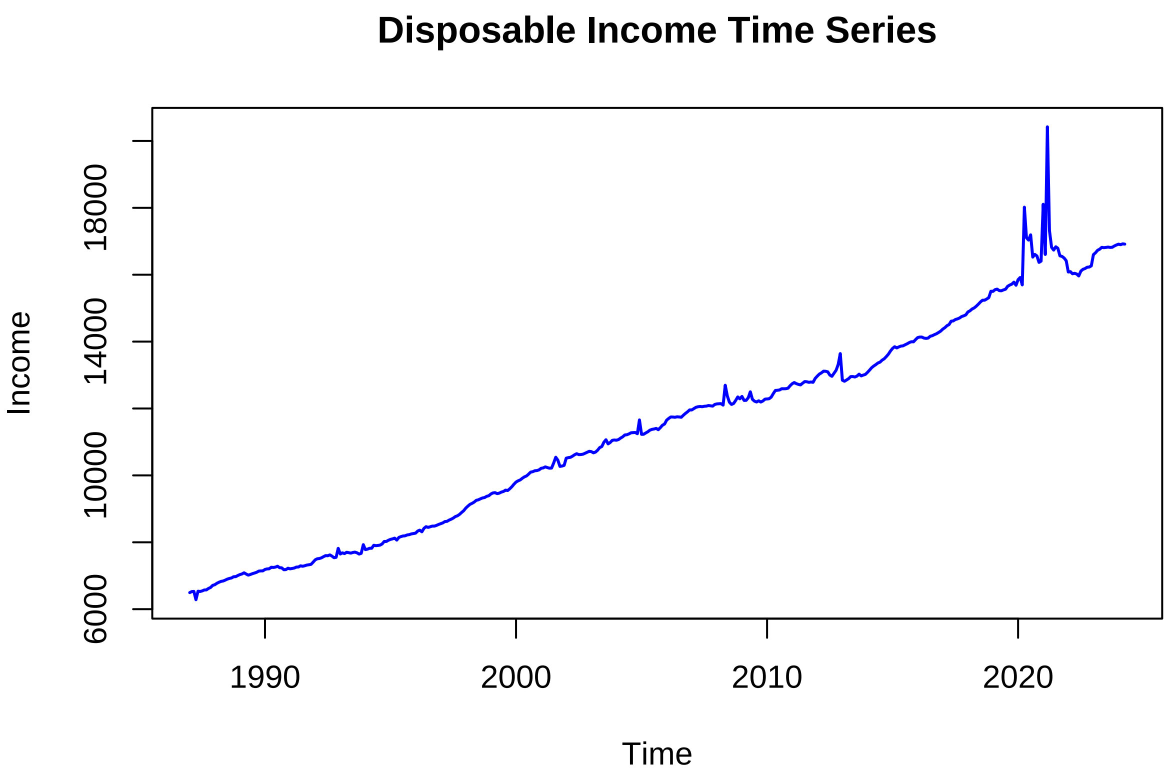
Project Analysis

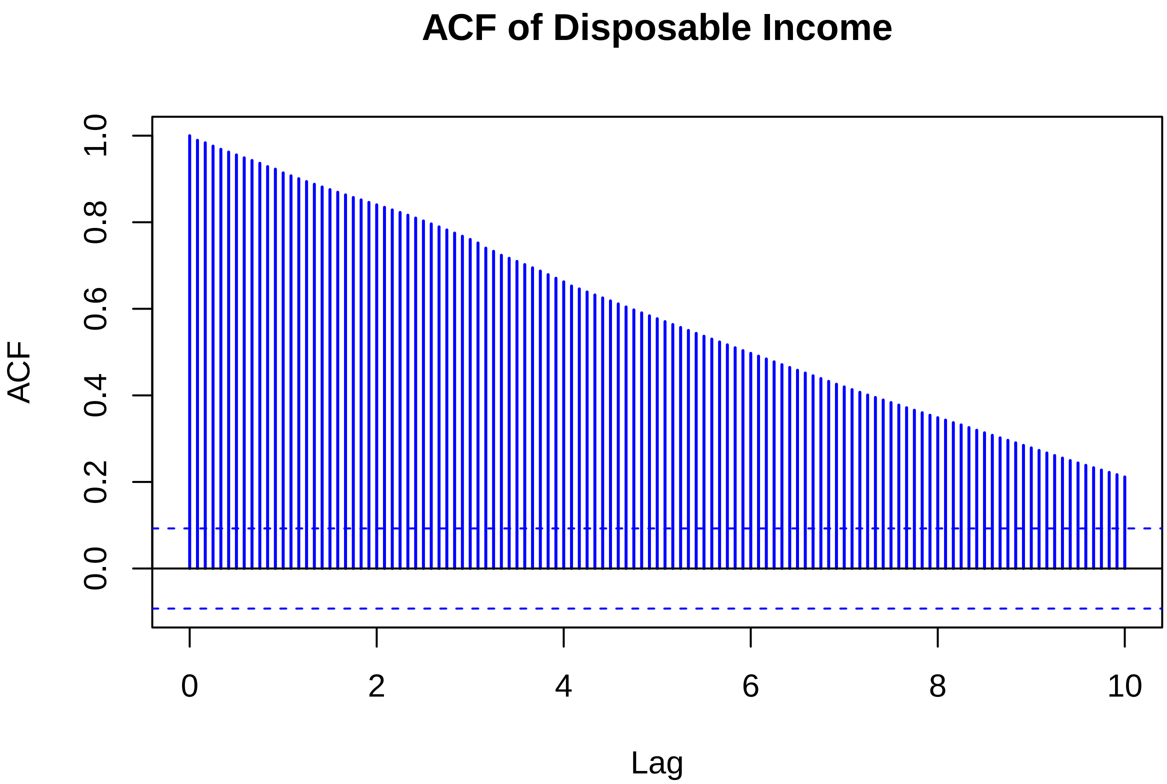
2025-03-14

Exploratory Analysis

Are there specific characteristics of the time series representing disposable income that contribute most to the predictability of the time series?

To begin our analysis, we will examine the disposable personal income time series data as well as its ACF plot.





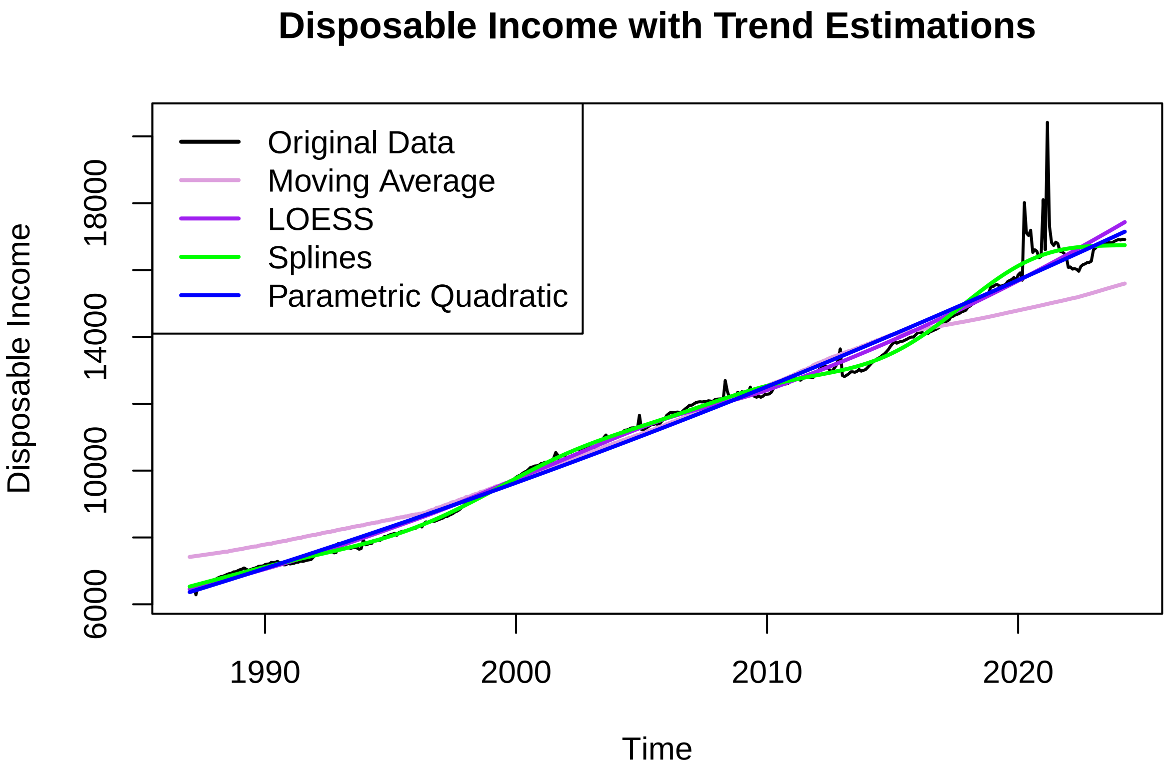
We can see that there is a clear upward trend indicating the data is non-stationary. More specifically, the mean is changing over time, which violates one of the assumptions of stationarity. There a sharp spikes sen in the series indicating a potential external shock, which in this case aligns with what we know to be the COVID-19 pandemic.

Looking at the ACF plot, we see that the ACF plot do not sharply drop to zero, as we would expect in a stationary series. This tells us that the past income values strongly influence future income values, which intuitively lines up with what we would expect.

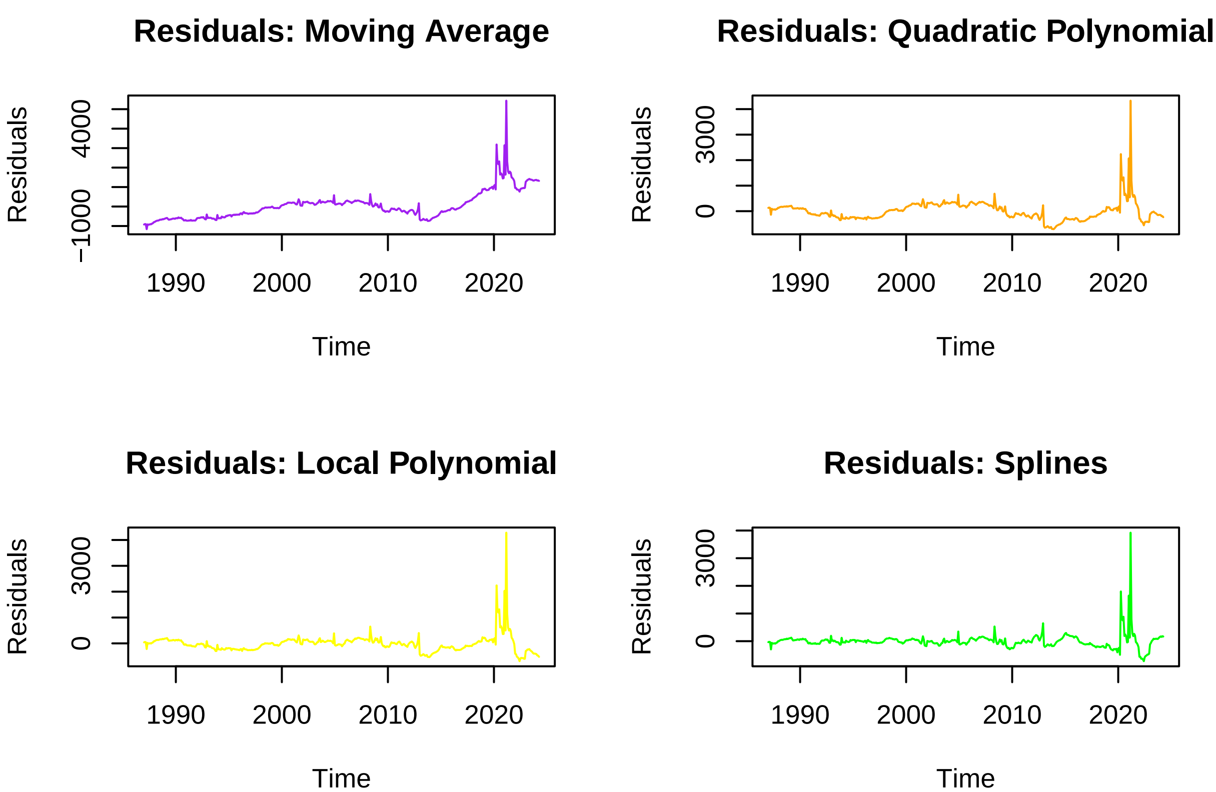
We can confirm non stationarity using a formal Augmented Dickey-Fuller (ADF) test below:

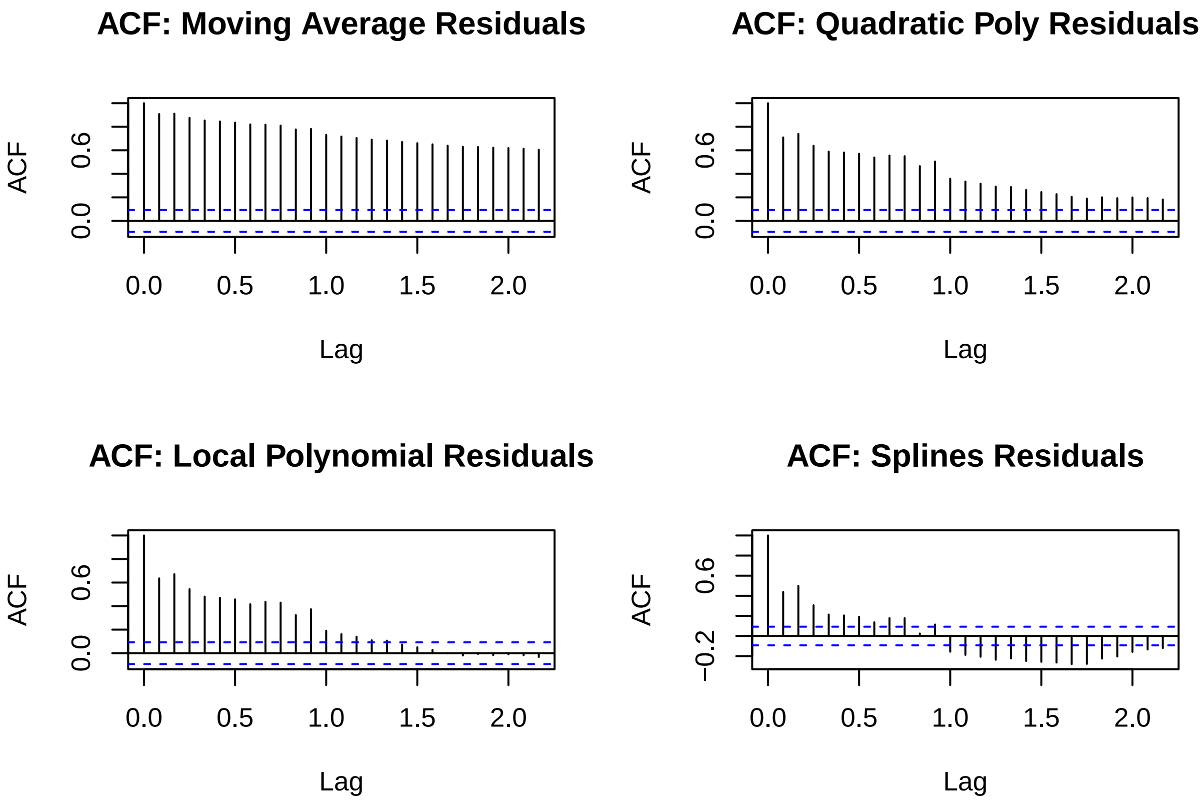
```
##
##  Augmented Dickey-Fuller Test
##
## data:  income_ts
## Dickey-Fuller = -2.6625, Lag order = 7, p-value = 0.2977
## alternative hypothesis: stationary
```

As expected, we see a p-value well above 0.05, indicating the series is non-stationary and will need differencing if it is to be fit to an ARIMA model. Before this, we will evaluate the fit of Moving Average, LOESS, Splines, and Quadratic Polynomial models.



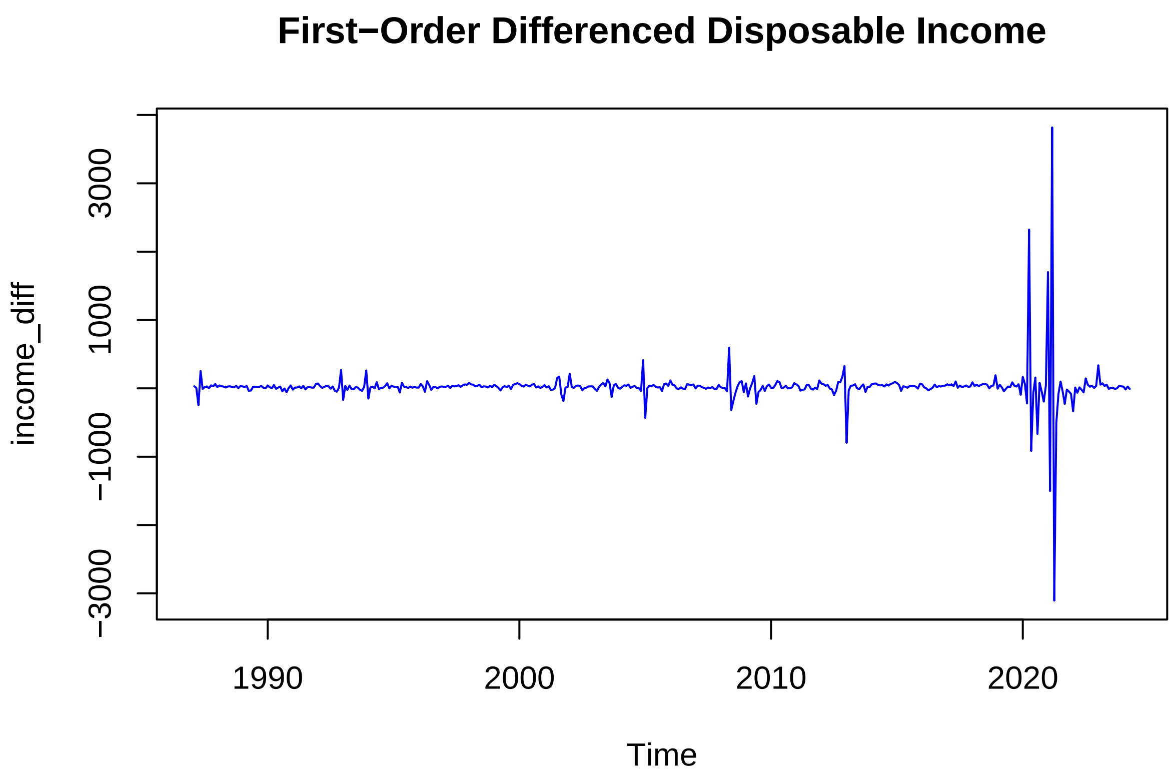
The models generally appear to fit the data well, but fails to capture sudden shocks. In this case, the SPLINES model appears to follow the data best. Next, we will examine the residuals of the models.

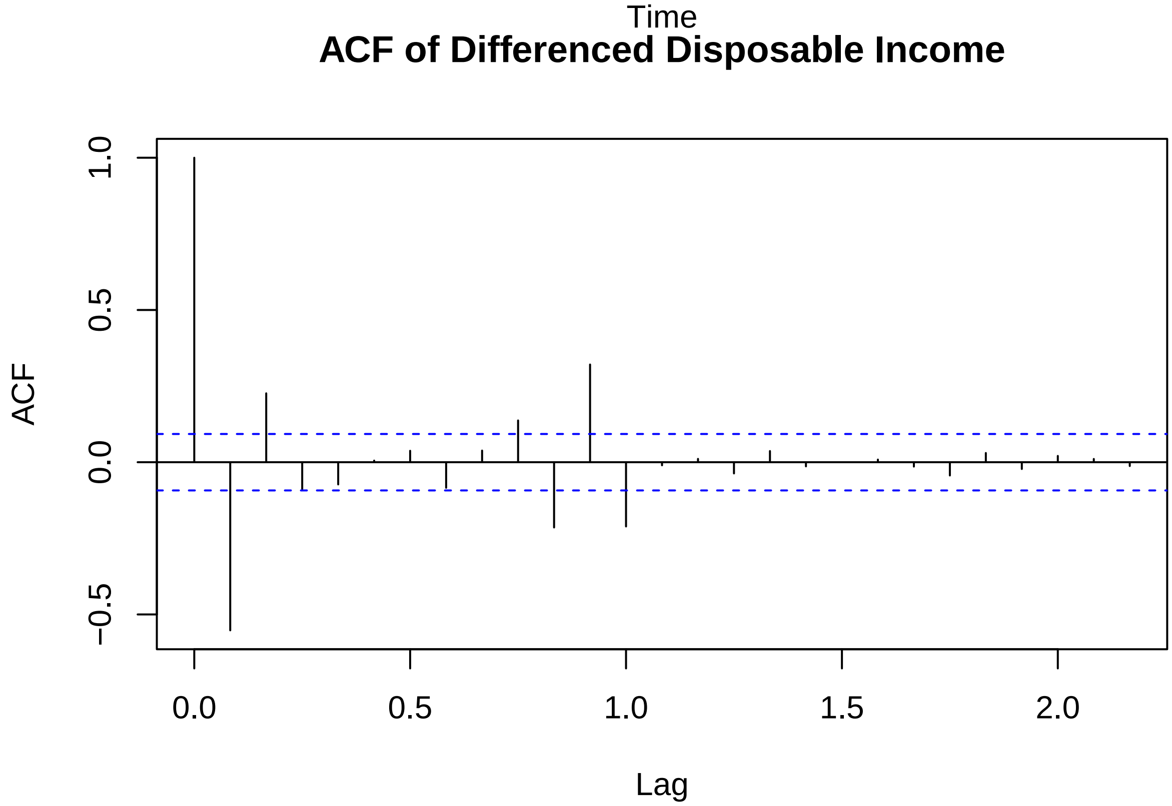




All four residual plots are showing large spikes around 2020, which suggests that none of the trend models were able to fully capture the shock effect. This shock is likely due to the COVID-19 pandemic. Out of the different models selected the Local Polynomial and Splines models appear to be the most stable, but in order to better explain the variations in income, we will likely need a different model such as ARIMA or SARIMA time series models.

In order to remove trend, we will perform first-order differncing.



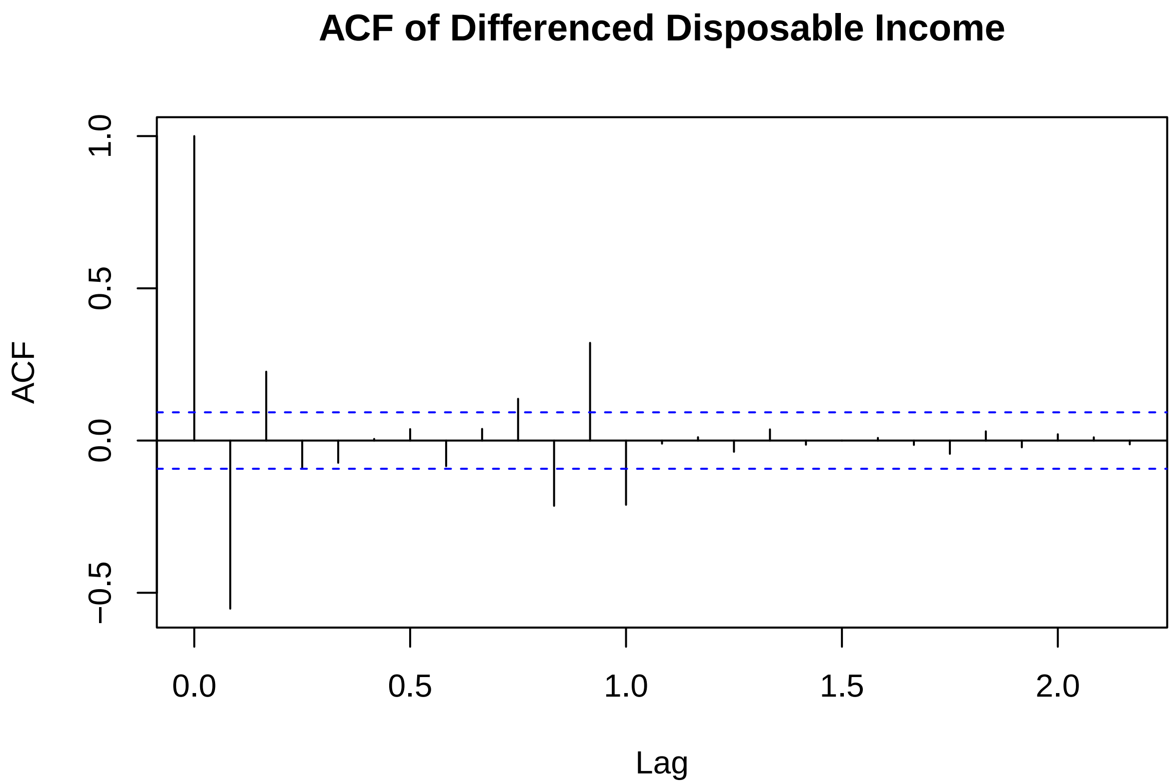


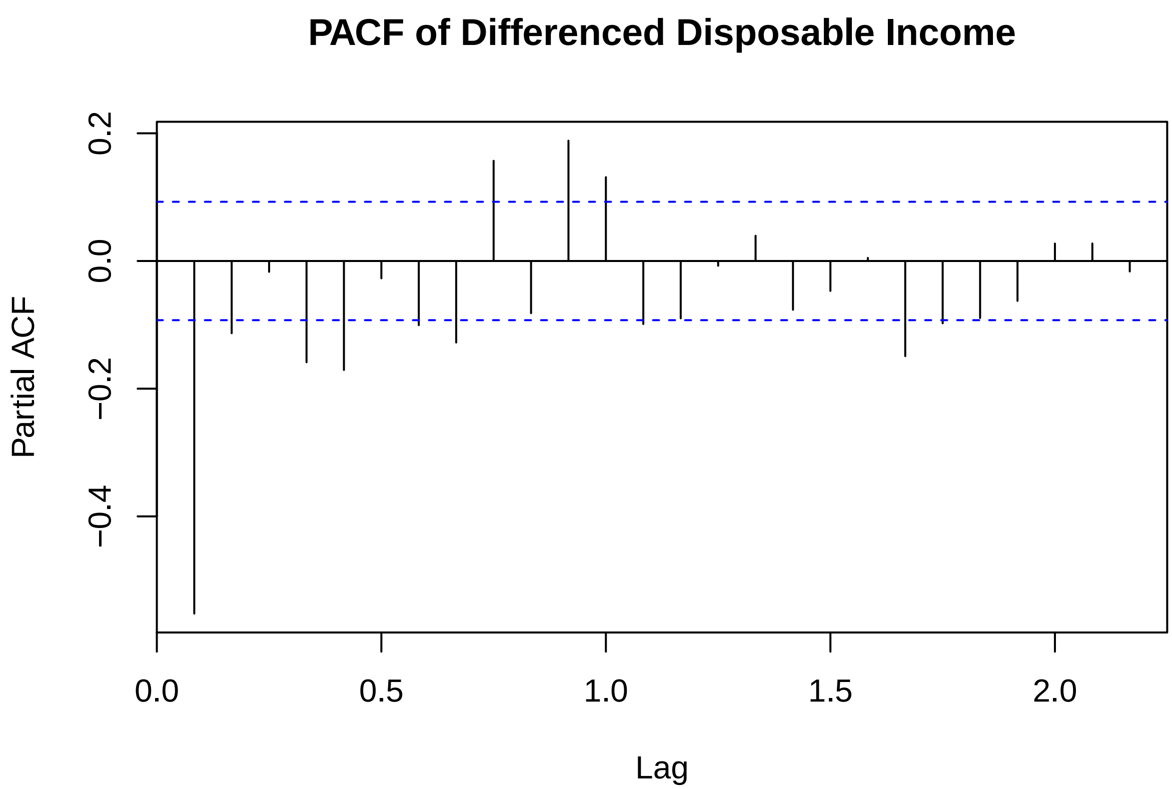

```
##
##  Augmented Dickey-Fuller Test
##
## data:  income_diff
## Dickey-Fuller = -11.362, Lag order = 7, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

After performing first-order differencing, we can visually see the trend has been removed from the series. The ADF confirms that the series is not stationary given the p-value of 0.01 is less than 0.05.

Before fitting an ARIMA model, we will check AR(p) and MA(q) terms using the PACF and ACF plots of the differenced series. If the PACF plot cuts off after lag k we will suggest an AR(p) model, if the ACF plot cuts off after lag m we will suggest an MA(q) model, and if both the ACF and PACF plot tail off we will suggest an ARMA model.

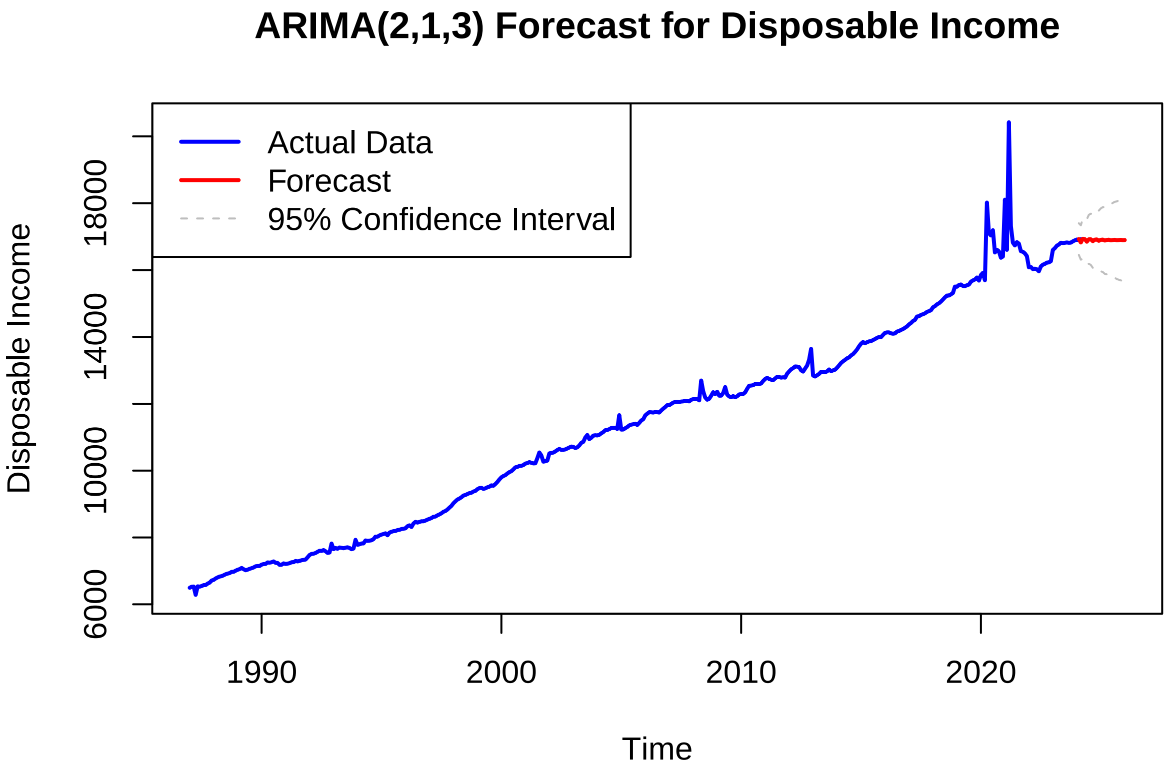


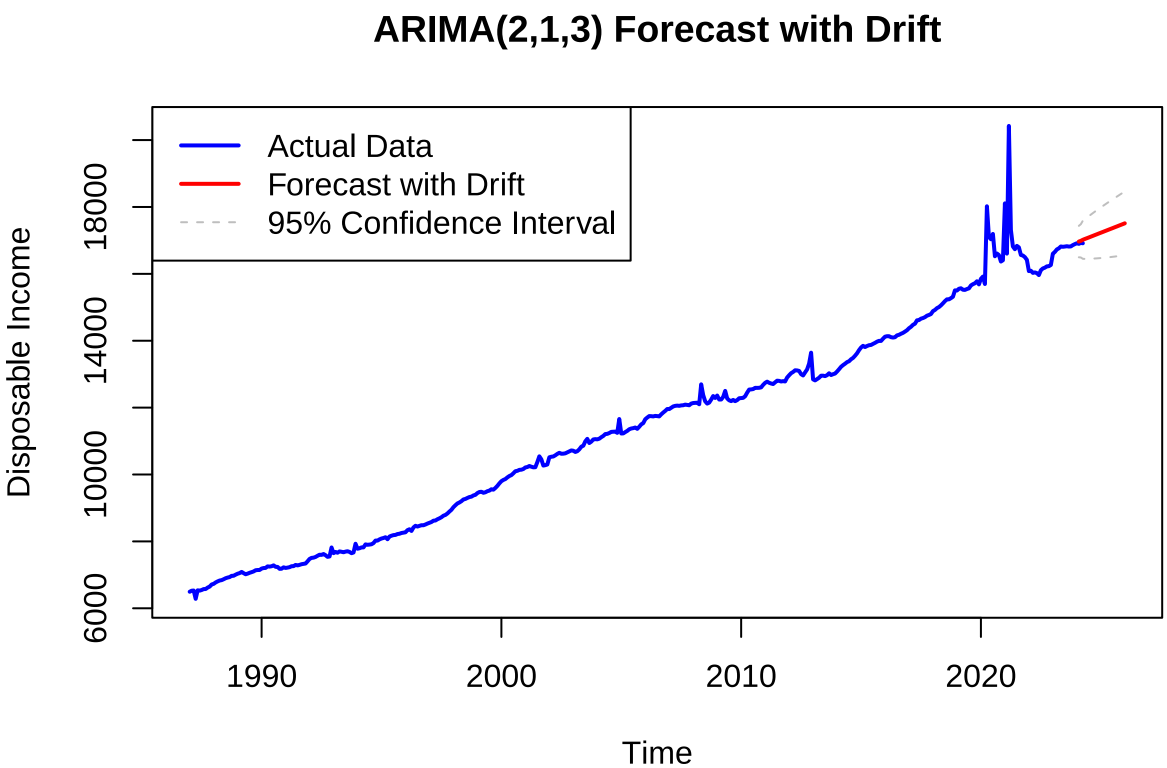


Examining the ACF plot, we see that the spikes drop off quickly to near zero after lag 1, suggesting we may want to use q = 1 (MA(1)). The PACF shows a significant spike at lag 1 and then also has some drop off, so we will use a value of p = 1 (AR(1)).

Based on these plots, we will start with an ARIMA model ARIMA(1,1,1)(p=1, d=1, q=1).

```
## Best Manual ARIMA Model: ARIMA( 2 , 1 , 3 )

##
## Call:
## arima(x = income_ts, order = c(p, d_values, q))
##
## Coefficients:
##           ar1      ar2     ma1     ma2      ma3
##       -0.8208  -0.7756  0.2464  0.5394  -0.4575
## s.e.   0.0590   0.0530  0.0653  0.0454   0.0527
##
## sigma^2 estimated as 58267:  log likelihood = -3087.29,  aic = 6186.58
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE      MAPE     MASE
## Training set 45.39018 241.1153 102.8327 0.4008022 0.8114639 1.259431
##                     ACF1
## Training set -0.04334266

## Manual RMSE: 241.1153

## Manual MAE: 102.8327

## Manual MAPE: 0.8114639 %
```


ARIMA(2,1,3) Forecast for Disposable Income
Actual Data
Forecast
95% Confidence Interval
Disposable Income
6000
10000
14000
18000
1990
2000
2010
2020
Time


ARIMA(2,1,3) Forecast with Drift
Actual Data
Forecast with Drift
95% Confidence Interval
Disposable Income
6000
10000
14000
18000
1990
2000
2010
2020
Time