A screenshot of a computer code

Description automatically generated

A graph with blue squares

Description automatically generated

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are instrumental in identifying the parameters of an ARIMA (AutoRegressive Integrated Moving Average) model. The parameters are typically denoted as ARIMA(p, d, q), where:

- `p` is the number of lag observations included in the model, also known as the lag order.

- `d` is the number of times that the raw observations are differenced, also known as the degree of differencing.

- `q` is the size of the moving average window, also known as the order of the moving average.

To determine these parameters, you would usually look at both the ACF and PACF plots, but based on the ACF plot you've provided:

1. The `d` parameter refers to the number of differences required to make the time series stationary. Since you've used the `diff` method in your code (as indicated by `df\_new['diff\_Log\_Real']`), you have already differenced the series at least once. If this differencing was sufficient to achieve stationarity, then `d` would be set to 1. If you had to difference the series more than once, `d` would be equal to the number of differencings needed.

2. For the `p` parameter, you'd typically look at the PACF plot to identify where the plot cuts off (meaning where the PACF values drop to zero beyond a certain lag). This would be your `p` value. Since we do not have the PACF plot, it's not possible to estimate `p` accurately from the ACF plot alone. ACF can sometimes help in identifying the MA component (q) by showing a sharp cut-off, while the PACF would help identify the AR component (p) by showing a gradual decay.

3. For the `q` parameter, you would generally look for where the ACF plot cuts off after a few lags. In your ACF plot, the first lag shows a significant negative correlation, and subsequent lags taper off. There's no clear cut-off in the ACF plot, which might suggest a more complex model or that the series does not have a strong MA component. However, if we had to estimate, we might start with a `q` of 1 or 2 given the first lag is quite significant and then test the model to see if it improves with higher lags.

Given this information, a starting point for your ARIMA model could be ARIMA(p, 1, q), where `p` and `q` would need to be identified, possibly through iterative fitting and testing or by looking at the PACF plot for `p`. It’s important to use both ACF and PACF plots together, as well as criteria like the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), for a more accurate determination of the ARIMA parameters. You may also want to perform a stationarity test such as the Augmented Dickey-Fuller test to confirm the order of differencing `d`.

A screenshot of a computer

Description automatically generated

A graph with blue squares

Description automatically generated

The Partial Autocorrelation Function (PACF) plot helps determine the order of the autoregressive (AR) part of an ARIMA model, denoted by 'p'. The 'p' value is typically where the PACF plot crosses the significance boundary (which is not shown on your plot but is often represented by blue shaded areas or lines at roughly ±1.96/√n, where n is the sample size) for the last time before the values fall within it for an extended number of lags.

In the PACF plot you've provided, it appears that the partial autocorrelations at lags 1 and 2 are significant (since they are notably above the rest of the lags, assuming significance limits would be around them), and all subsequent lags fall close to zero or within the expected significance bounds. The lag at which the PACF cuts off (significant spike followed by non-significant values) is typically used for the 'p' parameter.

Therefore, you might start by considering an AR(2) model, meaning the p value would be 2. This suggests that the present value of the series is related to its previous two values. However, in practice, you would need to consider the significance limits, which are not visible on the plot provided. If the significance limits are within the range where the PACF values for lag 1 and 2 fall, then p=2 would be a good starting point.

It is also common to test a few different 'p' values around the one suggested by the PACF plot, then use a model selection criterion such as AIC or BIC to choose the best model. Remember that ARIMA modeling is often an iterative process, where the choice of 'p', 'd', and 'q' is refined based on the model's performance and diagnostic checks.