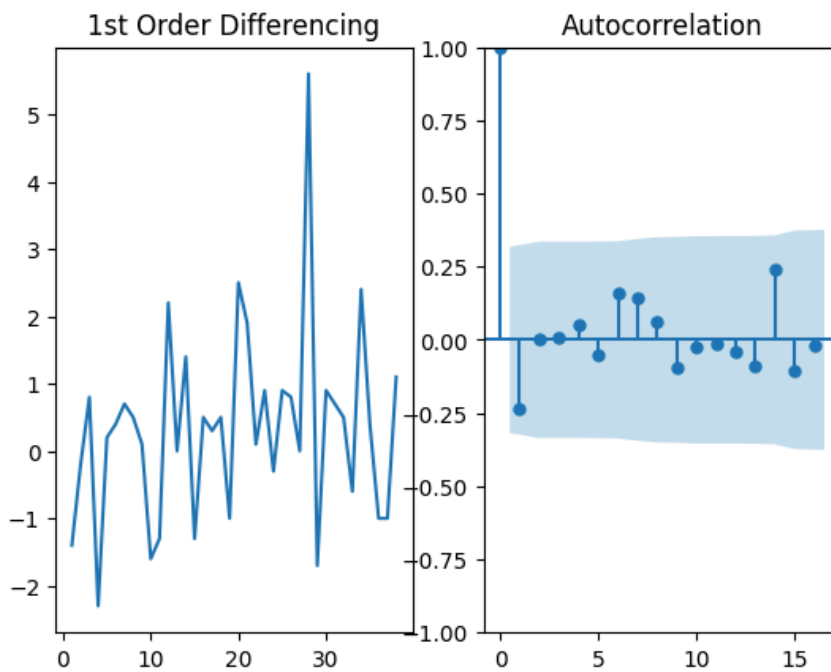
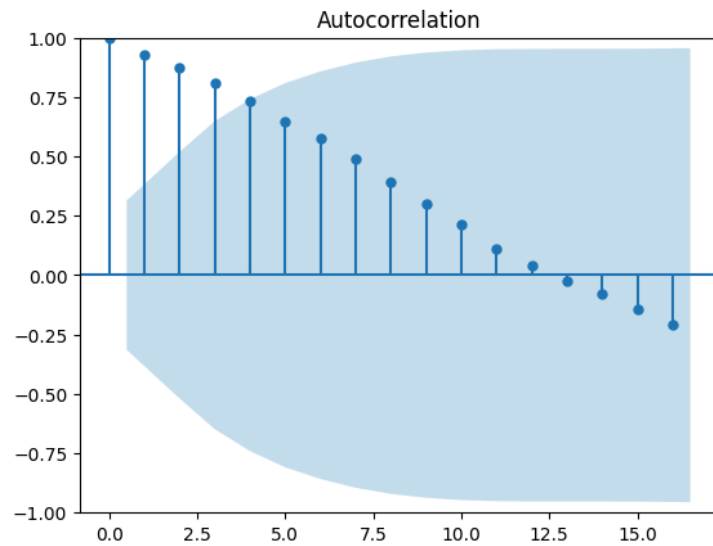
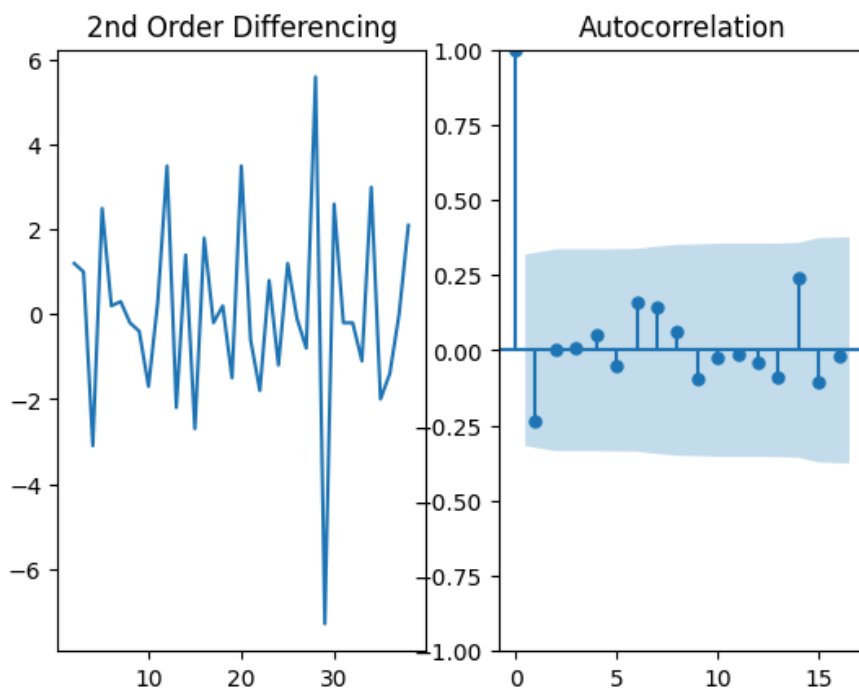


Choosing best ARIMA parameters





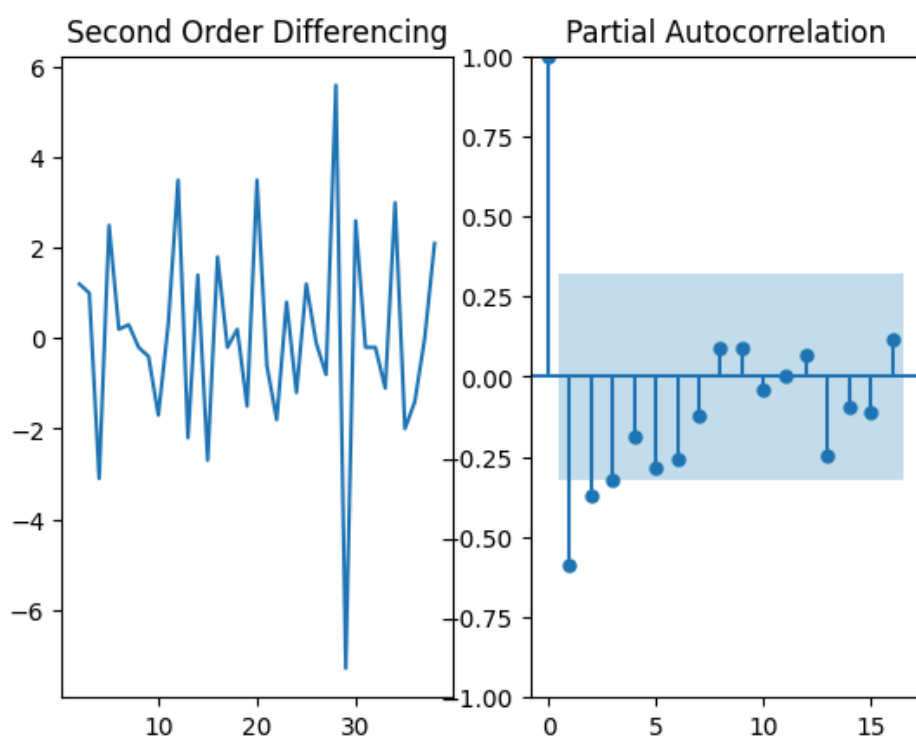
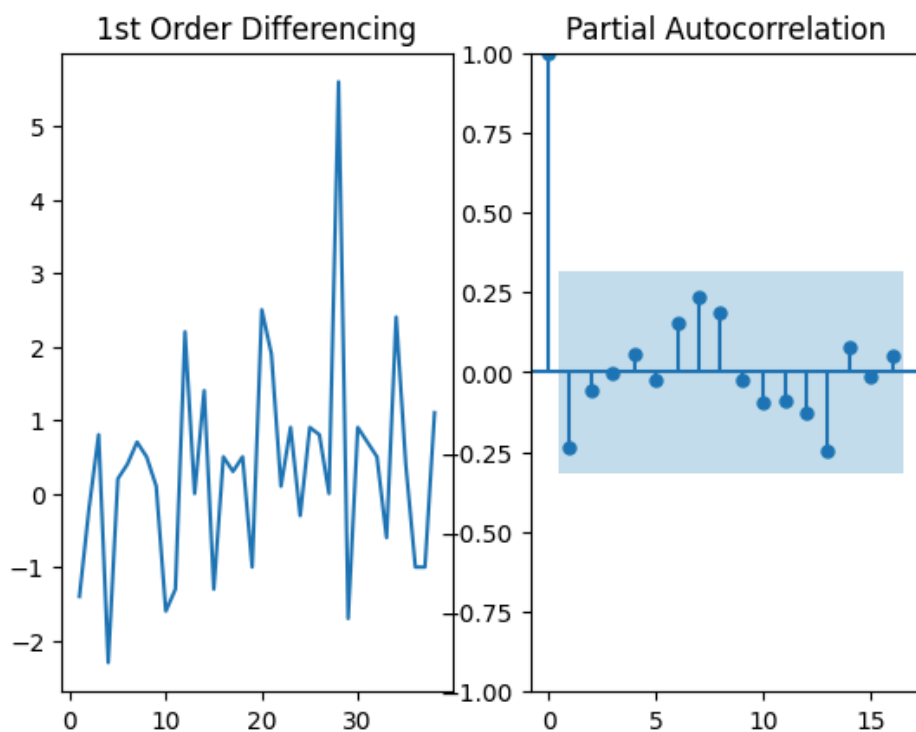
Testing if our data is stationary or not stationary

Augmented Dickey-Fuller Test, with first and second differencing included

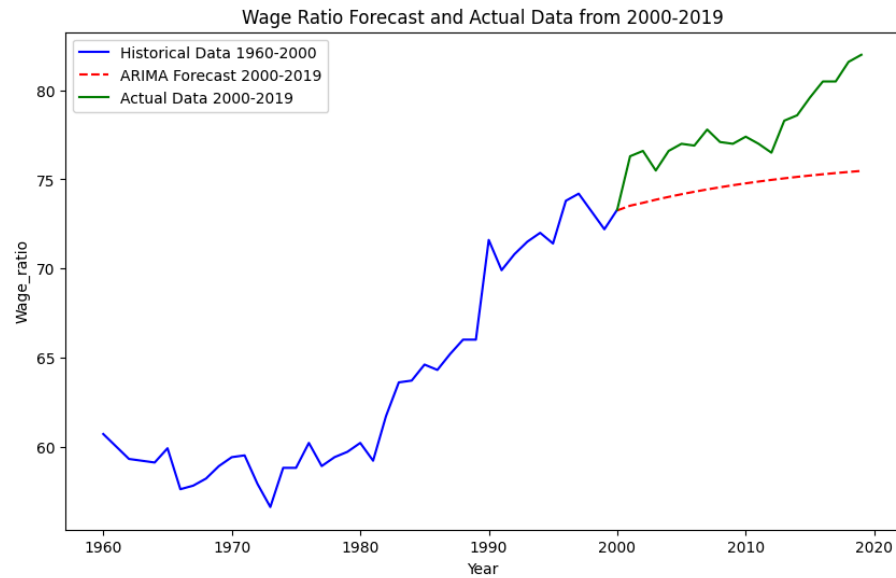
```
p value: 0.029983770167539794
p value: 1.6134859992644947e-11
p value: 0.00010345091848076797
```

KPSS Test

```
(np.float64(0.207634429618148),
 np.float64(0.013137088893194504),
 4,
 {'10%': 0.119, '5%': 0.146, '2.5%': 0.176, '1%': 0.216})
```



Now actually using the ARIMA model



Summary of model fit

SARIMAX Results

Dep. Variable:	Wage_ratio	No. Observations:	39
Model:	ARIMA(2, 1, 1)	Log Likelihood	-66.589
Date:	Tue, 22 Oct 2024	AIC	141.179
Time:	22:08:22	BIC	147.729
Sample:	0	HQIC	143.509
	- 39		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.6482	0.587	1.103	0.270	-0.503	1.800
ar.L2	0.2686	0.233	1.155	0.248	-0.187	0.725
ma.L1	-0.8464	0.591	-1.433	0.152	-2.004	0.312
sigma2	1.9354	0.411	4.709	0.000	1.130	2.741

Ljung-Box (L1) (Q):	0.18	Jarque-Bera (JB):	19.30
Prob(Q):	0.67	Prob(JB):	0.00
Heteroskedasticity (H):	2.04	Skew:	1.11
Prob(H) (two-sided):	0.21	Kurtosis:	5.70

Calculating MSE, RMSE, and create new dataframe with predicted and actual wage ratios, as well as the corresponding residuals

MSE = 12.6607

RMSE = 3.5582

Predicted, Actual Wage Ratio, and Residuals for 2000-2019:

	Year	Predicted_Wage_Ratio	Actual_Wage_Ratio	Residuals
0	2000	73.256089	73.3	0.043911
1	2001	73.523076	76.3	2.776924
2	2002	73.684339	76.6	2.915661
3	2003	73.860579	75.5	1.639421
4	2004	74.018129	76.6	2.581871
5	2005	74.167588	77.0	2.832412
6	2006	74.306782	76.9	2.593218
7	2007	74.437149	77.8	3.362851
8	2008	74.559038	77.1	2.540962
9	2009	74.673060	77.0	2.326940
10	2010	74.779707	77.4	2.620293
11	2011	74.879459	77.0	2.120541
12	2012	74.972761	76.5	1.527239
13	2013	75.060031	78.3	3.239969
14	2014	75.141659	78.6	3.458341
15	2015	75.218009	79.6	4.381991
16	2016	75.289422	80.5	5.210578
17	2017	75.356219	80.5	5.143781
18	2018	75.418696	81.6	6.181304
19	2019	75.477135	82.0	6.522865

Line plot of the residuals of the predicted wage ratios from the ARIMA model for 2001-2019

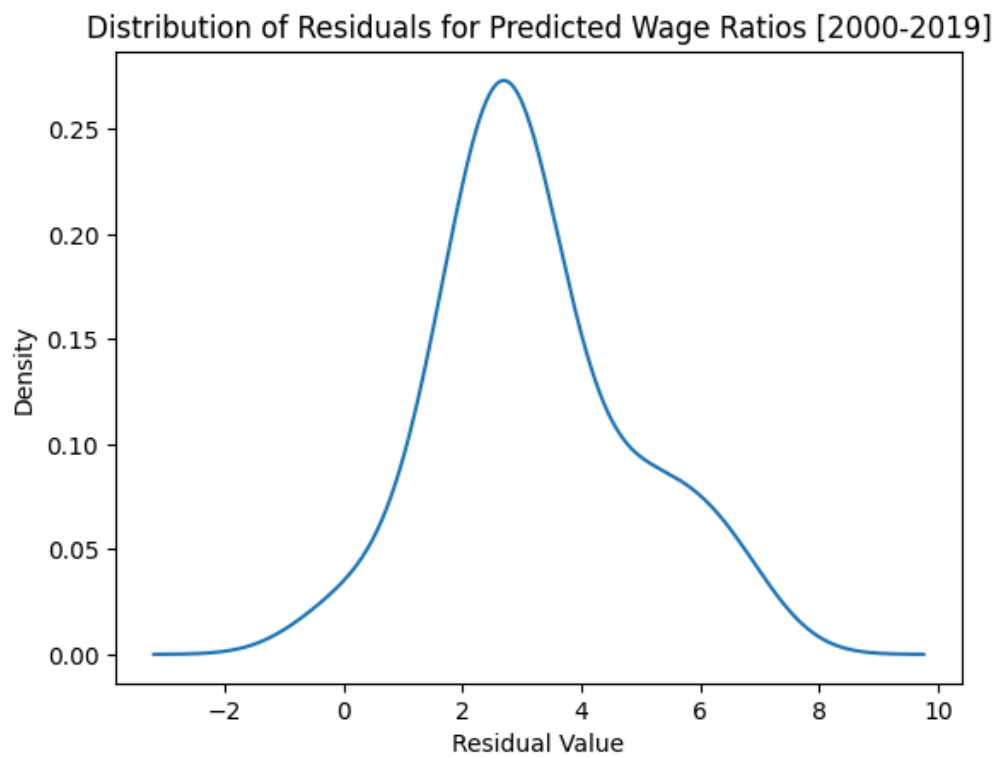


Summary statistics of the residuals

count	20.000000
mean	3.201054
std	1.594064
min	0.043911
25%	2.487456
50%	2.804668
75%	3.689254
max	6.522865

Name: Residuals, dtype: float64

Kernel density plot for residuals



Plotting the actual and predicted wage ratios for 2001-2019

