Forecasting Singapore's Inflation Using LASSO: A Comparison with Traditional Time Series Models

Stefano Zurita
EC665 – Forecasting and Time Series
Professor Stephen Snudden

Abstract

This paper evaluates the performance of LASSO regression for forecasting monthly inflation in Singapore. Using data from 1990 to 2024, the model is compared against a Random Walk with Drift and an ARIMA(1,0,1) benchmark. Forecasts are generated using expanding window estimation and evaluated over the 2010–2024 period. Performance is assessed in terms of forecast accuracy and directional predictive power. The LASSO model outperforms both benchmarks in reducing forecast errors and shows meaningful improvements in predicting the direction of monthly inflation changes. The results highlight the usefulness of LASSO in settings with many candidate predictors.

1. Introduction

Inflation in Singapore reflects both domestic and external conditions. Domestic indicators such as industrial production, retail sales, and money supply interact with external drivers like exchange rates, oil prices, and global financial markets. The presence of many potential predictors complicates model selection and raises concerns about overfitting. This paper examines whether applying regularized regression to a high-dimensional predictor set improves the forecast accuracy of monthly CPI relative to conventional benchmarks.

Penalized regression methods such as LASSO have gained attention for their ability to handle many predictors while automatically selecting the most relevant ones. Medeiros and Vasconcelos (2016) apply adaptive LASSO (adaLASSO) to macroeconomic forecasts and show that it outperforms standard models when model complexity is controlled using information criteria such as the Bayesian Information Criterion (BIC). Despite these advances, applications of LASSO-type methods to small open economies—and specifically to CPI forecasting—remain limited.

This study applies LASSO regression to forecast Singapore's monthly CPI and compares its performance to a Random Walk with Drift and an ARIMA(1,0,1) model. The goal is to evaluate whether including a diverse set of domestic and international indicators improves one-step-ahead inflation forecasts. Forecasts are generated using an expanding window approach, with monthly data from 1990 to 2024. The evaluation period runs from 2010 to 2024 and yields 180 one-step-ahead forecasts. Forecast performance is assessed using the Mean Squared Forecast Error (MSFE), the Diebold and Mariano (1995) test for predictive accuracy, and the Pesaran and Timmermann (2009) test for directional performance.

The LASSO model achieves the lowest MSFE among the three models evaluated. The improvement over the Random Walk with Drift is statistically significant, as confirmed by the Diebold-Mariano test (p-value = 0.000). In terms of directional accuracy, LASSO correctly predicts the sign of monthly inflation 62.8% of the time, with a PT p-value of 0.019, indicating statistical significance. ARIMA and the Random Walk perform slightly worse in directional accuracy, and neither passes the Pesaran-Timmermann test. These results suggest that LASSO improves both forecast precision and the ability to anticipate the direction of inflation changes.

This project contributes to the literature by applying regularized regression methods to inflation forecasting in a small open economy. It provides a benchmark for future work in similar settings and illustrates the value of data-driven variable selection in improving forecast accuracy without increasing model complexity.

2. Model and Methods

This project compares three models for forecasting Singapore's monthly inflation: a Random Walk with Drift, an ARIMA(1,0,1), and a penalized regression method known as the Least Absolute Shrinkage and Selection Operator (LASSO). The forecasts are generated using an expanding window, and the models are evaluated based on predictive accuracy and directional performance.

LASSO improves forecast performance when there is a large number of predictors or when multicollinearity is present. It selects variables automatically by penalizing the absolute size of the coefficients, forcing some to zero. This reduces overfitting and improves generalization. Applications in macroeconomic forecasting, as shown in Medeiros and Vasconcelos (2016) and Giannone et al. (2021), have found that LASSO can outperforms traditional forecsting models.

The LASSO estimator minimizes the following objective function:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2N} \sum_{t=1}^{N} (y_t - X_t \beta')^2 + \lambda \sum_{j=1}^{p} \omega_j |\beta_j| \right\} \tag{1}$$

where y_t represents the log-difference in CPI between t and $t+1$, calculated as $\ln(\text{CPI}_{t+1}) - \ln(\text{CPI}_t)$. The vector X_t contains predictors observed at time t, λ is a tuning parameter chosen via cross-validation, and ω_j are variable-specific weights. The LASSO model produces a forecast of y_{t+1}, which corresponds to the expected monthly inflation rate in log-difference form. This forecast is then converted to the predicted level of CPI using the identity:

$$\hat{\text{CPI}}_{t+1} = \exp(\hat{y}_{t+1}) \cdot \text{CPI}_t.$$

This transformation ensures that the prediction is expressed on the same scale as the actual CPI values, allowing for a direct comparison between the forecasted and observed levels in levels.

To evaluate forecast performance, the LASSO model is compared to two benchmarks using one-step-ahead forecasts. The benchmark specifications are:

Random Walk with Drift:

$$\hat{y}_{t+1} = y_t + \mu + \varepsilon_{t+1} \tag{2}$$

where $y_t = \ln(\mathrm{CPI}_t)$, and μ is the estimated monthly drift.

ARIMA(1,0,1):

$$\hat{y}_{t+1} = \mu + \phi y_t + \theta \hat{\varepsilon}_t \tag{3}$$

where $y_t = \Delta \ln(\mathrm{CPI}_t)$ represents the monthly log-difference in CPI.

All model forecasts are transformed to levels using either $\exp(\hat{y}_{t+1})$ or $\exp(\hat{y}_{t+1}) \cdot \mathrm{CPI}_t$, depending on the model. This allows forecasted values to be directly compared to actual CPI levels.

The forecasts from all three models are evaluated using the following criteria. First, accuracy is measured using the Mean Squared Forecast Error (MSFE). Second, the Diebold and Mariano (1995) test with Newey-West standard errors is used to assess whether differences in forecast accuracy are statistically significant. Third, directional performance is evaluated using the Success Ratio, defined as the share of months in which the model correctly predicts the direction of inflation. The Pesaran and Timmermann (2009) test is used to assess the statistical significance of directional predictability.

3. Data

The dataset consists of monthly observations from January 1990 to December 2024. The target variable is Singapore's Consumer Price Index (CPI), obtained from Statistics Singapore.[1] Predictor variables include both domestic and international macroeconomic indicators commonly used in empirical inflation forecasting. Domestic variables include the USD/SGD exchange rate, oil prices, the Straits Times Index (STI), retail sales, industrial production, and M2 money supply. International indicators include the SPY Index and Chinese export growth, which capture external financial conditions and regional trade exposure. China is Singapore's largest trading partner, and fluctuations in Chinese exports directly affect Singapore's economy (OEC World, 2024).

[1] Source: Statistics Singapore, Monthly Consumer Price Index.

The choice of variables is consistent with prior literature. Parrado (2024) shows that Singapore uses the exchange rate as its main monetary policy tool to stabilize inflation. Bürgi et al. (2023) find that oil prices are strong predictors of inflation expectations. Pinem et al. (2023) document that stock index movements, including the STI, have short-run effects on price levels. Retail sales and industrial production are standard macroeconomic indicators Potdar and Kinnerkar (2017) show that industrial production improves inflation forecasts. The SPY Index is included to capture global financial trends, particularly from the U.S. (Liu, 2023). M2 is used as a proxy for monetary conditions; Barkan et al. (2023) demonstrate its relevance in modern forecasting models based on neural networks forecasting.

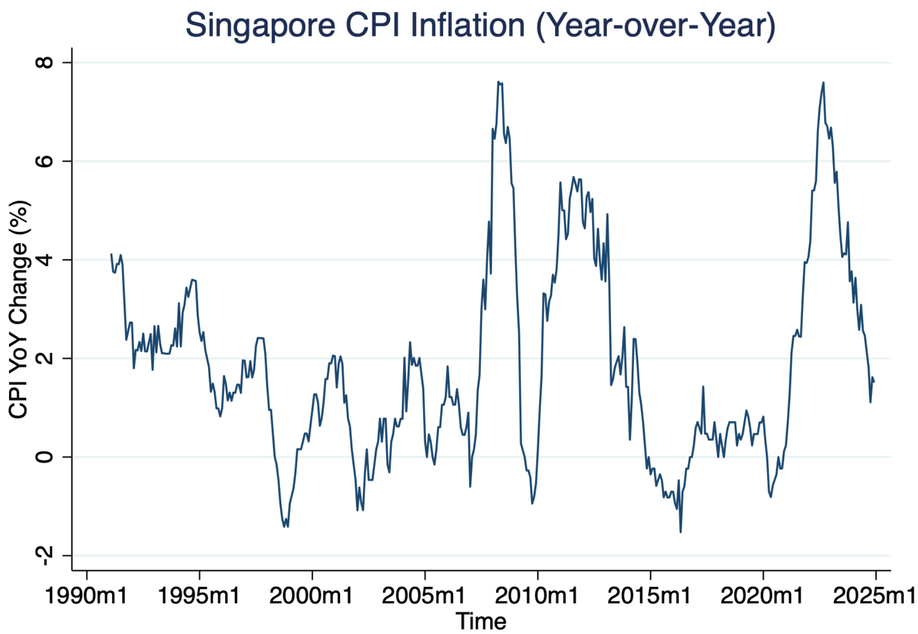

Figure 1: Singapore CPI Inflation (Year-over-Year), 1990–2024

Note: This figure shows the annual percentage change in Singapore's Consumer Price Index. Data are from Statistics Singapore and calculated as the 12-month log-difference of CPI levels.

All variables are nominal and transformed by taking natural logarithms followed by first differences to obtain stationary monthly growth rates. This approach stabilizes variance and approximates percentage changes. For example, the transformation applied to the exchange rate is:

$$\Delta \ln(\text{Exchange Rate})_t = \ln(\text{Exchange Rate}_t) - \ln(\text{Exchange Rate}_{t-1})$$

The same transformation is applied to all predictor variables and the CPI. Each transformed series contains 419 observations after removing the initial missing values due to differencing.

Stationarity was tested using the Augmented Dickey-Fuller (ADF) test on the level of each variable. In all cases, the test failed to reject the presence of a unit root, which justifies the use of first differences. Results are provided in Appendix Table A3.

Table 1 reports summary statistics for the stationary variables used in the LASSO model. All variables are expressed in log-differences and multiplied by 100 to approximate monthly percentage changes. These transformed series serve as the input for the forecasting models.

Table 1: Summary Statistics for Stationary Variables Used in the LASSO Model

Variable	**Mean (%)**	**Std. Dev. (%)**	**Min (%)**	**Max (%)**
$\Delta \ln$(CPI)	0.152	0.453	-1.650	2.059
$\Delta \ln$(Exchange Rate)	-0.080	1.263	-5.518	5.545
$\Delta \ln$(Oil Prices)	0.297	10.022	-55.491	46.911
$\Delta \ln$(STI Index)	0.219	6.026	-27.365	24.856
$\Delta \ln$(Retail Sales)	0.180	11.656	-47.063	34.282
$\Delta \ln$(Industrial Production)	0.481	11.125	-27.967	34.556
$\Delta \ln$(SPY Index)	0.688	4.295	-18.564	11.942
$\Delta \ln$(M2)	0.637	1.280	-6.844	17.145
$\Delta \ln$(Chinese Exports)	1.139	19.219	-121.463	83.127

Note: The data frame starts at January 1990 and ends at December 2024. All variables are log-differenced and scaled by 100 to represent monthly percentage changes. Sources include Statistics Singapore (CPI, Retail Sales, SPY), Yahoo Finance(STI Index), Investing.com (Exchange rate), Trading Economics (M2, Chinese Exports, Industrial Production), and the U.S. Energy Information Administration (Oil Prices).

The dataset includes variables related to domestic activity, global markets, and monetary policy. While some data are obtained from commercial platforms, they are commonly used in applied macroeconomics and offer consistent time-series coverage. Future extensions could explore the use of real-time vintages or regime-switching models to account for structural changes.

4. Results

This section presents the forecast evaluation based on accuracy, directional performance, and statistical significance across the models. Forecast accuracy is evaluated using the Mean Squared Forecast Error (MSFE), and relative performance is measured against a Random Walk with Drift. Statistical significance of differences in predictive accuracy is assessed using the Diebold and Mariano (1995) test with Newey-West standard errors. Directional accuracy is evaluated based on the proportion of months in which the model correctly predicts the sign of inflation. Significance is assessed using the Pesaran and Timmermann (2009) test of directional predictability.

Table 2: Forecast Evaluation: MSFE Ratios, Directional Accuracy, and Statistical Tests

Model	**MSFE Ratio**	**DM (p-value)**	**Directional Accuracy**	**PT (p-value)**
Random Walk with Drift	0.879	0.004	0.611	1.000
ARIMA(1,0,1)	0.875	0.002	0.606	0.418
LASSO	0.834	0.000	0.628	0.019

Note: This table reports one-step-ahead forecast performance for monthly CPI over the evaluation period 2010M1–2024M12. The MSFE ratio is computed by dividing the model's Mean Squared Forecast Error (MSFE) by that of the Random Walk with Drift benchmark. The DM p-value corresponds to the one-sided Diebold-Mariano test (Diebold and Mariano, 1995), which tests the null hypothesis of equal predictive accuracy. Directional Accuracy measures the proportion of months where the model correctly predicts the sign of inflation changes. The PT p-value is from the Pesaran-Timmermann test (Pesaran and Timmermann, 2009), which tests the null that the model has no ability to predict the direction of changes better than chance.

Among the three models, LASSO delivers the lowest MSFE ratio (0.834), indicating the most accurate point forecasts. The Diebold-Mariano test strongly rejects the null of equal forecast accuracy between LASSO and the benchmark, with a p-value of 0.000. ARIMA also outperforms the benchmark, with an MSFE ratio of 0.875 and a p-value of 0.002, indicating statistically significant improvements.

In terms of directional accuracy, LASSO correctly predicts the sign of monthly inflation 62.8% of the time. The corresponding p-value from the Pesaran and Timmermann test is 0.019, providing statistical evidence of directional predictability. By comparison, ARIMA and the Random Walk achieve slightly lower accuracy rates, but neither yields statistically significant results.

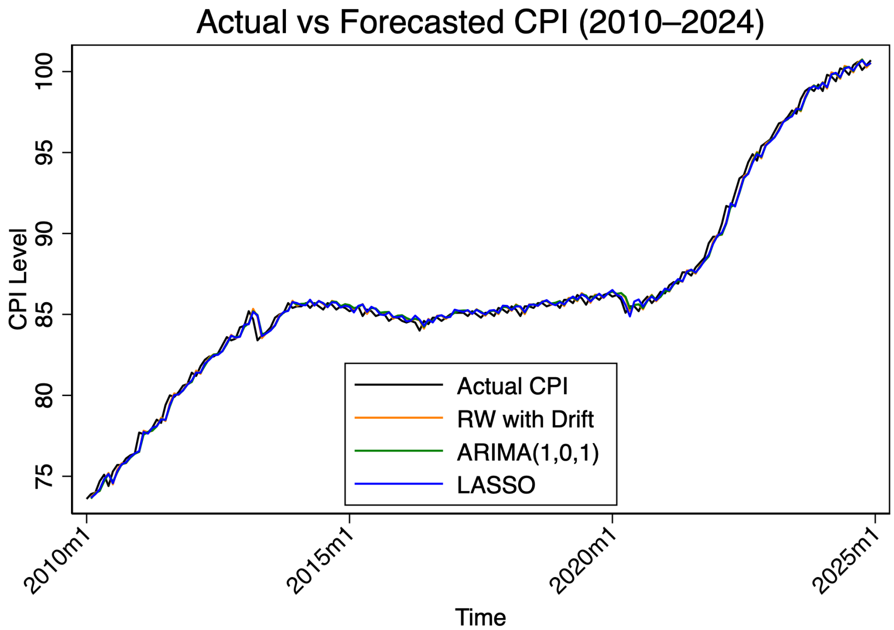

Figure 2: Actual and Forecasted CPI Levels: 2010M1–2024M12

Notes: One-step-ahead forecasts of monthly CPI levels over the evaluation period. Forecasts are computed from models estimated recursively through expanding windows and transformed to levels from log-differences. The actual CPI is shown in black, with model forecasts from the Random Walk with Drift (orange), ARIMA(1,0,1) (green), and LASSO (blue).

As shown in Figure 2, all three models closely track the actual CPI path over the evaluation window, with minimal divergence. Showing the complexity of choosing the best model to forecast CPI.

5. Conclusion

This paper evaluates the performance of LASSO regression for forecasting monthly inflation in Singapore using a set of domestic and international predictors. The results show that LASSO produces lower forecast errors than both benchmark models and performs better in predicting the direction of inflation changes. These gains are observed over a 15-year out-of-sample period, using expanding window estimation. The findings are consistent with prior work, such as Medeiros and Vasconcelos (2016) and Giannone et al. (2021), which show that regularized regression methods can outperform standard time-series models when variable selection is important.

These results are relevant for policymakers, forecasters, and researchers working with large predictor sets and limited sample sizes. The LASSO framework provides a disciplined way to reduce dimensionality without requiring subjective model selection. However, this study uses revised data and assumes a stable relationship between predictors and inflation. Future research could explore whether the results hold in real-time settings, examine potential structural breaks, or extend the approach to disaggregated inflation components.

References

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., and Koenigstein, N. (2023). Forecasting cpi inflation components with hierarchical recurrent neural networks. *International Journal of Forecasting*, 39(3):1145–1162.

Bürgi, C., Srivastava, P., and Whelan, K. (2023). Oil prices and inflation forecasts. *Econstor*.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3):253–63.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Liu, Y. (2023). Time series modeling and forecasting of us consumer price index using wti crude oil price. *ResearchGate*.

Medeiros, M. C. and Vasconcelos, G. F. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138:50–52.

OEC World (2024). Singapore - country profile - oec. `https://oec.world/en/profile/country/sgp`.

Parrado, E. (2024). Time to rethink exchange rate orthodoxy for open economies. *Financial Times*.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Pinem, D., Ariani, M., and Desmintari (2023). Analysis of global stock index, inflation and interest rates on the indonesia stock exchange joint stock price index. *International Journal of Research in Business and Social Science*, 12(3):308–317.

Potdar, K. and Kinnerkar, R. (2017). A non-linear autoregressive neural network model for forecasting indian index of industrial production. In *2017 IEEE Region 10 Symposium (TENSYMP)*, pages 1–5.

Appendix

Table A3: ADF Test Results for Variables in Levels

Variable	ADF Stat.	1% Crit.	5% Crit.	10% Crit.	p-value
Consumer Price Index (CPI)	0.384	-3.447	-2.874	-2.570	0.9809
Exchange Rate (USD/SGD)	-1.632	-3.447	-2.874	-2.570	0.4663
Oil Prices (Monthly Avg)	-1.782	-3.447	-2.874	-2.570	0.3896
STI Index	-1.904	-3.447	-2.874	-2.570	0.3304
Retail Sales Index	-1.441	-3.447	-2.874	-2.570	0.5625
Industrial Production Index	0.676	-3.447	-2.874	-2.570	0.9893
SPY Index	2.271	-3.447	-2.874	-2.570	0.9989
M2 Money Supply	2.249	-3.447	-2.874	-2.570	0.9989
Chinese Exports	1.269	-3.447	-2.874	-2.570	0.9964

Notes: This table shows Augmented Dickey-Fuller (ADF) test results for the level values of all series used in the LASSO model. ADF tests include 12 lags and test for a unit root without drift. The null hypothesis is that the series has a unit root. Critical values are from MacKinnon (1996). All p-values are approximate. Full transformations and results for first-differences are used in the forecasting exercise.