Forecasting U.S. Inflation – Managing Seasonal Patterns

Alicja Kalwat V13043, Jan Muchorowski V13034

December 8, 2025

Abstract

In this paper we try to find the best possible way to forecast inflation in the United States using only econometric methods - namely SARIMA models. We use monthly data from 1978 to 2025 divided into two parts: the first 80% of observations are in-sample, used to calibrate model and find the right parameters, and the rest 20% is used to check forecasts. We use rolling-window approach with two SARIMA models given by BIC and AIC. They both perform significantly better than the benchmarks - Random Walk and Seasonal Random Walk, reducing the Root Mean Squared Forecast Error by 21% compared to Random Walk.

1 Introduction

Inflation is an economic phenomenon that affects all of us. The speed of the overall increase of prices is connected to the costs of living in a particular country, and that is why it is also so interesting to forecast. As we may expect, it is not an easy thing to do. One of the main factors that causes such difficulties is seasonality. During the year we can observe months with higher costs for different kinds of goods, and drops appearing after those spikes. Usually economists are using seasonally adjusted data to not have to worry about those issues. In this paper we want to use raw data coming from the Federal Reserve Bank of St. Louis and try to forecast the U.S. inflation using SARIMA models, which includes also managing those complex, seasonal patterns.

2 Data

The data we will be analyzing comes from https://fred.stlouisfed.org and describes the Consumer Price Index for All Urban Consumers (All Items in U.S. City Average). The data has monthly frequency, with units of percent change from year ago, and is not seasonally adjusted, like previously said. The time series starts in January 1978 and ends in September 2025, which gives us 573 observations. We will use the last 20% of the data as the pseudo out-of-sample to evaluate forecasts, and we will work on the rest, the data from January 1978 to March 2016 as the in-sample. Below, on the Figure 1 we can see the plot of our time series before any transformations, but already divided into two parts.

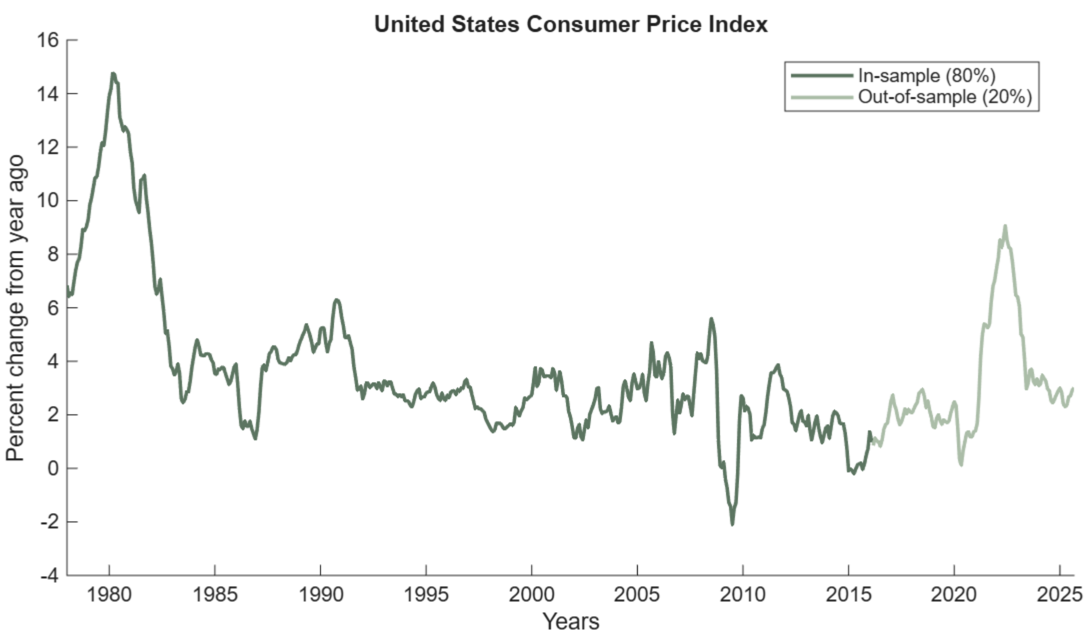

Figure 1: Time series

2.1 Stationarity

Let us look closer in our in-sample data. In Figure 1 showing the inflation values over the years we can see that the time series is definitely non-stationary. It drifts away from the mean for long periods of time, which is a characteristic of a random-walk process. In this case let us check for the presence of the unit root - to do that we will use the Augmented Dickey-Fuller test. Since we see no clear trend and the mean is different from 0, we are choosing the variant with no trend, but with the intercept. The test gives us the p-value = 0.398. In this case we definitely cannot reject the null hypothesis, which tells us that the series indeed has a unit root.

It is important to note that, theoretically speaking, if there is a functioning Central Bank, inflation should be a mean-reverting process, moving towards the inflation target in the medium to long term. It would obviously imply stationarity. However, in case of this exercise we want to forecast inflation, and as such we are not really interested in the phenomenon in the economic sense. For the purpose of this project, the inflation is seen as just a stochastic process.

Another thing we can check to further support the presence of a unit root is the ACF plot (Figure 2). It shows a slow decay of the lagged correlations, which is consistent with a unit root behaviour.

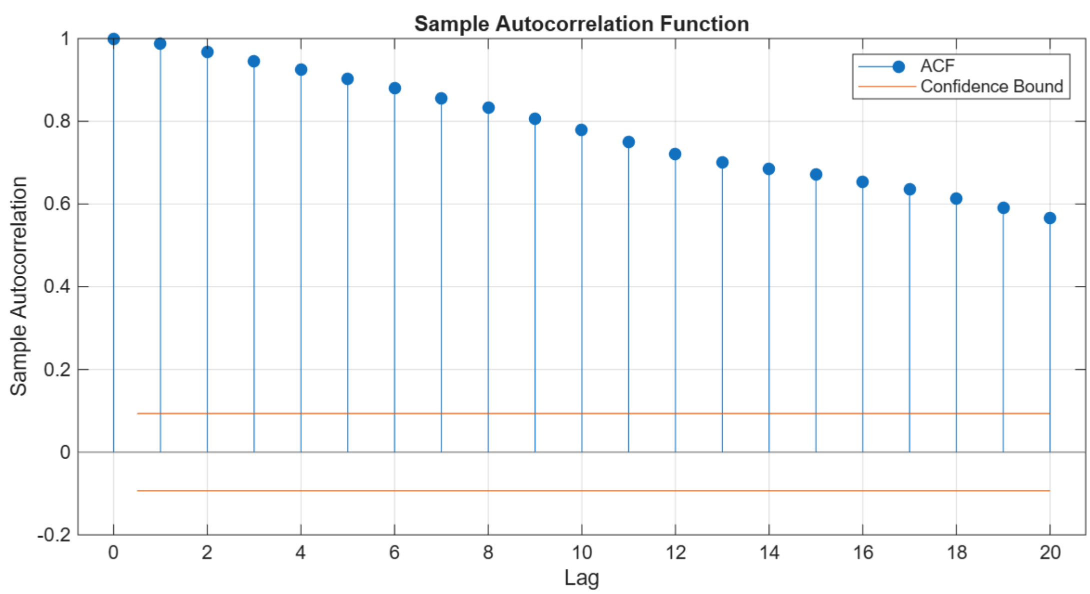

Figure 2: ACF plot

2.2 First-order differencing

In order to make the series stationary we transform the series by taking the first-order differences:

$$\Delta(y_t) = y_t - y_{t-1}.$$

Now, by looking at the plot of $\Delta(y_t)$ (Figure 3) we could assume the series is stationary - the persistent deviations from the mean were effectively removed. Let us make sure about that by doing the Augmented Dickey-Fuller test once again. Now the p-value is equal to 0.001 - we don't have a unit root anymore.

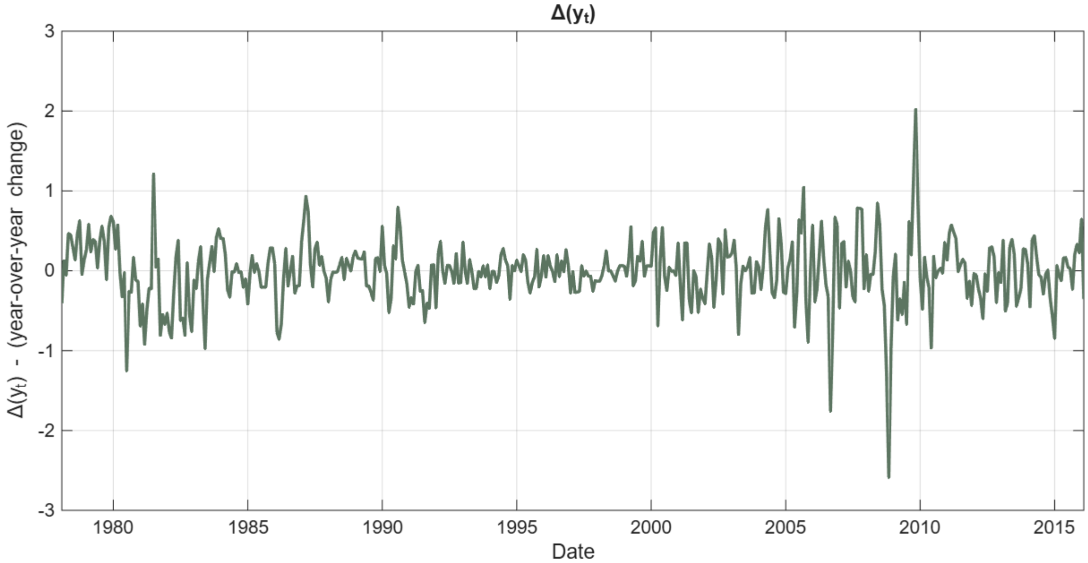

Figure 3: First-order difference of inflation

3 Methods and Models

Since we know that $\Delta(y_t)$ is stationary, we can look at this variable's autocorrelation and partial autocorrelation plot in order to observe patterns that help us identify an appropriate SARIMA model.

For example, there is a big spike at the first lag in ACF, and then a rapid drop - that could suggest us to choose a non-seasonal MA(1). There are also significant spikes at the first two lags in PACF - that means a non-seasonal AR(2) would also be a good option.

When it comes to the seasonal patterns, we can observe a strong spike at lag 12 in ACF and slow decay at seasonal lags in PACF - that is a typical seasonal MA(1) behaviour. On the other hand, there is also a slightly significant spike at lag 24 in ACF - it could also suggest a possible seasonal MA(2).

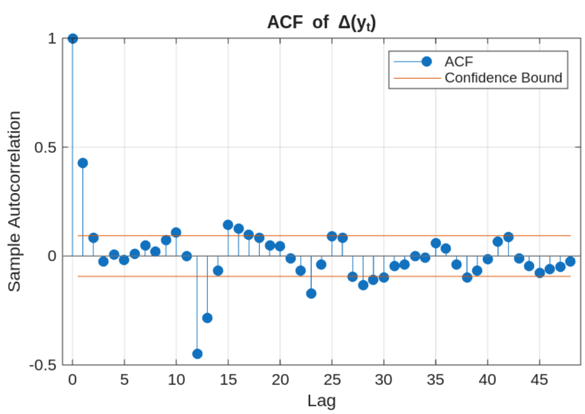

Figure 4: ACF plot

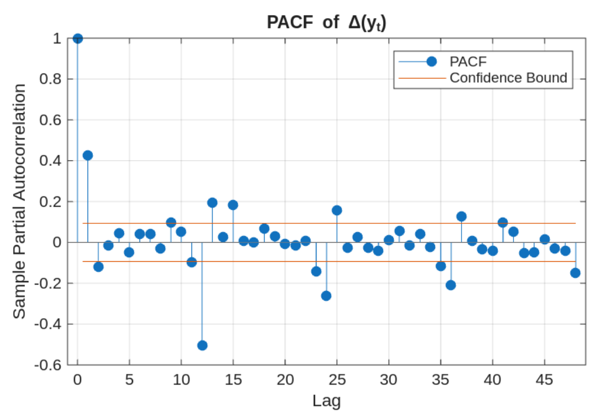

Figure 5: PACF plot

3.1 Benchmark models

Before estimating the complex SARIMA models, let us first create the benchmark, which will serve as a reference point for evaluating the results. Since we know that y_t, the inflation, is a time series with a unit root, our benchmark could be simply a Random Walk, or even a Seasonal Random Walk - since our data has the seasonal component. These simple benchmarks make sense from an economic point of view, and these models will be the ones that we aim to outperform with our SARIMA models.

- **Random Walk:**

$$(1-L)y_t = u_t \quad \Longrightarrow \quad y_t = y_{t-1} + u_t$$

- **Seasonal Random Walk:**

$$(1-L^{12})y_t = u_t \quad \Longrightarrow \quad y_t = y_{t-12} + u_t$$

3.2 Model selection

For our model we chose SARIMA models, to which we had to fit 4 parameters.
The basic SARIMA$(p,d,q)(P,D,Q)_s$ model can be described as:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-\sum_{j=1}^{P}\Phi_j L^{sj})(1-L)^d(1-L^s)^D y_t = \mu + (1+\sum_{k=1}^{q}\theta_k L^k)(1+\sum_{j=1}^{Q}\Theta_j L^{sj})\varepsilon_t \tag{1}$$

Our choice for d and D can be made using a unit root test, which we have already performed in section 2. Because the Augmented Dickey-Fuller test did show a unit root in the first difference in y_t and no unit root for Δy_t, we can choose $d = 1$ and $D = 0$. The remaining parameters are p, q, P, and Q. We can try to select the appropriate model by examining the ACF and PACF like we did in the beginning of the section, however, it is still only a basic analysis, and might not be enough for such complex data.

To choose the best fitting model, we decided to use information criteria. The most popular ones are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). In order to see their performance, we decided to test for the right parameters using both AIC and BIC, and compare resulting SARIMA models for their accuracy. In the tables below we can see 3 best models selected by two information criteria.

p	d	q	P	D	Q	BIC
1	0	2	2	0	1	263.15
2	0	3	2	0	1	268.00
2	0	2	2	0	1	268.66

Table 1: Top 3 SARIMA models sorted by BIC

p	d	q	P	D	Q	AIC
2	0	3	2	0	1	230.86
2	0	3	2	0	2	231.33
3	0	3	2	0	1	232.78

Table 2: Top 3 SARIMA models sorted by AIC

As we can see, the information criteria have given us somewhat different results. From the tables above we can observe that the Bayesian Information Criterion selected the model with 7 parameters (including the intercept). The other choices have more parameters, and slightly higher BIC value. What is interesting is the fact that the second-best model according to BIC is the top choice according to AIC. It was also expected for AIC to provide more parameters; the penalty for the number of parameters in BIC is larger than in AIC. It it also worth mentioning that there are clear similarities with the ACF and PACF plots of $\Delta(y_t)$: all selected models include seasonal components as well as non-seasonal autoregressive and moving average ones.

As a final result of conducting the two tests, two models have been chosen. The resulting models are:

- Best by BIC: SARIMA$(1, 0, 2)(2, 0, 1)_{12}$
- Best by AIC: SARIMA$(2, 0, 3)(2, 0, 1)_{12}$

We decided to test our data using both. The resulting models have been used first to estimate and then to forecast the data. It is important to note, that this result was achieved using differentiated data - as such, for the later use we will set $d = 1$, the rest of parameters unchanged.

3.3 Residual diagnostics

Before using our models for forecasting, we can check the properties of in-sample residuals. We can see that, unlike in the ACF plot of $\Delta(y_t)$, the previously strong significant autocorrelations have mostly disappeared (Figures 6 and 7). However, we still have some minor, but significant lagged correlations left. It means that there is still some behaviour that could be explained by the models, but ours are already complex, and we don't want to add even more parameters, also trying to avoid overfitting.

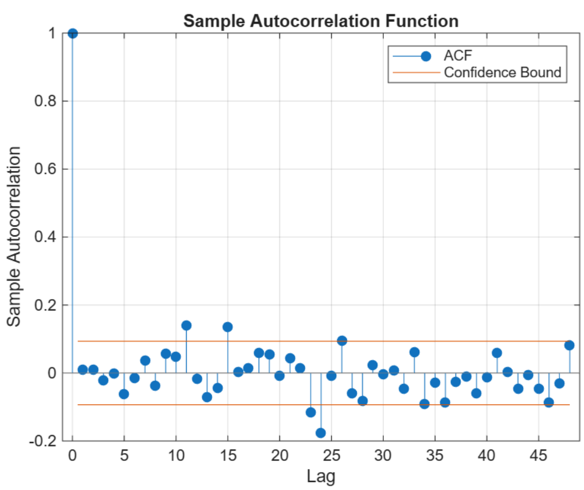

Figure 6: Model selected by BIC

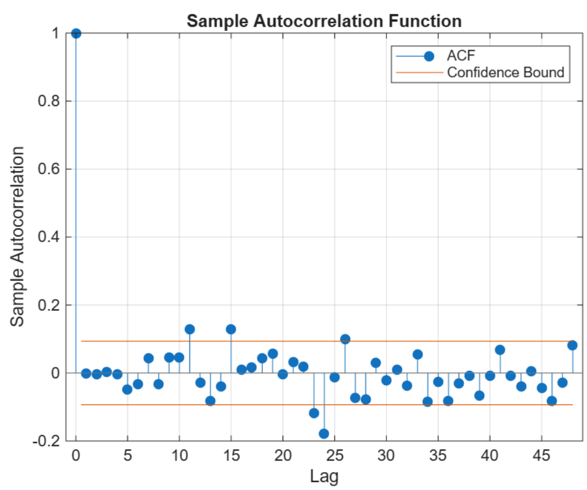

Figure 7: Model selected by AIC

4 Forecasting

SARIMA models create worse forecasts the more advanced into future the forecast is. Therefore, it is best to reduce the steps of forecast as much as possible. To do so, we have chosen a minimal number of steps - 1 - as our basis. We have also decided to adopt a rolling window approach - we can then discard older observation in favour of more recent (and therefore more important to our forecast) ones. Therefore, we estimated the model each time and created forecasts one step ahead.

4.1 Model estimation and calculating forecasts

We used four forecasts (two SARIMA models - SARIMA$(1,1,2)(2,0,1)_{12}$ and SARIMA$(2,1,3)(2,0,1)_{12}$, Seasonal Random Walk and Random Walk) and in each step had to forecast both SARIMA models - the SRW and RW do not need estimation. To do this, we simply used built-in functions of MATLAB.

Our estimated SARIMA models for the variable y_t representing inflation, take the following forms:

- SARIMA$(1,1,2)(2,0,1)_{12}$

$$(1-\phi L)(1-\sum_{j=1}^{2}\Phi_j L^{12j})(1-L)y_t = \mu + (1+\sum_{k=1}^{2}\theta_k L^k)(1+\Theta L^{12})\varepsilon_t \tag{2}$$

- SARIMA$(2,1,3)(2,0,1)_{12}$

$$(1-\sum_{i=1}^{2}\phi_i L^i)(1-\sum_{j=1}^{2}\Phi_j L^{12j})(1-L)y_t = \mu + (1+\sum_{k=1}^{3}\theta_k L^k)(1+\Theta L^{12})\varepsilon_t \tag{3}$$

After estimating for each step, we do 4 forecasts. The Random Walk and Seasonal Random Walk forecasts are simple - we just use the previous observations, while for SARIMA models we use estimated data to get a forecast. Results can be seen in the Figure 8 below.

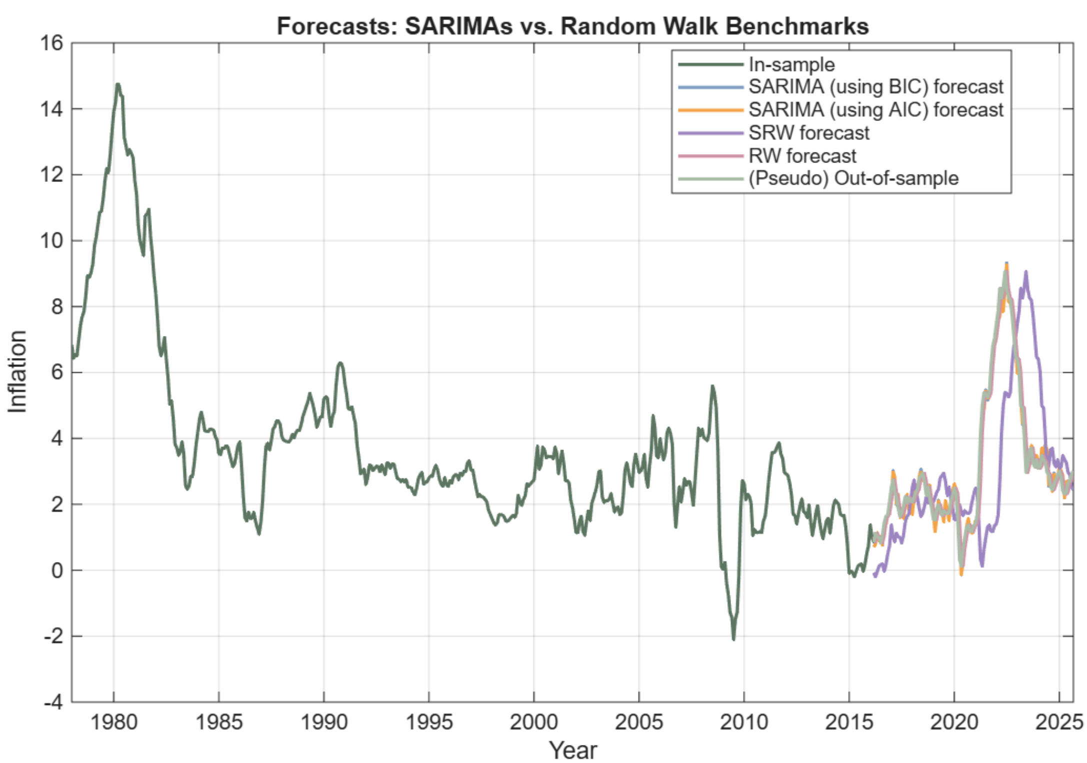

Figure 8: Forecasts compared with out-of-sample observations

4.2 Comparing predictive accuracy

Our first and foremost way of comparing accuracy of forecasts is calculating the root mean square forecast error using our forecasts and out-of-sample observations:

$$RMSFE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_{T+t} - \hat{y}_{T+t})^2} \tag{4}$$

The smaller the root mean square error is, the better our forecast is. The next table shows the result of our calculations.

Model	RMSFE
$\mathrm{SARIMA}(1,1,2)(2,0,1)_{12}$	0.30827
$\mathrm{SARIMA}(2,1,3)(2,0,1)_{12}$	0.30635
Seasonal Random Walk	2.4123
Random Walk	0.39368

Table 3: Root mean square forecast errors for our 4 forecasts

The model using $\mathrm{SARIMA}(2,1,3)(2,0,1)_{12}$ (which we got from AIC) has been the best, reducing RMSFE by 21% compared to our main benchmark - Random Walk. Seasonal Random Walk performs really badly - which should not come as a surprise, looking at the data. SRW is only good, when the seasonal part is dominating and there are no sudden changes. The inflation rise after COVID hit SRW with 12-months lag, increasing substantially its RMSFE.

The difference between our models and RW is not that large, and between SARIMA models themselves is quite small. We therefore wanted to test if the reduction in RMSFE is statistically significant. To test it we used Diebold-Mariano test, the results of which can be seen in the table below.

Models compared	P-value
$\mathrm{SARIMA}(1,1,2)(2,0,1)_{12}$ vs RW	0.022889
$\mathrm{SARIMA}(2,1,3)(2,0,1)_{12}$ vs RW	0.020301
$\mathrm{SARIMA}(2,1,3)(2,0,1)_{12}$ vs $\mathrm{SARIMA}(1,1,2)(2,0,1)_{12}$	0.37947

Table 4: P-value of Diebold-Mariano test

As we can see, we can reject the hypothesis that RW and either of SARIMA models have the same predictive power (because p-value < 0.05) - our models are significantly better. However, the two models themselves appear to be of the same class - we cannot reject the null hypothesis that their predictive power is the same.

In the end we can summarize our work by saying that our two SARIMA models were able to accurately predict both periods of high volatility and stability, and thanks to this, they were able to outperform the benchmarks.