)}(\cdot)$ is estimated using a training sample and evaluated out-of-sample. The goal of the MLCI is to convert the information contained in the high-dimensional x_t into a *single composite statistic* that is predictive of y_{t+h} .

We consider a set of learners that is sufficiently broad to capture standard benchmark performance and documented ML improvements in macroeconomic forecasting: (i) penalized linear models (LASSO), which provide sparse variable selection and are a standard approach in high-dimensional forecasting (Masini et al., 2023; Tibshirani, 1996); and (ii) tree-based ensembles (Random Forest, XGBoost, LightGBM), which capture nonlinearities and interactions and have demonstrated gains in macro forecasting contexts (Breiman, 2001; Chen and Guestrin, 2016; Goulet Coulombe et al., 2022; Ke et al., 2017). These choices are motivated by evidence that nonlinearities and regularization are key mechanisms behind ML forecasting gains (Goulet Coulombe et al., 2022).

To retain interpretability, we use SHAP values to compute consistent feature attributions for tree-based learners (Lundberg and Lee, 2017). For each trained model, we compute global importance measures $\phi_j^{(y,h)}$ (e.g., mean absolute SHAP contributions). The horizon- and target-specific MLCI is then defined as an importance-weighted composite:

$$\text{MLCI}_t^{(y,h)} = \sum_{j \in \mathcal{S}_{y,h}} \tilde{\phi}_j^{(y,h)} x_{j,t}, \quad \tilde{\phi}_j^{(y,h)} = \frac{\max\{\phi_j^{(y,h)}, 0\}}{\sum_{\ell \in \mathcal{S}_{y,h}} \max\{\phi_\ell^{(y,h)}, 0\}},$$

where $\mathcal{S}_{y,h}$ denotes the selected subset of predictors (e.g., top- k by importance or those selected by LASSO).

SHAP values are used in this chapter not only as a machine-learning interpretation tool, but also as a way to make the proposed index economically meaningful. A central concern with machine-learning forecasts in macroeconomics is that improved predictive accuracy may come at the cost of interpretability. For economists, especially those working in policy institutions, a forecast is more useful when it can be connected to recognizable economic mechanisms. SHAP values help address this concern by identifying which variables contribute most to the model’s predictions and by showing whether those contributions are positive or negative for a given forecast. This is particularly relevant for central bank economists and other policy analysts who consume forecasts as part of a broader decision-making process. A central bank economist may care not only whether inflation is forecast to rise, but also whether the forecast is being driven by labor-market tightness, financial conditions, survey expectations, commodity-sensitive indicators, or measures of real activity. SHAP values provide a structured way to decompose the model’s forecast into contributions from individual predictors. This allows the economist to assess whether the forecast is consistent with an economically plausible narrative, whether it reflects broad-based macroeconomic pressure, or whether it is driven by a small number of volatile indicators.

In this dissertation, SHAP values also serve a second purpose: they provide the weighting logic used to construct the machine-learning-based composite leading indicator. Rather than treating the machine-learning model as a black box, the SHAP-based decomposition identifies

the indicators that repeatedly contribute to forecast performance across countries, targets, and horizons. The resulting MLCI is therefore not only evaluated by its forecast accuracy, but also by the economic content of the variables that receive the greatest importance. This makes the index more transparent and more useful for economists who need to understand why a data-driven indicator performs well or poorly.

While SHAP values provide a consistent measure of feature importance, they do not directly correspond to structural coefficients. To construct a stable composite index, we aggregate SHAP values across the training sample and normalize them to obtain relative importance weights. This approach allows us to retain interpretability while approximating a linear representation of the predictive model.

This construction differs from the OECD CLI in two ways. First, the component set is selected based on predictive relevance in a data-rich environment rather than a fixed basket (Gyomai and Guidetti, 2012). Second, the resulting composite is **target- and horizon-specific**, allowing the information content to differ between inflation, unemployment, and industrial production, and across h .

3.3.2 Forecasting and Model Comparison

We evaluate three forecasting specifications: (i) a baseline model that excludes any composite index (e.g., an ARIMA benchmark), (ii) a model augmented with the OECD CLI, and (iii) a model augmented with the MLCI. Forecast performance is measured by the root mean squared forecast error (rMSFE),

$$\text{rMSFE}(h) = \sqrt{\frac{1}{T_h} \sum_{t=1}^{T_h} (y_{t+h} - \hat{y}_{t+h})^2},$$

computed on a common out-of-sample period for all models and horizons. Using rMSFE is standard in forecasting evaluation and places greater weight on large forecast errors, which are particularly costly around recessions and high-volatility episodes.

To assess whether differences in accuracy are statistically meaningful, we apply Diebold–

Mariano tests of equal predictive accuracy based on squared-error loss differentials (Diebold and Mariano, 1995). For multi-step horizons, we use HAC-robust variance estimates to account for serial correlation in forecast errors, following standard practice in forecast evaluation.

3.3.3 Forecast Encompassing Test

An important distinction is whether MLCI simply improves accuracy relative to CLI in an average-loss sense or whether it contains *incremental predictive information* beyond the CLI. To test incremental information, we conduct forecast encompassing tests by estimating:

$$y_{t+h} = \alpha + \beta_1 \hat{y}_{\text{CLI},t+h} + \beta_2 \hat{y}_{\text{MLCI},t+h} + \varepsilon_t.$$

The forecast encompassing test provides a stringent evaluation of the informational content of the MLCI relative to the OECD CLI. Unlike simple forecast comparisons, encompassing tests determine whether one model contains all relevant predictive information from another.

This is particularly important in this context, as the MLCI is designed to generalize the information contained in the CLI rather than simply compete with it. Under this framework, if β_2 is statistically significant while β_1 is not, then the MLCI encompasses the CLI (the CLI adds no additional information once MLCI is included). If both coefficients are significant, the two forecasts contain complementary information. If only β_1 is significant, the CLI encompasses the MLCI. This test provides a formal basis for whether MLCI should replace or complement CLI in forecasting practice. This approach builds on the seminal contribution of (Fair and Shiller, 1990; Harvey et al., 1998).

3.4 Results and Discussion

Table 3.1 summarizes the forecast comparison between the OECD CLI and the machine learning composite indicator (MLCI) for the United States at the six-month horizon. The results reveal substantial heterogeneity across targets. For CPI, the MLCI provides only a modest reduction in MSE relative to the CLI, and the encompassing regression does not

Table 3.1: *United States forecast comparison at horizon $h = 6$: CLI versus MLCI*

Target	% Improve MSE	Adj. R^2 gain	$\hat{\beta}_{MLCI}$	p -value
Inflation	1.86	-0.0007	-0.5318	0.4888
Industrial Production	-23.75	-0.0003	0.0607	0.3732
Unemployment	62.14	0.5945	0.8083	0.0000

Notes: Percentage improvement in MSE is computed as $100 \times (MSE_{CLI} - MSE_{MLCI})/MSE_{CLI}$, so positive values indicate better performance of the MLCI relative to the OECD CLI. The adjusted R^2 gain is the difference between the encompassing regression including both CLI and MLCI forecasts and the regression using the CLI forecast alone.

indicate meaningful incremental explanatory power. For industrial production, the MLCI performs worse than the CLI, both in terms of forecast accuracy and adjusted R^2 contribution. In contrast, the unemployment results are markedly different: the MLCI reduces MSE by more than 60% relative to the CLI and produces a large increase in adjusted R^2 , with a statistically significant coefficient in the encompassing regression. These findings suggest that the gains from the MLCI are concentrated in specific targets rather than being uniform across macroeconomic variables.

To illustrate why the MLCI improves forecast performance in some settings but not in others, Figures 3.3 and 3.2 plot forecast errors over time for representative cases in which the MLCI respectively outperforms and underperforms the OECD CLI. The top panels compare the forecast errors from the two models directly, while the bottom panels plot the difference in squared forecast errors, $e_{CLI,t}^2 - e_{MLCI,t}^2$. Positive values therefore indicate periods in which the MLCI delivers lower squared forecast error than the CLI.

In the unemployment case, the improvement from the MLCI is concentrated in specific episodes in which the CLI errors become large, suggesting that the machine-learning composite captures variation that is particularly relevant during periods of labor-market adjustment. In contrast, in the industrial production case the squared-error differential is frequently negative, indicating that the additional flexibility of the MLCI does not translate into better forecasting performance and may instead introduce noise. These figures reinforce the broader result that the value of the MLCI is target-specific rather than uniform across macroeconomic variables.

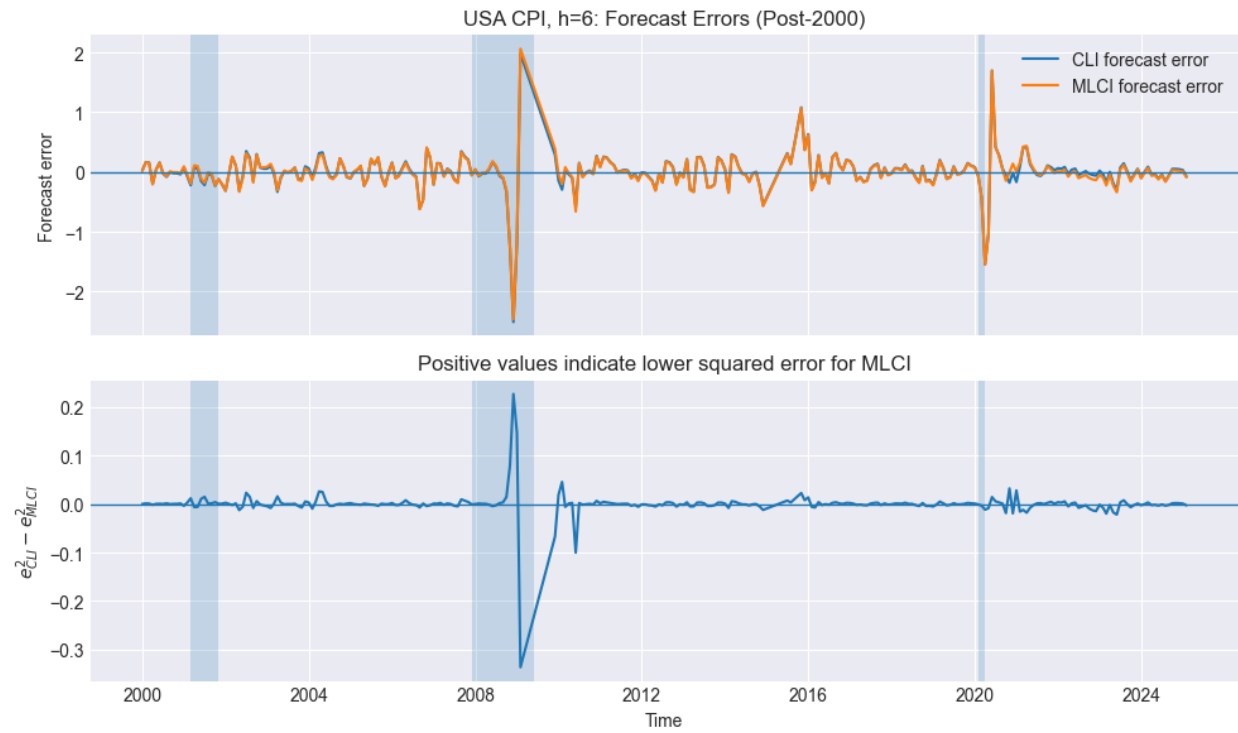


Figure 3.1: *Forecast error comparison for a representative case in which the MLCI performance is mixed relative to the OECD CLI. Negative values in the lower panel indicate periods in which the OECD CLI yields lower squared forecast error than the MLCI.*

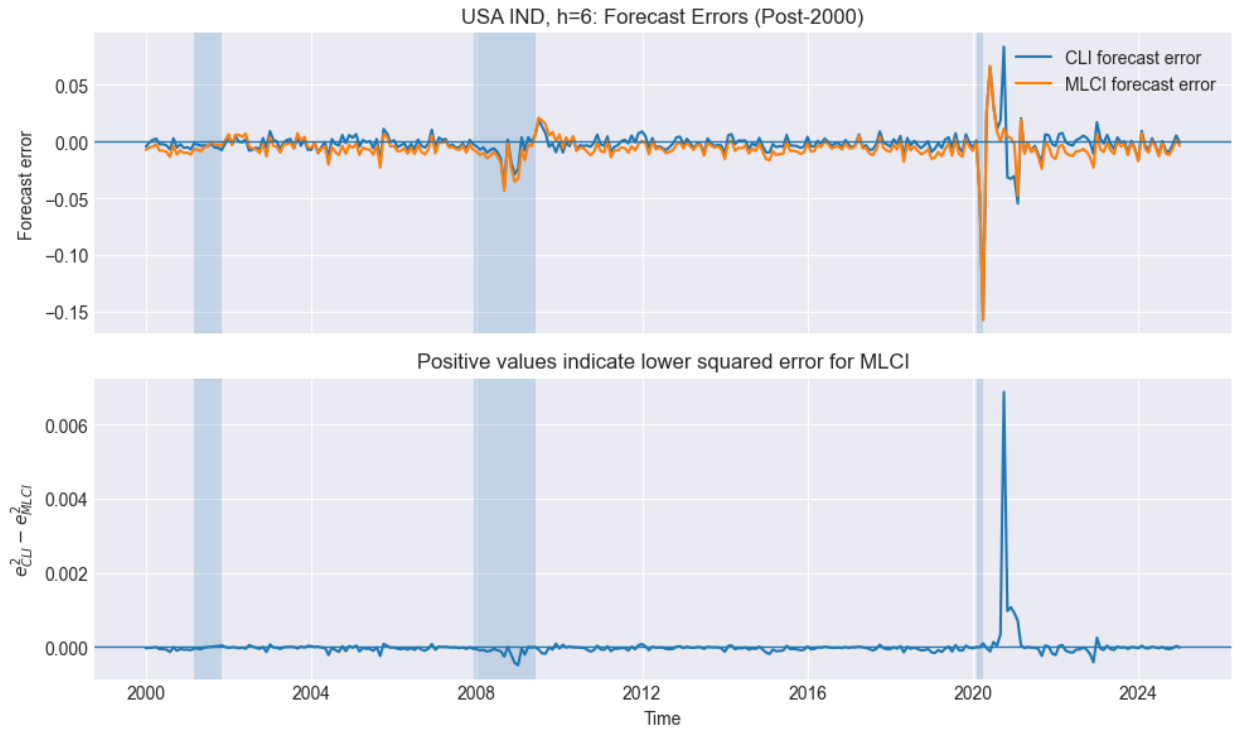


Figure 3.2: *Forecast error comparison for a representative case in which the MLCI underperforms the OECD CLI. Negative values in the lower panel indicate periods in which the OECD CLI yields lower squared forecast error than the MLCI.*

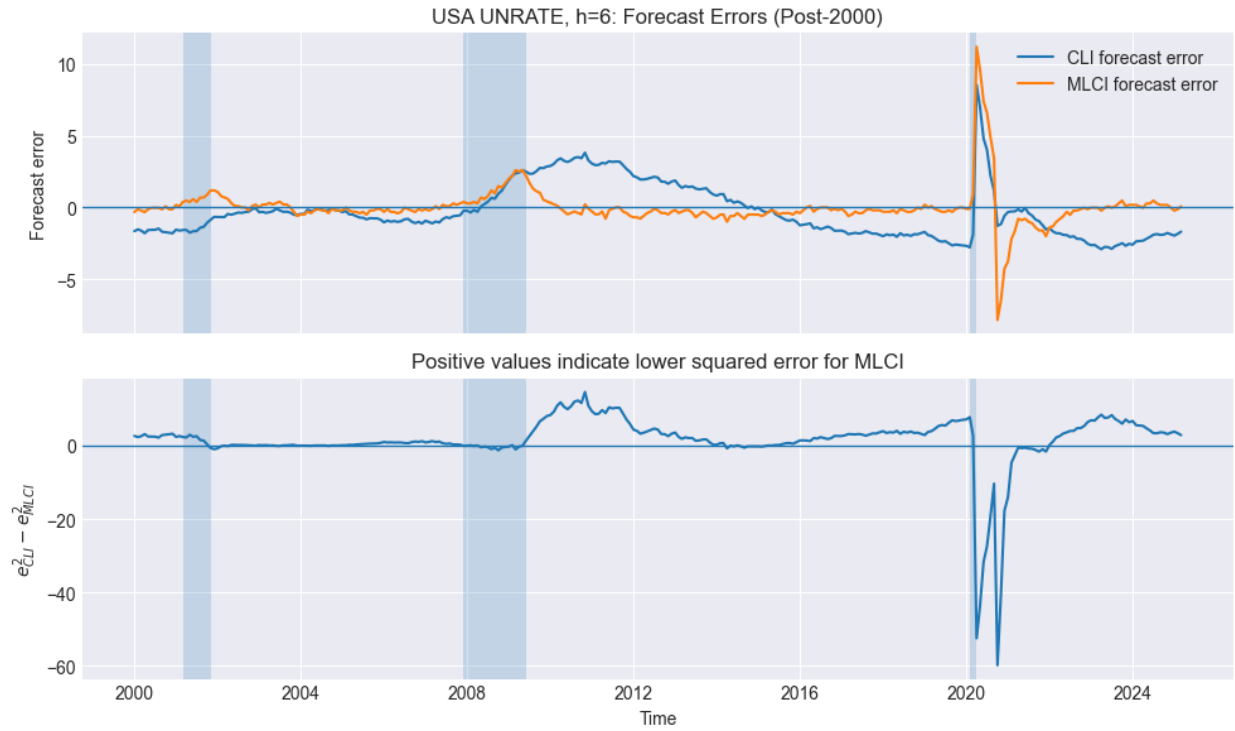


Figure 3.3: *Forecast error comparison for a representative case in which the MLCI outperforms the OECD CLI. The lower panel plots the difference in squared forecast errors, so positive values indicate periods in which the MLCI improves forecast accuracy.*

A notable result is the relatively weak and inconsistent performance of the MLCI for CPI inflation forecasting. In many cases, the MLCI does not significantly outperform the OECD CLI or even simpler benchmarks. This finding is consistent with the broader macroeconomic forecasting literature, which has repeatedly documented the difficulty of predicting inflation, particularly at short horizons. (Stock and Watson, 2007) Inflation dynamics are often influenced by idiosyncratic shocks, policy interventions, and measurement noise, which reduce the predictive content of leading indicators (Faust and Wright, 2013; Stock and Watson, 2007). As a result, even flexible machine learning approaches may yield limited gains in this domain. Additionally, it appears that improvements from machine learning and composite indicators are more likely to arise in variables with stronger cyclical structure, such as unemployment and industrial production, rather than inflation.

3.4.1 Forecast Accuracy Across Models and Horizons

Table 3.3 reports the share of significant Diebold–Mariano tests when comparing forecasts generated with the machine-learning composite leading indicator (MLCI) against forecasts using the OECD CLI. Consistent with the broader summary results, MLCI delivers the most reliable improvements for unemployment across countries and horizons. For CPI inflation, gains are modest and uneven, while for industrial production the evidence is mixed, with improvements concentrated in a subset of countries and horizons. Overall, the evidence suggests that MLCI most reliably adds incremental predictive content beyond the CLI for labor-market outcomes.

Table 3.2 summarizes the performance of the machine learning composite indicator (MLCI) relative to the OECD CLI across countries, targets, and forecast horizons. The results reveal substantial heterogeneity across targets. For CPI, the MLCI generally does not improve forecast accuracy, with negative median improvements in MSE across all horizons. For industrial production, the gains are modest and inconsistent, with only limited evidence of statistically significant contributions. A further explanation for the weak forecast performance is the relatively short effective sample available for inflation forecasting. Although the monthly

Table 3.2: *Summary of MLCI performance relative to the OECD CLI across countries*

Target- h	Wins	Med. $\% \Delta \text{MSE}$	Mean. $\% \Delta \text{MSE}$	$\Delta \text{Adj. } R^2 > 0$	$p < 0.05$
Inflation-1	3/9	-0.25	0.01	44.4	22.2
Inflation-3	5/9	0.04	-0.47	44.4	22.2
Inflation-6	5/9	0.18	-0.30	44.4	22.2
Industrial Production-1	2/9	-5.36	-5.02	55.6	33.3
Industrial Production-3	5/9	3.44	1.94	55.6	22.2
Industrial Production-6	5/9	3.80	-0.72	44.4	0.0
Unemployment-1	8/9	80.74	55.52	100.0	100.0
Unemployment-3	8/9	79.35	50.89	100.0	88.9
Unemployment-6	6/9	62.14	44.79	88.9	88.9

Notes: “Wins” denotes the number of countries in which the MLCI achieves a lower MSE than the OECD CLI. $\% \Delta \text{MSE}$ is computed as $100 \times (MSE_{CLI} - MSE_{MLCI}) / MSE_{CLI}$. Positive values indicate improved performance of the MLCI. The last two columns report the share of countries with a positive adjusted R^2 gain and statistically significant MLCI coefficients ($p < 0.05$) in the encompassing regression.

data begin in 2000, the usable sample is reduced by transformations, lag construction, forecast horizons, rolling-window estimation, and the need to reserve observations for out-of-sample evaluation. This limitation is especially important for machine-learning methods, which generally require large amounts of stable and informative training data to estimate nonlinear relationships reliably. In small macroeconomic samples, flexible models such as random forests, XGBoost, and LightGBM may fit noise rather than persistent economic structure, particularly when the target variable is volatile or subject to regime changes. [Chen and Guestrin \(2016\)](#) This issue is likely more severe for inflation than for some other macroeconomic targets because inflation dynamics changed substantially over the sample period. The post-2000 period includes the Great Recession, a prolonged low-inflation environment, the COVID-19 shock, supply-chain disruptions, and the subsequent inflation surge. These episodes create instability in the relationship between predictors and future inflation. As a result, the available sample may be too short for the machine-learning models to distinguish durable predictive signals from episode-specific relationships. Therefore, weak forecast performance should not be interpreted only as evidence that the selected indicators lack predictive content. It may also reflect the difficulty of applying data-hungry forecasting methods to short and structurally unstable macroeconomic samples.

In contrast, the results for unemployment are striking. The MLCI outperforms the CLI in nearly all country cases, with large median reductions in MSE ranging from roughly 62% to 81% across horizons. These improvements are accompanied by consistently positive gains in adjusted R^2 and statistically significant coefficients in the encompassing regressions. This pattern suggests that the machine learning composite captures predictive information for unemployment that is not fully incorporated in the traditional CLI.

The patterns documented in Table 3.2 provide insight into the conditions under which the machine learning composite indicator (MLCI) delivers improvements over the OECD CLI. The gains from the MLCI are not uniform across targets, but instead appear to depend on the underlying data-generating process of each macroeconomic variable.

For inflation, the results suggest that the information captured by the OECD CLI is already sufficient to explain much of the predictable variation, consistent with the relatively stable and well-understood relationships underlying price dynamics. In this setting, the additional flexibility of the MLCI does not translate into meaningful improvements in forecasting performance.

In contrast, the results for unemployment indicate that the MLCI captures predictive relationships that are not fully incorporated in the traditional CLI. Labor market dynamics are influenced by a broader and more complex set of factors, including nonlinear interactions, structural changes, and varying adjustment speeds across sectors. The machine learning framework is better suited to capture these high-dimensional and potentially nonlinear relationships, which may explain the substantial and consistent gains observed across countries and forecast horizons.

These findings highlight a key contribution of this chapter: data-driven composite indicators constructed using machine learning can outperform traditional leading indicators when the underlying economic relationships are complex and not well approximated by linear combinations of a fixed set of variables. The MLCI addresses this limitation by allowing the data to determine both the relevant predictors and their functional form, leading to substantial gains in forecasting performance for variables such as unemployment.

Table 3.3: *Share of significant DM tests by comparison, target, and horizon*

target	h	sig-share
Inflation	1	0.333
Inflation	3	0.333
Inflation	6	0.000
Industrial Production	1	0.125
Industrial Production	3	0.125
Industrial Production	6	0.125
Unemployment	1	0.750
Unemployment	3	0.625
Unemployment	6	0.625

Appendix C summarizes out-of-sample rMSFE by country, target, and horizon for (i) ARIMA, (ii) ARIMAX with the OECD CLI, and (iii) ARIMAX with the ML-weighted CLI (MLCI). Three main findings emerge.

First, augmenting the baseline with composite information reduces forecast errors in most cases. The gains are largest for unemployment and industrial production and are front-loaded at short horizons ($h=1$).

Second, replacing the fixed-weight CLI with the ML-weighted index (MLCI) delivers additional reductions in rMSFE. In several country–target pairs (e.g., CPI at $h=3$ and ind at $h=1$), the MLCI improvement is statistically significant by Diebold–Mariano tests (see Appendix C).

Third, benefits vary by country. Data-rich economies exhibit clearer and more consistent gains, whereas smaller samples and noisier indicators attenuate improvements and statistical significance.

3.4.2 Horizon Profiles

Across countries, the relative advantage of CLI/MLCI is strongest at $h=1$, persists at $h=3$, and generally narrows by $h=6$. This pattern is consistent with composite indicators capturing near-term cyclical information that diffuses as the forecast horizon expands.

3.4.3 Model Ranking

Among ML approaches, gradient boosting (LightGBM/XGBoost) generally attains the lowest rMSFE, followed by Random Forest, with LASSO competitive against ARIMA. The MLCI—constructed via ML weighting—tends to outperform the fixed CLI within otherwise comparable forecasting setups.

3.4.4 Statistical Significance

Diebold–Mariano (DM) tests corroborate the economic significance of the rMSFE patterns. Short-horizon gains (particularly for Industrial Production at $h=1$ and CPI at $h=3$) are frequently significant at conventional levels, while longer-horizon differences are smaller and less often significant. Full DM statistics and p -values are reported in [Appendix C](#).

3.4.5 Robustness and Sensitivity

Results are qualitatively unchanged under rolling vs. expanding estimation windows, alternative initial sample sizes, and plausible hyperparameter grids (learning rates, tree depth/leaves, LASSO penalty sequences). Alternative transformations (e.g., alternative seasonal filters, log levels vs. differences for CPI/Industrial Production, level vs. change for Unemployment) deliver similar rankings. Using MAE instead of RMSE yields similar model rankings. Clark–West-type adjustments ([Clark and West, 2007](#)) and MSE-adjusted DM tests for nested models provide concordant inference.

3.4.6 Discussion and Implications

The evidence indicates (i) OECD CLIs contain useful predictive information in many contexts; (ii) data-driven reweighting can improve accuracy without sacrificing transparency; and (iii) ML-constructed composites can deliver further gains and sometimes *encompass* the CLI. For forecasters, this supports using the CLI as a baseline while augmenting it with adaptable, explainable ML composites, particularly for horizons and targets where gains are largest. The stronger performance of tree-based ensembles is consistent with recent findings that nonlinearities are a central source of ML gains in macroeconomic forecasting (Goulet Coulombe et al., 2022; Masini et al., 2023), and with evidence that ML methods improve inflation forecasting in data-rich environments (Medeiros et al., 2021; Naghi et al., 2024). A complete overview of the results for each country can be found in Appendix C.

3.4.7 Limitations

Several limitations should be considered when interpreting the results. First, the analysis is constrained by the length and quality of the available macroeconomic time series. Although monthly data provide more observations than quarterly data, the effective sample is substantially reduced after constructing lags, forecast horizons, rolling estimation windows, and out-of-sample evaluation periods. This is particularly important for machine-learning models, which typically perform best when trained on large datasets with stable relationships between predictors and outcomes. In this setting, the relatively short sample may limit the ability of flexible models to identify persistent predictive patterns.

Second, the sample period includes several major structural disruptions, including the Great Recession, the COVID-19 pandemic, and the inflationary period that followed. These events may have changed the relationship between leading indicators and macroeconomic outcomes. A model trained on one regime may not perform well in another, especially when the underlying economic mechanisms differ across periods. This limitation is important because poor forecast performance may reflect structural instability rather than the absence of useful information in the predictors.

Third, the analysis relies on revised data rather than a fully real-time data environment. In practice, economists and policymakers make forecasts using data that are available at the time, often before later revisions. If revised data differ substantially from real-time releases, the estimated forecast performance may not fully represent the performance that would have been achieved in real-time policy settings. Future work could extend the analysis using real-time vintages where available.

Fourth, the machine-learning-based composite indicators are designed to improve forecast accuracy, but forecast accuracy is only one criterion for evaluating a leading indicator. Policy institutions also value stability, transparency, ease of communication, and robustness across time. A data-driven index that performs well in one sample may be less useful if its composition changes frequently or if its signal is difficult to explain to policymakers. For this reason, the results should be interpreted as evidence on predictive performance rather than as a complete replacement for expert-designed indicators.

Finally, the results may not generalize equally across countries, target variables, and forecast horizons. Countries differ in data quality, economic structure, financial development, and the reliability of leading indicators. Similarly, inflation, unemployment, and industrial production follow different dynamics and may respond to different predictors. The country-specific and target-specific nature of the results suggests that machine-learning-based indicators may be most useful as complements to existing forecasting tools rather than universal replacements for the OECD CLI.

3.4.8 Conclusion of Results

Across countries, targets, and horizons, the CLI improves forecasting accuracy in a sizable minority of cases; reweighting of components provides additional improvements; and the MLCI frequently delivers the strongest accuracy and adds information beyond the CLI. These findings motivate the design of adaptive, explainable indicator systems for policy and business-cycle monitoring. Overall, composite information improves forecasting accuracy; data-driven reweighting improves it further. Gains are largest at short horizons and

for unemployment/industrial production, with statistically significant improvements in a nontrivial share of country–target pairs. The MLCI is a flexible, explainable composite that often outperforms both baseline and fixed-weight CLIs. It adds statistically significant information in many settings and identifies country-specific drivers of future macro outcomes. An important direction for future research is to further investigate which specific features and nonlinear mechanisms drive the gains observed in unemployment forecasting, and whether similar patterns emerge in other macroeconomic contexts.

Bibliography

- Serkan Aras and Paulo J.G. Lisboa. Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, 207:117982, 2022. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2022.117982>. URL <https://www.sciencedirect.com/science/article/pii/S0957417422012106>.
- G. Araujo and W. Gaglianone. Machine learning methods for inflation forecasting: Evidence from brazil. Banco Central do Brasil Working Paper 527, Banco Central do Brasil, 2020. URL <https://www.bcb.gov.br/pec/wps/ingl/wps527.pdf>.
- Kosala Bandara, Christoph Bergmeir, and Hansika Hewamalage. Forecasting monthly unemployment rates using seasonal autoregressive integrated moving average models and artificial neural networks. In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.
- J. M. Bates and C. W. J. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969. doi: 10.1057/jors.1969.103.
- Ben S. Bernanke and Jean Boivin. Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546, April 2003. ISSN 0304-3932. doi: 10.1016/S0304-3932(03)00024-2. URL <https://www.sciencedirect.com/science/article/pii/S0304393203000242>.
- George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5th edition, 2015. Foundational reference for ARIMA models.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Markus Buckmann, Andreas Joseph, Dave Robertson, and Filippo Vigorito. An interpretable machine learning workflow with an application to economic forecasting. Staff Working Paper 984, Bank of England, 2022. URL <https://www.bankofengland.co.uk/working-paper/2022/an-interpretable-machine-learning-workflow-with-an-application-to-economic-forecasting>

Arthur F. Burns and Wesley C. Mitchell. *Measuring Business Cycles*. National Bureau of Economic Research, New York, 1946.

Maximo Camacho and Diego Palmieri. Evaluating the oecd’s main economic indicators at anticipating recessions. *Journal of Forecasting*, 40(1):99–111, 2021. doi: 10.1002/for.2708.

Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, pages 785–794, New York, NY, USA, August 2016. Association for Computing Machinery. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785. URL <https://dl.acm.org/doi/10.1145/2939672.2939785>.

J. E. Choi et al. Vector shap values for machine learning time series forecasting. *Journal of Forecasting*, 2025. Proposes SHAP-style attribution adapted for time-series forecasting with lag vectors.

Todd E. Clark and Kenneth D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311, 2007. doi: 10.1016/j.jeconom.2006.05.023.

Dean Croushore. Frontiers of Real-Time Data Analysis. *Journal of Economic Literature*, 49(1):72–100, March 2011. ISSN 0022-0515. doi: 10.1257/jel.49.1.72. URL <https://www.aeaweb.org/articles?id=10.1257/jel.49.1.72>.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995. doi: 10.1080/07350015.1995.10524523.

- Francis X. Diebold and Glenn D. Rudebusch. Forecasting Output With the Composite Leading Index: A Real-Time Analysis. *Journal of the American Statistical Association*, 86(415):603–610, 1991. ISSN 0162-1459. doi: 10.2307/2290388. URL <https://www.jstor.org/stable/2290388>. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. doi: 10.2307/1912773.
- Ray C. Fair and Robert J. Shiller. Comparing information in forecasts from econometric models. *The American Economic Review*, 80(3):375–389, 1990.
- Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 1996. ISBN 9780412983210.
- Jon Faust and Jonathan H. Wright. Forecasting inflation. *Handbook of Economic Forecasting*, 2:2–56, 2013.
- Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840, 2005. doi: 10.1198/016214504000002050.
- Claudia Goldin. Interview with victor zarnowitz. Interview transcript, Department of Economics, Harvard University and the NBER, December 2001. New York City, December 11, 2001.
- Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964, 2022. ISSN 1099-1255. doi: 10.1002/jae.2910. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2910>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2910>.

- Clive W. J. Granger and Paul Newbold. *Forecasting Economic Time Series*. Academic Press, San Diego, 2 edition, 1986.
- Gábor Gyomai and Elisabetta Guidetti. The oecd system of composite leading indicators. OECD Statistics Working Papers 2012/01, OECD Publishing, 2012.
- James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989. doi: 10.2307/1912559.
- David I. Harvey, Stephen J. Leybourne, and Paul Newbold. Tests for Forecast Encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259, 1998. ISSN 0735-0015. doi: 10.2307/1392581. URL <https://www.jstor.org/stable/1392581>. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- Patrick C. Higgins. GDPNow: A Model for GDP 'Nowcasting'. *SSRN Journal*, 2014. ISSN 1556-5068. doi: 10.2139/ssrn.2580350. URL <http://www.ssrn.com/abstract=2580350>.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- Philip Klein and Geoffrey Moore. The leading indicator approach to economic forecasting - retrospect and prospect. *Journal of Forecasting*, 2:119–135, 1983.
- Kajal Lahiri and Alex Monokroussos. Nowcasting us gdp: The role of leading indicators. *Journal of Forecasting*, 32(4):339–353, 2013.
- Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.
- Ricardo P. Masini, Marcelo C. Medeiros, and Eduardo F. Mendes. Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1):76–111, 2023. ISSN 1467-

6419. doi: 10.1111/joes.12429. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12429>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12429>.
- Michael W. McCracken. Asymptotics for out-of-sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752, 2007. doi: 10.1016/j.jeconom.2006.07.019.
- Marcelo C Medeiros, Giovanna Vasconcelos, Alisson F Veiga, and Eduardo Zilberman. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119, 2021.
- M. Messaoudi and H. Khoudmi. Forecasting gdp growth using machine learning: Evidence from algeria. *International Journal of Sustainable Development and Planning*, 19(8): 1233–1244, 2024. doi: 10.18280/ijstdp.190812.
- N. Mirza et al. Machine learning versus traditional models for inflation forecasting in emerging economies. *International Review of Financial Analysis*, 94:103344, 2024. doi: 10.1016/j.irfa.2024.103344.
- Sendhil Mullainathan and Jann Spiess. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106, May 2017. ISSN 0895-3309. doi: 10.1257/jep.31.2.87. URL <https://www.aeaweb.org/articles?id=10.1257%2Fjep.31.2.87&ref=ds-econ>.
- A. Naghi et al. Do random forests beat traditional benchmarks in u.s. inflation forecasting? *Journal of Applied Econometrics*, 39(7):1189–1212, 2024. doi: 10.1002/jae.3148.
- Rickard Nyman and Paul Ormerod. Predicting Economic Recessions Using Machine Learning Algorithms, January 2017. URL <http://arxiv.org/abs/1701.01428>. arXiv:1701.01428 [q-fin].
- Adrian Pagan and Aman Ullah. *Nonparametric Econometrics*. Cambridge University Press, 1999. ISBN 9780521635673.
- Christopher A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

A. Smalter Hall. Using machine learning to forecast unemployment: An elastic net approach. *Federal Reserve Bank of Kansas City Economic Review*, 2018. URL <https://www.kansascityfed.org/research/economic-review/using-machine-learning-to-forecast-unemployment/>.

James H. Stock and Mark W. Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30, 1996. doi: 10.1080/07350015.1996.10524636.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002a.

James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002b.

James H. Stock and Mark W. Watson. Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33, 2007.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Allan Timmermann. Forecast combinations. In Graham Elliott, Clive W. J. Granger, and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, chapter 4, pages 135–196. Elsevier, 2006. doi: 10.1016/S1574-0706(05)01004-9.

Appendix A

CLI Compositions by Country

Table A.1: *OECD Composite Leading Indicator Components for Canada*

Component Series (Unit)	Source
Deflated Monetary Aggregate M1 sa (cad)	Statistics Canada
Housing starts large cities sa (number)	Statistics Canada
Manufacturing survey - confidence indicator (US...)	Institute of Supply Management USA
Consumer survey - confidence indicator	Conference Board Canada
Spread of interest rates (% p.a.)	Statistics Canada
Ratio of inventories to shipments (ratio) inverted	Statistics Canada
Share prices: S&P/TSX composite index (OECD bas...	Statistics Canada

Table A.2: *OECD Composite Leading Indicator Components for Italy*

Component Series (Unit)	Source
Consumer survey - confidence indicator sa (% ba...	European Commission
Manufacturing survey - Order books: level sa (%...	European Commission
Manufacturing survey - production: future tende...	European Commission
Deflated orders for total manufactured goods (v...	ISTAT
CPI All items (2010=10) inverted	ISTAT
Imports from Germany c.i.f. (USD)	ISTAT

Table A.3: *OECD Composite Leading Indicator Components for Japan*

Component Series (Unit)	Source
Small Business Survey: sales tendency (% balance)	Japan Finance Corporation
Ratio of inventories to shipments sa (OECD base...	Statistics Japan
Monthly hours worked: manufacturing sa (OECD ba...	Statistics Japan
Ratio loans to deposits sa (%) inverted	Bank of Japan
Share prices: TOPIX index (OECD base year = 100)	Bank of Japan
Spread of interest rates (% p.a.)	Bank of Japan
ITS imports/exports (volume) sa (OECD base year...	Japan Customs
Work started for dwellings sa (OECD base year =...	Ministry of Construction

Table A.4: *OECD Composite Leading Indicator Components for Korea*

Component Series (Unit)	Source
Manufacturing survey - business situation: futu...	Bank of Korea
Share prices: KOSPI index (OECD base year = 100)	Bank of Korea
Stocks of manufactured investment goods (volume...	National Statistical Office
Inventory circulation indicator (manufacturing)	National Statistical Office
Spread of interest rates (% p.a.)	National Statistical Office
Net Barter Terms of trade (OECD base year = 100)	National Statistical Office

Table A.5: *OECD Composite Leading Indicator Components for Mexico*

Component Series (Unit)	Source
Manufacturing survey - employment: tendency sa ...	Bank of Mexico
Manufacturing survey - production: tendency sa ...	Bank of Mexico
Consumer survey - current major purchases sa (%...	Bank of Mexico
Domestic PPI Industrial activities	Bank of Mexico
Employees -manufacturing (D) sa (%)	INEGI
CPI Household equipment and maintenance (OECD b...	INEGI
ITS Imports f.o.b. total sa	INEGI
IPC share price index	Mexico Stock Exchange
Real effective exchange rates - CPI Based (OECD...	OECD

Table A.6: *OECD Composite Leading Indicator Components for Spain*

Component Series (Unit)	Source
Manufacturing survey - rate of capacity utiliza...	European Commission
Construction survey - employment: future tenden...	European Commission
Consumer survey - confidence indicator sa (% ba...	European Commission
CPI Services less housing (OECD base year = 100...	Bank of Spain
Share prices: IGBM general index (OECD base yea...	Bank of Spain
Passenger car registrations (OECD base year = 1...	Asociación Española de Fabricantes

Table A.7: *OECD Composite Leading Indicator Components for Turkey*

Component Series (Unit)	Source
Manufacturing survey - finished goods stocks: l...	Bank of the Republic of Türkiye (TCMB)
Manufacturing survey - new domestic orders: ten...	Bank of the Republic of Türkiye (TCMB)
Manufacturing survey - employment: future tende...	Bank of the Republic of Türkiye (TCMB)
Services survey – demand evolution: future tend...	European Commission
Prospects for exports (manuf.) (% balance)	Bank of the Republic of Türkiye (TCMB)
Discounted treasury auction interest rate (% p...	Bank of the Republic of Türkiye (TCMB)
Production of electricity sa (GW hours)	State Institute of Statistics
ITS Import c.i.f. of intermediate goods sa (USD)	State Institute of Statistics

Table A.8: *OECD Composite Leading Indicator Components for United Kingdom*

Component Series (Unit)	Source
Consumer survey - confidence indicator sa (% ba...	UK GfK
Manufacturing survey - production: future tende...	CBI
Passenger car registrations sa (number)	Society of Motor Manufacturers &
Retail Prices Index - All Items Index (January ...	Office for National Statistics
GBP Interbank LIBOR 3 months delayed (% p.a.) i...	Bank of England
Share prices: FTSE LOCAL UK (£) index (OECD bas...	London Stock Exchange

Table A.9: *OECD Composite Leading Indicator Components for Unites States of America*

Component Series (Unit)	Source
Consumer survey - confidence indicator sa (norm...	Survey of consumers – University of Michigan
Manufacturing survey - confidence indicator (% ...	US Institute of Supply Management
Work started for dwellings sa (number)	Bureau of the Census
Net new orders - durable goods sa (USD)	Bureau of the Census
Share prices: NYSE composite (OECD base year = ...	New York Stock Exchange
Weekly hours worked: manufacturing sa (hours)	Bureau of Labor Statistics
Spread of interest rates (% p.a.)	Federal Reserve

Appendix B

pyOECD

```
def search_oecd(search_string = '') -> pd.DataFrame:
    """
    Search for a string in the OECD dataflows and return the
        results.
    Parameters:
    search_string (str): The string to search for in the dataflows
        .
    Returns:
    pd.DataFrame: A DataFrame containing the dataflows that match
        the search string.
    """
```

```
# Suppress everything below ERROR
logging.getLogger("sdmx").setLevel(logging.ERROR)
```

```

oecd = sdmx.Client("OECD")
flow_msg = oecd.dataflow()
# Get the dataflows
dataflows = pd.DataFrame(sdmx.to_pandas(flow_msg.dataflow))

# Search for the string in names and descriptions
result = dataflows[dataflows[0].str.contains(search_string.
    strip(), case=False)]
# If no results found, search in the index
if result.empty:
    # Search in the index
    result = dataflows[dataflows.index.str.contains(
        search_string.strip(), case=False)]
# If still no results, try referencing the first index
    directly if it is a fullid
    if result.empty:
        try:
            # try referencing the first index directly if it
            is a fullid
            result = pd.DataFrame([dataflows.loc[search_string
                .strip()]])

            print("This is a fullid of a dataflow")
            # result = pd.DataFrame([result])
            # If able to find, return that line of the
            dataframe

```

```

        except:

            # If not able to find, return None

            result = pd.DataFrame([])

            print("No results found")

    return result

def get_dataset_id(dataset_name):
    """
    """
    """Get the id of a dataset in the OECD database."""
    Parameters:
    dataset_name (str): The name of the dataset.
    Returns:
    str: The id of the dataset.

    """

    oecd = sdmx.Client("OECD")

    # Get the dataflows
    flow_msg = oecd.dataflow()

    # Get the dataset id
    dataset_id = flow_msg.dataflow[dataset_name].id

    return dataset_id

```

```

def get_var_dimensions(dataset_id):
    """
    Get the keys for a given dataset id.
    Parameters:
    dataset_id (str): The id of the dataset.
    Returns:
    list: A list of the keys for the dataset.

    """
    return get_variables(get_full_dataset_id(dataset_id)).
        structure[0].dimensions.components


def get_dataset_name(dataset_id):
    """
    Get the name of a dataset in the OECD database.
    Parameters:
    dataset_id (str): The id of the dataset.
    Returns:
    str: The name of the dataset in the OECD.

    """

    oecd = sdmx.Client("OECD")

    # Get the dataflows
    flow_msg = oecd.dataflow()

    # Get the dataset name

```

```

dataset_name = flow_msg.dataflow[dataset_id].name

return dataset_name

def get_full_dataset_id(dataset_name):
    """
    Get the full id of a dataset in the OECD database.
    Parameters:
    dataset_name (str): The name of the dataset (e.g. Composite
        Leading Indicators) or a short/simple ID (e.g. DF_CLI).
    Returns:
    str: The full id of the dataset in the OECD.
    """

    found = search_oecd(dataset_name)

    # if found is None or found.empty:
    #     # Maybe it is a short/simple ID (e.g. DF_CLI)
    #     # Try to find the ID in the dataflows

    if len(found) > 1:
        # If not found, return None
        raise ValueError(f"Multiple datasets found for {
            dataset_name}. Please be more specific.")
    elif len(found) == 0:
        # If not found, return None
        raise ValueError(f"No dataset found for {dataset_name}.
```

```

        Please check the name or ID.")
    else:
        return found.index[0]

def get_variables(dataflow_id):
    """
    Get the variables for a given dataflow id (for the proper id,
        use the search_oecd function).
    Parameters:
    dataflow_id (str): The id of the dataflow.
    Returns:
    pd.DataFrame: A DataFrame containing the variables for the
        dataflow.
    """

    # search = search_oecd(dataflow_id)
    # # Get the dataflow id
    # print(search)
    oecd = sdmx.Client("OECD");
    flow_msg = oecd.dataflow();

    # print(oecd.dataflow(resource=flow_msg.dataflow[dataflow_id])
    # )
    try:
        return oecd.dataflow(resource=flow_msg.dataflow[
            dataflow_id])
    except:

```

```

    try:
        fullid = get_full_dataset_id(dataflow_id)
        return oecd.dataflow(resource=flow_msg.dataflow[fullid
                                ])
    except:
        var = get_dataset_id(dataflow_id)
        return var

def help_query(dataset_id):
    """
    Get the help for a given dataset id.
    Parameters:
    dataset_id (str): The id of the dataset.
    Returns:
    pd.DataFrame: A DataFrame containing the help for the dataset.
    """

    # Get the dataflows
    var = get_variables(dataset_id)

    # build a dictionary with the dimensions and their values
    dimensions = {}

    for i in var.structure[0].dimensions.components:

        index = var.structure[0].dimensions.components.index(i)

```

```

        possible = var.structure[0].dimensions.components[index].
            local_representation.enumerated

    if possible is not None:
        dimensions[i.id] = possible.items.keys()
    else:
        dimensions[i.id] = None

    return dimensions

def get_data(dataset_id, key={}, params = dict(
    dimensionAtObservation="AllDimensions")):
    """
    Get the data for a given dataset id and parameters.
    Parameters:
    dataset_id (str): The id of the dataset.
    key (dict): The key for the dataset.
    params (dict): The parameters for the dataset.
    Returns:
    pd.DataFrame: A DataFrame containing the data for the dataset.
    """

    oecd = sdmx.Client("OECD")

    try:
        var = get_variables(dataset_id)
        dataset_id_temp = dataset_id

```

```

except:
    try:
        # Try to get the dataset id from the dataset name
        dataset_id_temp = get_dataset_id(dataset_id)
        var = get_variables(dataset_id_temp)
    except:
        try:
            # If not found, try to get the dataset id from the
            full id
            dataset_id_temp = get_full_dataset_id(dataset_id)
            var = get_variables(dataset_id_temp)
        except:
            raise ValueError(f"Dataset {dataset_id} not found.
                               Please check the ID given.")

# Need to make sure that the key and params are in the correct
    format and have the mandatory fields
# Check if arguments in key exist in the dataset
for i in key:

    index = var.structure[0].dimensions.components.index(i)
    possible = var.structure[0].dimensions.components[index].
        local_representation.enumerated.items.keys()
    # if value is not in the list of possible values, raise an
        error
    # print(key)

```

```

# key is a dictionary , so we need to check if the current
    key is a list
if isinstance(key[i], list):
    # Check if all values in the list are in the possible
        values
    for j in key[i]:
        if j not in possible:
            raise ValueError(f"{j} is not a valid value
                for {i}. Possible values are {possible}.")
# if key is not a list , check if it is in the possible
    values
elif isinstance(key[i], str):
    # Check if the value is in the possible values
    if key[i] not in possible:
        raise ValueError(f"{key[i]} is not a valid value
            for {i}. Possible values are {possible}.")

# Check if arguments in params exist in the dataset

# Get the var structure
# var = get_variables(dataset_id_temp)
# Get the dataflow id
dsd = var.structure[0]
var_id = var.dataflow[0].id

# # Get the data

```

```

data_msg = oecd.data(resource_id=var_id, key=key, params=
    params, dsd=dsd)
# Have to add the other possible frequencies in the future
try:
    freq_temp = key["FREQ"]

    if key["FREQ"] == 'M':
        freq = 'MS'
    elif key["FREQ"] == 'Q':
        freq = 'QS'
    elif key["FREQ"] == 'A':
        freq = 'YS'
    else:
        freq = ''
except:
    freq = ''

## Step 4: Convert to pandas DataFrame
## Convert to pandas DataFrame with REF_AREA as part of the
    index
df = sdmx.to_pandas(data_msg.data)
df = df.reset_index()
# convert the time period to datetime
df['TIME_PERIOD'] = pd.to_datetime(df['TIME_PERIOD'])
df['TIME_PERIOD'].freq = freq
# Convert the index to the correct frequency

df = df.set_index(['REF_AREA', 'TIME_PERIOD'])

```

```
df.sort_index(level=0, ascending=True, inplace=True)

return df
}
```

Appendix C

Results by Country-Variable

C.1 Additional Results for Chapter 1

This appendix presents supplementary results for the evaluation of the OECD Composite Leading Indicator forecasting exercise discussed in Chapter 1. The results correspond to the data, model estimation, and forecast evaluation procedures described in Sections 1.2 and 1.3. The appendix includes additional country-level, target-specific, and horizon-specific results that support the main forecast comparison between models estimated with and without the OECD CLI.

The appendix tables and figures provide the full empirical basis for the summary findings reported in the main text. They report out-of-sample forecast accuracy measures, relative model performance, and statistical forecast comparison results across countries, target variables, and forecast horizons. These additional results make it possible to assess whether the value of the OECD CLI varies systematically by country, macroeconomic target, and forecast horizon while keeping the main chapter focused on the broader patterns and interpretation.

Canada

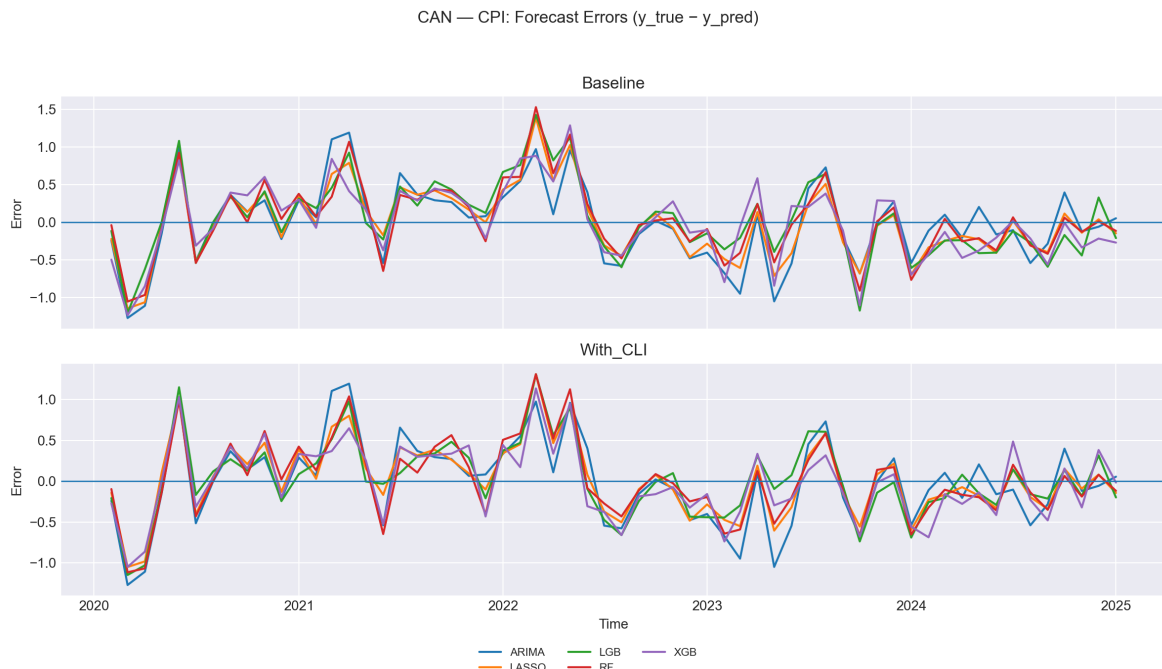


Figure C.1: *Chapter 1 - Canada CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*

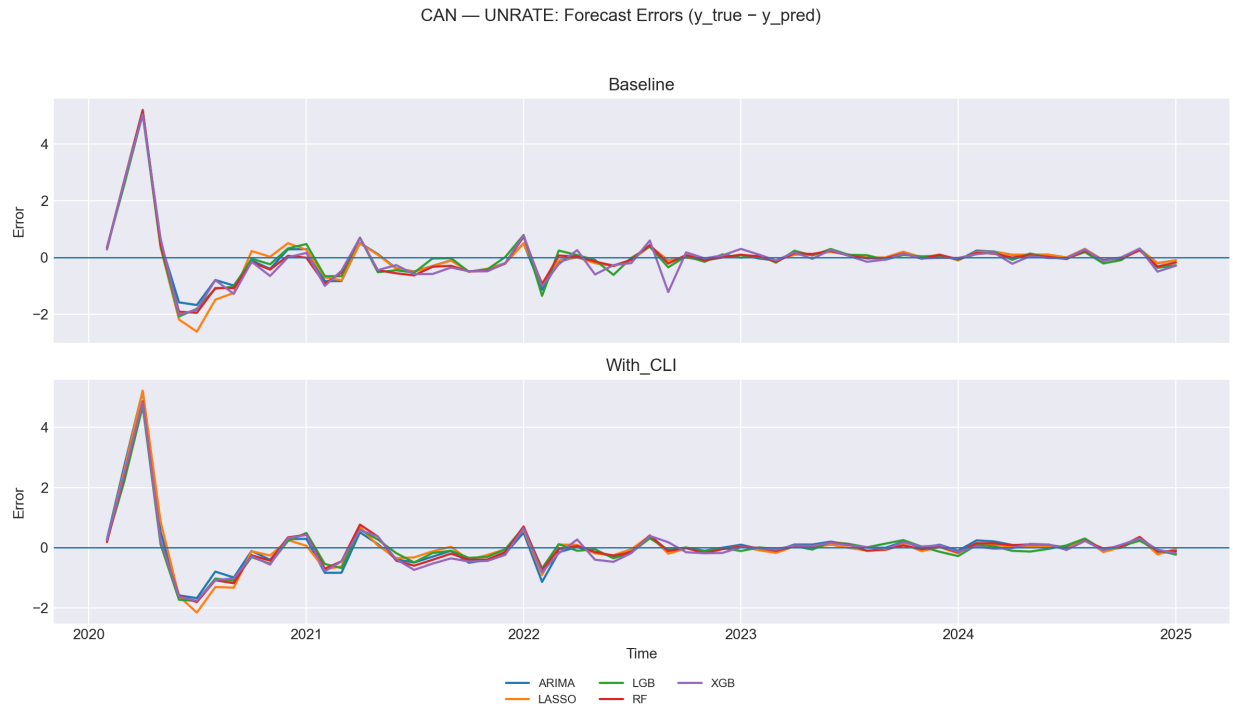


Figure C.2: Chapter 1 - Canada Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

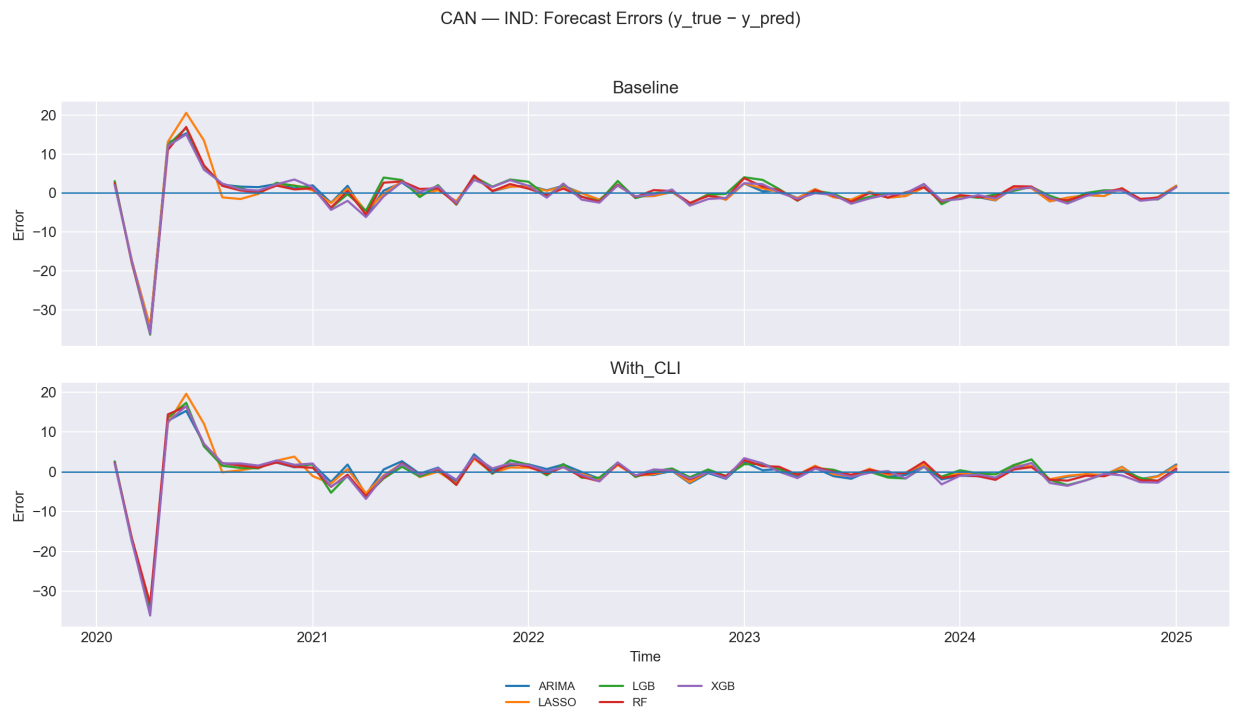


Figure C.3: Chapter 1 - Canada Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Spain

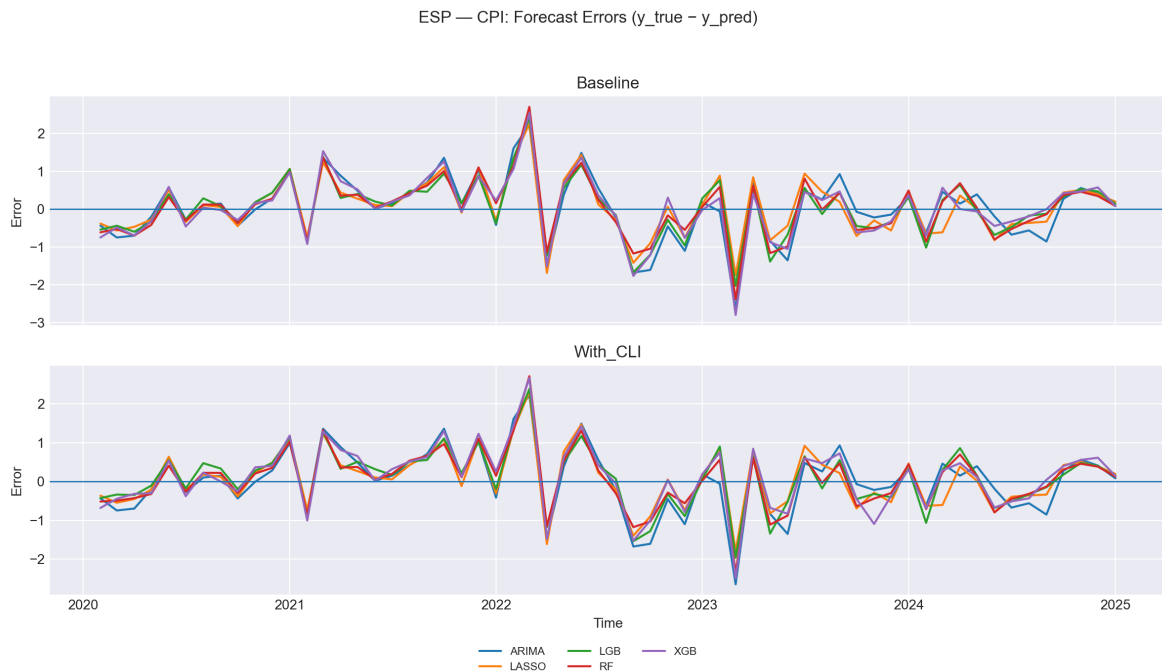


Figure C.4: Chapter 1 - Spain CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

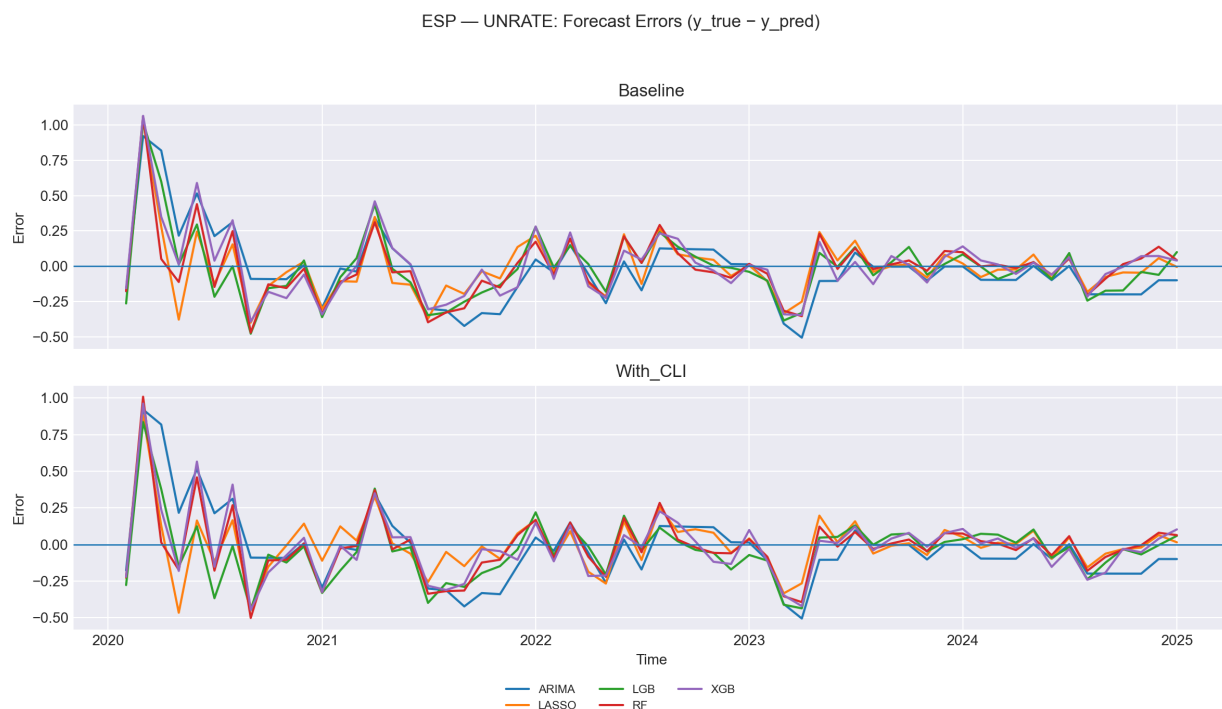


Figure C.5: Chapter 1 - Spain Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

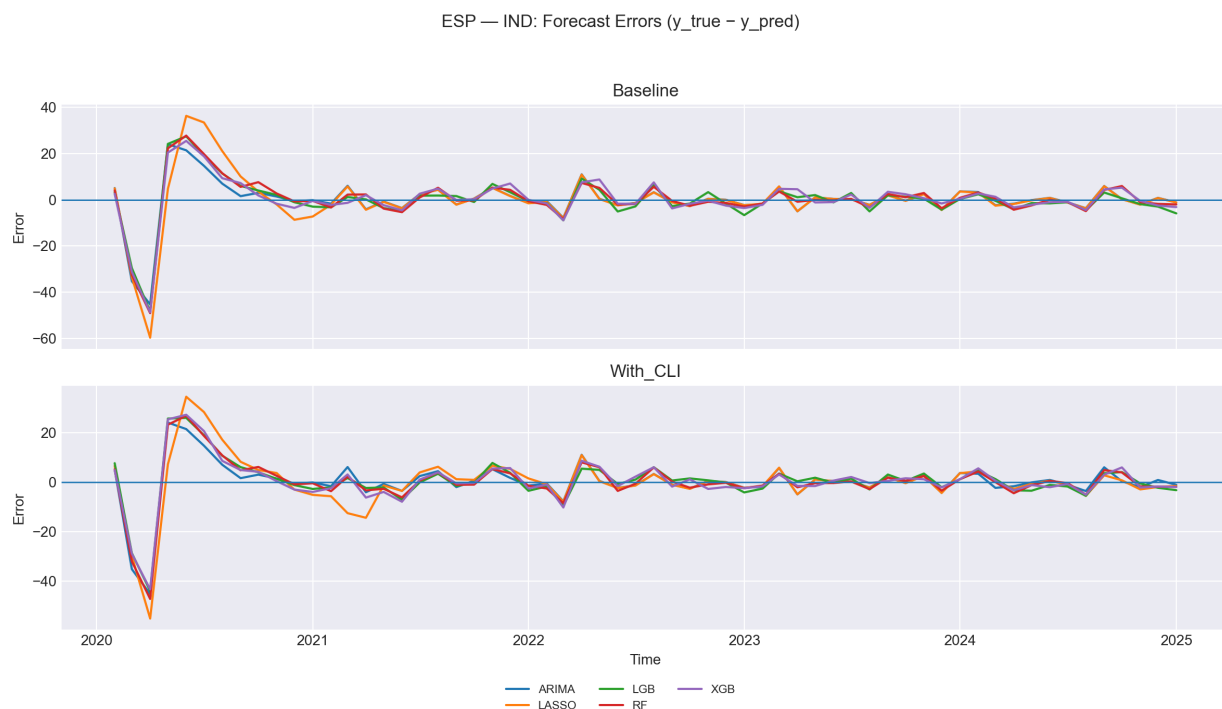


Figure C.6: Chapter 1 - Spain Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Great Britain

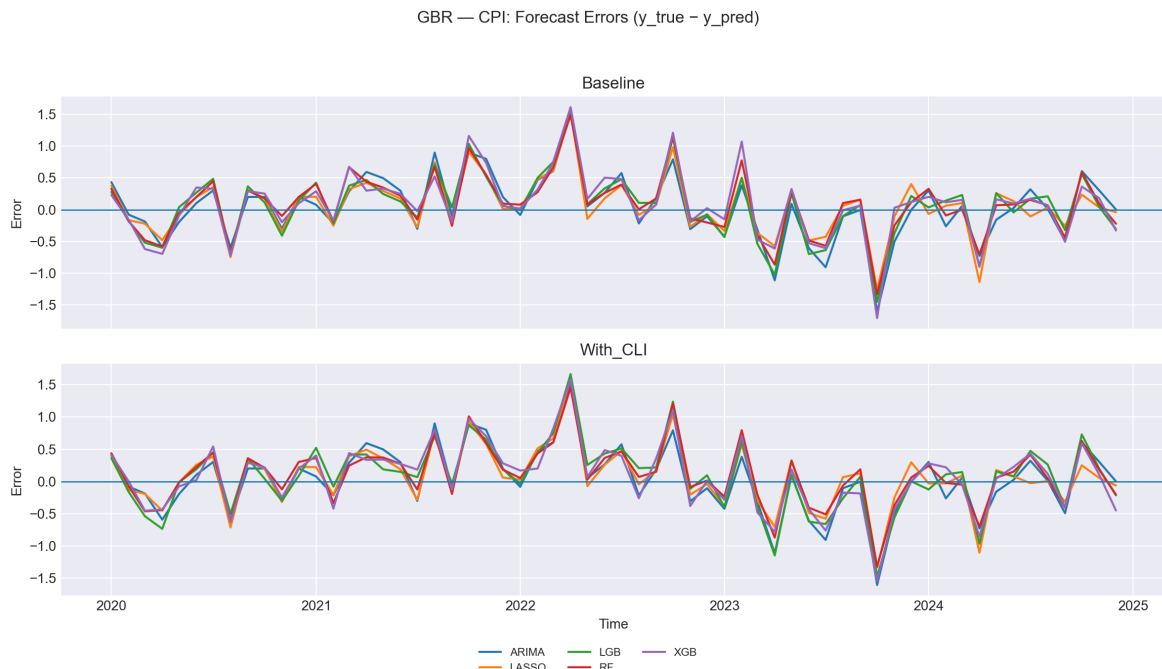


Figure C.7: Chapter 1 - Great Britain CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

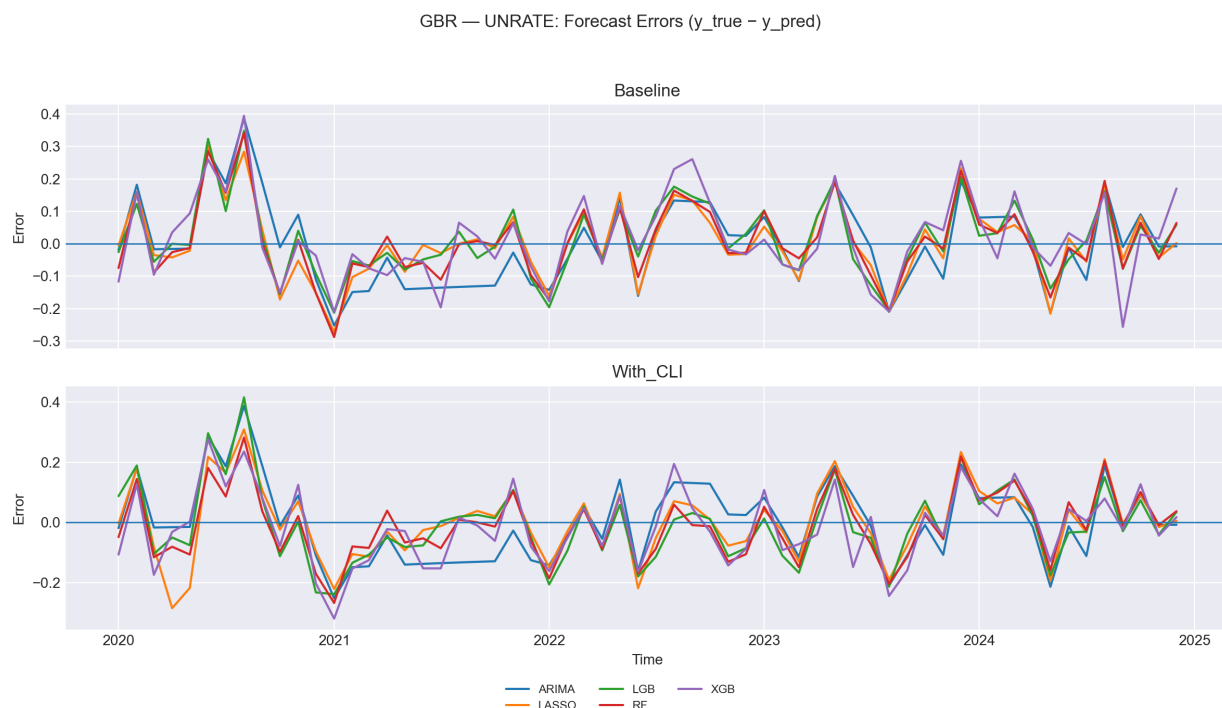


Figure C.8: Chapter 1 - Great Britain Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

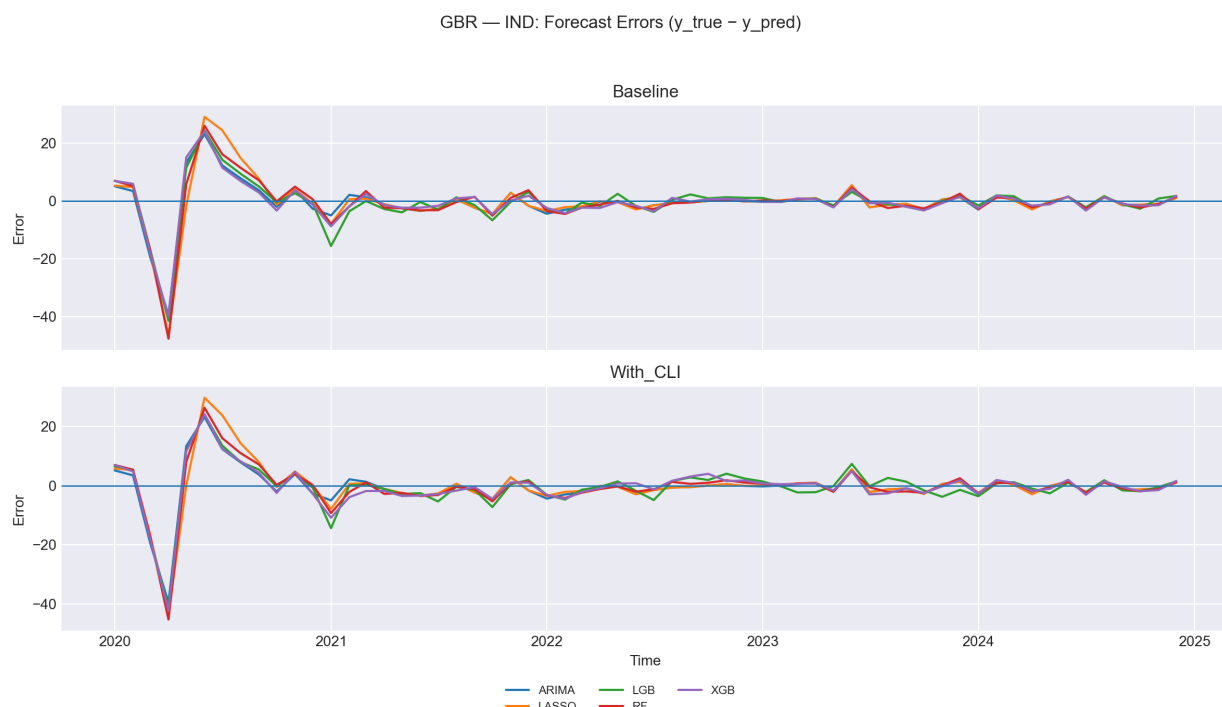


Figure C.9: Chapter 1 - Great Britain Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Italy

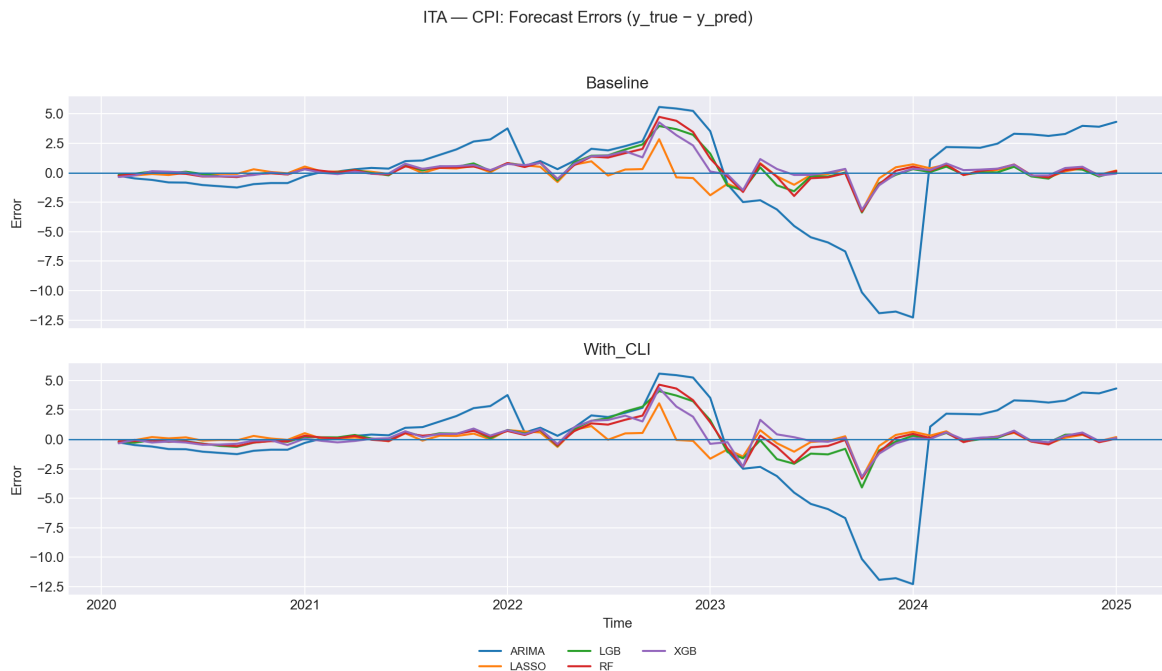


Figure C.10: *Chapter 1 - Italy CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.*

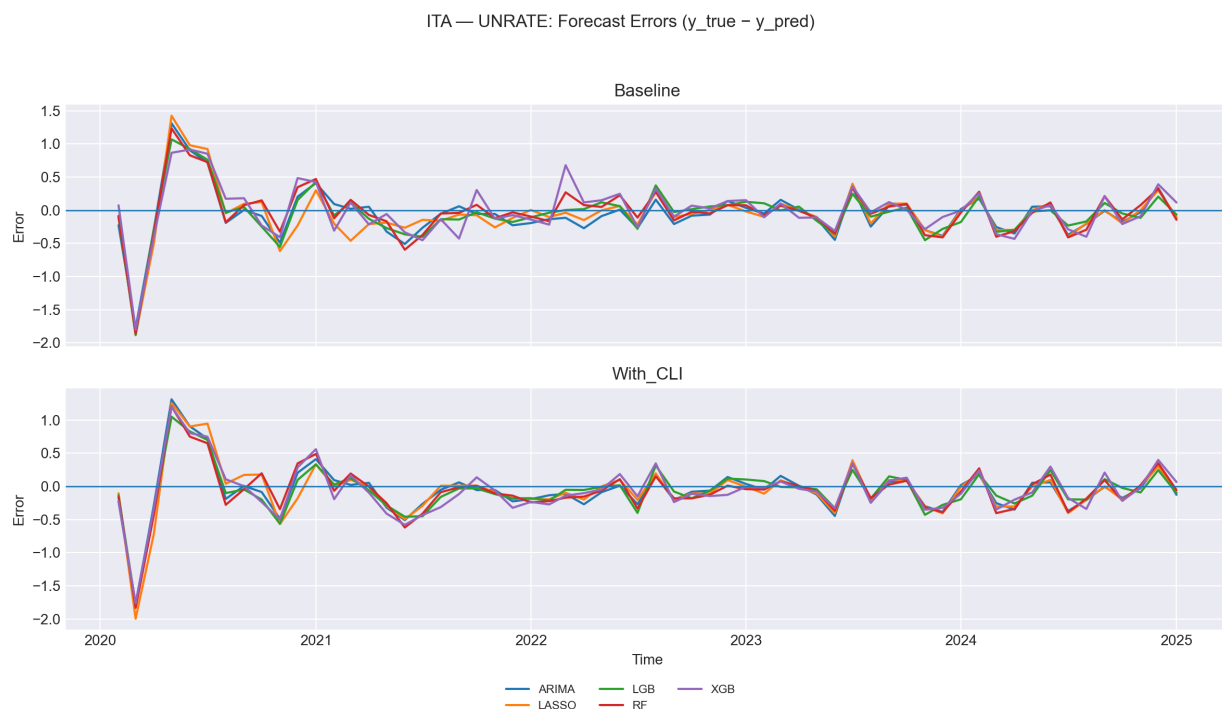


Figure C.11: Chapter 1 - Italy Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.



Figure C.12: Chapter 1 - Italy Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Japan

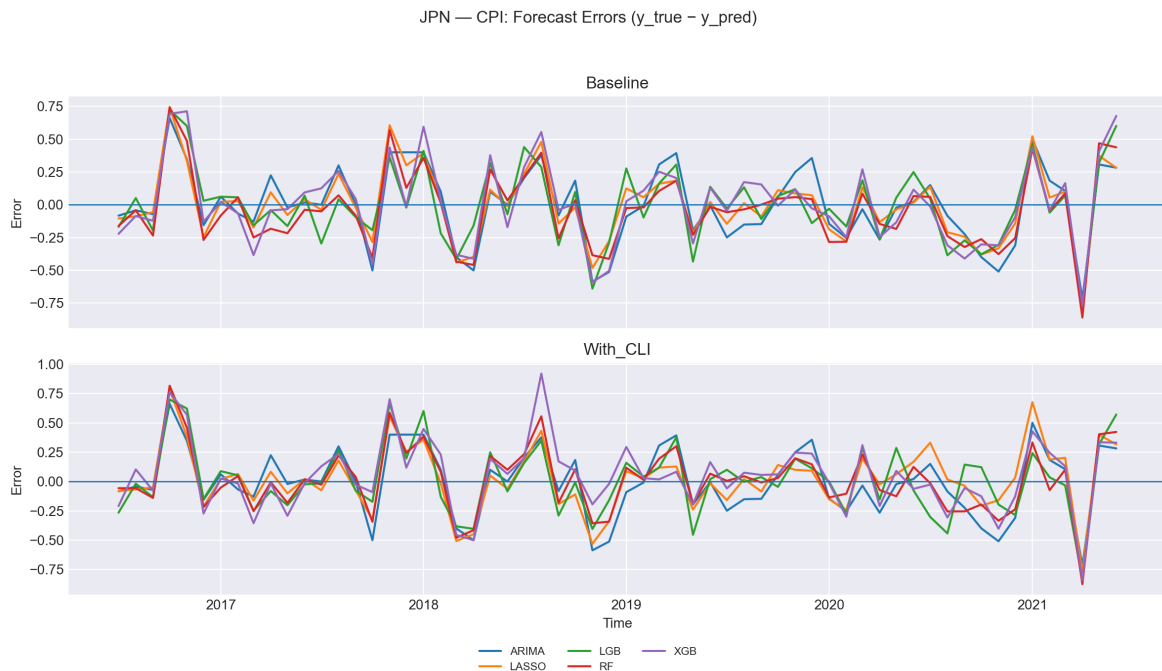


Figure C.13: Chapter 1 - Japan CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

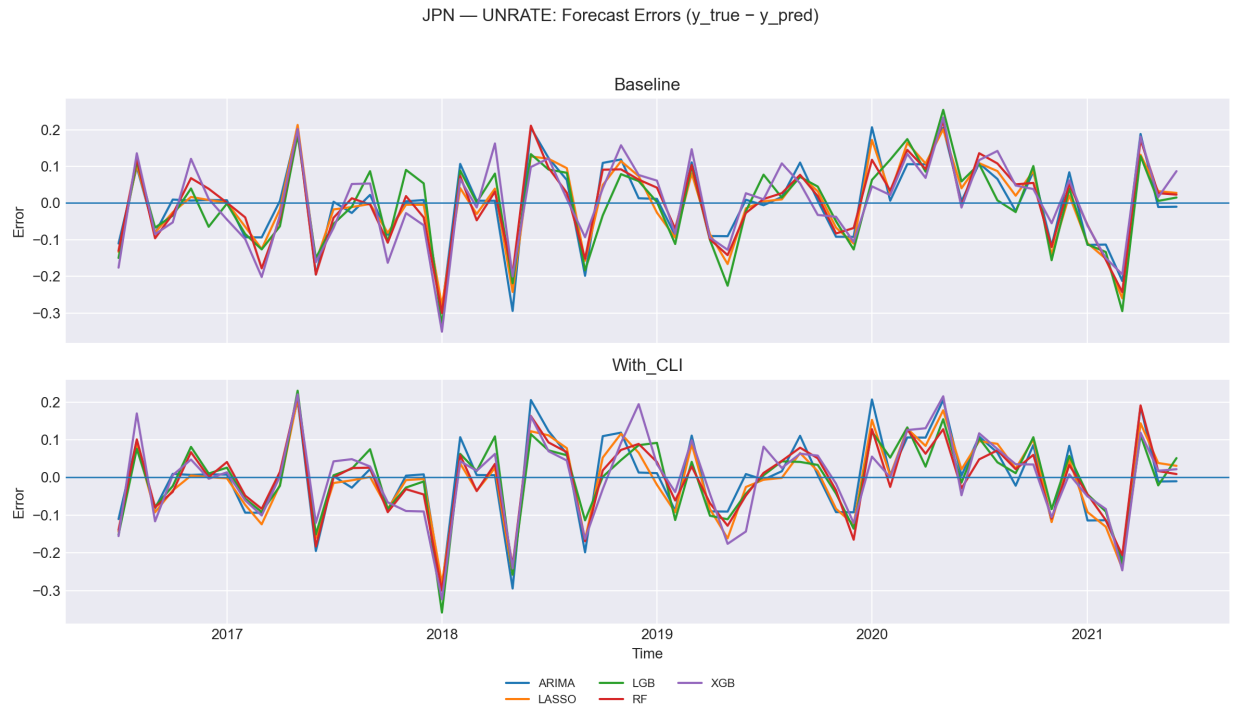


Figure C.14: Chapter 1 - Japan Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.



Figure C.15: Chapter 1 - Japan Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

South Korea

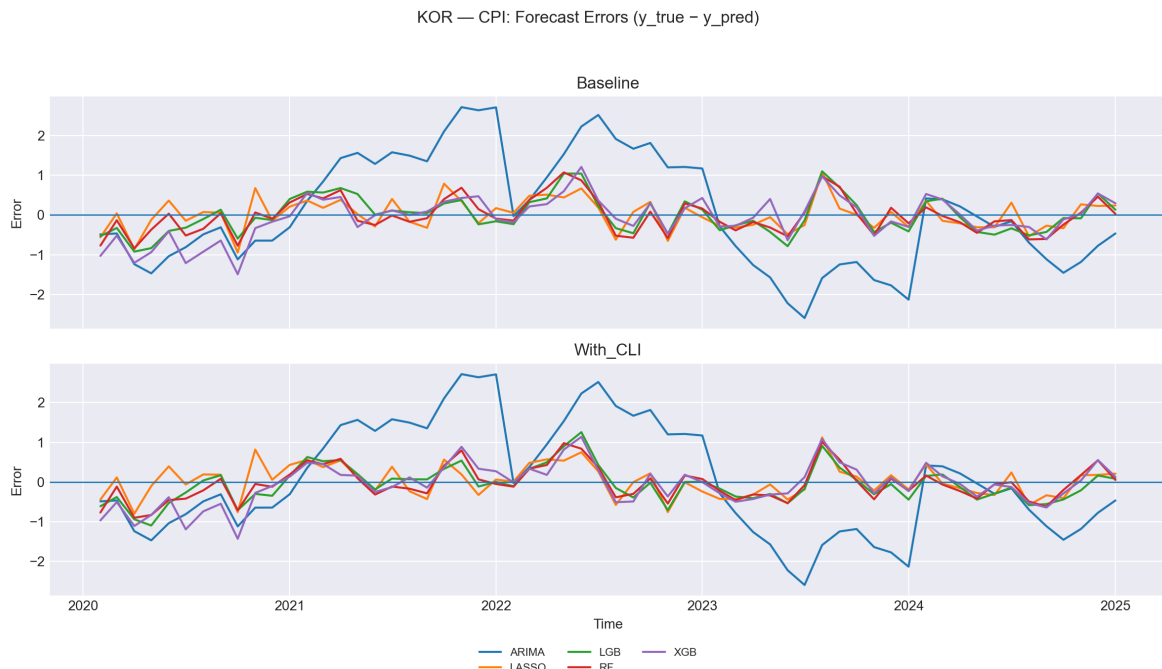


Figure C.16: Chapter 1 - South Korea CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

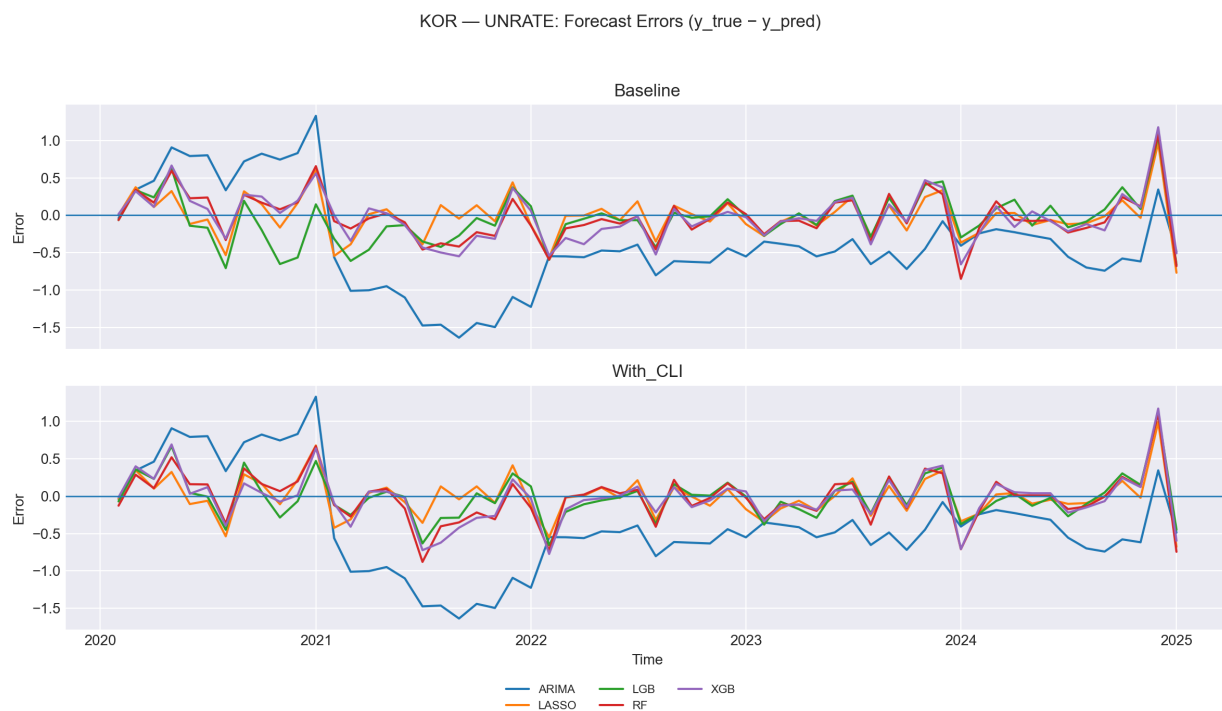


Figure C.17: Chapter 1 - South Korea Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

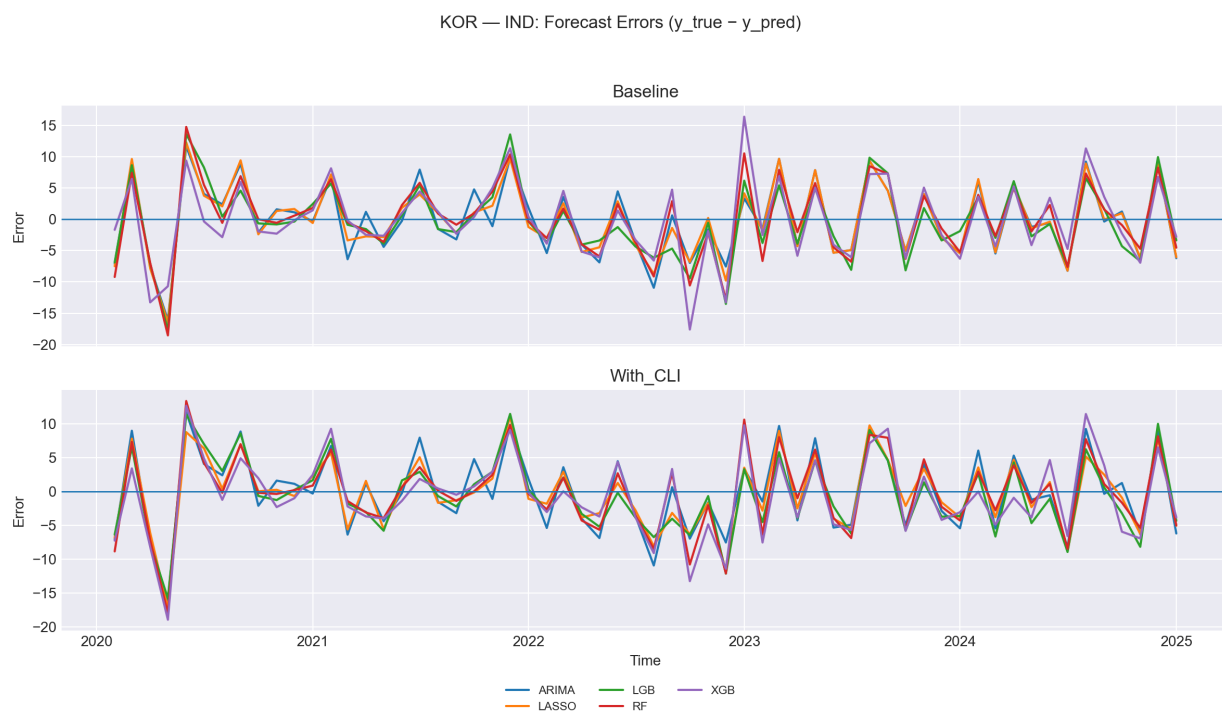


Figure C.18: Chapter 1 - South Korea Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Mexico

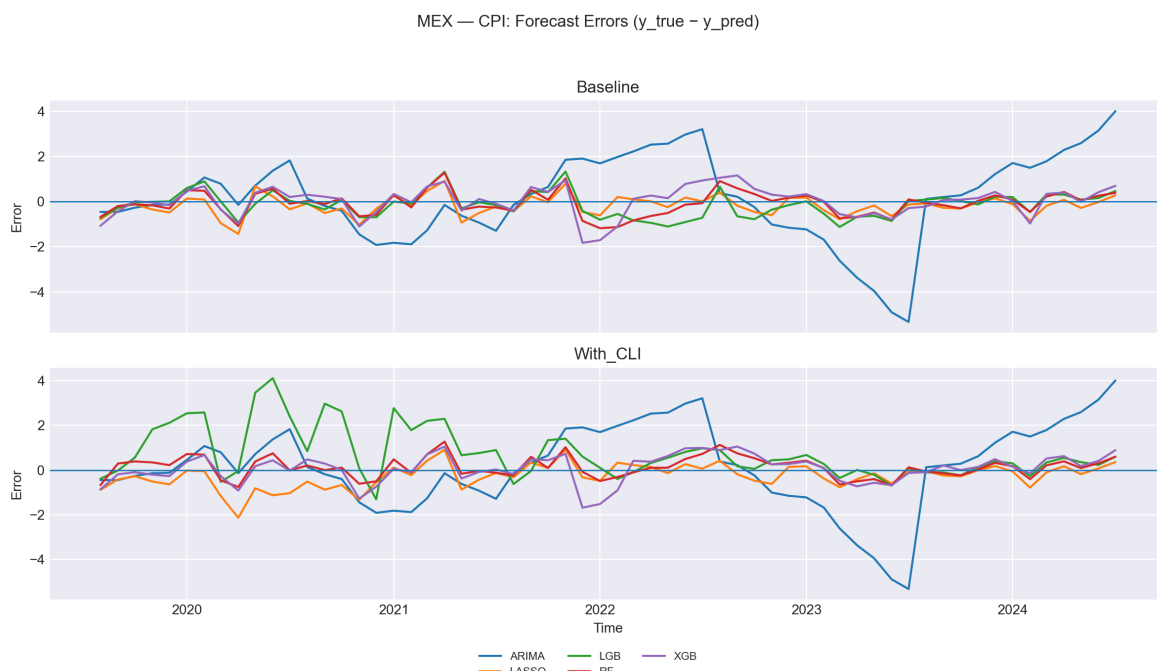


Figure C.19: Chapter 1 - Mexico CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

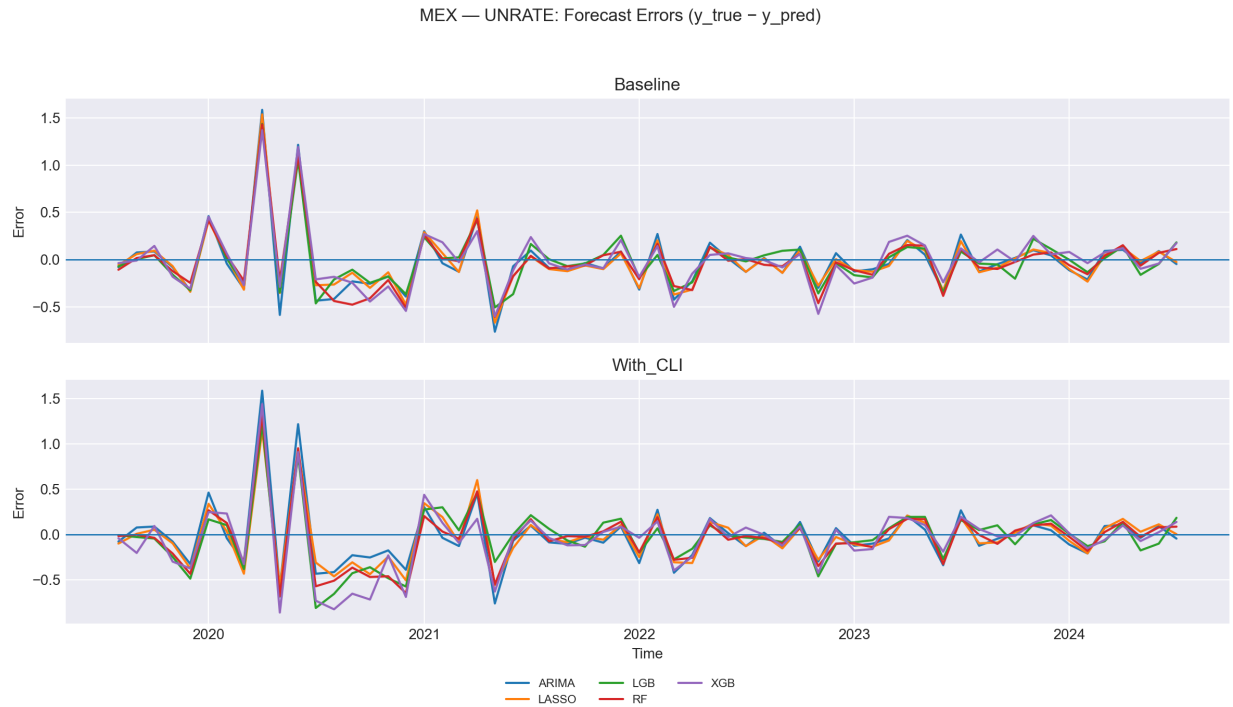


Figure C.20: Chapter 1 - Mexico Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.



Figure C.21: Chapter 1 - Mexico Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

Turkey

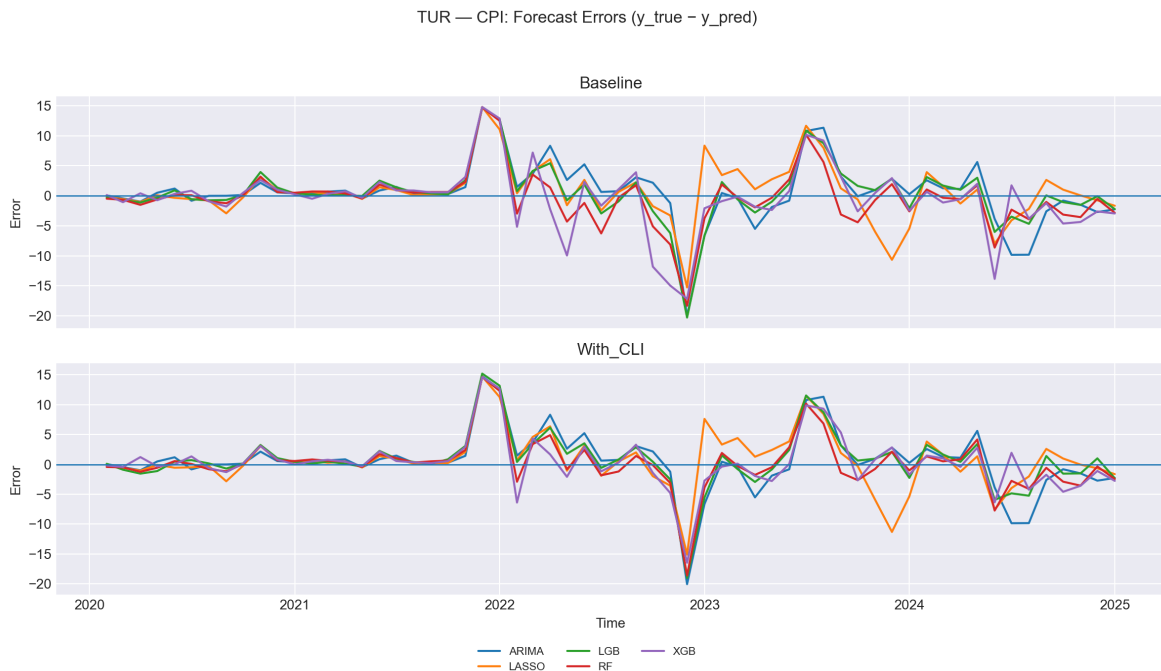


Figure C.22: Chapter 1 - Turkey CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

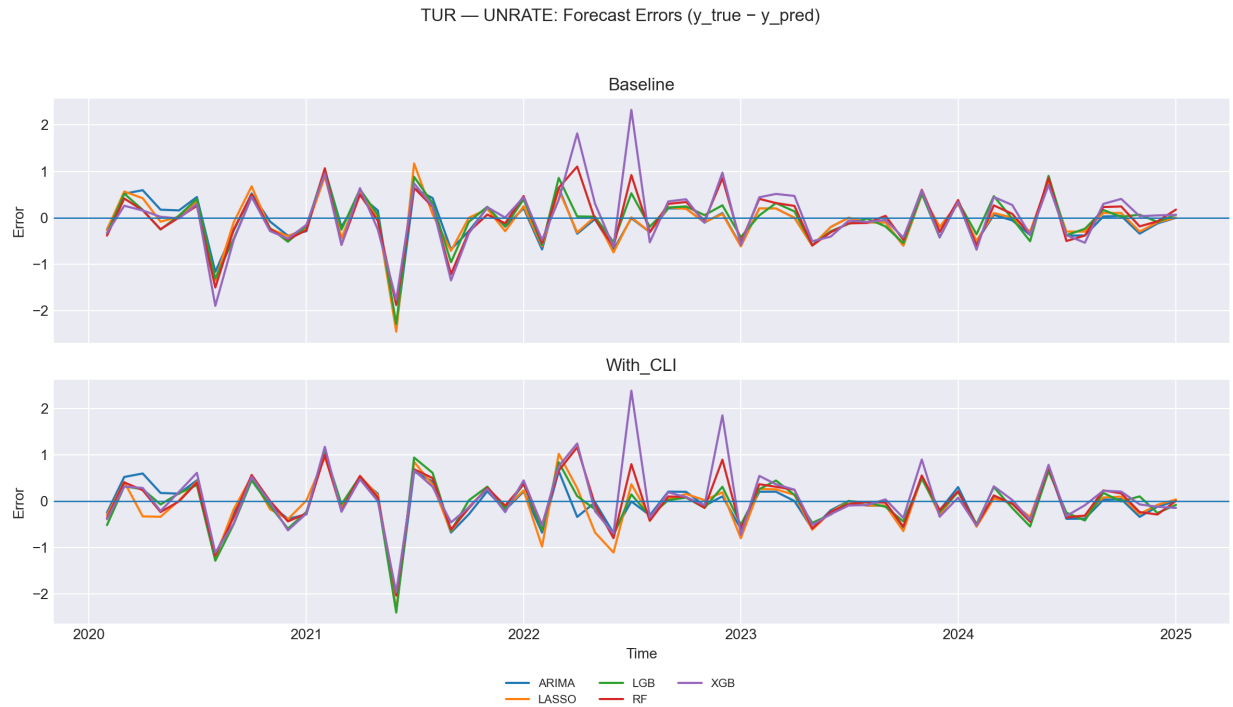


Figure C.23: Chapter 1 - Turkey Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

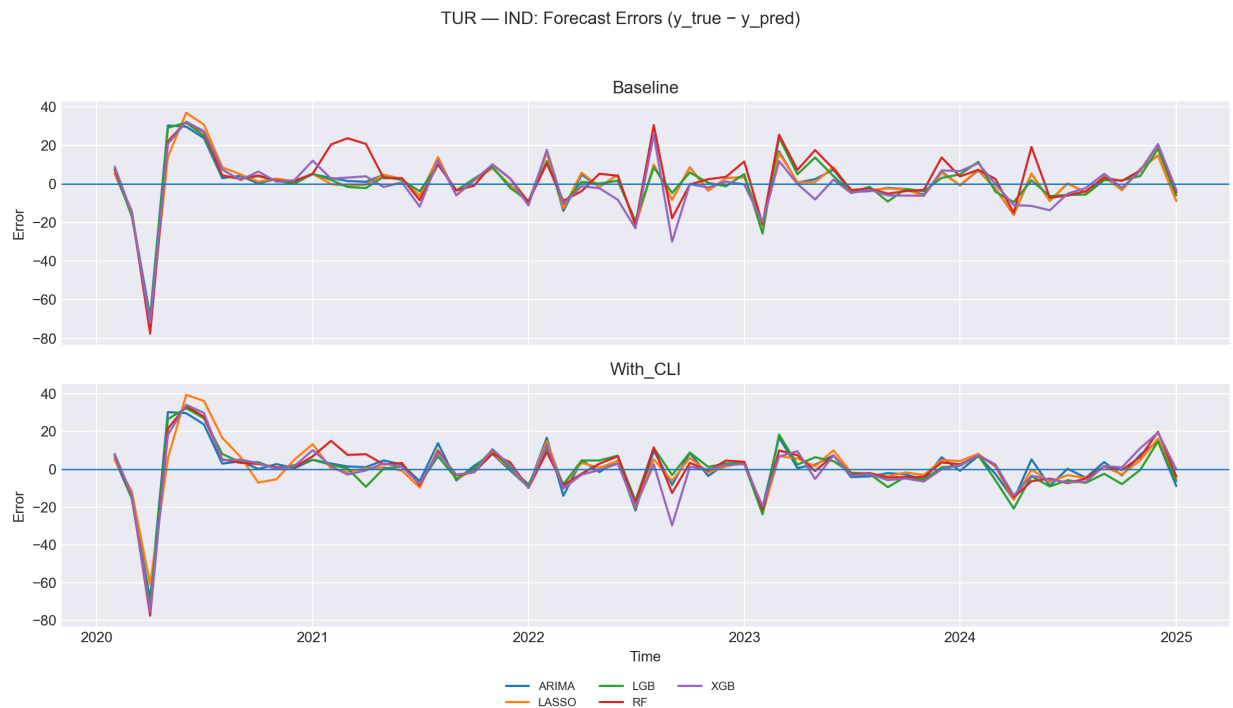


Figure C.24: Chapter 1 - Turkey Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

United States

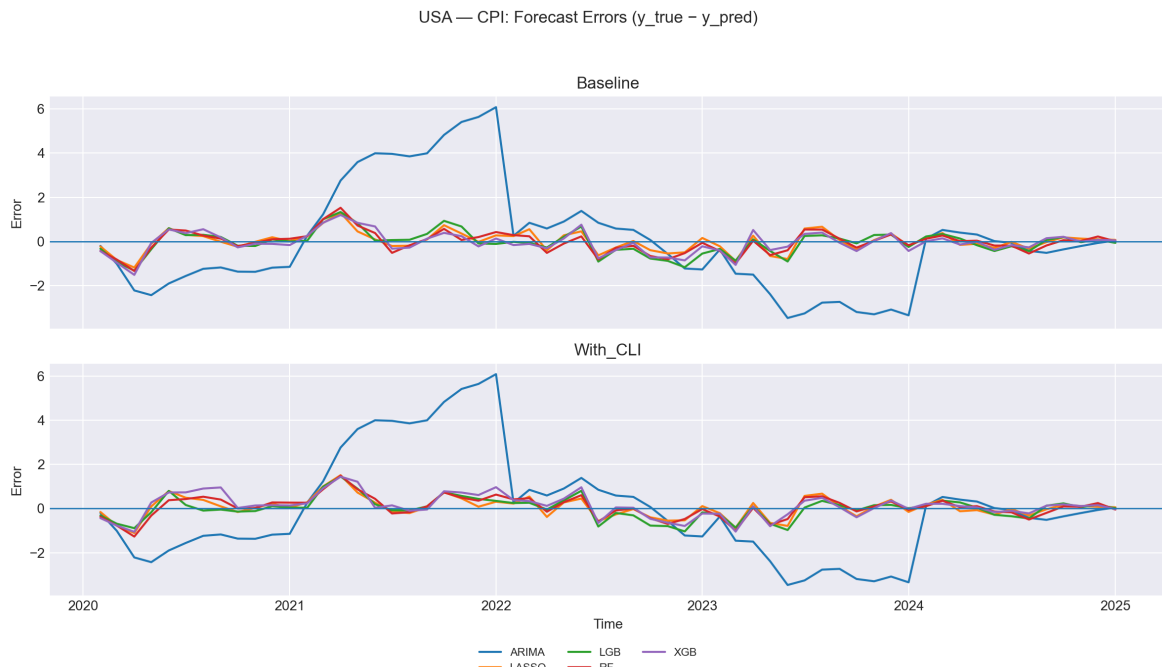


Figure C.25: Chapter 1 - United States CPI Inflation: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

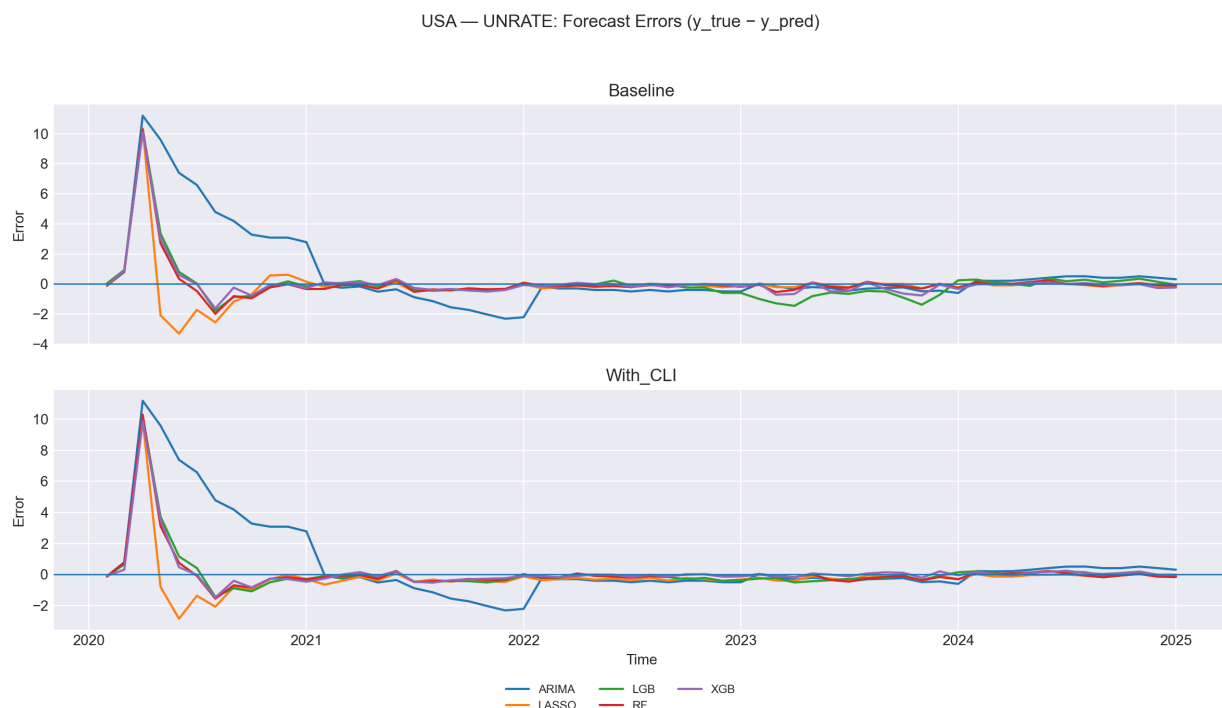


Figure C.26: Chapter 1 - United States Unemployment: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.



Figure C.27: Chapter 1 - United States Industrial Production: forecast errors (two-panel) by model. Top: Baseline. Bottom: With CLI.

C.2 Additional Forecast Accuracy Results for Chapter 2

This appendix reports additional country-level forecast accuracy results for the reweighted OECD Composite Leading Indicator analysis discussed in Sections 2.3 and 2.4. Each figure compares out-of-sample rMSFE values across the benchmark time-series model, the machine-learning models, and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate better forecast accuracy.

Table C.1: *Diebold–Mariano tests for CAN (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-1.715 (0.086)
Inflation	6	LASSO	7.35 (0.0)**
Inflation	6	LGB	3.04 (0.002)**
Inflation	6	RF	3.807 (0.0)**
Inflation	6	XGB	3.198 (0.001)**
Unemployment	6	ARIMAX_MLCLI	-1.922 (0.055)
Unemployment	6	LASSO	5.707 (0.0)**
Unemployment	6	LGB	3.099 (0.002)**
Unemployment	6	RF	3.837 (0.0)**
Unemployment	6	XGB	3.13 (0.002)**
Industrial Production	6	ARIMAX_MLCLI	0.453 (0.65)
Industrial Production	6	LASSO	13.535 (0.0)**
Industrial Production	6	LGB	8.169 (0.0)**
Industrial Production	6	RF	8.847 (0.0)**
Industrial Production	6	XGB	8.476 (0.0)**

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Canada at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

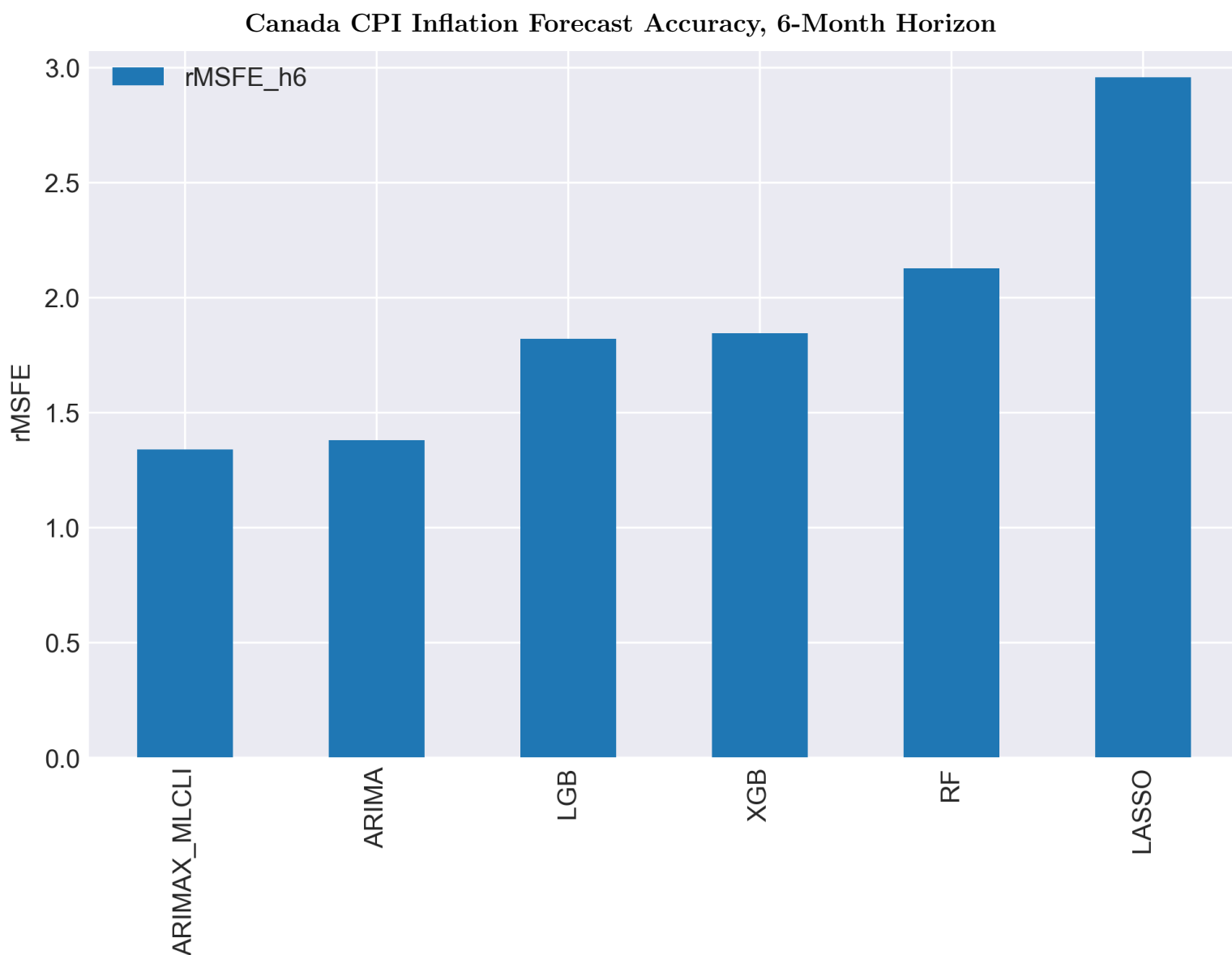


Figure C.28: Chapter 2 - Forecast accuracy results for Canada CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

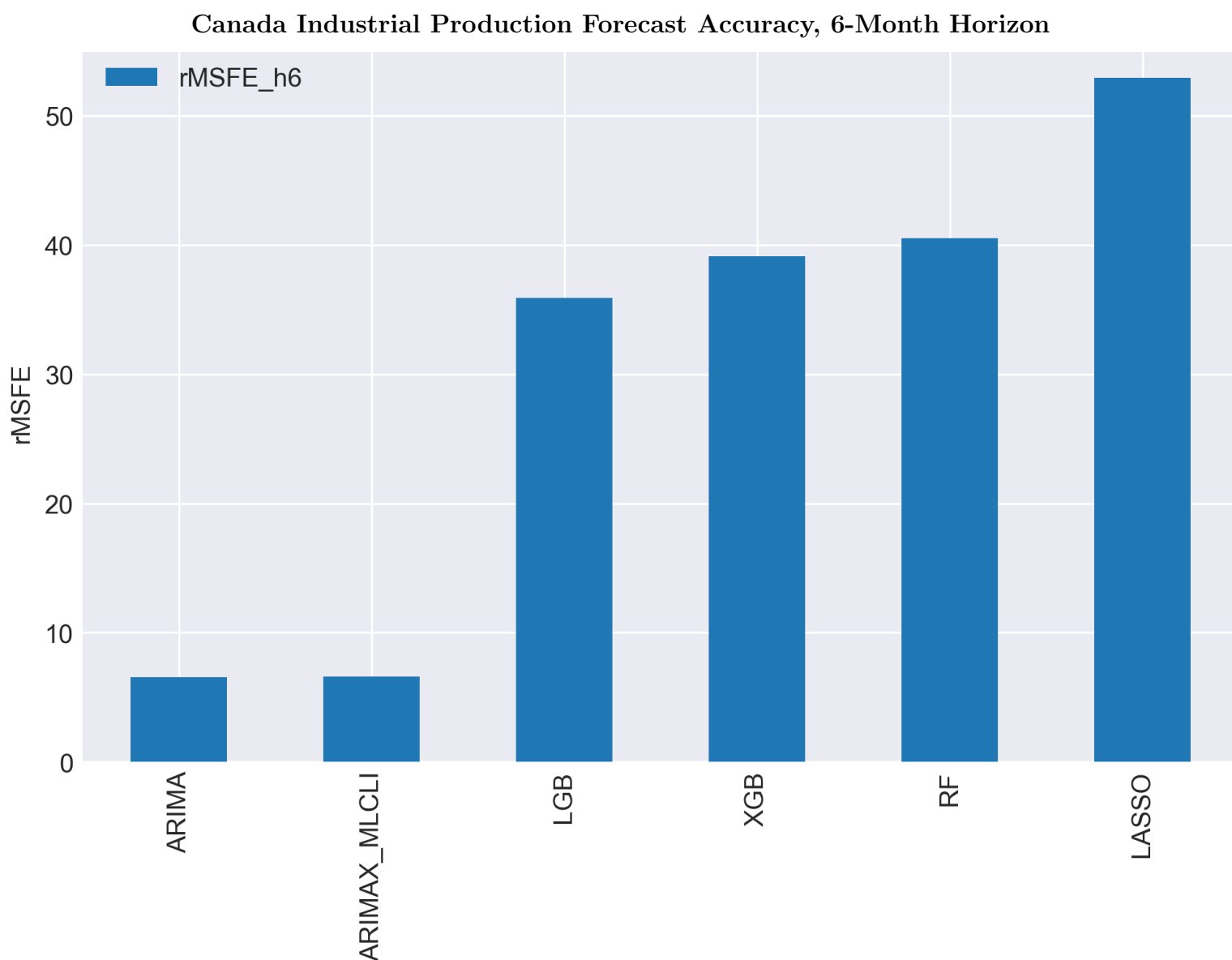


Figure C.29: Chapter 2 - Forecast accuracy results for Canada Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

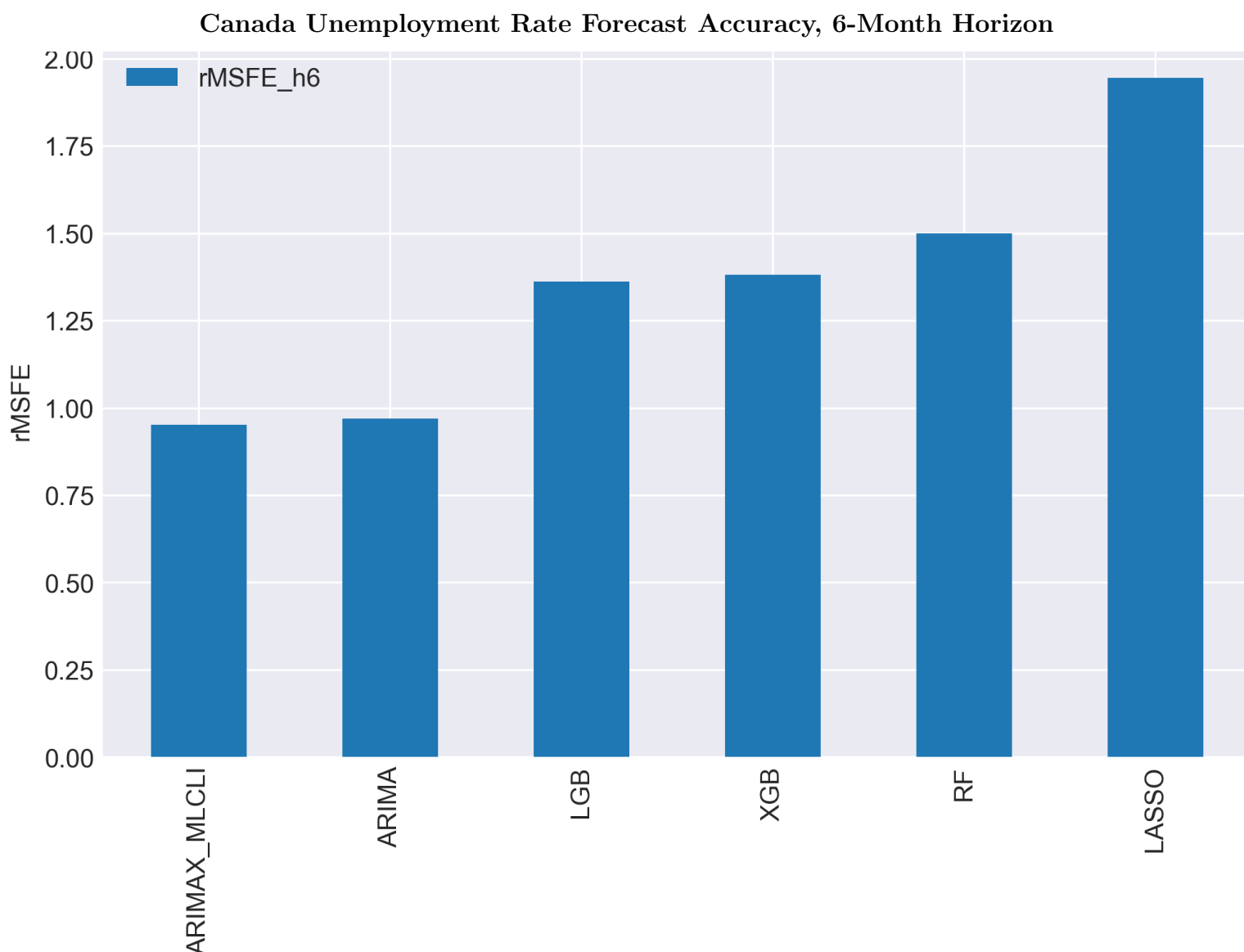


Figure C.30: Chapter 2 - Forecast accuracy results for Canada Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.2: *Diebold–Mariano tests for ESP (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	0.805 (0.421)
Inflation	6	LASSO	2.635 (0.008)**
Inflation	6	LGB	-1.197 (0.231)
Inflation	6	RF	0.594 (0.552)
Inflation	6	XGB	-0.353 (0.724)
Unemployment	6	ARIMAX_MLCLI	-2.218 (0.027)**
Unemployment	6	LASSO	6.263 (0.0)**
Unemployment	6	LGB	3.535 (0.0)**
Unemployment	6	RF	5.001 (0.0)**
Unemployment	6	XGB	4.29 (0.0)**
Industrial Production	6	ARIMAX_MLCLI	-3.115 (0.002)**
Industrial Production	6	LASSO	5.713 (0.0)**
Industrial Production	6	LGB	1.364 (0.172)
Industrial Production	6	RF	2.595 (0.009)**
Industrial Production	6	XGB	1.77 (0.077)

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Spain at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

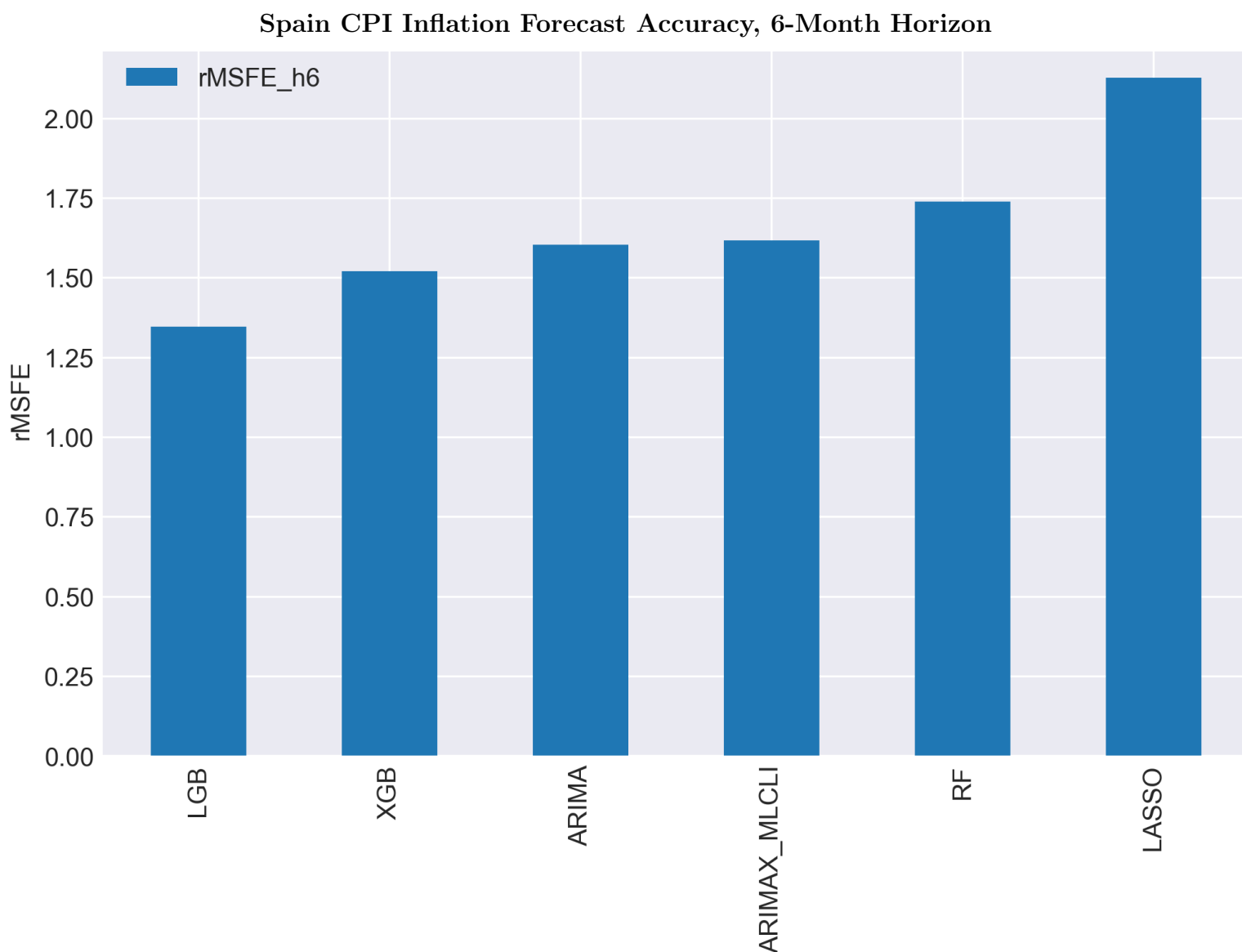


Figure C.31: Chapter 2 - Forecast accuracy results for Spain CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

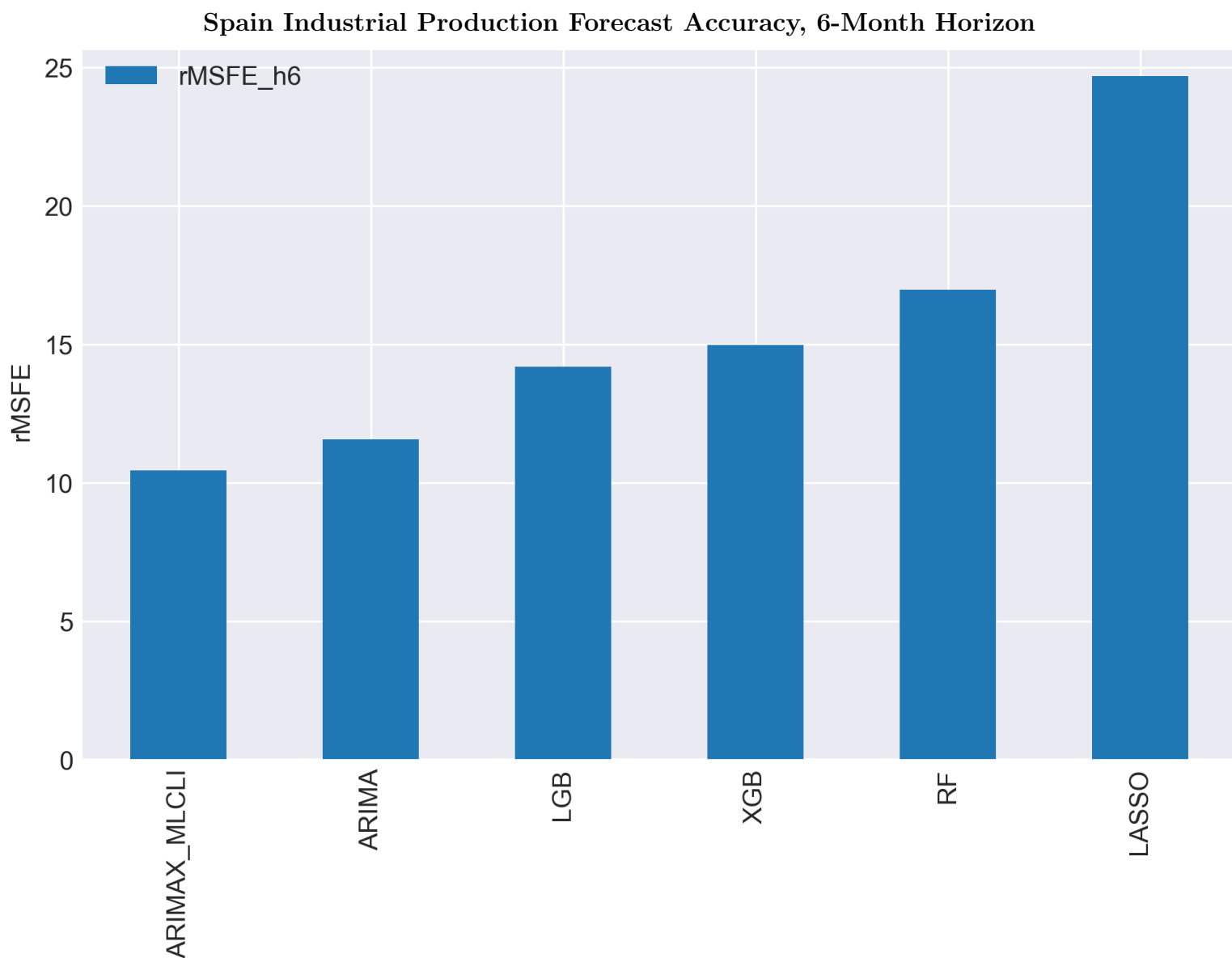


Figure C.32: Chapter 2 - Forecast accuracy results for Spain Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

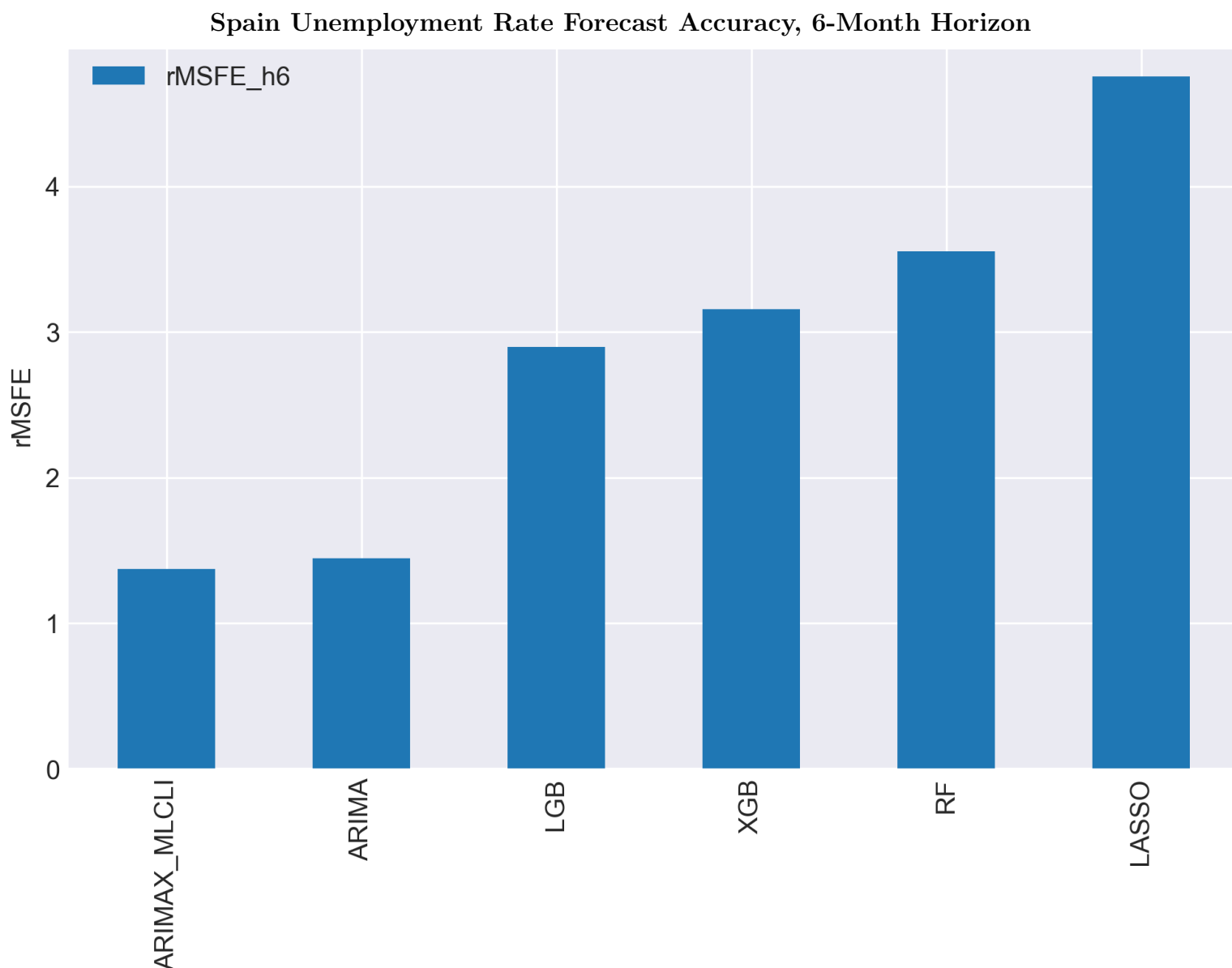


Figure C.33: Chapter 2 - Forecast accuracy results for Spain Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.3: *Diebold–Mariano tests for GBR (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-2.449 (0.014)**
Inflation	6	LASSO	5.183 (0.0)**
Inflation	6	LGB	-0.231 (0.817)
Inflation	6	RF	1.28 (0.201)
Inflation	6	XGB	-0.688 (0.491)
Unemployment	6	ARIMAX_MLCLI	-0.217 (0.828)
Unemployment	6	LASSO	5.894 (0.0)**
Unemployment	6	LGB	4.845 (0.0)**
Unemployment	6	RF	4.666 (0.0)**
Unemployment	6	XGB	5.067 (0.0)**
Industrial Production	6	ARIMAX_MLCLI	1.515 (0.13)
Industrial Production	6	LASSO	8.26 (0.0)**
Industrial Production	6	LGB	6.819 (0.0)**
Industrial Production	6	RF	7.37 (0.0)**
Industrial Production	6	XGB	7.173 (0.0)**

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Great Britain at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

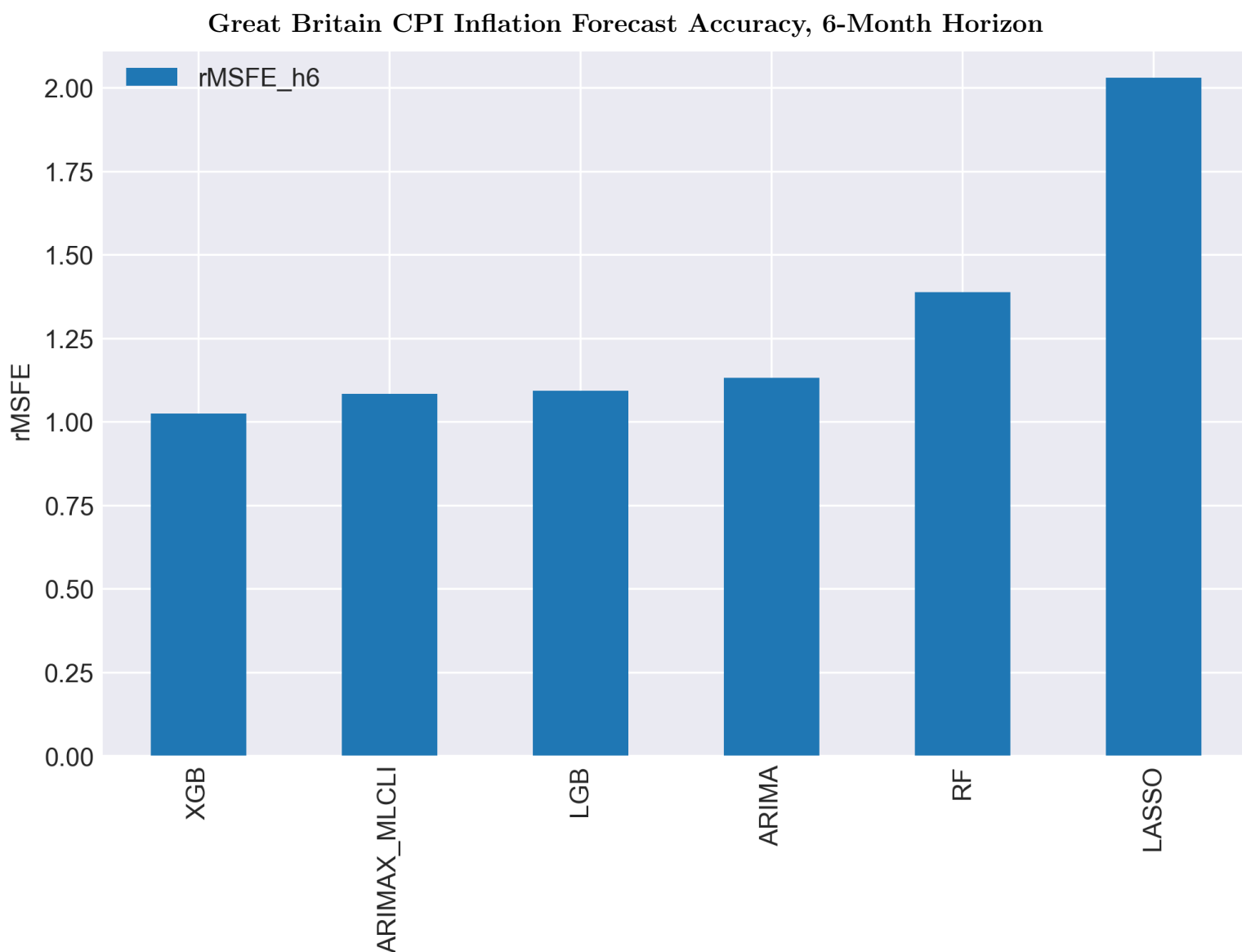


Figure C.34: Chapter 2 - Forecast accuracy results for Great Britain CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

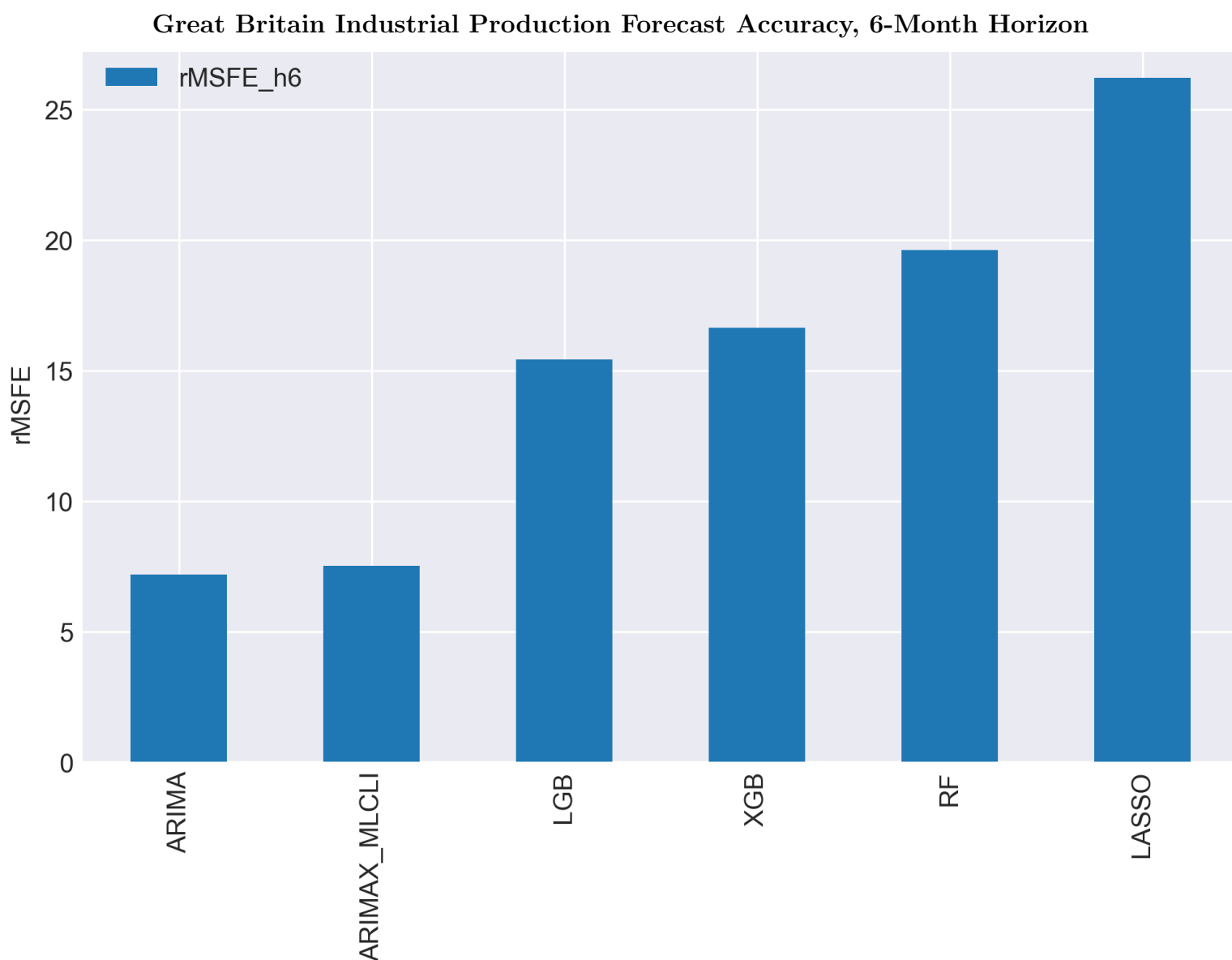


Figure C.35: Chapter 2 - Forecast accuracy results for Great Britain Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

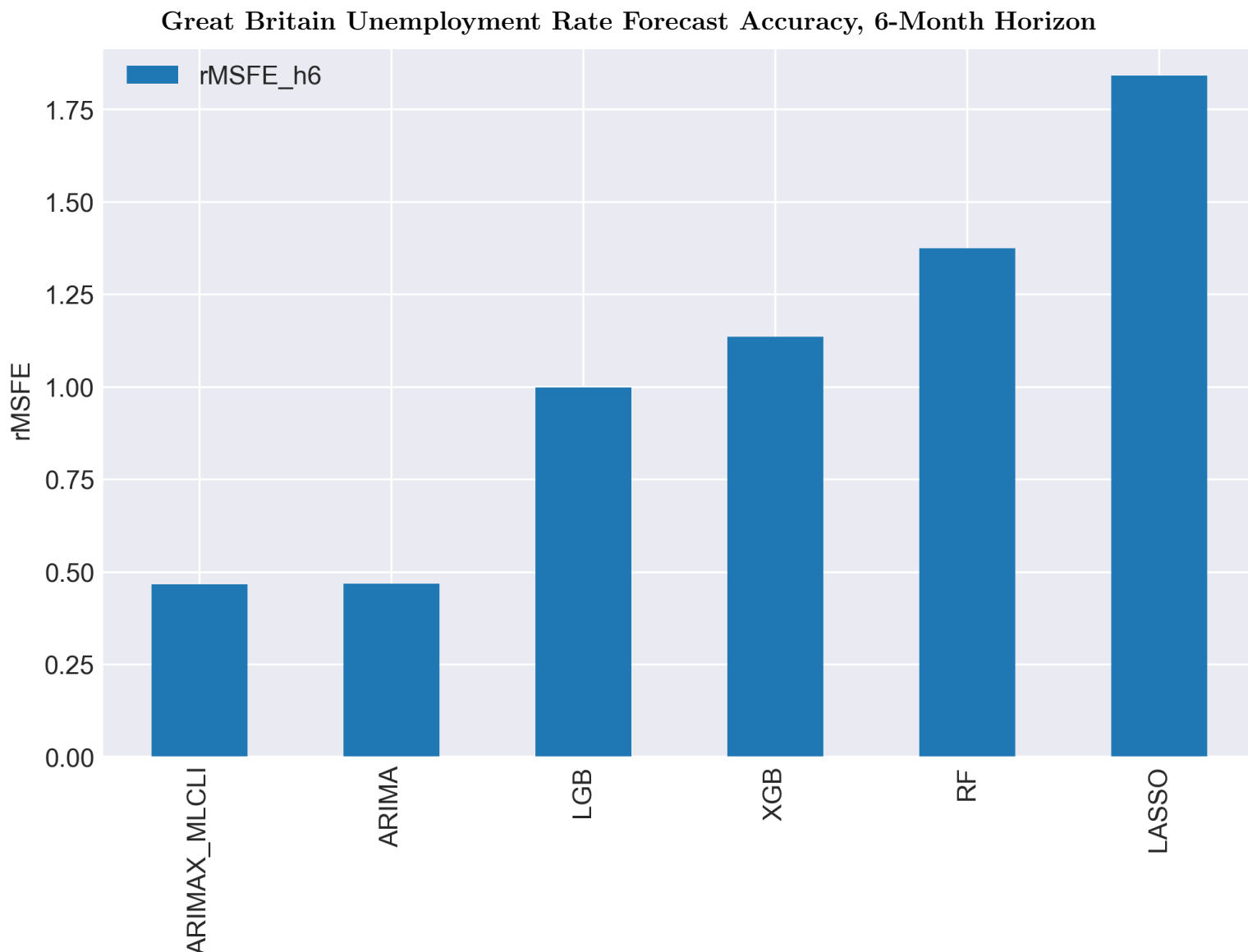


Figure C.36: Chapter 2 - Forecast accuracy results for Great Britain Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.4: *Diebold–Mariano tests for ITA (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-2.847 (0.004)**
Inflation	6	LASSO	11.873 (0.0)**
Inflation	6	LGB	4.496 (0.0)**
Inflation	6	RF	4.984 (0.0)**
Inflation	6	XGB	5.087 (0.0)**
Unemployment	6	ARIMAX_MLCLI	-1.612 (0.107)
Unemployment	6	LASSO	6.946 (0.0)**
Unemployment	6	LGB	4.67 (0.0)**
Unemployment	6	RF	5.14 (0.0)**
Unemployment	6	XGB	4.422 (0.0)**
Industrial Production	6	ARIMAX_MLCLI	-1.481 (0.139)
Industrial Production	6	LASSO	7.951 (0.0)**
Industrial Production	6	LGB	3.729 (0.0)**
Industrial Production	6	RF	4.176 (0.0)**
Industrial Production	6	XGB	4.604 (0.0)**

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Italy at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

Italy CPI Inflation Forecast Accuracy, 6-Month Horizon

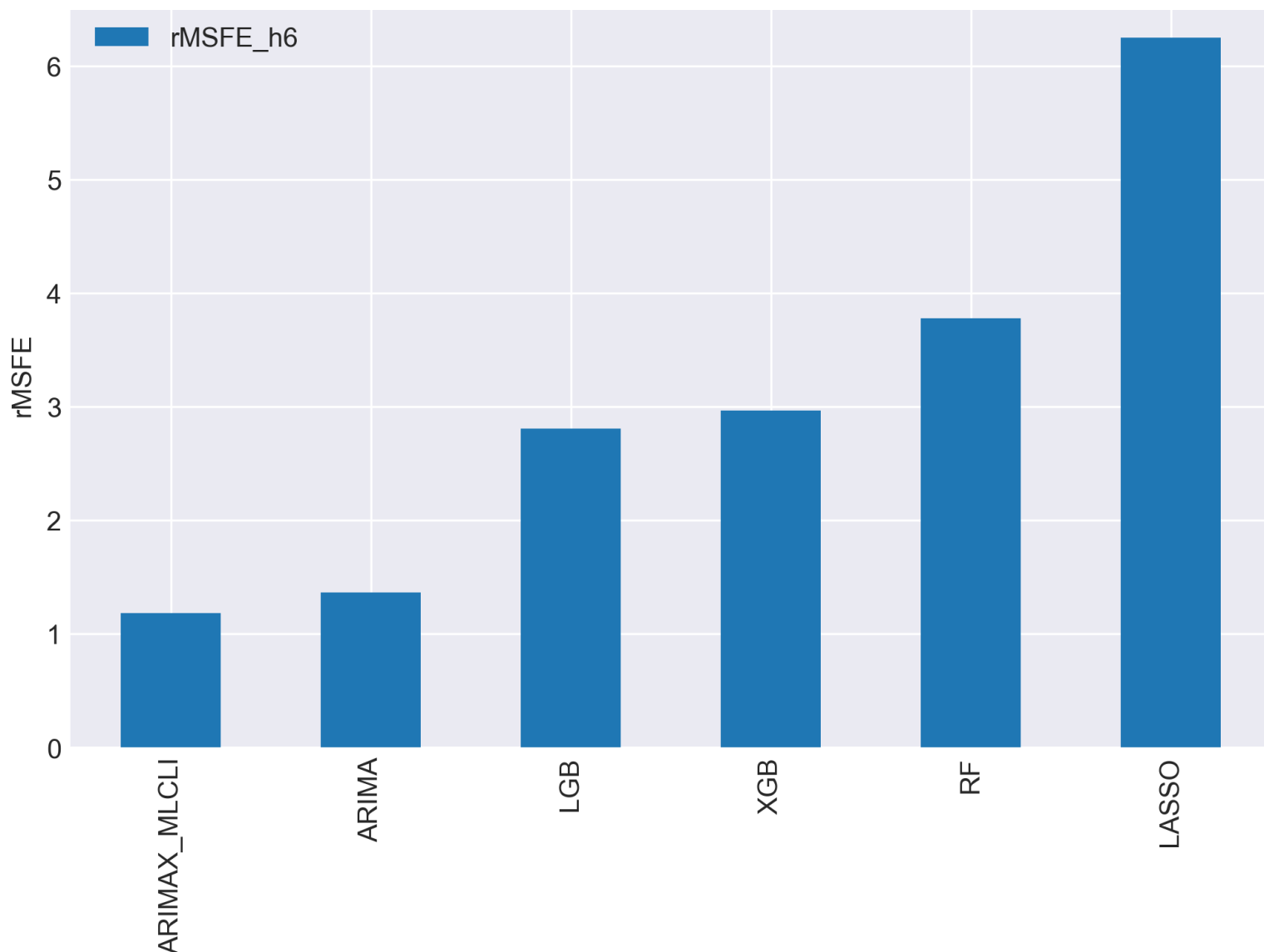


Figure C.37: Chapter 2 - Forecast accuracy results for Italy CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

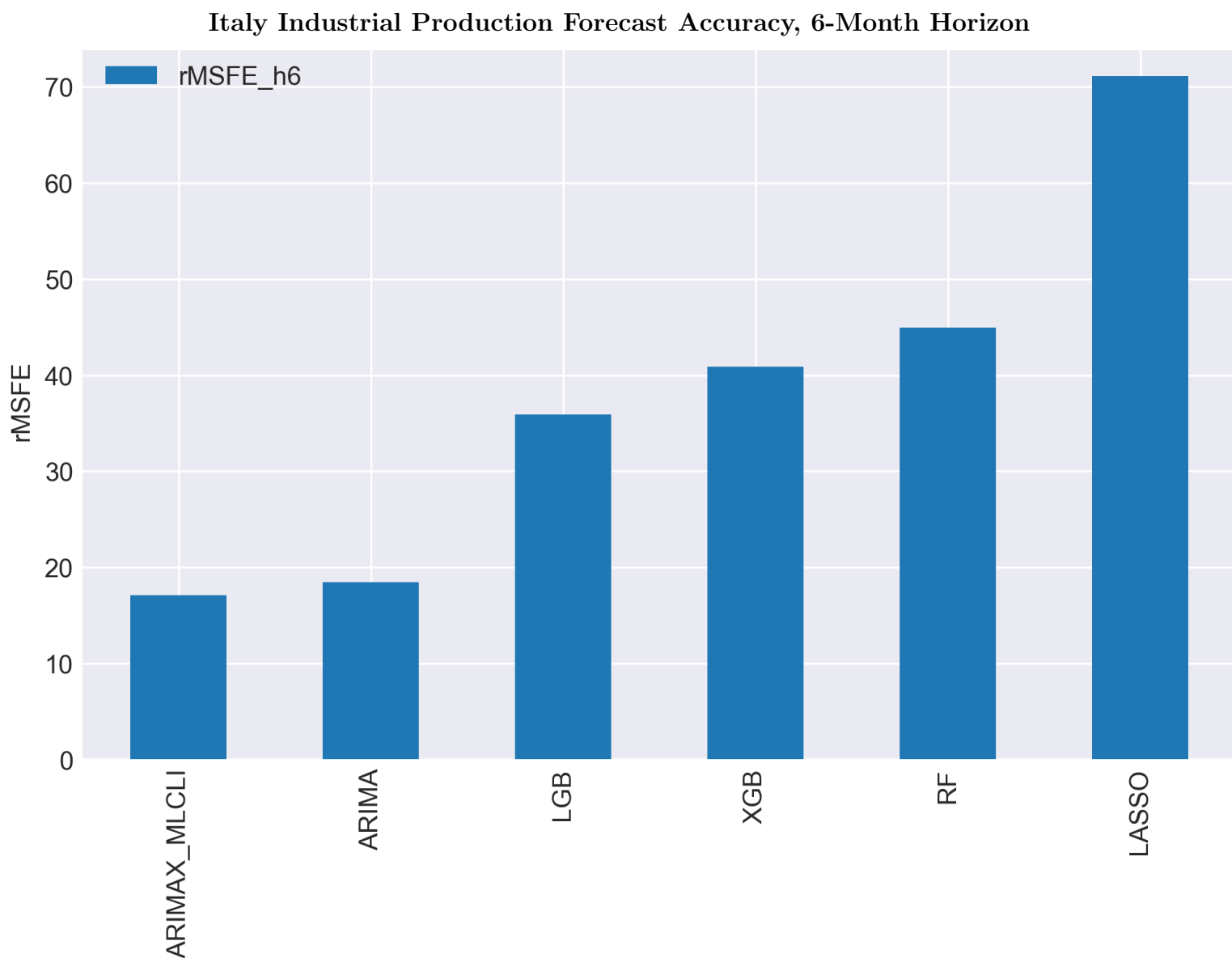


Figure C.38: Chapter 2 - Forecast accuracy results for Italy Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

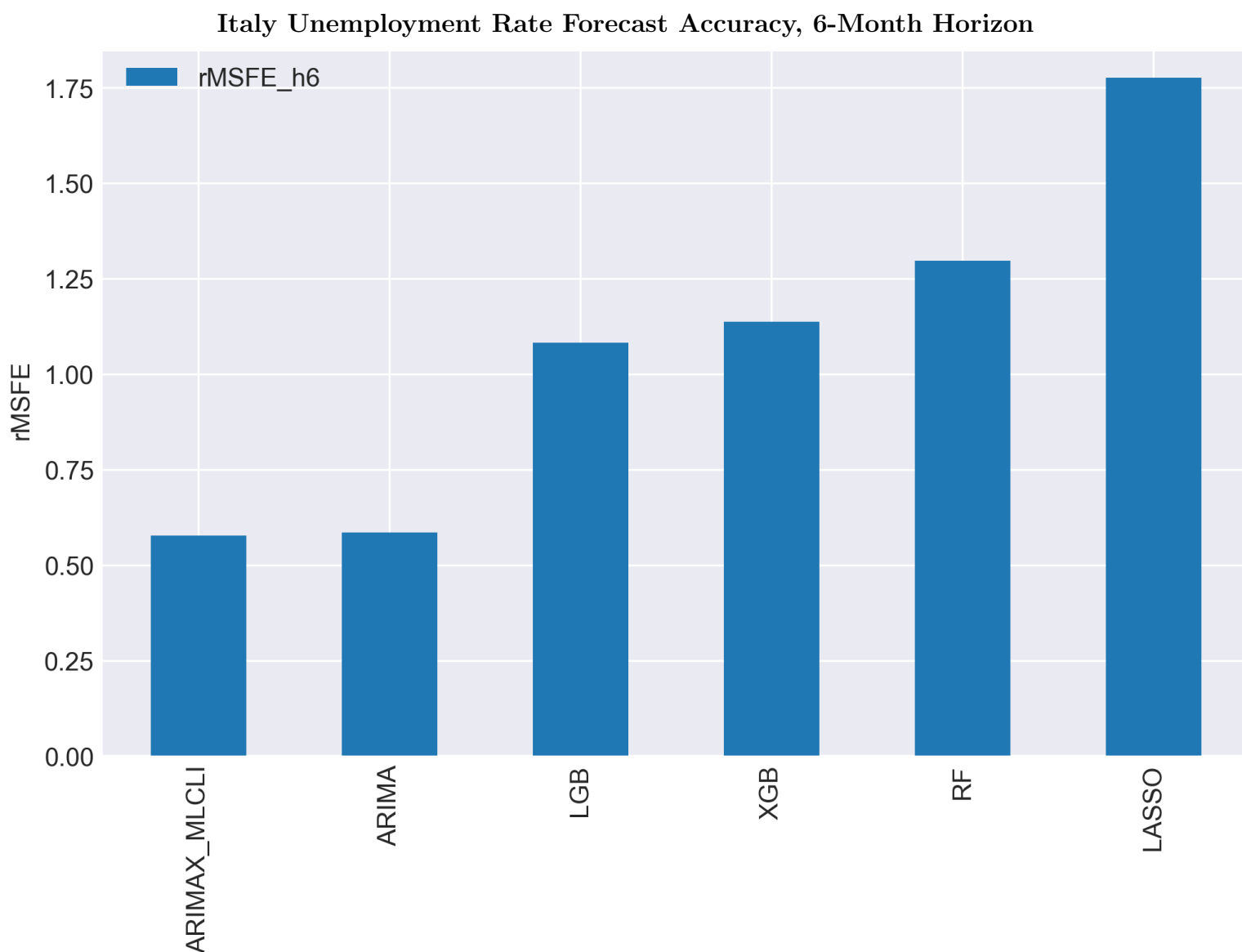


Figure C.39: Chapter 2 - Forecast accuracy results for Italy Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.5: *Diebold–Mariano tests for JPN (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-0.83 (0.406)
Inflation	6	LASSO	2.007 (0.045)**
Inflation	6	LGB	-3.372 (0.001)**
Inflation	6	RF	-1.835 (0.067)
Inflation	6	XGB	-3.077 (0.002)**
Unemployment	6	ARIMAX_MLCLI	-1.803 (0.071)
Unemployment	6	LASSO	6.837 (0.0)**
Unemployment	6	LGB	5.103 (0.0)**
Unemployment	6	RF	5.639 (0.0)**
Unemployment	6	XGB	4.932 (0.0)**
Industrial Production	6	ARIMAX_MLCLI	-1.371 (0.17)
Industrial Production	6	LASSO	1.796 (0.073)
Industrial Production	6	LGB	-1.679 (0.093)
Industrial Production	6	RF	-0.777 (0.437)
Industrial Production	6	XGB	-1.114 (0.265)

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Japan at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

Japan CPI Inflation Forecast Accuracy, 6-Month Horizon

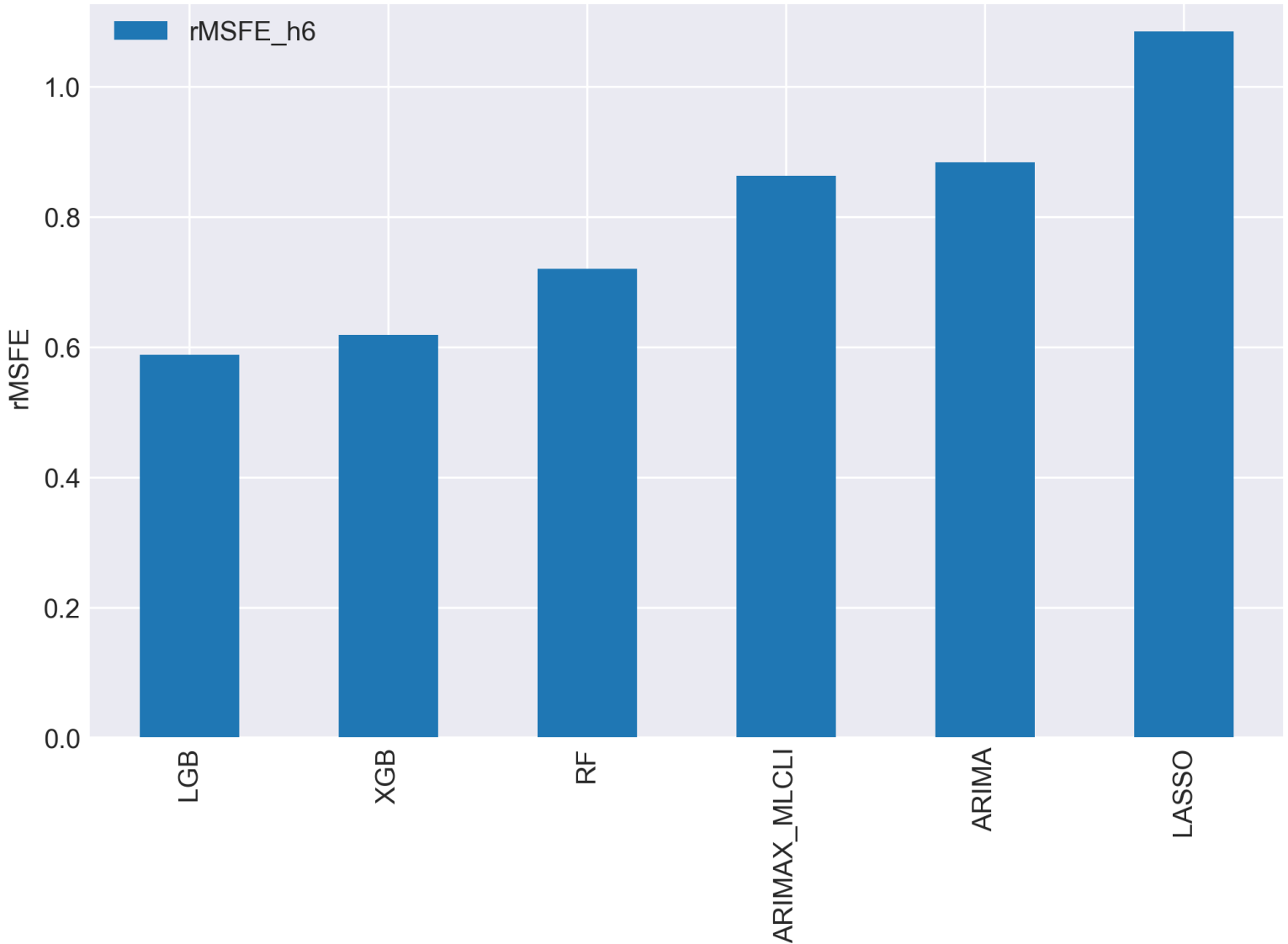


Figure C.40: Chapter 2 - Forecast accuracy results for Japan CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

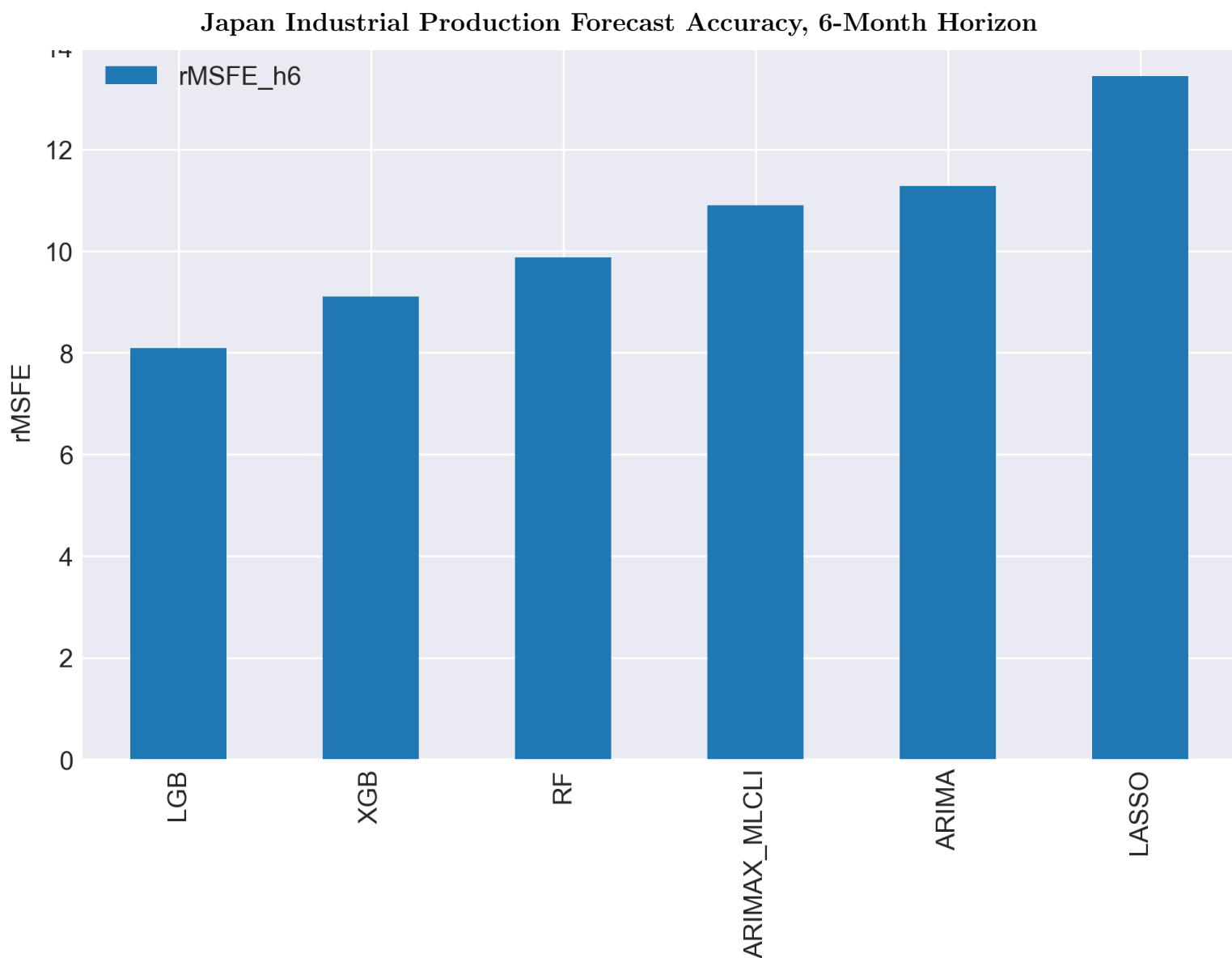


Figure C.41: Chapter 2 - Forecast accuracy results for Japan Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

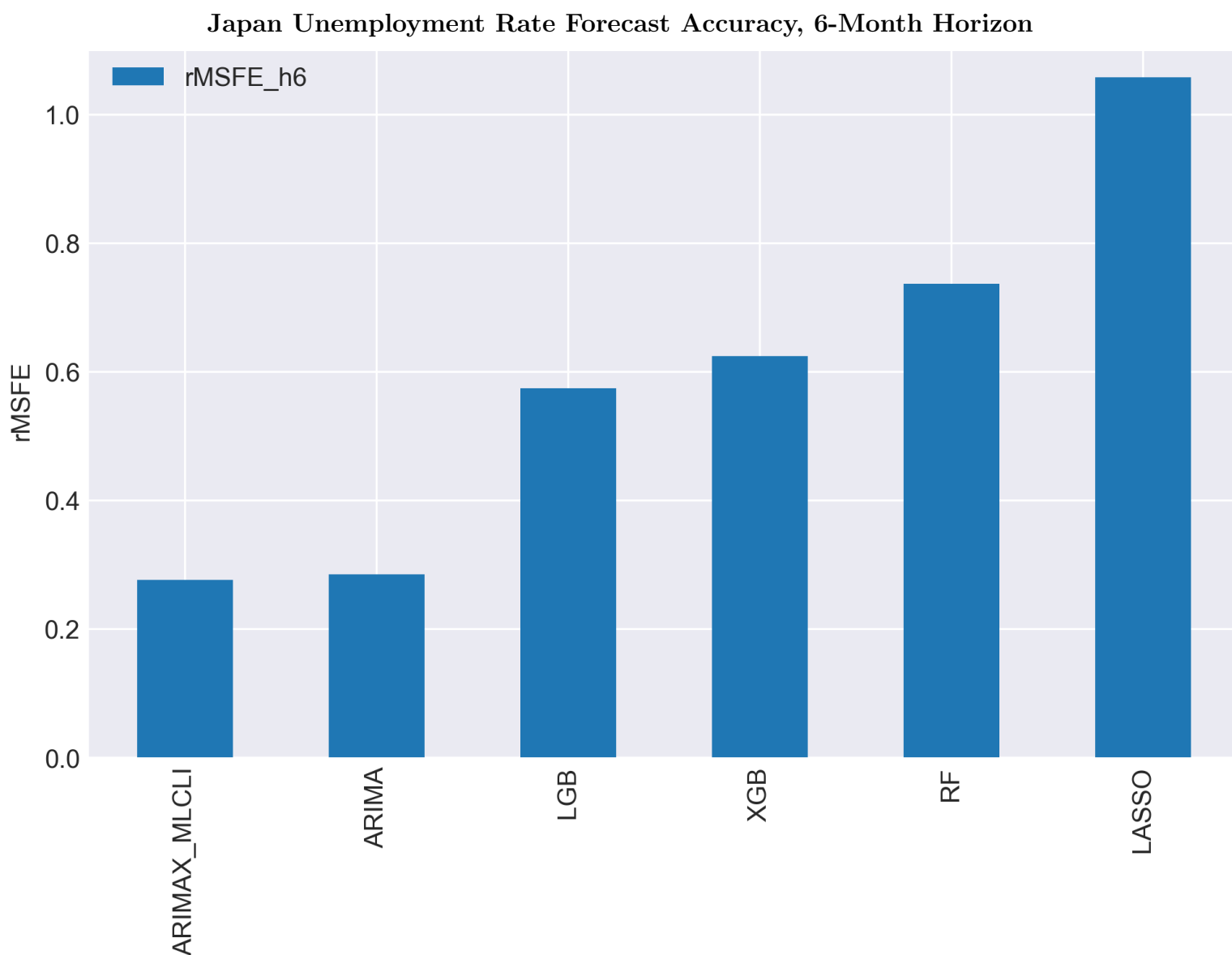


Figure C.42: Chapter 2 - Forecast accuracy results for Japan Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.6: *Diebold–Mariano tests for KOR (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	0.751 (0.453)
Inflation	6	LASSO	2.515 (0.012)**
Inflation	6	LGB	-1.262 (0.207)
Inflation	6	RF	-2.275 (0.023)**
Inflation	6	XGB	-3.477 (0.001)**
Unemployment	6	ARIMAX_MLCLI	-1.696 (0.09)
Unemployment	6	LASSO	3.615 (0.0)**
Unemployment	6	LGB	3.465 (0.001)**
Unemployment	6	RF	2.102 (0.036)**
Unemployment	6	XGB	2.338 (0.019)**
Industrial Production	6	ARIMAX_MLCLI	-0.533 (0.594)
Industrial Production	6	LASSO	11.259 (0.0)**
Industrial Production	6	LGB	5.743 (0.0)**
Industrial Production	6	RF	6.561 (0.0)**
Industrial Production	6	XGB	6.593 (0.0)**

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for South Korea at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

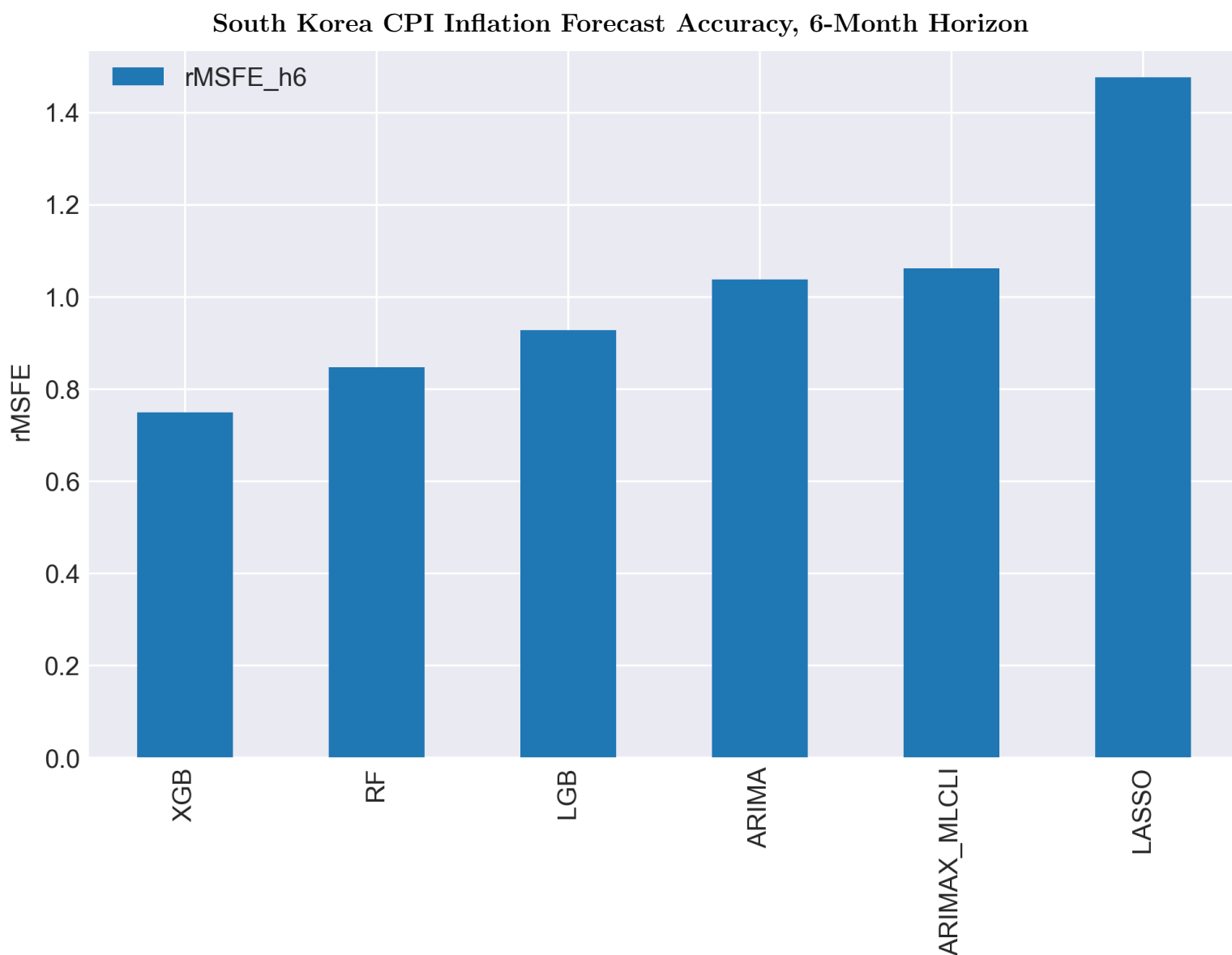


Figure C.43: Chapter 2 - Forecast accuracy results for South Korea CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

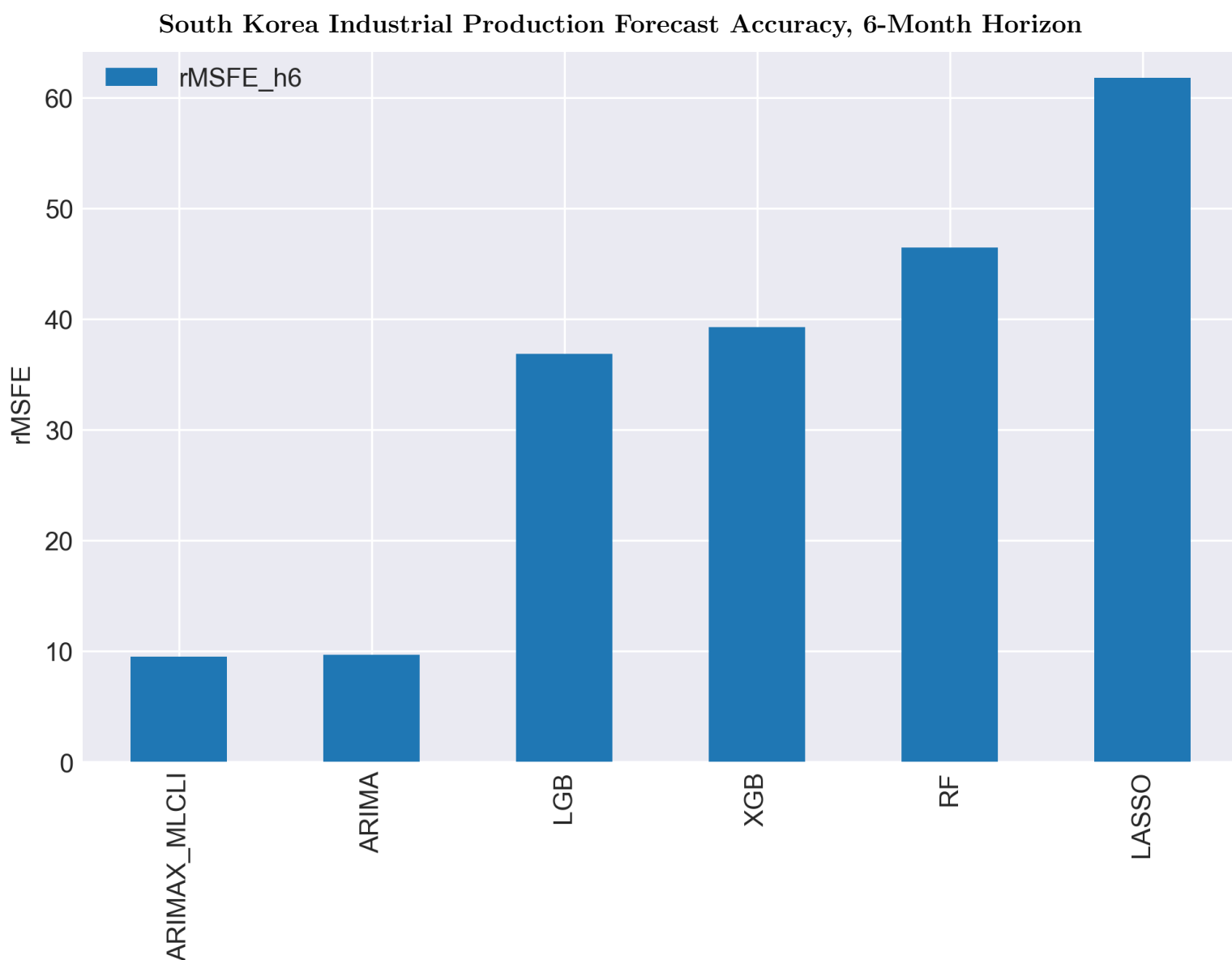


Figure C.44: Chapter 2 - Forecast accuracy results for South Korea Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

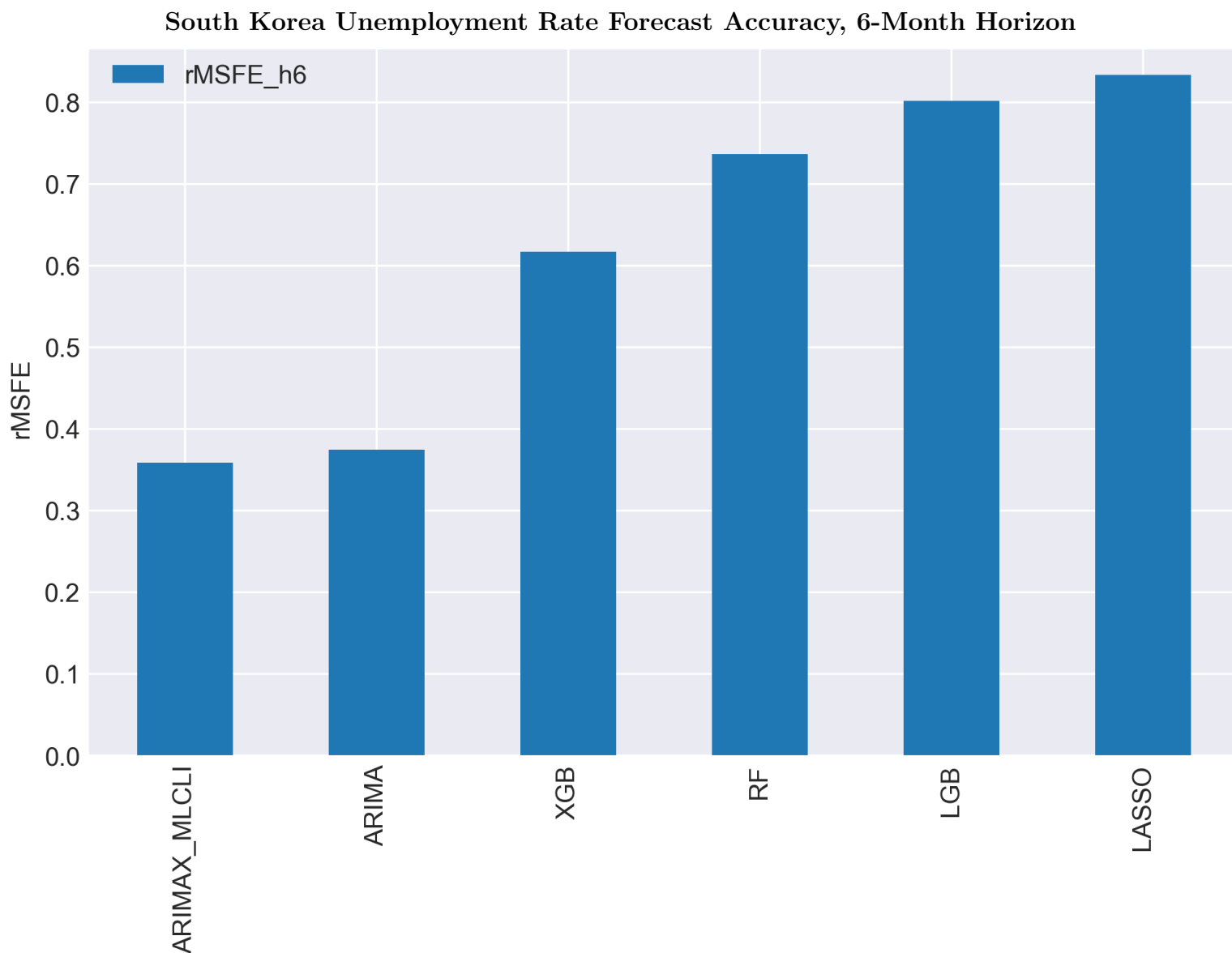


Figure C.45: Chapter 2 - Forecast accuracy results for South Korea Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.7: *Diebold–Mariano tests for MEX (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	0.394 (0.694)
Inflation	6	LASSO	-0.342 (0.732)
Inflation	6	LGB	-3.412 (0.001)**
Inflation	6	RF	-2.475 (0.013)**
Inflation	6	XGB	-3.653 (0.0)**
Unemployment	6	ARIMAX_MLCLI	-1.936 (0.053)
Unemployment	6	LASSO	3.367 (0.001)**
Unemployment	6	LGB	0.047 (0.962)
Unemployment	6	RF	2.286 (0.022)**
Unemployment	6	XGB	0.068 (0.946)
Industrial Production	6	ARIMAX_MLCLI	-0.371 (0.711)
Industrial Production	6	LASSO	3.788 (0.0)**
Industrial Production	6	LGB	-0.176 (0.86)
Industrial Production	6	RF	1.264 (0.206)
Industrial Production	6	XGB	-0.504 (0.614)

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Mexico at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

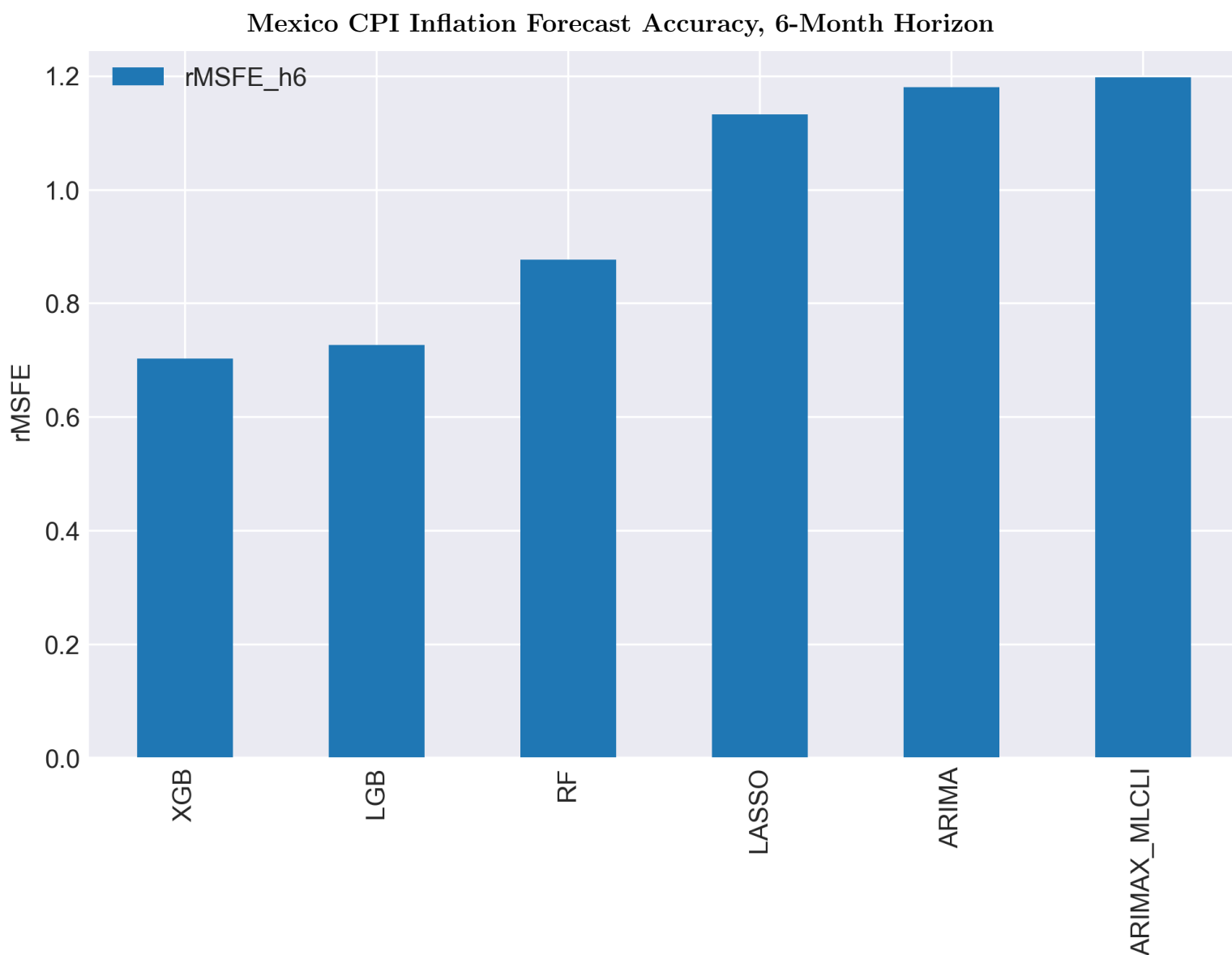


Figure C.46: Chapter 2 - Forecast accuracy results for Mexico CPI inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

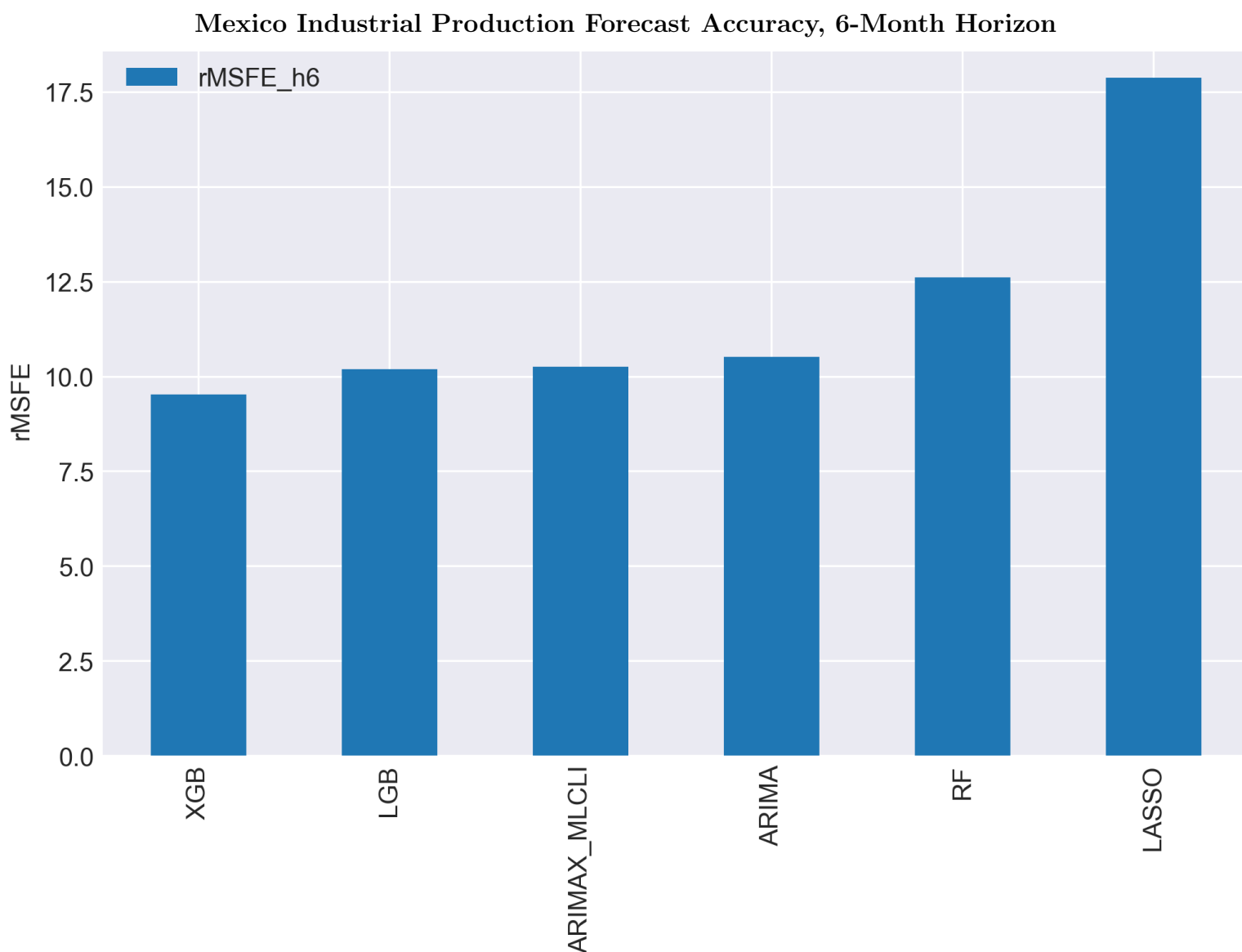


Figure C.47: Chapter 2 - Forecast accuracy results for Mexico Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

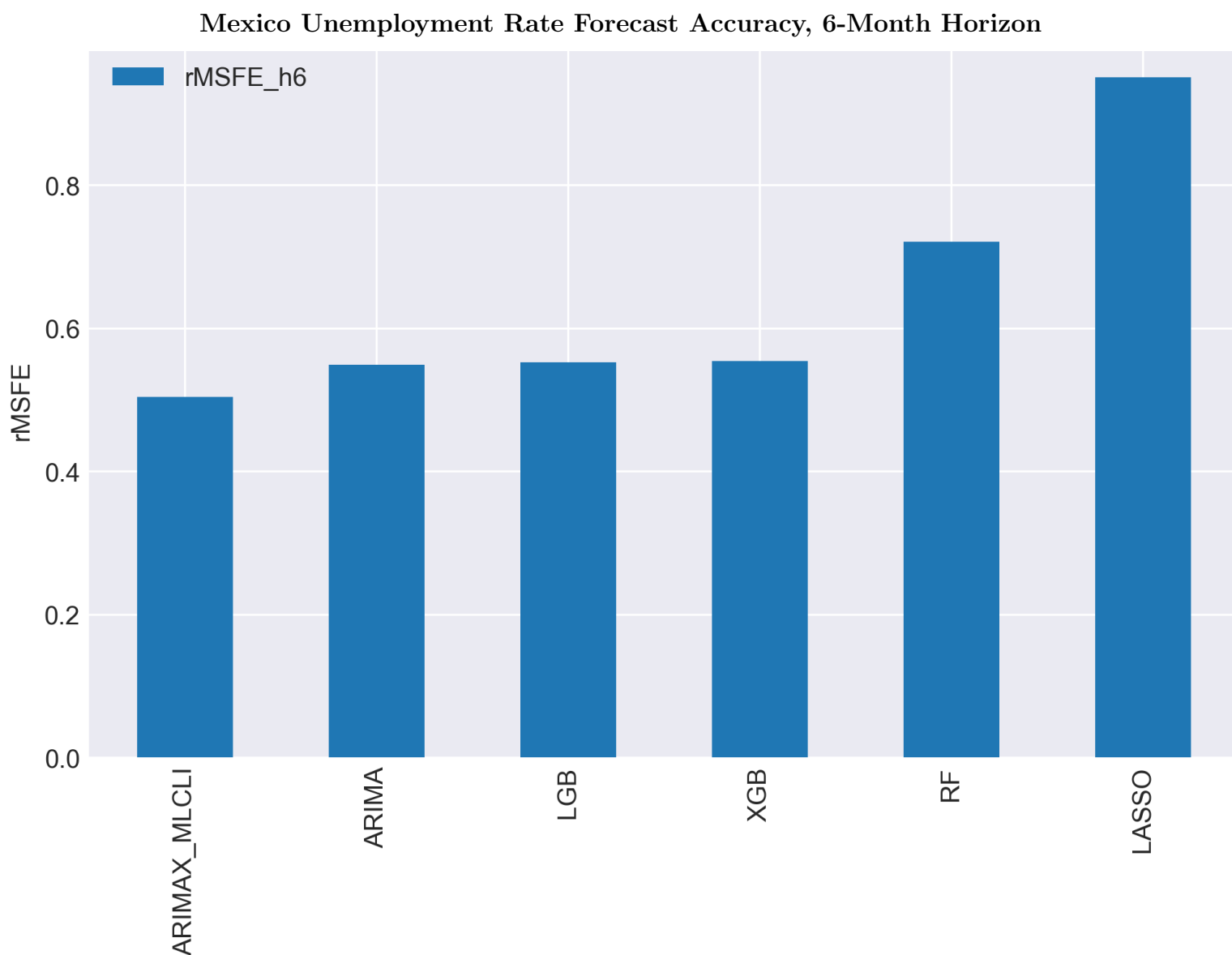


Figure C.48: Chapter 2 - Forecast accuracy results for Mexico Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.8: *Diebold–Mariano tests for TUR (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-1.469 (0.142)
Inflation	6	LASSO	-0.248 (0.804)
Inflation	6	LGB	-1.467 (0.142)
Inflation	6	RF	-0.852 (0.394)
Inflation	6	XGB	-1.35 (0.177)
Unemployment	6	ARIMAX_MLCLI	-2.618 (0.009)**
Unemployment	6	LASSO	1.001 (0.317)
Unemployment	6	LGB	-2.301 (0.021)**
Unemployment	6	RF	-0.889 (0.374)
Unemployment	6	XGB	-2.046 (0.041)**
Industrial Production	6	ARIMAX_MLCLI	-1.284 (0.199)
Industrial Production	6	LASSO	2.517 (0.012)**
Industrial Production	6	LGB	1.944 (0.052)
Industrial Production	6	RF	2.687 (0.007)**
Industrial Production	6	XGB	1.906 (0.057)

*Additional Chapter 2 Diebold–Mariano forecast comparison tests for Turkey at the 6-month forecast horizon. The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level. These results correspond to the forecast evaluation procedure described in Section 2.3.*

Turkey CPI Inflation Forecast Accuracy, 6-Month Horizon

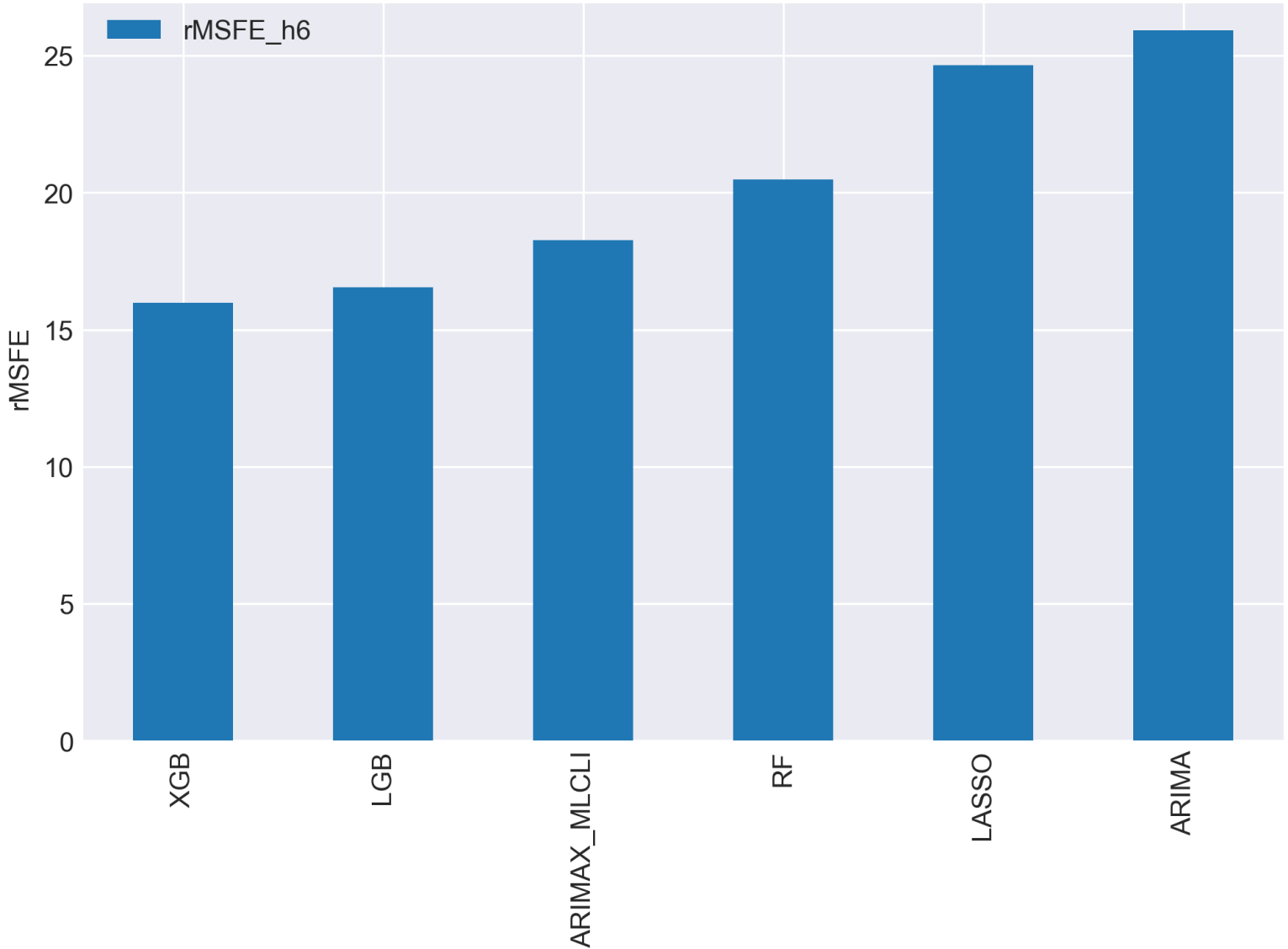


Figure C.49: Chapter 2 - Forecast accuracy results for Turkey CPI Inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

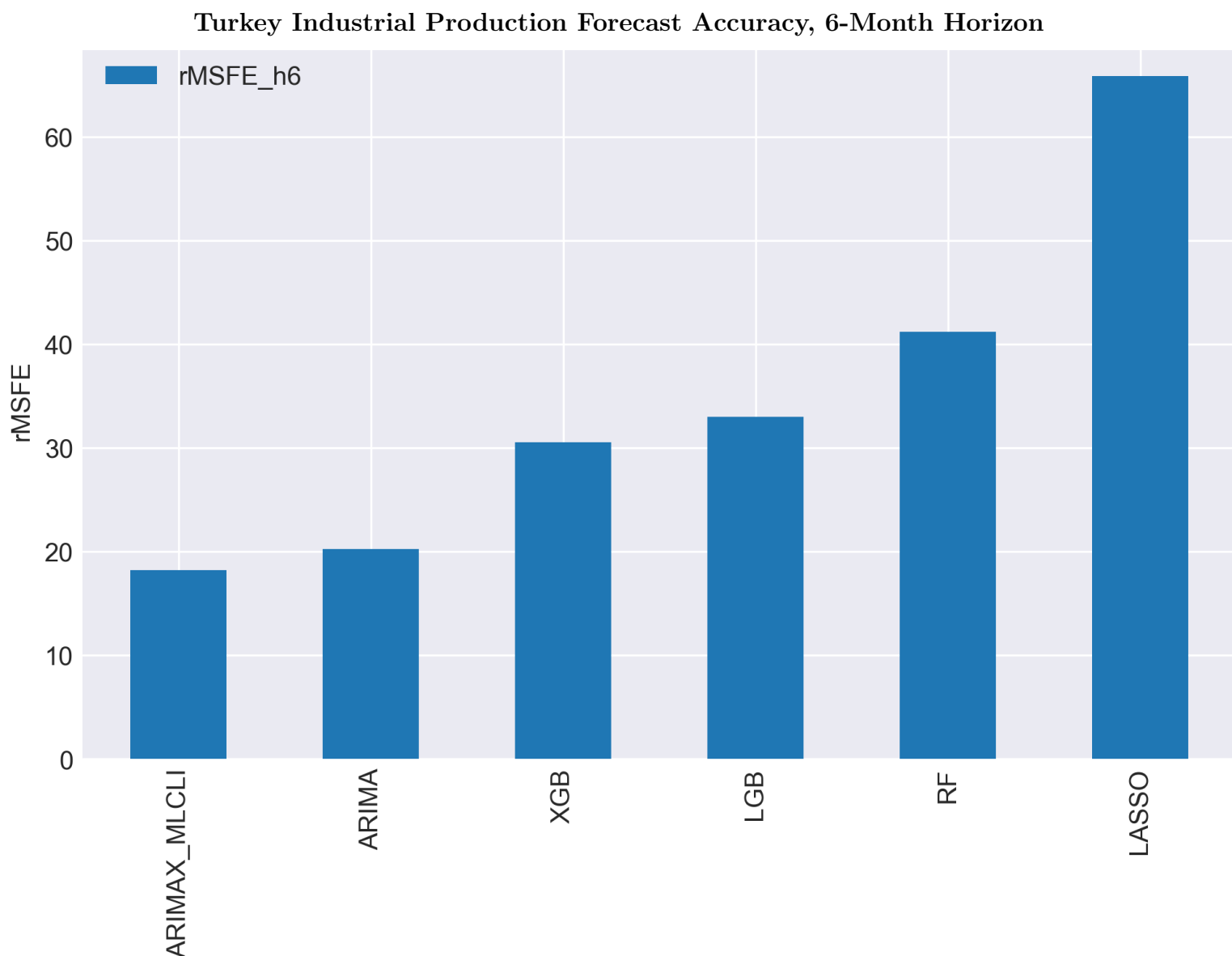


Figure C.50: Chapter 2 - Forecast accuracy results for Turkey Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error ($rMSFE$) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower $rMSFE$ values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

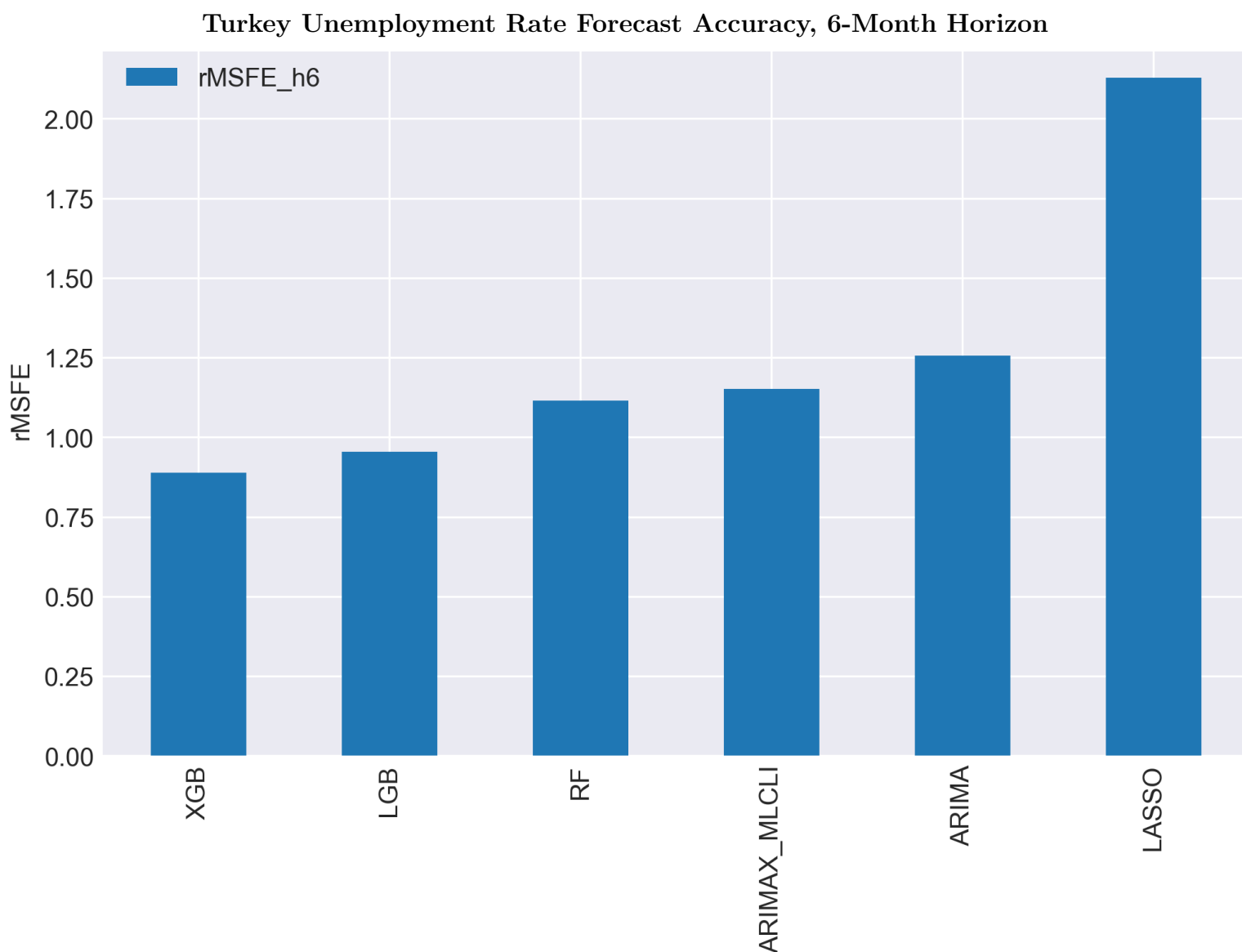


Figure C.51: Chapter 2 - Forecast accuracy results for Turkey Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

Table C.9: *Diebold–Mariano tests for USA (DM statistic with p-value)*

target	h	comparison	DM (p)
Inflation	6	ARIMAX_MLCLI	-0.442 (0.658)
Inflation	6	LASSO	5.892 (0.0)**
Inflation	6	LGB	1.978 (0.048)**
Inflation	6	RF	2.717 (0.007)**
Inflation	6	XGB	1.502 (0.133)
Unemployment	6	ARIMAX_MLCLI	-0.169 (0.866)
Unemployment	6	LASSO	3.048 (0.002)**
Unemployment	6	LGB	-1.019 (0.308)
Unemployment	6	RF	-0.502 (0.616)
Unemployment	6	XGB	-0.89 (0.374)
Industrial Production	6	ARIMAX_MLCLI	-0.735 (0.463)
Industrial Production	6	LASSO	11.543 (0.0)**
Industrial Production	6	LGB	7.764 (0.0)**
Industrial Production	6	RF	8.007 (0.0)**
Industrial Production	6	XGB	8.002 (0.0)**

Note: The table reports DM statistics and p-values for pairwise forecast accuracy tests across inflation, unemployment, and industrial production. The comparison column identifies the model tested against the benchmark forecast used in Chapter 2. P-values are reported in parentheses, and ** denotes statistical significance at the 5 percent level.

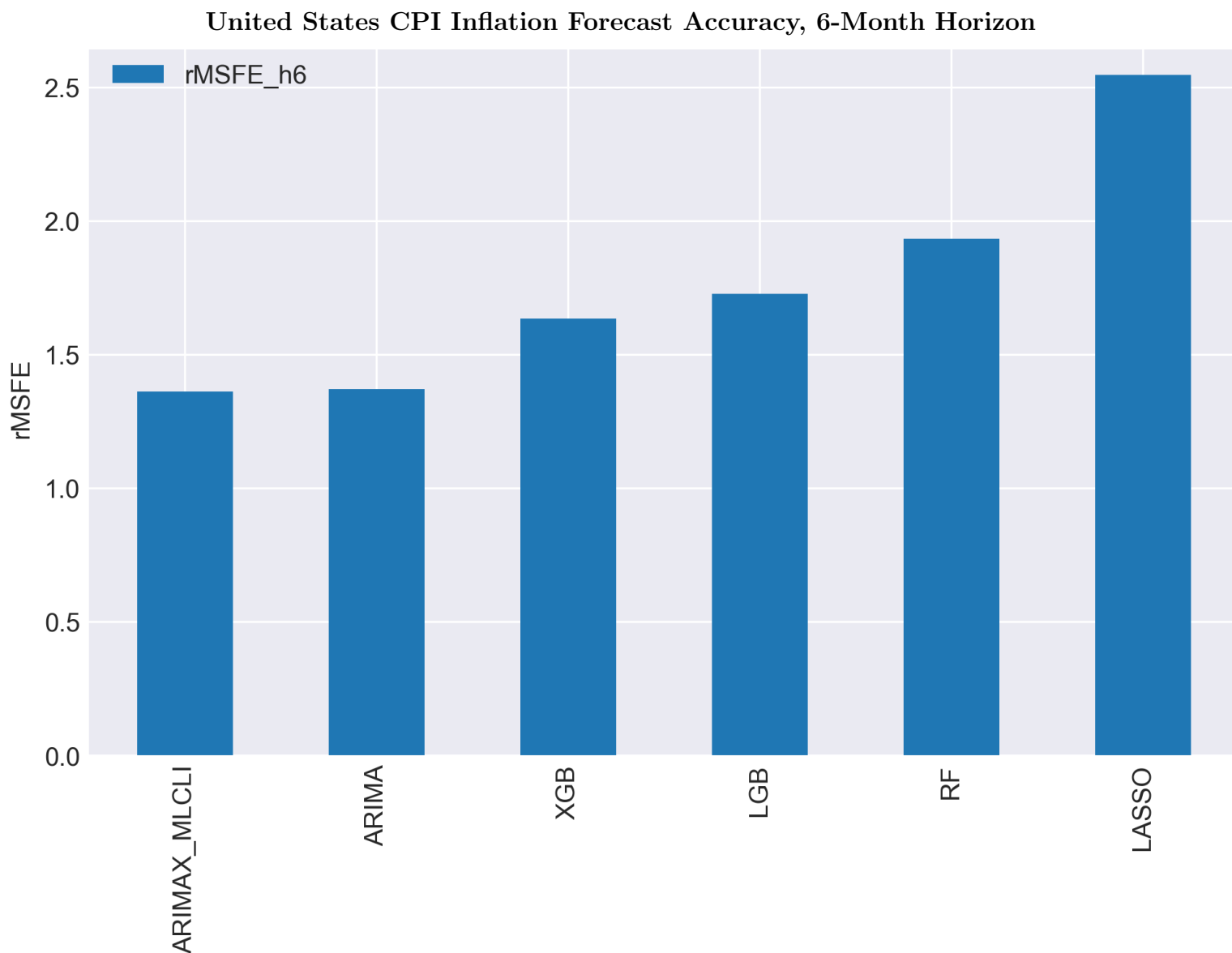


Figure C.52: Chapter 2 - Forecast accuracy results for the United States CPI Inflation model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

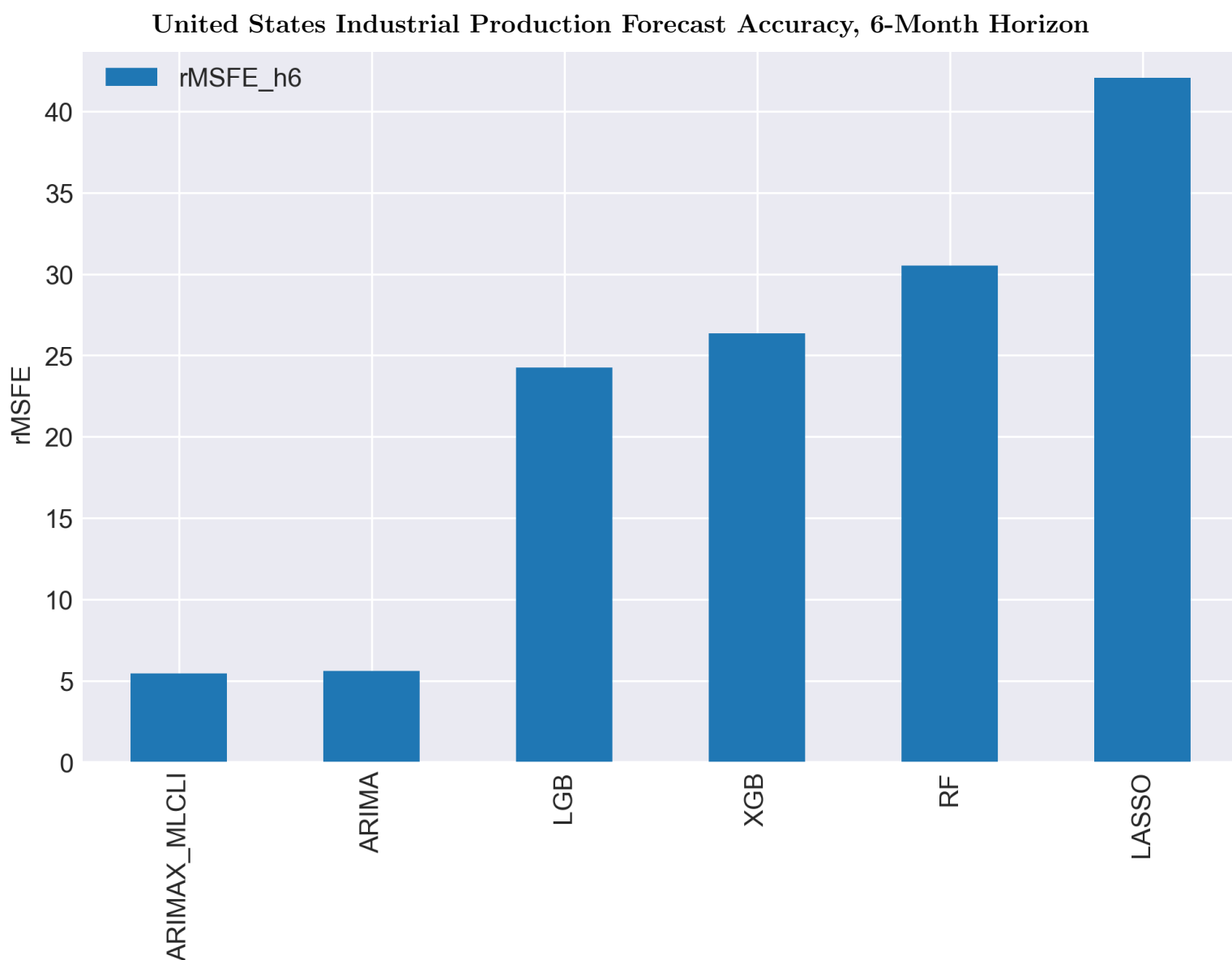


Figure C.53: Chapter 2 - Forecast accuracy results for the United States Industrial Production model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

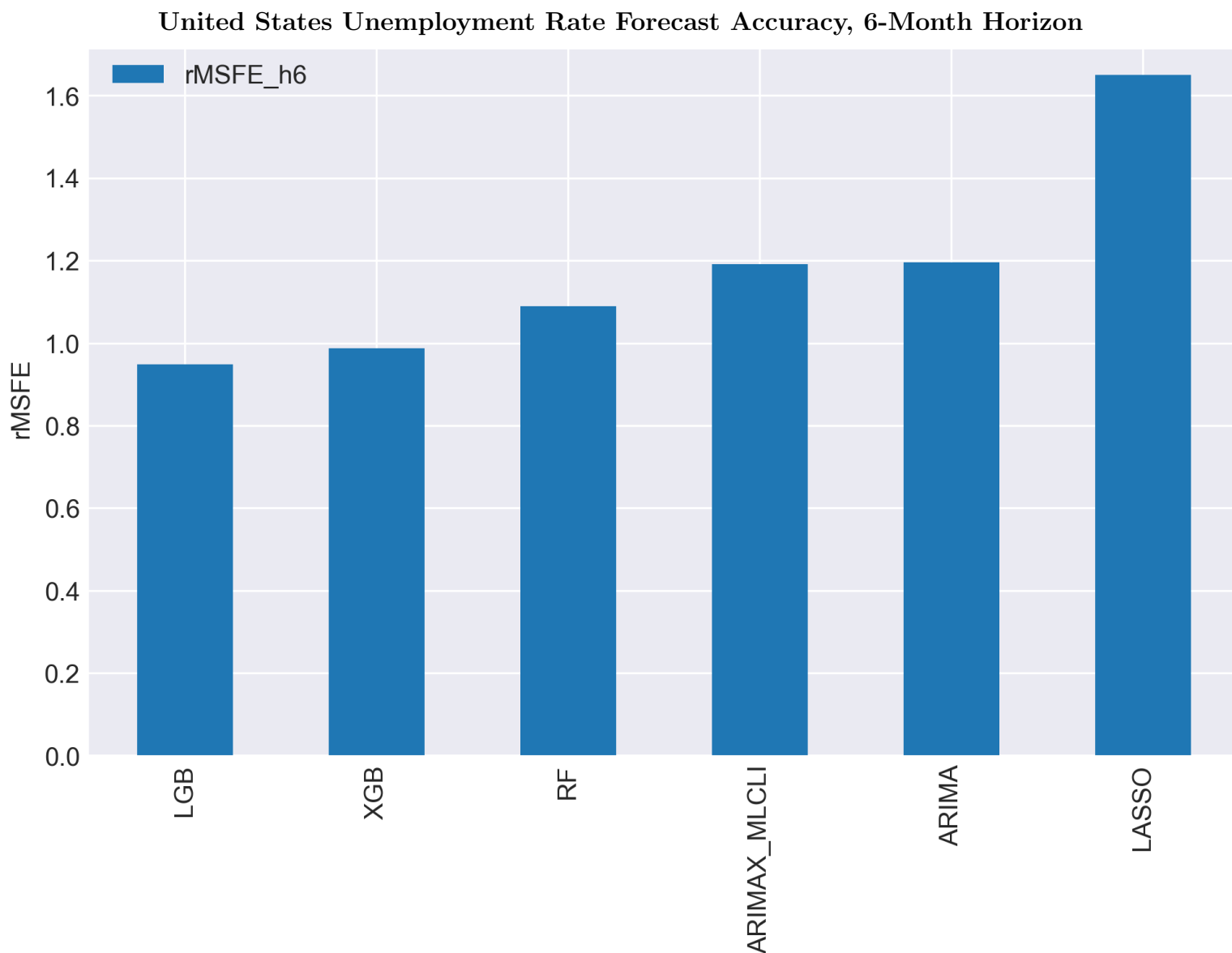


Figure C.54: Chapter 2 - Forecast accuracy results for the United States Unemployment Rate model at the six-month forecast horizon. The figure reports out-of-sample root mean squared forecast error (rMSFE) values across ARIMA, LASSO, random forest (RF), LightGBM (LGB), XGBoost (XGB), and the ARIMAX model using the machine-learning-reweighted CLI. Lower rMSFE values indicate more accurate forecasts. These results correspond to the model comparison procedure described in Section 2.3 and the robustness discussion in Section 2.4.

C.3 Additional Results for Chapter 3

This appendix presents supplementary results for the machine-learning-based composite leading indicator analysis discussed in Chapter 3. The results correspond to the indicator construction, forecast evaluation, and model comparison procedures described in Sections 3.3 and 3.4. The appendix includes the full country-target-horizon forecast comparison table and the target-specific machine-learning-derived weights used to construct the new composite leading indicators.

The full results table reports the complete set of model comparisons across countries, target variables, and forecast horizons. These results provide the detailed empirical basis for the summary findings reported in the main text. The weight tables report the normalized weights assigned to the selected predictors in each machine-learning-based composite indicator. These weights clarify which variables drive the construction of the new indicators and provide additional evidence on the economic content of the machine-learning forecasts.

Table C.10: Chapter 3 - Forecast comparison across countries, targets, and horizons: baseline AR, OECD CLI, and MLCI.

Country	Target- h	% Δ MSE	Δ Adj. R^2	p -value	Winner
Canada	Inflation-1	0.38	-0.0000	0.3230	MLCI
Canada	Inflation-3	1.68	-0.0001	0.3290	MLCI
Canada	Inflation-6	0.63	0.0006	0.2362	MLCI
Canada	Unemployment-1	97.25	0.8640	0.0000	MLCI
Canada	Unemployment-3	90.00	0.7827	0.0000	MLCI
Canada	Unemployment-6	81.00	0.6909	0.0000	MLCI
Canada	Industrial Production-1	-10.34	0.0028	0.0756	CLI
Canada	Industrial Production-3	-3.38	0.0015	0.1479	CLI
Canada	Industrial Production-6	-7.38	-0.0014	0.9932	CLI
Spain	Inflation-1	-0.25	0.0177	0.0058	CLI
Spain	Inflation-3	0.04	0.0214	0.0027	MLCI
Spain	Inflation-6	-0.29	0.0100	0.0308	CLI
Spain	Unemployment-1	54.03	0.5445	0.0000	MLCI
Spain	Unemployment-3	54.35	0.5437	0.0000	MLCI
Spain	Unemployment-6	53.31	0.5334	0.0000	MLCI
Spain	Industrial Production-1	-9.21	0.0185	0.0024	CLI
Spain	Industrial Production-3	11.99	0.0075	0.0414	MLCI
Spain	Industrial Production-6	9.16	-0.0008	0.4110	MLCI
Great Britain	Inflation-1	5.06	0.0028	0.0798	MLCI
Great Britain	Inflation-3	-1.18	0.0006	0.2266	CLI

Continued on next page

Table C.10: Chapter 3 - Country-level forecast comparison
of the MLCI relative to the OECD CLI.

Country	Target- h	% Δ MSE	Δ Adj. R^2	p -value	Winner
Great Britain	Inflation-6	0.18	0.0010	0.1892	MLCI
Great Britain	Unemployment-1	80.74	0.3479	0.0000	MLCI
Great Britain	Unemployment-3	80.63	0.3620	0.0000	MLCI
Great Britain	Unemployment-6	79.57	0.3651	0.0000	MLCI
Great Britain	Industrial Production-1	-2.78	0.0003	0.2725	CLI
Great Britain	Industrial Production-3	8.07	-0.0002	0.3592	MLCI
Great Britain	Industrial Production-6	8.73	-0.0013	0.8483	MLCI
Italy	Inflation-1	0.08	0.0048	0.0407	MLCI
Italy	Inflation-3	0.96	0.0031	0.0791	MLCI
Italy	Inflation-6	0.78	-0.0002	0.3510	MLCI
Italy	Unemployment-1	98.36	0.9848	0.0000	MLCI
Italy	Unemployment-3	95.18	0.9505	0.0000	MLCI
Italy	Unemployment-6	91.52	0.9098	0.0000	MLCI
Italy	Industrial Production-1	4.61	-0.0002	0.3618	MLCI
Italy	Industrial Production-3	6.36	-0.0012	0.6716	MLCI
Italy	Industrial Production-6	3.95	-0.0012	0.6558	MLCI
Japan	Inflation-1	-0.85	-0.0018	0.7375	CLI
Japan	Inflation-3	-0.57	-0.0021	0.9387	CLI
Japan	Inflation-6	-0.88	-0.0021	0.9813	CLI
Japan	Unemployment-1	87.10	0.5668	0.0000	MLCI
Japan	Unemployment-3	86.44	0.5717	0.0000	MLCI

Continued on next page

Table C.10: Chapter 3 - Country-level forecast comparison
of the MLCI relative to the OECD CLI.

Country	Target- h	% Δ MSE	Δ Adj. R^2	p -value	Winner
Japan	Unemployment-6	85.14	0.5733	0.0000	MLCI
Japan	Industrial Production-1	-8.31	-0.0004	0.4252	CLI
Japan	Industrial Production-3	-4.29	-0.0008	0.5604	CLI
Japan	Industrial Production-6	-1.16	0.0017	0.1321	CLI
South Korea	Inflation-1	-0.04	-0.0018	0.5482	CLI
South Korea	Inflation-3	-0.61	-0.0027	0.8511	CLI
South Korea	Inflation-6	1.38	-0.0028	0.9934	MLCI
South Korea	Unemployment-1	6.20	0.4201	0.0000	MLCI
South Korea	Unemployment-3	4.53	0.3850	0.0000	MLCI
South Korea	Unemployment-6	-2.11	0.2606	0.0000	CLI
South Korea	Industrial Production-1	-5.36	-0.0016	0.5276	CLI
South Korea	Industrial Production-3	-6.01	-0.0027	0.9329	CLI
South Korea	Industrial Production-6	-5.98	0.0003	0.2964	CLI
Mexico	Inflation-1	-3.59	-0.0016	0.5912	CLI
Mexico	Inflation-3	-8.28	0.0071	0.0385	CLI
Mexico	Inflation-6	-5.52	0.0252	0.0004	CLI
Mexico	Unemployment-1	9.27	0.2944	0.0000	MLCI
Mexico	Unemployment-3	3.19	0.2377	0.0000	MLCI
Mexico	Unemployment-6	-5.97	0.1594	0.0000	CLI
Mexico	Industrial Production-1	-2.24	0.0155	0.0043	CLI
Mexico	Industrial Production-3	3.44	0.0011	0.2213	MLCI

Continued on next page

Table C.10: Chapter 3 - Country-level forecast comparison
of the MLCI relative to the OECD CLI.

Country	Target- h	% Δ MSE	Δ Adj. R^2	p -value	Winner
Mexico	Industrial Production-6	3.80	0.0007	0.2487	MLCI
Turkey	Inflation-1	-0.40	-0.0026	0.9920	CLI
Turkey	Inflation-3	0.87	-0.0025	0.8462	MLCI
Turkey	Inflation-6	-0.84	-0.0008	0.4052	CLI
Turkey	Unemployment-1	-26.25	0.0404	0.0023	CLI
Turkey	Unemployment-3	-35.64	0.0108	0.0726	CLI
Turkey	Unemployment-6	-41.52	-0.0051	0.9459	CLI
Turkey	Industrial Production-1	2.36	-0.0018	0.5855	MLCI
Turkey	Industrial Production-3	10.96	0.0039	0.1135	MLCI
Turkey	Industrial Production-6	6.19	0.0037	0.1194	MLCI
USA	Inflation-1	-0.35	0.0007	0.2223	CLI
USA	Inflation-3	2.86	-0.0001	0.3432	MLCI
USA	Inflation-6	1.86	-0.0007	0.4888	MLCI
USA	Unemployment-1	92.99	0.8499	0.0000	MLCI
USA	Unemployment-3	79.35	0.7308	0.0000	MLCI
USA	Unemployment-6	62.14	0.5945	0.0000	MLCI
USA	Industrial Production-1	-13.96	0.0624	0.0000	CLI
USA	Industrial Production-3	-9.66	0.0124	0.0009	CLI
USA	Industrial Production-6	-23.75	-0.0003	0.3732	CLI

Table C.11: *Canada chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.062441	0.110882	0.042774
Composite consumer confidence	0.070118	0.083441	0.058882
Commodity prices	0.097118	0.072774	0.038774
Employment, total	0.058774	0.067441	0.106118
Exports of goods and services	0.082774	0.096441	0.043774
Imports of goods and services	0.079441	0.072774	0.036774
Long-term interest rates	0.053441	0.049882	0.087774
Short-term interest rates / call money rate	0.048118	0.044441	0.062774
Monetary aggregate, narrow money	0.091118	0.093774	0.125118
Orders inflow	0.061774	0.121441	0.032774
Production volume	0.094774	0.159441	0.049774
Retail sales	0.086774	0.074441	0.146774
Share prices	0.060774	0.055774	0.029774
Unemployment level	0.084774	0.065774	0.175441
Unemployment rate	0.081880	0.068095	0.188306

Table C.12: *Spain chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.061114	0.113227	0.043118
Composite consumer confidence	0.069227	0.084118	0.059441
Commodity prices	0.096441	0.073118	0.038227
Employment, total	0.057882	0.067118	0.105774
Exports of goods and services	0.082118	0.097227	0.043118
Imports of goods and services	0.078882	0.072118	0.036441
Long-term interest rates	0.053118	0.049774	0.087118
Short-term interest rates / call money rate	0.047774	0.044118	0.062118
Monetary aggregate, narrow money	0.090227	0.094118	0.124882
Orders inflow	0.061118	0.122774	0.032118
Production volume	0.094118	0.160227	0.049118
Retail sales	0.086118	0.075118	0.147118
Share prices	0.060118	0.055118	0.029118
Unemployment level	0.084118	0.065118	0.174118
Unemployment rate	0.080887	0.067585	0.187233

Table C.13: *Great Britain chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.060882	0.111447	0.042118
Composite consumer confidence	0.068331	0.082774	0.057993
Commodity prices	0.094771	0.071882	0.037441
Employment, total	0.056882	0.066118	0.103227
Exports of goods and services	0.081003	0.095227	0.042441
Imports of goods and services	0.078114	0.071003	0.035774
Long-term interest rates	0.052661	0.048993	0.085774
Short-term interest rates / call money rate	0.047003	0.043662	0.061003
Monetary aggregate, narrow money	0.089118	0.092447	0.122774
Orders inflow	0.060118	0.120883	0.031441
Production volume	0.092774	0.157662	0.047662
Retail sales	0.084771	0.073662	0.144882
Share prices	0.059227	0.054118	0.028331
Unemployment level	0.082774	0.063774	0.171882
Unemployment rate	0.079771	0.065662	0.184551

Table C.14: *Italy chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.061337	0.112884	0.042991
Composite consumer confidence	0.069114	0.083227	0.058662
Commodity prices	0.095882	0.072441	0.037881
Employment, total	0.057331	0.066774	0.104552
Exports of goods and services	0.081774	0.096118	0.042883
Imports of goods and services	0.078552	0.071662	0.035991
Long-term interest rates	0.052884	0.049331	0.086552
Short-term interest rates / call money rate	0.047228	0.043991	0.061774
Monetary aggregate, narrow money	0.089662	0.093118	0.123441
Orders inflow	0.060774	0.121552	0.031662
Production volume	0.093227	0.158884	0.048114
Retail sales	0.085441	0.074118	0.145882
Share prices	0.059774	0.054662	0.028662
Unemployment level	0.083118	0.064227	0.172774
Unemployment rate	0.079902	0.066003	0.185191

Table C.15: *Japan chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.058221	0.104337	0.039118
Composite consumer confidence	0.066774	0.079552	0.056441
Commodity prices	0.091883	0.068771	0.036992
Employment, total	0.054662	0.063991	0.097882
Exports of goods and services	0.079554	0.093776	0.041227
Imports of goods and services	0.076993	0.069118	0.034552
Long-term interest rates	0.051112	0.047884	0.081993
Short-term interest rates / call money rate	0.045771	0.042663	0.059884
Monetary aggregate, narrow money	0.085993	0.089441	0.118227
Orders inflow	0.058774	0.114992	0.030771
Production volume	0.089882	0.151883	0.046882
Retail sales	0.081772	0.070991	0.139552
Share prices	0.058339	0.053882	0.027771
Unemployment level	0.079991	0.061883	0.168114
Unemployment rate	0.075279	0.063821	0.180594

Table C.16: *South Korea chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.063418	0.109527	0.041882
Composite consumer confidence	0.071204	0.082119	0.059337
Commodity prices	0.094661	0.071552	0.038220
Employment, total	0.058993	0.066407	0.102771
Exports of goods and services	0.083114	0.097338	0.043559
Imports of goods and services	0.079887	0.072996	0.036884
Long-term interest rates	0.052773	0.049881	0.084112
Short-term interest rates / call money rate	0.047662	0.044530	0.062408
Monetary aggregate, narrow money	0.088531	0.091662	0.121554
Orders inflow	0.061228	0.118774	0.032116
Production volume	0.092774	0.156882	0.049003
Retail sales	0.084339	0.073115	0.143227
Share prices	0.060112	0.055774	0.029441
Unemployment level	0.082441	0.063882	0.171998
Unemployment rate	0.078863	0.065561	0.183488

Table C.17: *Mexico chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite consumer confidence	0.058421	0.066907	0.041215
Commodity prices	0.083274	0.059318	0.033907
Employment, total	0.064118	0.072441	0.118906
Exports of goods and services	0.071503	0.083229	0.036442
Imports of goods and services	0.079662	0.068991	0.027885
Long-term interest rates	0.052004	0.047118	0.089731
Monetary aggregate, narrow money	0.094771	0.088532	0.142660
Production volume	0.088936	0.140227	0.052119
Retail sales	0.102553	0.081774	0.164382
Share prices	0.060317	0.055991	0.028104
Unemployment level	0.086441	0.064229	0.169308
Unemployment rate	0.058000	0.071243	0.095341

Table C.18: *Turkey chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite consumer confidence	0.166946	0.168491	0.118602
Commodity prices	0.109477	0.093330	0.091141
Exports of goods and services	0.110330	0.109538	0.044986
Imports of goods and services	0.142083	0.127281	0.147603
Monetary aggregate, narrow money	0.138301	0.125612	0.235001
Production volume	0.148460	0.271212	0.051739
Retail sales	0.184403	0.104536	0.310928

Table C.19: *USA chapter 3 weights by measure*

Measure	Inflation	Industrial Production	Unemployment
Composite business confidence	0.022968	0.091347	0.000007
Composite consumer confidence	0.016693	0.021033	0.000074
Commodity prices	0.082886	0.030362	0.000018
Employment expectations	0.048966	0.035254	0.000007
Employment, total	0.029398	0.022491	0.001109
Exports of goods and services	0.031117	0.018489	0.000018
Hourly earnings (manufacturing)	0.036613	0.046622	0.000010
Imports of goods and services	0.036609	0.026649	0.000001
Long-term interest rates	0.052556	0.028003	0.009695
Short-term interest rates / call money rate	0.034361	0.021544	0.000401
Composite leading indicator	0.032852	0.037831	0.000009
Monetary aggregate, broad money	0.044252	0.082964	0.000042
Monetary aggregate, narrow money	0.055385	0.052339	0.001430
Orders inflow	0.025243	0.108462	0.000003
Producer prices	0.067707	0.022253	0.002649
Industrial production	0.053993	0.094620	0.000001
Production volume	0.034390	0.035248	0.000021
Retail sales	0.055837	0.034676	0.003588
Share prices	0.029639	0.021363	0.000002
Passenger car registrations	0.046327	0.034547	0.000002
Retail trade turnover volume	0.045507	0.047251	0.000043
Unemployment level	0.061965	0.028052	0.465311
Unemployment rate	0.029774	0.038787	0.515202
Construction work started on dwellings	0.024962	0.019811	0.000358