# Problem 1 Bond Dataset:

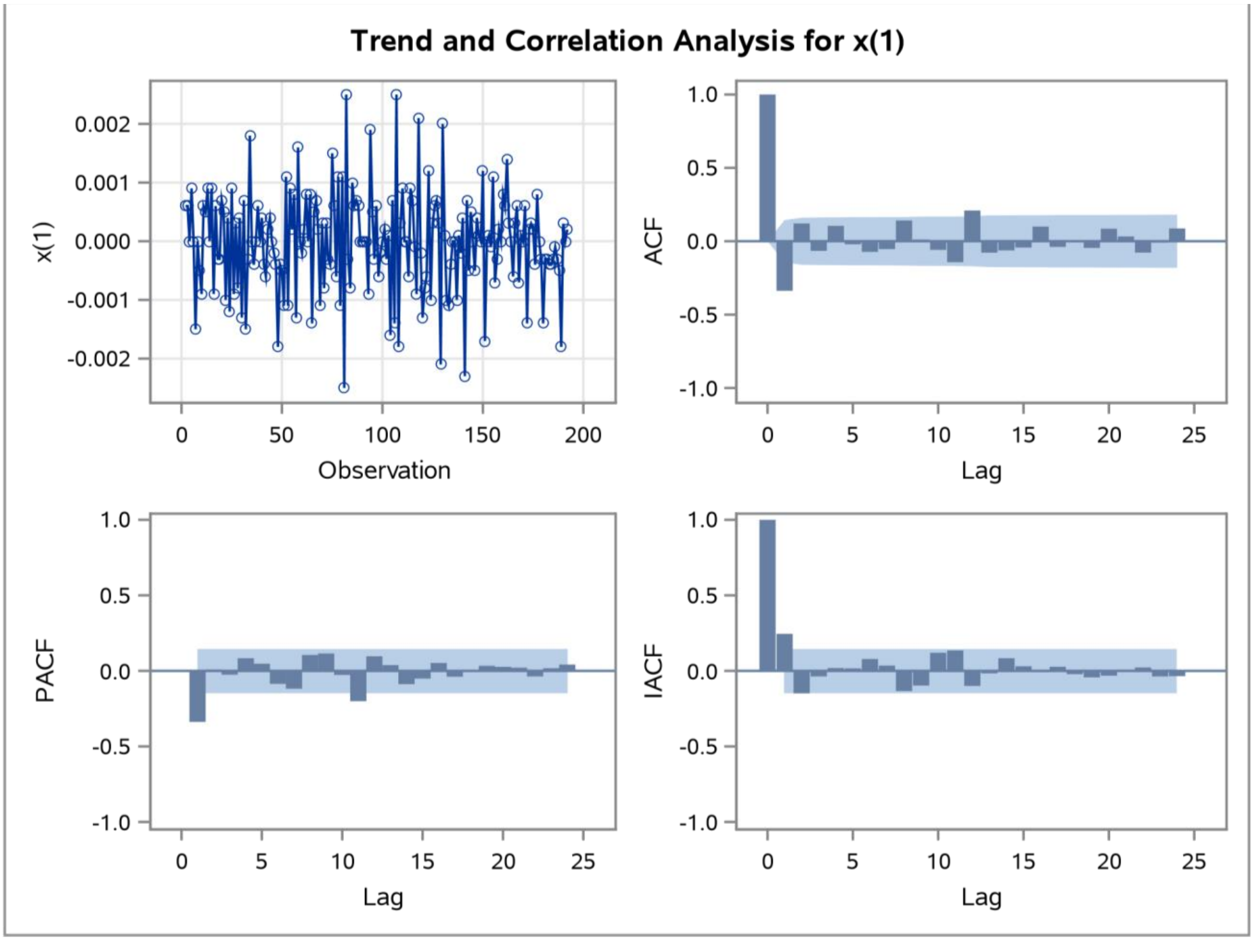
**For the Series X, I chose a ARMA(0,1,2) Model without a constant(NOCONSTANT).**

**I chose this model because the ACF of the data, after two lags**

**fell within the confidence interval, and the PACF of the differenced data exponentially decayed rapidly. This is behavior associated with a MA(q) process. Additionally, the ARMA(0,1,2) model yielded the highest AIC when compared to other models like ARMA(0,2,2) because the ARMA(0,2,2) was clearly over-differenced since the IACF decayed slowly.**

**After I decided on the model, and compared AIC's, I tested for normally distributed residuals. This was confirmed in the output of our estimation, the distribution takes on the bell shape.**
**Additionally, when the residuals are plotted against the predicted values, there is constant variance and seemingly zero correlation! All of this supports the conclusion that ARMA(0,1,2) is an adequate model .**

## ACF, PACF, IACF, Time Series Plot

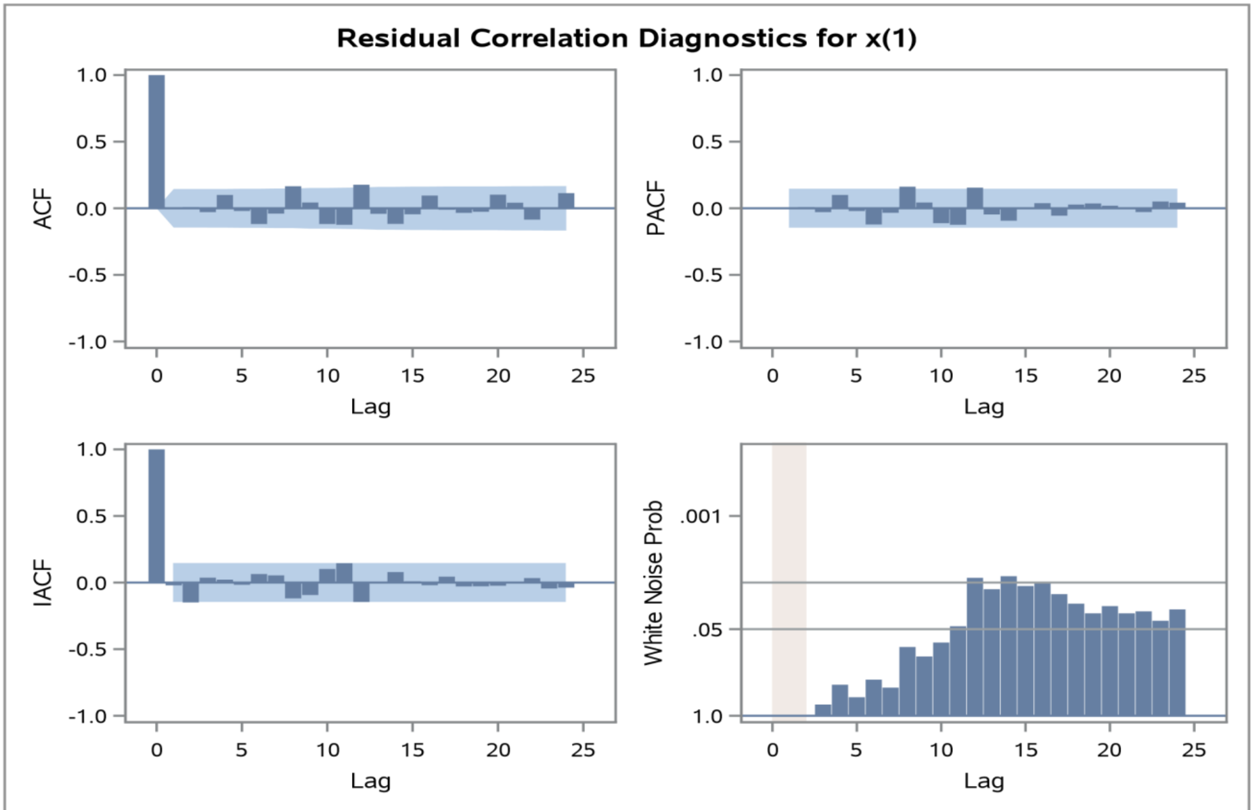


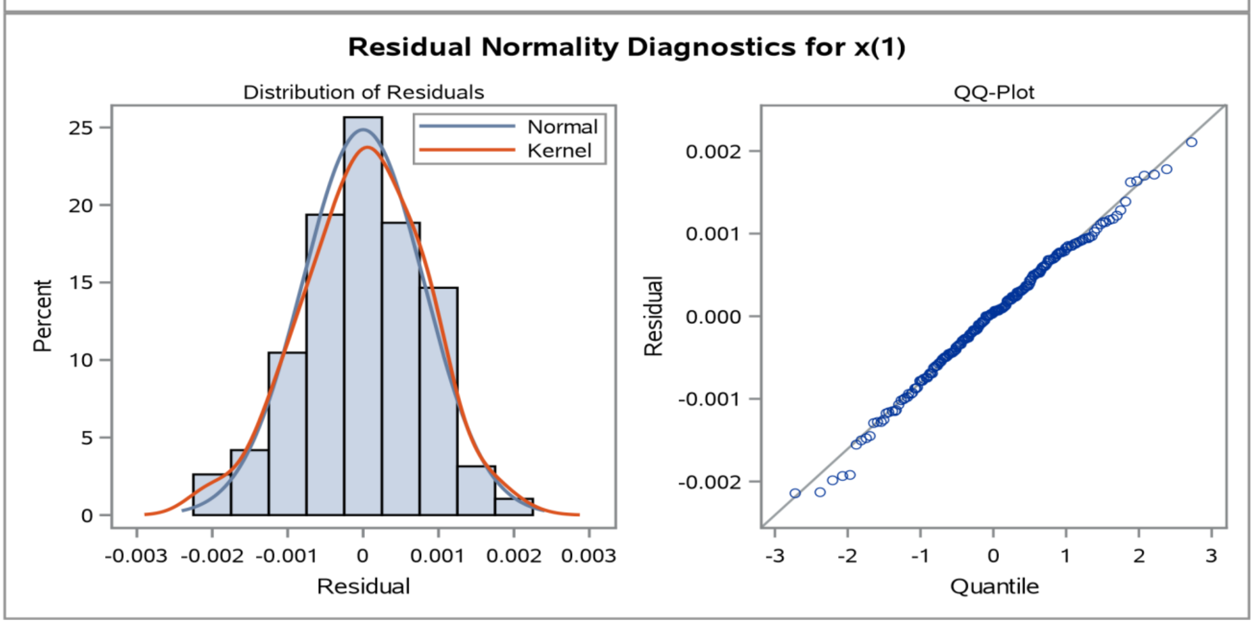

**This is the time series plots of the data after it's differenced once.**

**In the ACF after two lags, ACF stays within confidence band.**

**In PACF it is clearly sinusoidal dampening towards zero. The IACF quickly converges to zero telling us that we haven't over-differenced yet.**

## Residual Diagnostics




**The ACF and PACF of the residuals stays in the confidence band, and the distribution of the residuals is more or less normal. The QQ-Pot confirms this.**

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | -0.0000203 | 0.00004483 | -0.45 | 0.6512 | 0 |
| MA1,1 | 0.33943 | 0.07254 | 4.68 | <.0001 | 1 |
| MA1,2 | -0.10695 | 0.07257 | -1.47 | 0.1406 | 2 |

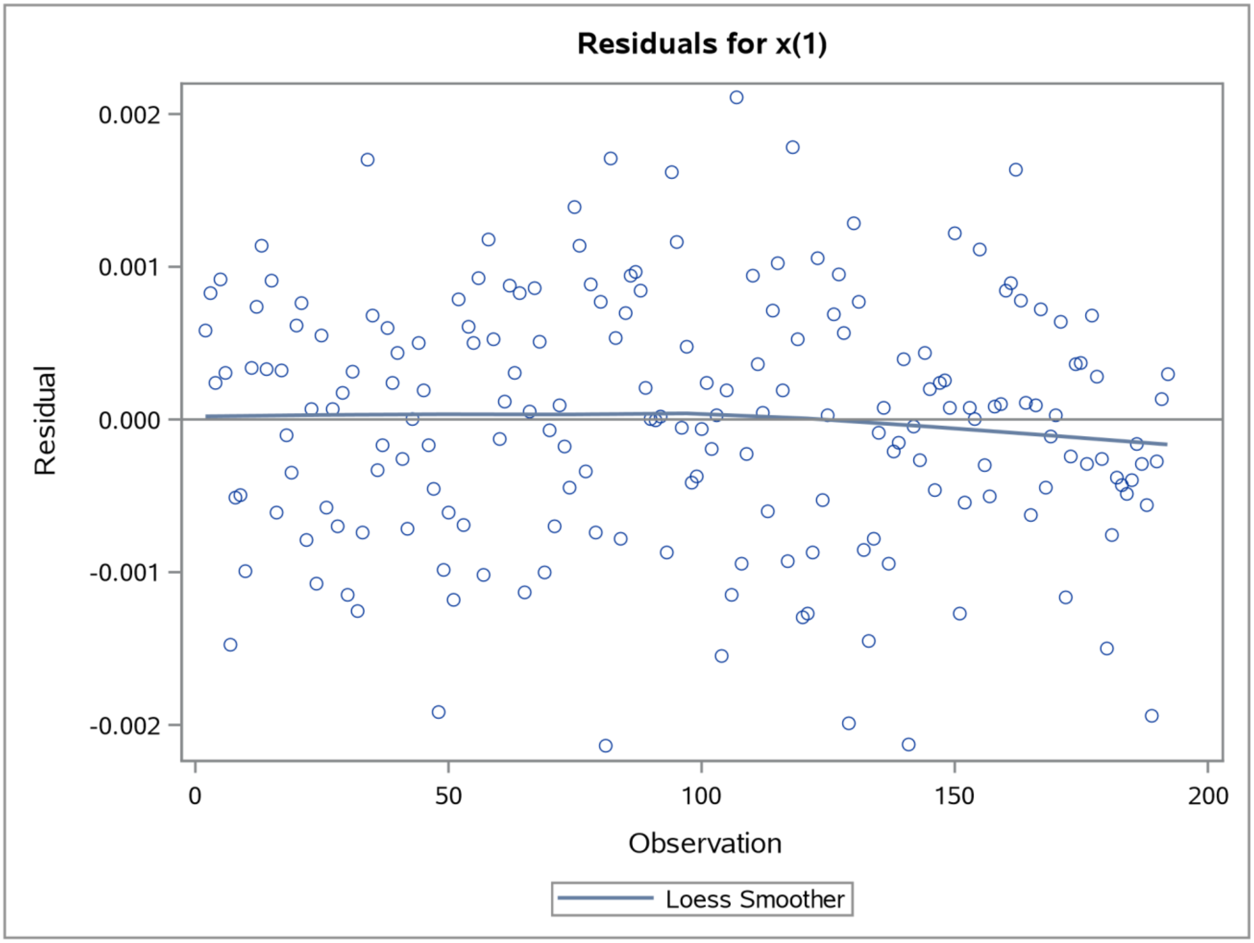
| | |
|---|---|
| Constant Estimate | -0.00002 |
| Variance Estimate | 6.507E-7 |
| Std Error Estimate | 0.000807 |
| AIC | -2175.69 |
| SBC | -2165.94 |
| Number of Residuals | 191 |

| Correlations of Parameter Estimates | | | |
|---|---|---|---|
| Parameter | MU | MA1,1 | MA1,2 |
| MU | 1.000 | -0.002 | -0.003 |
| MA1,1 | -0.002 | 1.000 | -0.307 |
| MA1,2 | -0.003 | -0.307 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 5.00 | 4 | 0.2869 | 0.000 | 0.008 | -0.029 | 0.100 | -0.020 | -0.118 |
| 12 | 23.69 | 10 | 0.0085 | -0.040 | 0.166 | 0.043 | -0.118 | -0.124 | 0.177 |
| 18 | 29.51 | 16 | 0.0207 | -0.041 | -0.117 | -0.045 | 0.095 | -0.011 | -0.034 |
| 24 | 36.74 | 22 | 0.0253 | -0.026 | 0.102 | 0.042 | -0.085 | -0.002 | 0.114 |
| 30 | 53.10 | 28 | 0.0029 | -0.016 | -0.252 | -0.049 | -0.060 | 0.014 | -0.054 |
| 36 | 65.54 | 34 | 0.0009 | 0.046 | 0.106 | 0.068 | -0.186 | 0.013 | 0.020 |

## Parameter estimates for our equation so that predictions may be possible!

## Residual Time Series Plot

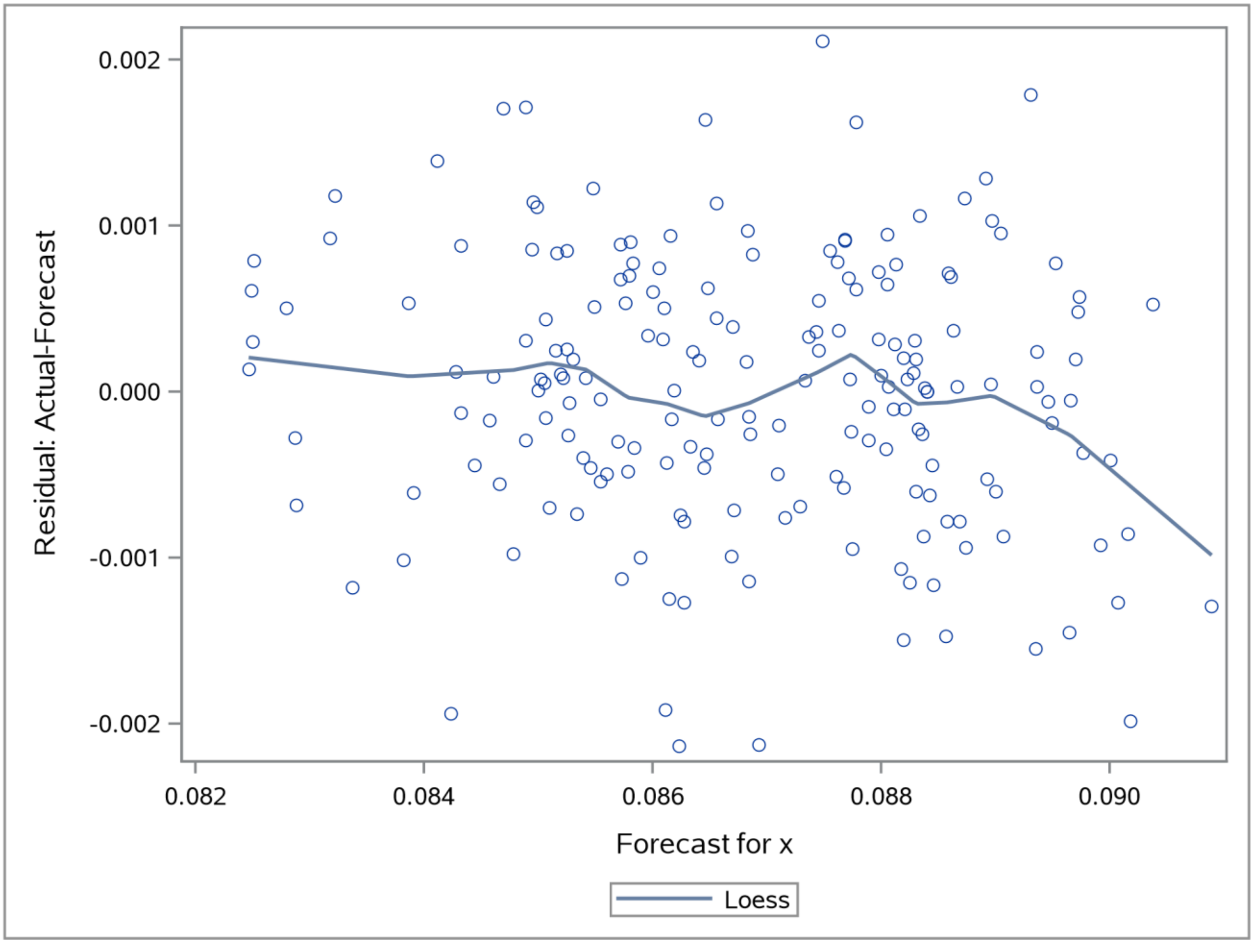



| Model for variable x | |
|---|---|
| **Estimated Mean** | -0.00002 |
| **Period(s) of Differencing** | 1 |

| Moving Average Factors | |
|---|---|
| **Factor 1:** | 1 - 0.33943 B**(1) + 0.10695 B**(2) |

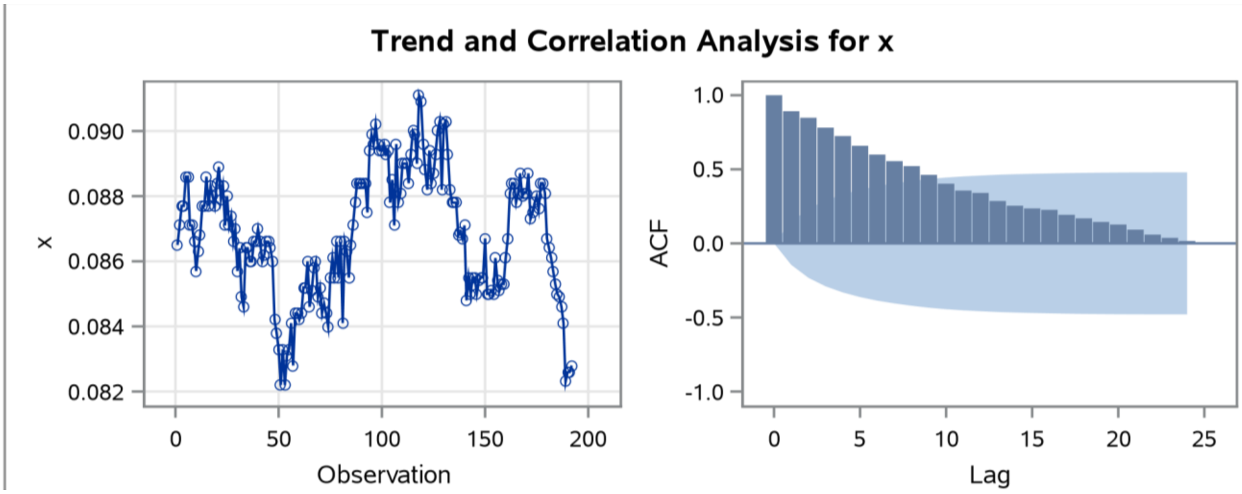
## This shows constant variance and mean.

## Residuals vs Predicted Values Plot



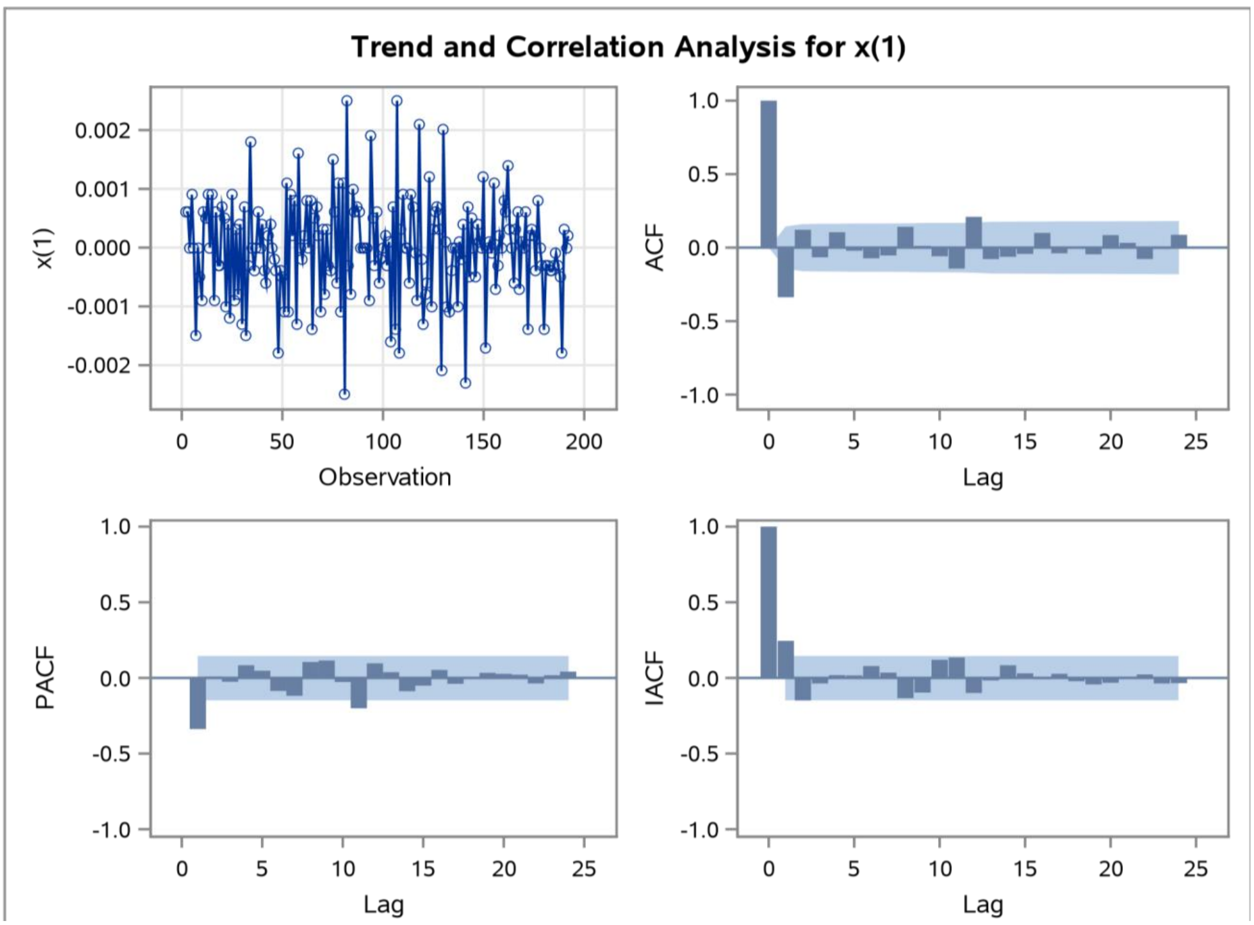

**This shows very little to no correlation between residuals and predicted values.**

## ***Raw Bond Data Plots***




**Here we see there is a lot of variability and the ACF tells us that this process is not stationary granted how slowly it decays to zero. We look for stationarity by differencing.**

## First Difference, with constant variance, now looks stationary.



# Problem 2: Geyser Dataset

**For the series X, I chose an ARMA(1,0,0) or an AR(1) Model with no constants(NOCONSTANT).**

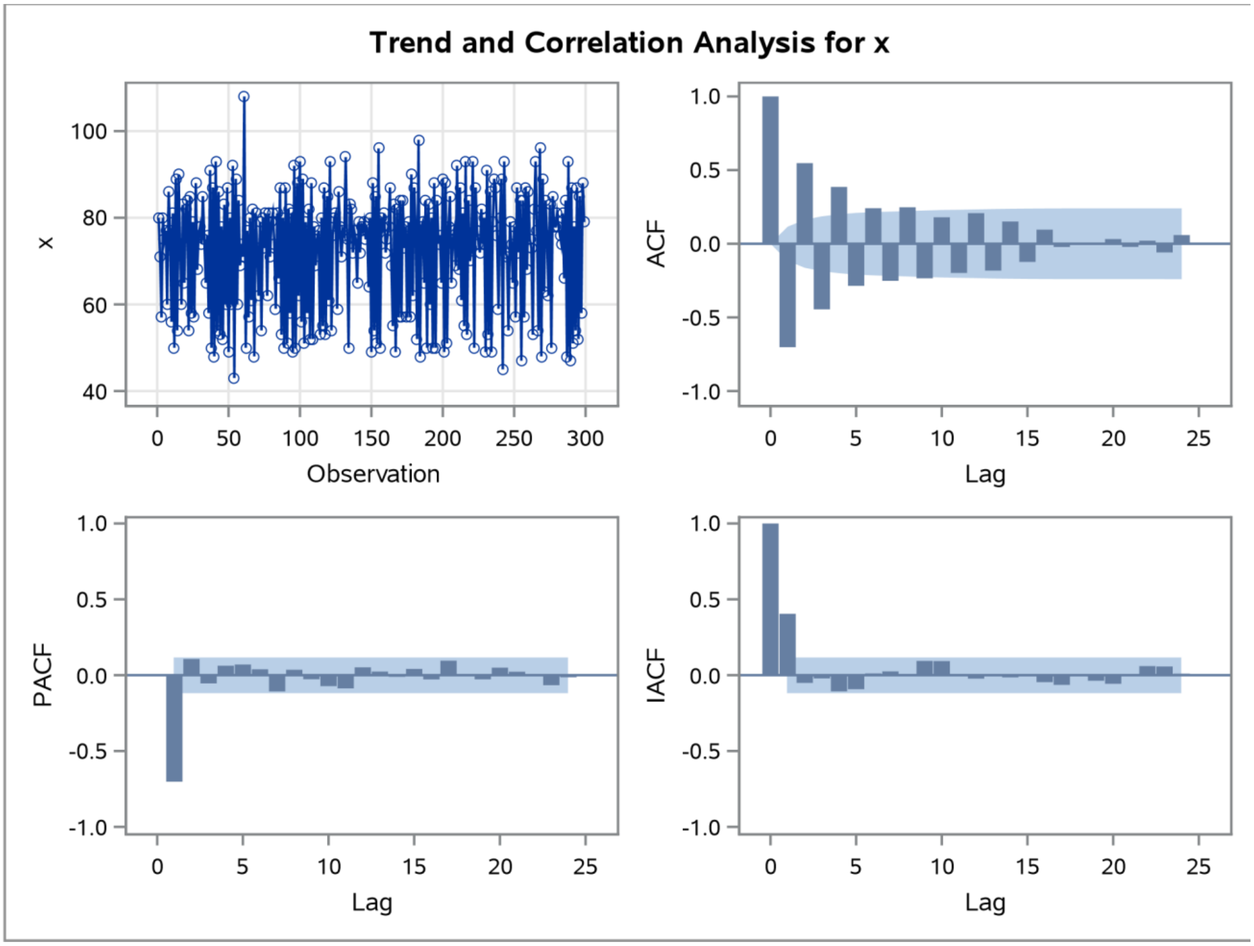
**I chose this model because from first glance the raw data Exhibits behavior we associate with AR(1) processes. The ACF decays and dampens simultaneously. Additionally, after 1 lag, the PACF falls within the confidence band which is indicative of an AR(1) process.**

**After guessing this, the diagnostics gave us confirmation. The ACF and the PACF of the residuals were mostly within their respective confidence bands.**

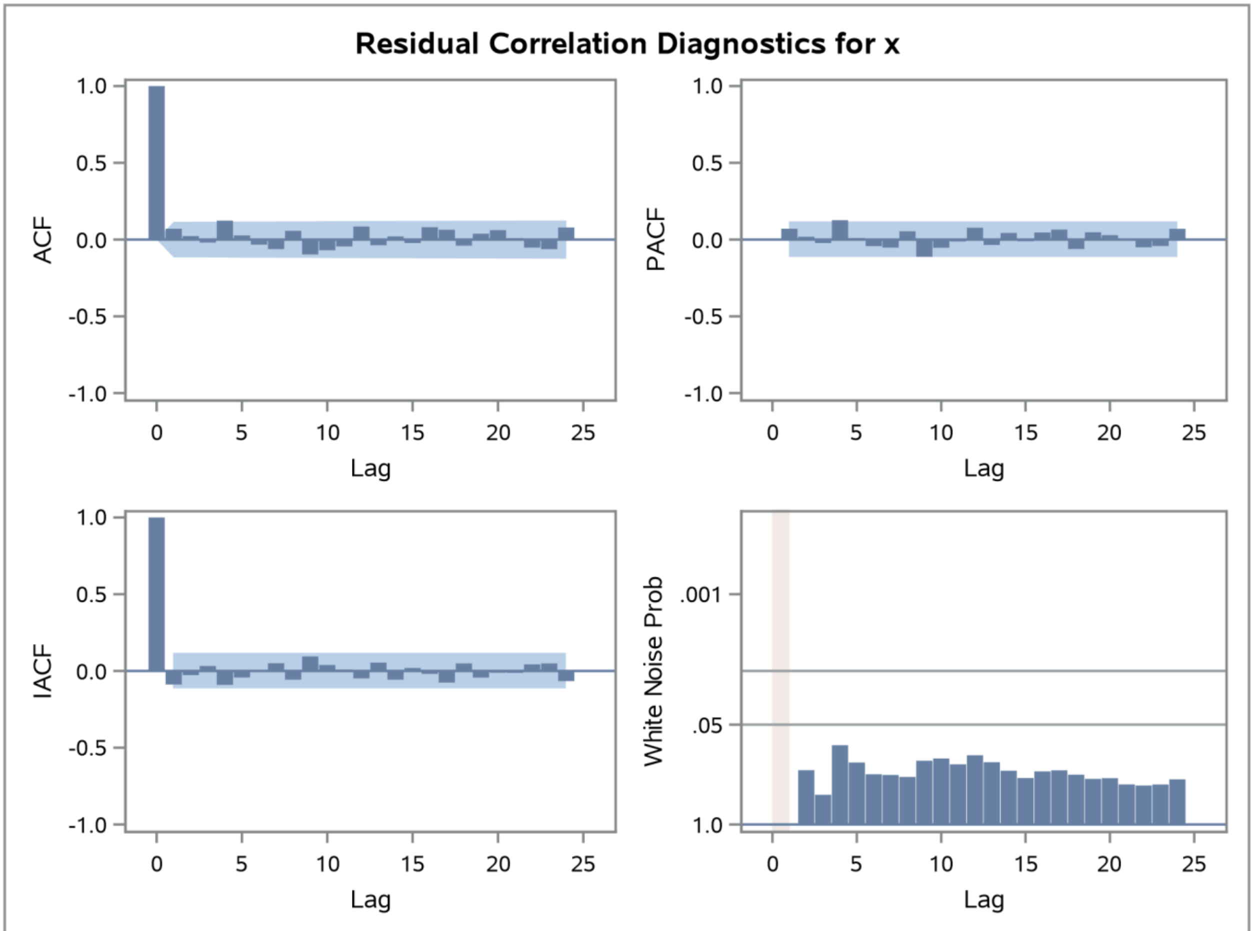
**The distribution of the residuals corresponded with a normal distribution and also the residuals fell neatly on a straight line as expected in a QQ Plot.**

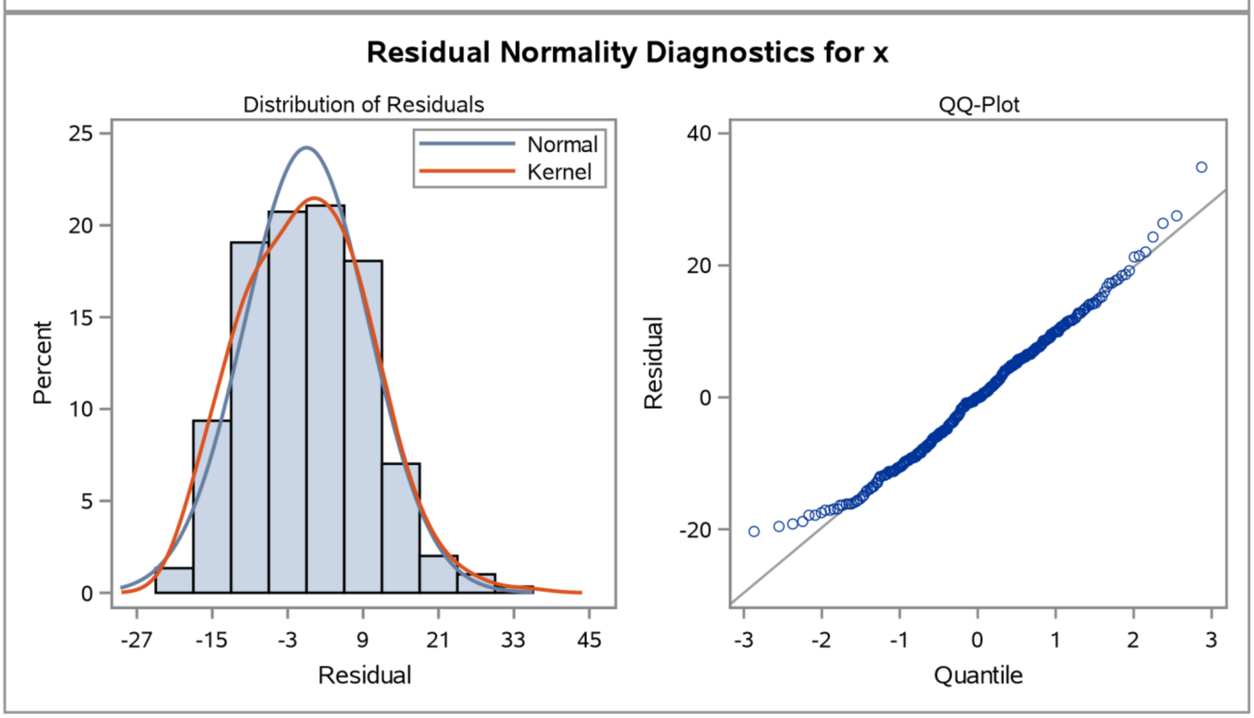
**The residuals also had constant variance and little to no correlation in the plot of the residuals vs the predicted values.**

# Time Series Plot, ACF, PACF, IACF



# Residual Diagnostics



| | |
|---|---|
| Constant Estimate | 122.9845 |
| Variance Estimate | 97.95699 |
| Std Error Estimate | 9.897322 |
| AIC | 2221.969 |
| SBC | 2229.37 |
| Number of Residuals | 299 |

| Correlations of Parameter Estimates | | |
|---|---|---|
| Parameter | MU | AR1,1 |
| MU | 1.000 | 0.001 |
| AR1,1 | 0.001 | 1.000 |

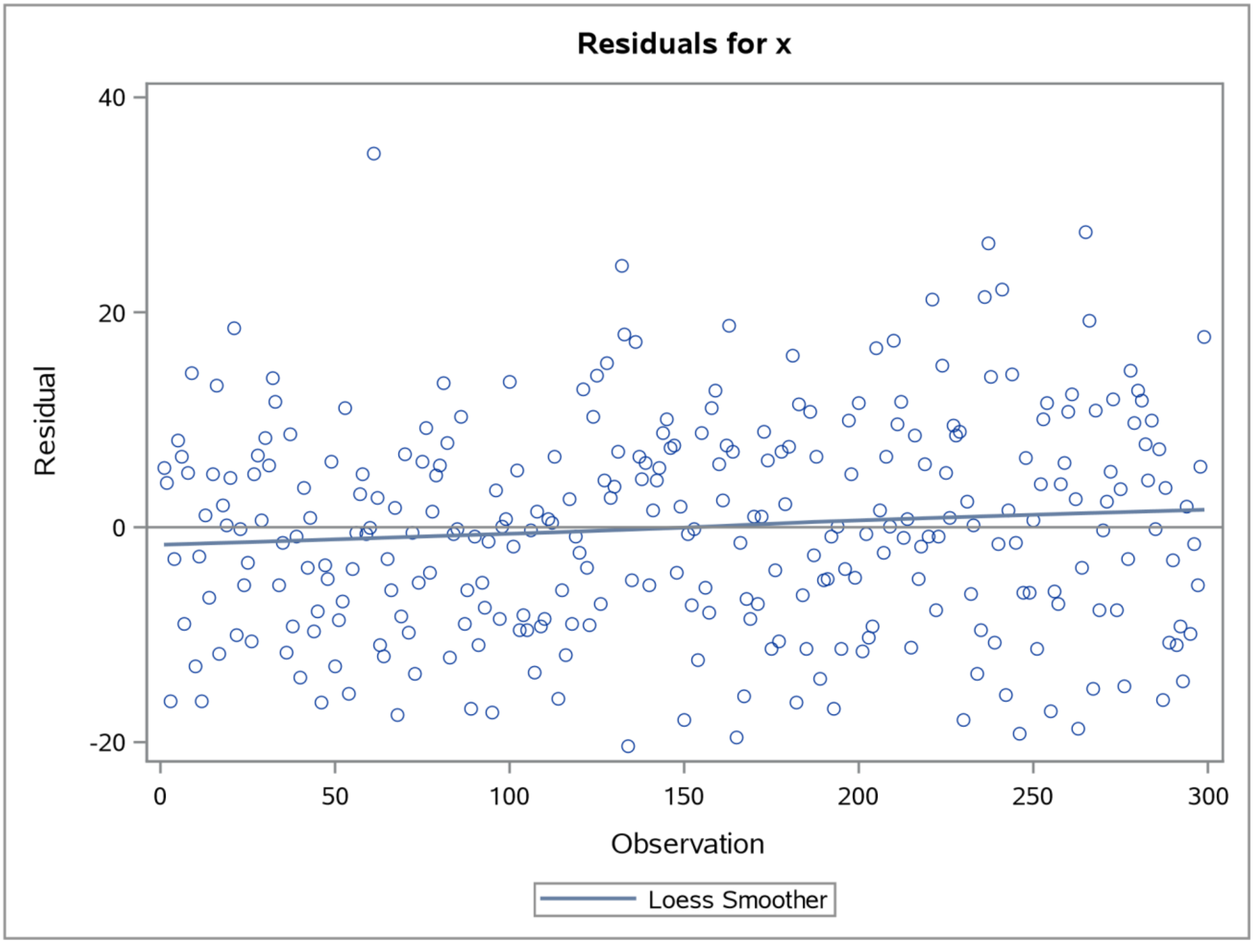
| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 6.98 | 5 | 0.2220 | 0.071 | 0.023 | -0.019 | 0.123 | 0.027 | -0.033 |
| 12 | 16.45 | 11 | 0.1254 | -0.062 | 0.057 | -0.097 | -0.069 | -0.044 | 0.085 |
| 18 | 21.04 | 17 | 0.2245 | -0.037 | 0.020 | -0.022 | 0.081 | 0.064 | -0.039 |
| 24 | 26.95 | 23 | 0.2581 | 0.038 | 0.062 | 0.010 | -0.052 | -0.062 | 0.079 |
| 30 | 30.63 | 29 | 0.3832 | 0.028 | -0.048 | 0.014 | 0.055 | -0.027 | -0.063 |
| 36 | 37.07 | 35 | 0.3737 | -0.009 | -0.032 | 0.002 | -0.063 | -0.088 | -0.079 |
| 42 | 38.71 | 41 | 0.5729 | 0.022 | 0.004 | -0.008 | -0.007 | 0.060 | 0.021 |
| 48 | 44.26 | 47 | 0.5868 | 0.021 | -0.050 | 0.059 | -0.035 | 0.078 | -0.044 |

**Parameter Estimates so we may begin predictions.**

**The ACF and PACF of the residuals behave as expected! With residual ACF and PACF being mostly within confidence band.**

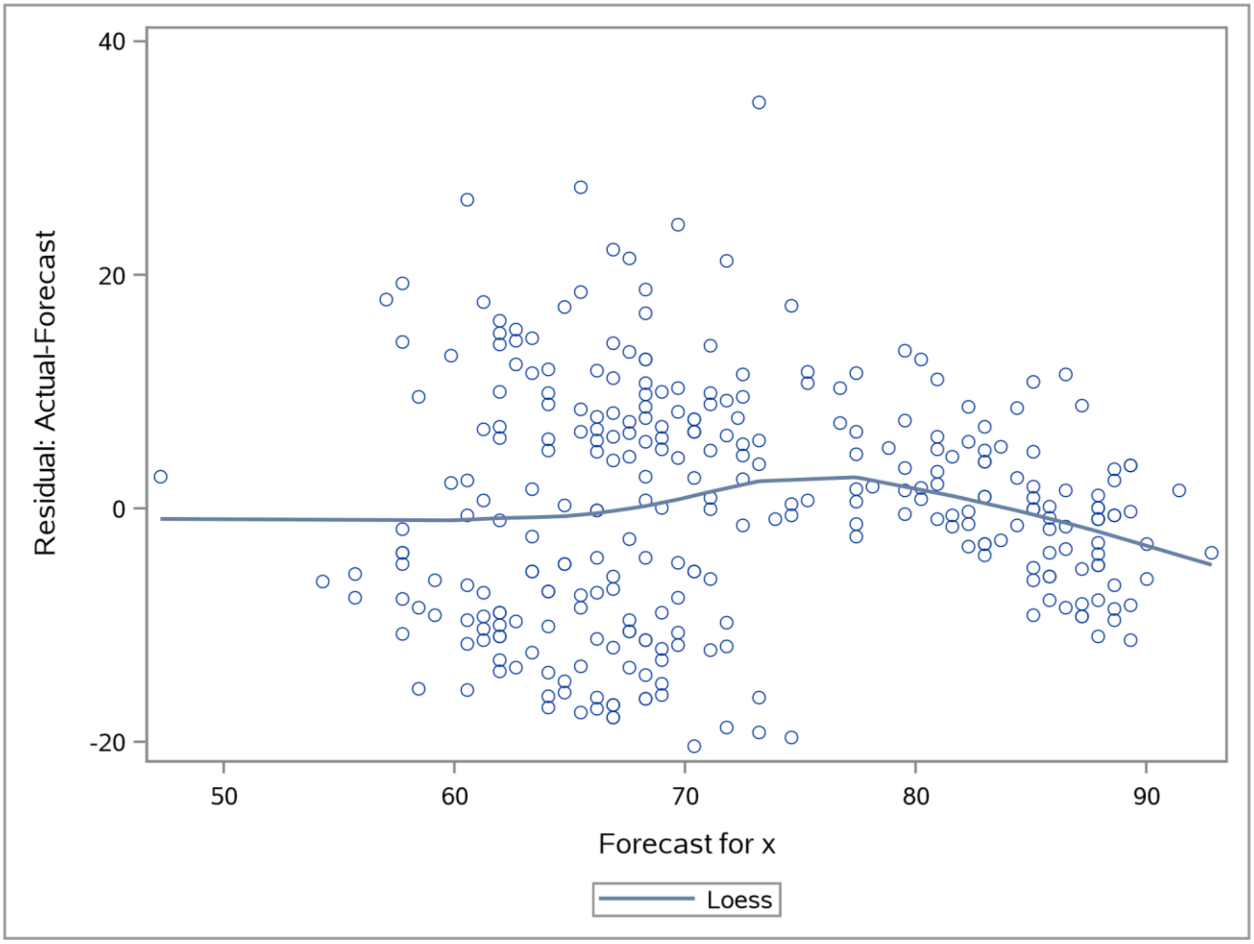
**Residuals also follow a normal distribution nicely.**

# Residual Time Series Plot




## We see random clutter that signals constant variance.

## Residuals Vs Predicted Values

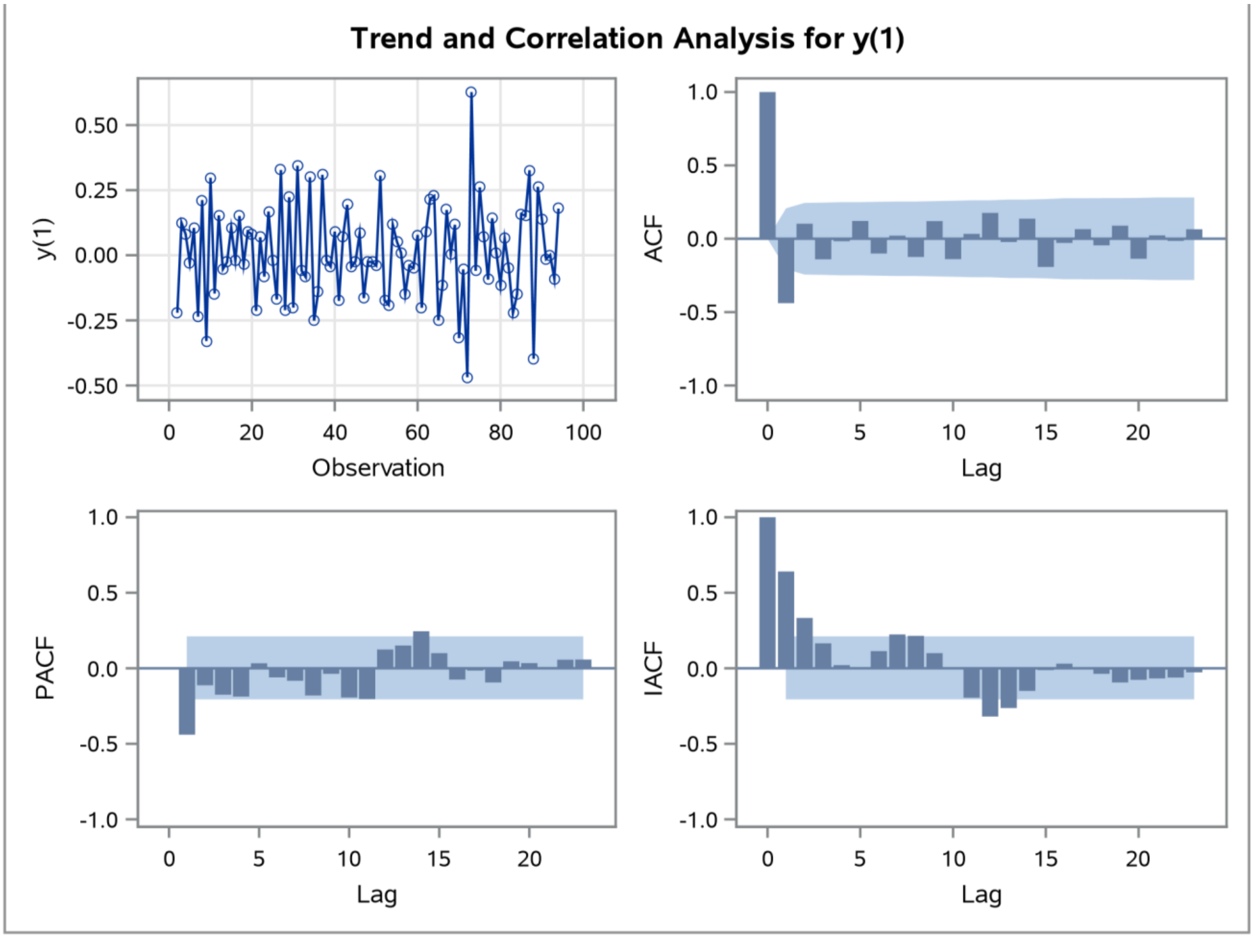



## Little to no correlation!

# Problem 3 Repair Dataset:

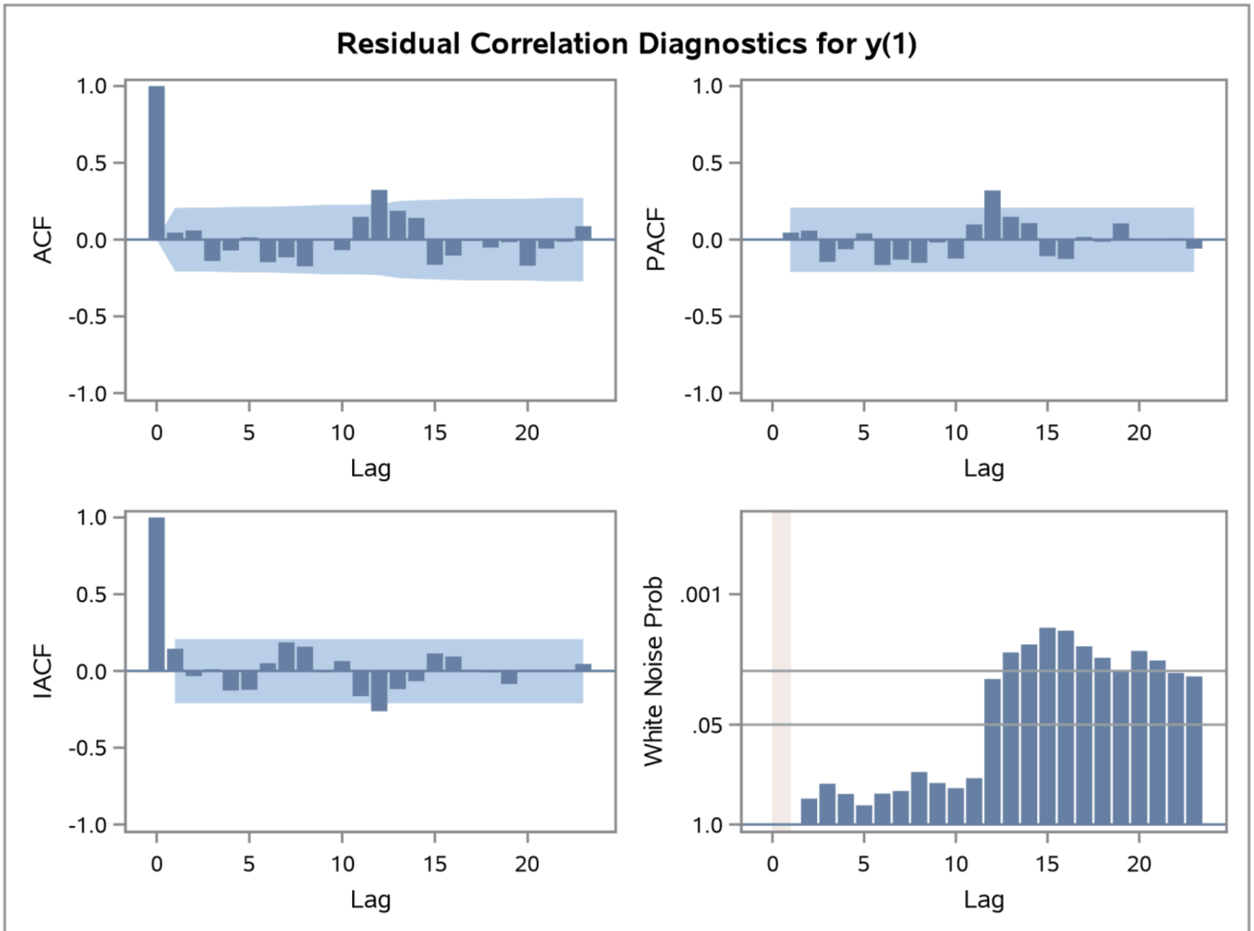
**For the series Y = Log(X) I chose an ARMA(0,1,2) model with no constant (NOCONSTANT). I chose this model because the ACF of the time series after lag 1 stayed within the confidence band, and the PACF of the time series displayed a sinusoidal dampening behavior indicative of a MA(1) process after differencing. I took this guess and applied an ARMA(0,1,2) process, and the results gave us normally distributed residuals in the distribution plot. The residuals also fell on the straight line in the QQ-Plot, thus fulfilling our assumptions with respect to normality of the residuals. In the Residual vs Predicted value plot, we see that the residuals seem to have constant variance and little to no correlation, which verify that our model choice is a good one.**

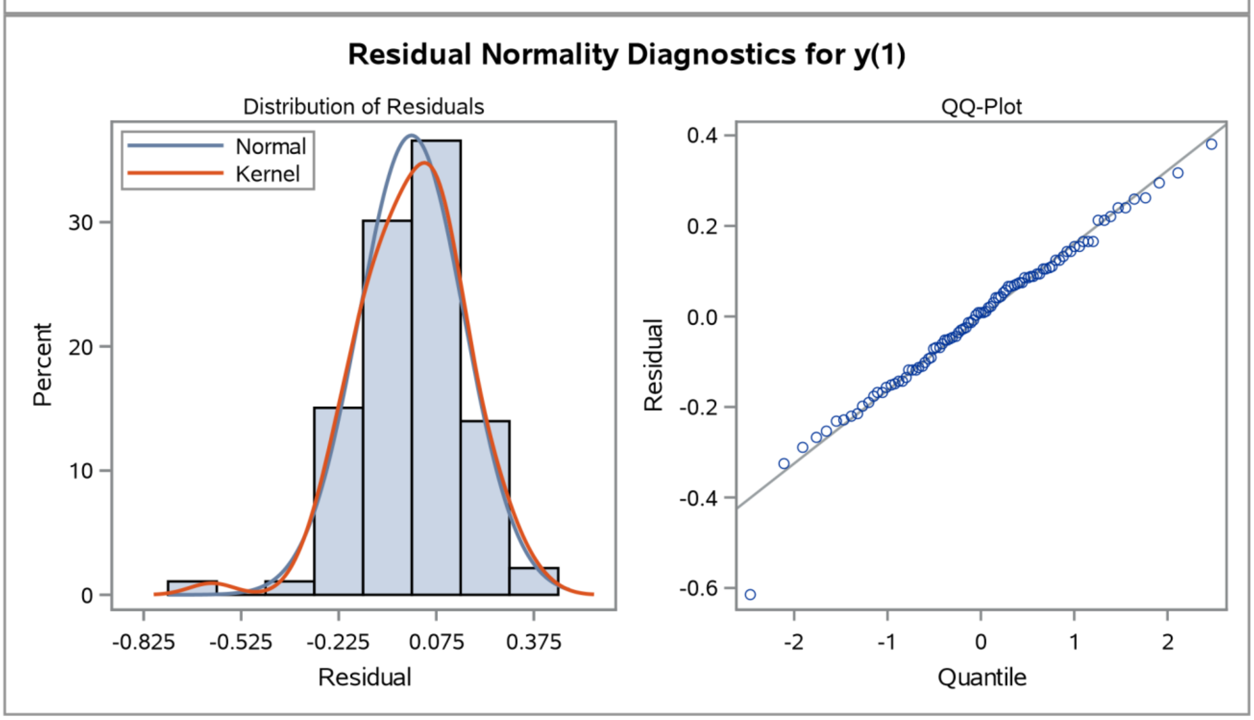
# ACF, PACF, IACF, Time Series Plot




| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 0.01177 | 0.0066874 | 1.76 | 0.0784 | 0 |
| MA1,1 | 0.60985 | 0.08404 | 7.26 | <.0001 | 1 |

# Residual Diagnostics



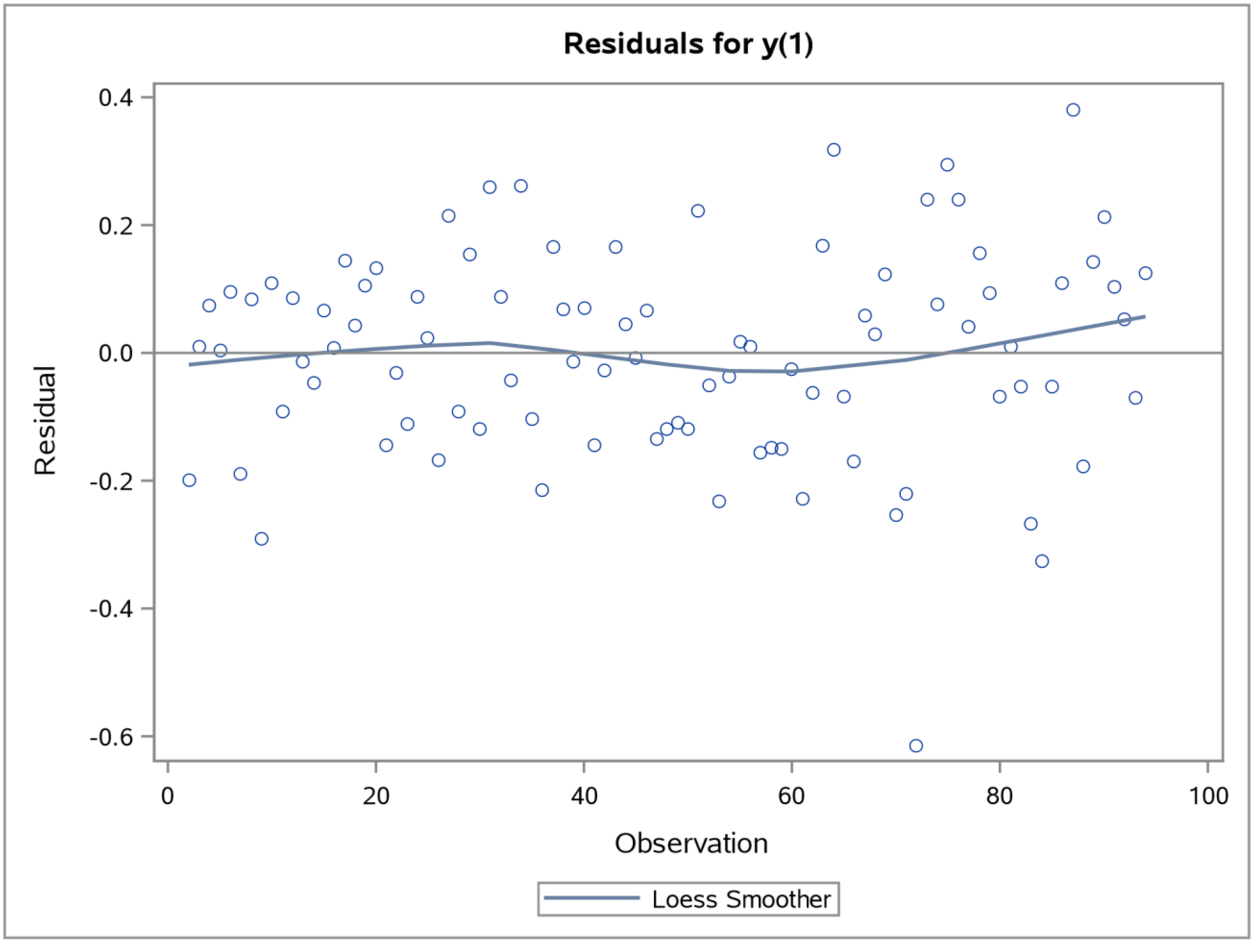
| | |
|---|---|
| **Constant Estimate** | 0.011769 |
| **Variance Estimate** | 0.026473 |
| **Std Error Estimate** | 0.162705 |
| **AIC** | -71.3764 |
| **SBC** | -66.3112 |
| **Number of Residuals** | 93 |

| Correlations of Parameter Estimates | | |
|---|---|---|
| **Parameter** | **MU** | **MA1,1** |
| **MU** | 1.000 | -0.020 |
| **MA1,1** | -0.020 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **To Lag** | **Chi-Square** | **DF** | **Pr > ChiSq** | **Autocorrelations** | | | | | |
| **6** | 5.16 | 5 | 0.3967 | 0.046 | 0.060 | -0.139 | -0.071 | 0.015 | -0.147 |
| **12** | 23.99 | 11 | 0.0128 | -0.115 | -0.174 | 0.004 | -0.068 | 0.149 | 0.324 |
| **18** | 34.71 | 17 | 0.0068 | 0.188 | 0.141 | -0.164 | -0.104 | -0.008 | -0.052 |
| **24** | 39.95 | 23 | 0.0156 | -0.016 | -0.169 | -0.058 | -0.013 | 0.088 | 0.050 |

**The residuals follow a normal distribution fairly well! As well as act as expected in the QQ-Plot.**

# Residual Time Series Plot

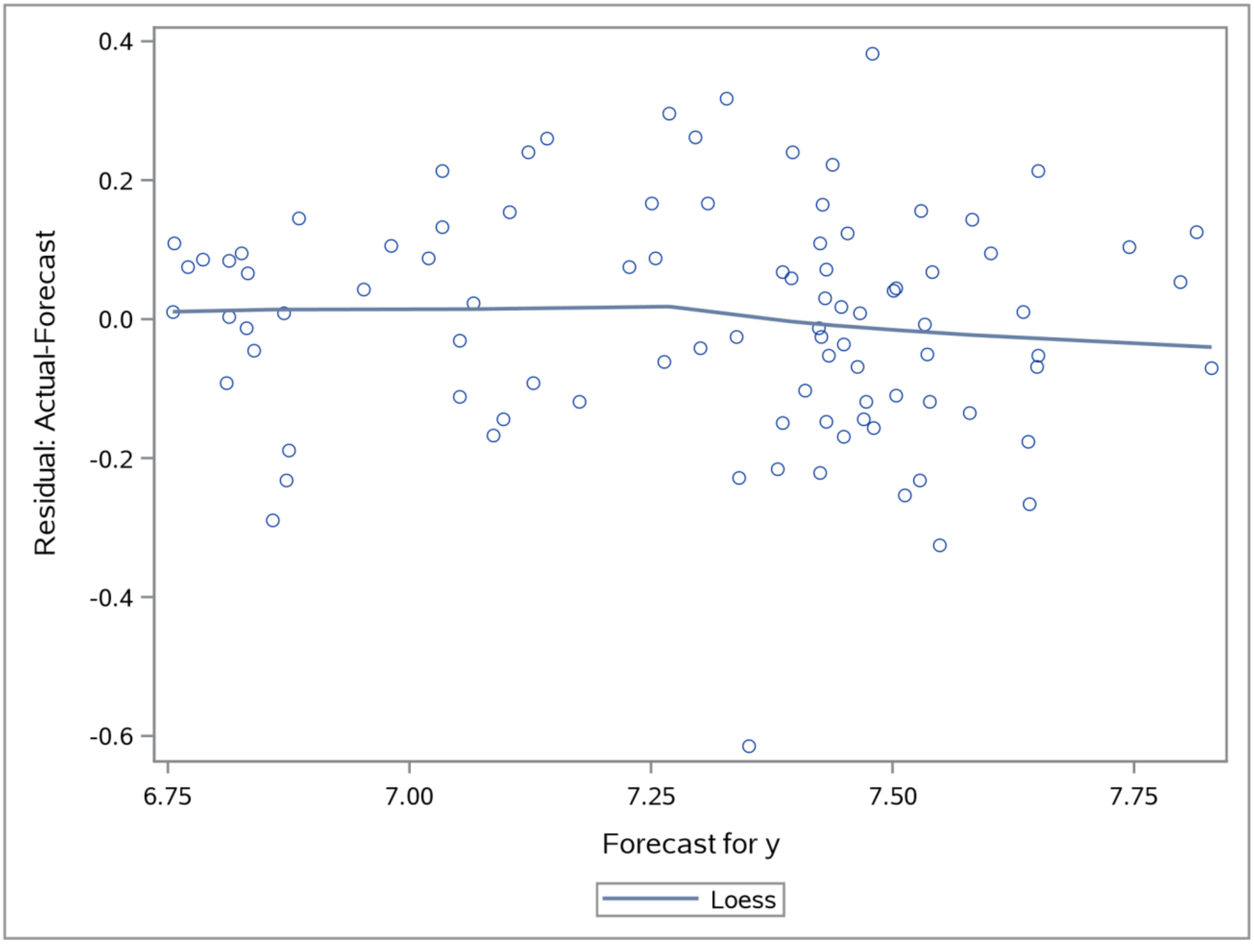



| Model for variable y | |
|---|---|
| **Estimated Mean** | 0.011769 |
| **Period(s) of Differencing** | 1 |

| Moving Average Factors | |
|---|---|
| **Factor 1:** | 1 - 0.60985 B**(1) |

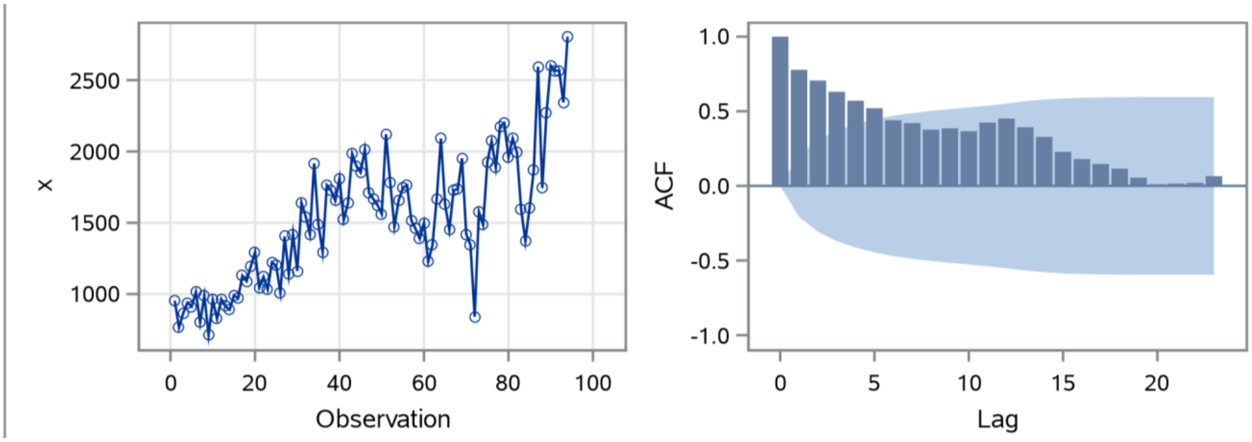
**Here we see there is constant variance.**

## Residuals vs Predicted Values



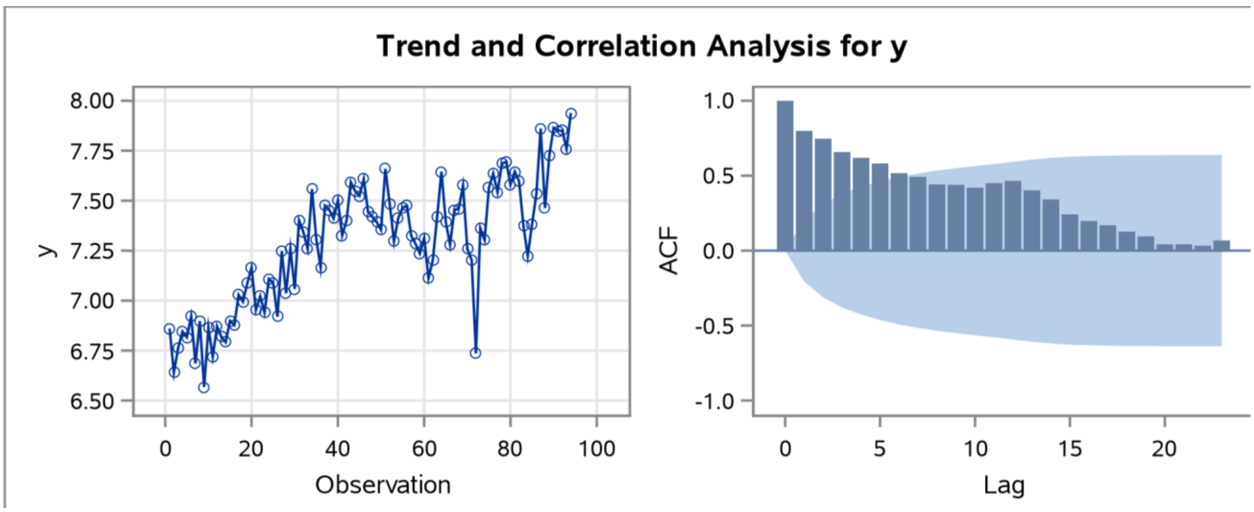

**We can observe there is constant variance with little to no correlation.**

## Raw Data Time Series

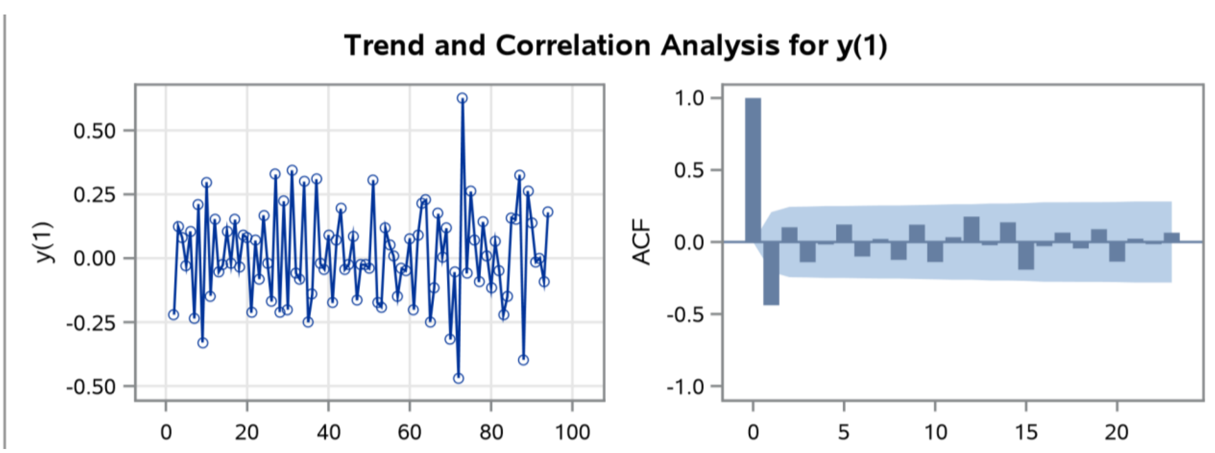



**Data is on a huge scale with lots of variety in very small intervals. Also is not stationary because ACF diminishes very slowly. This shows a positive trend which suggests a log transformation**

# After Log Transformation




# After First Difference

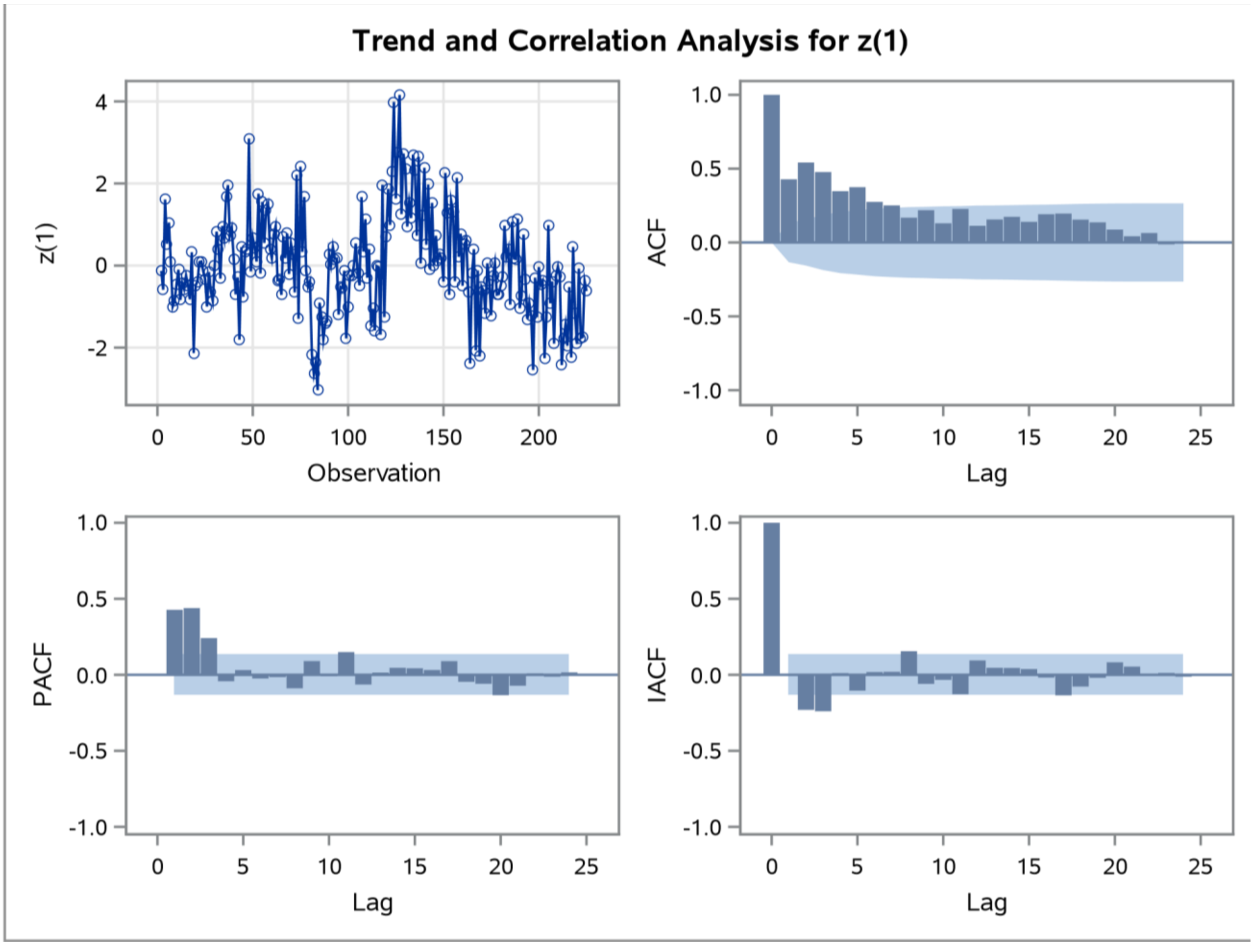


## Problem 4 Fake Dataset:

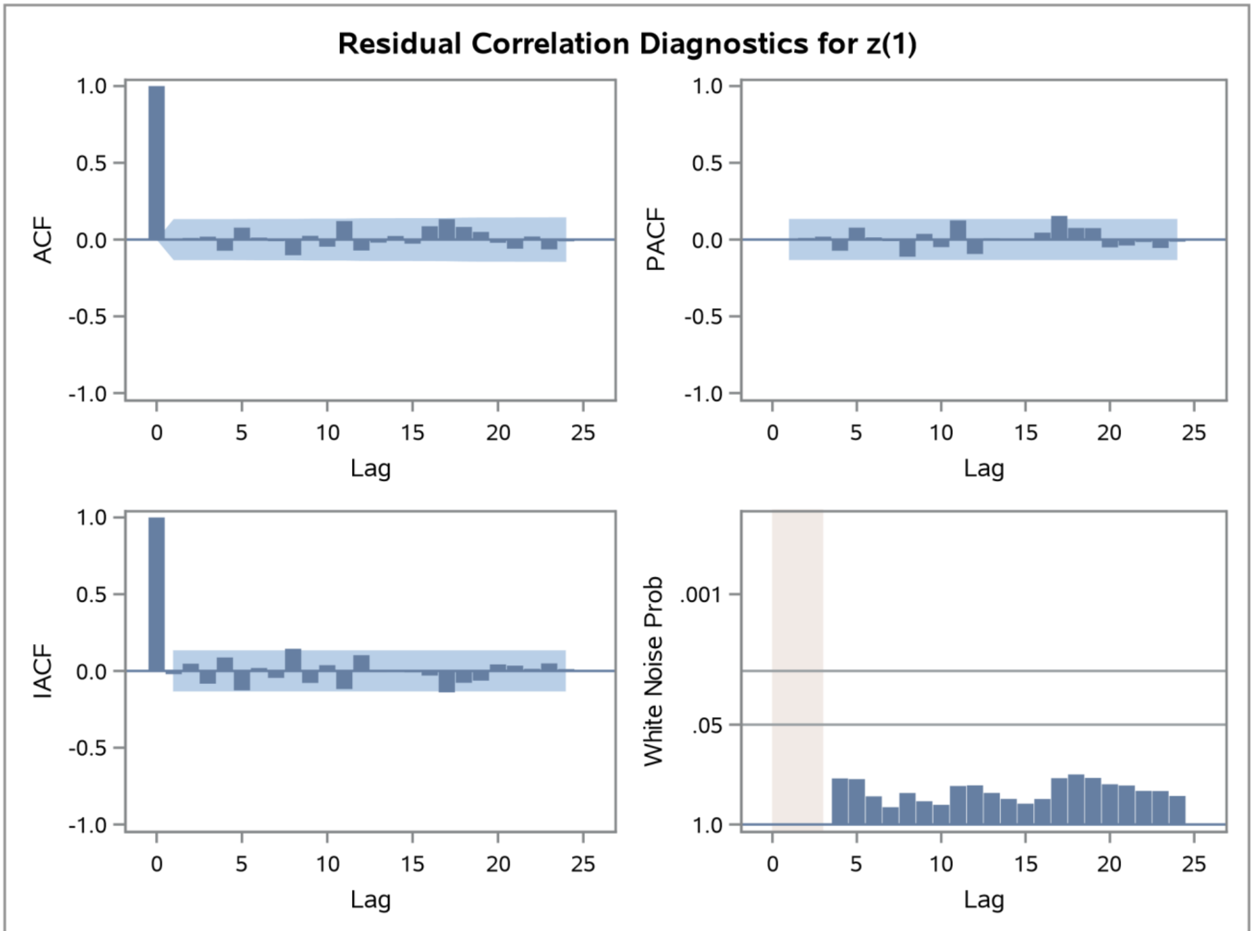
**For the series Y = square_root(X) I chose Arma(1,1,2) model without Constants(NOCONSTANT). I chose this model because it yielded the best AIC out of other qualifying models. It also gave us the best bell shape that was closest to a normal distribution with respect to the residuals. Additionally, the first difference was the best choice, because after second differencing the IACF showed signs of over-differencing by decaying to zero much more slowly than when we differenced once. Furthermore, the ACF and the PACF both converge to zero. The ACF does so exponentially, and the PACF approaches zero in through sinusoidal dampening. This indicates an ARMA(P,Q) process, but we cannot deduce the values of P,Q! Hence, knowing that we will rarely need processes with parameters P,Q greater than 2, we can set an upper bound on their values and try all possible ARMA(P,Q) processes and compare them using the AICs, this process led me to choose an ARMA(1,1,2) model which had the best AIC.**

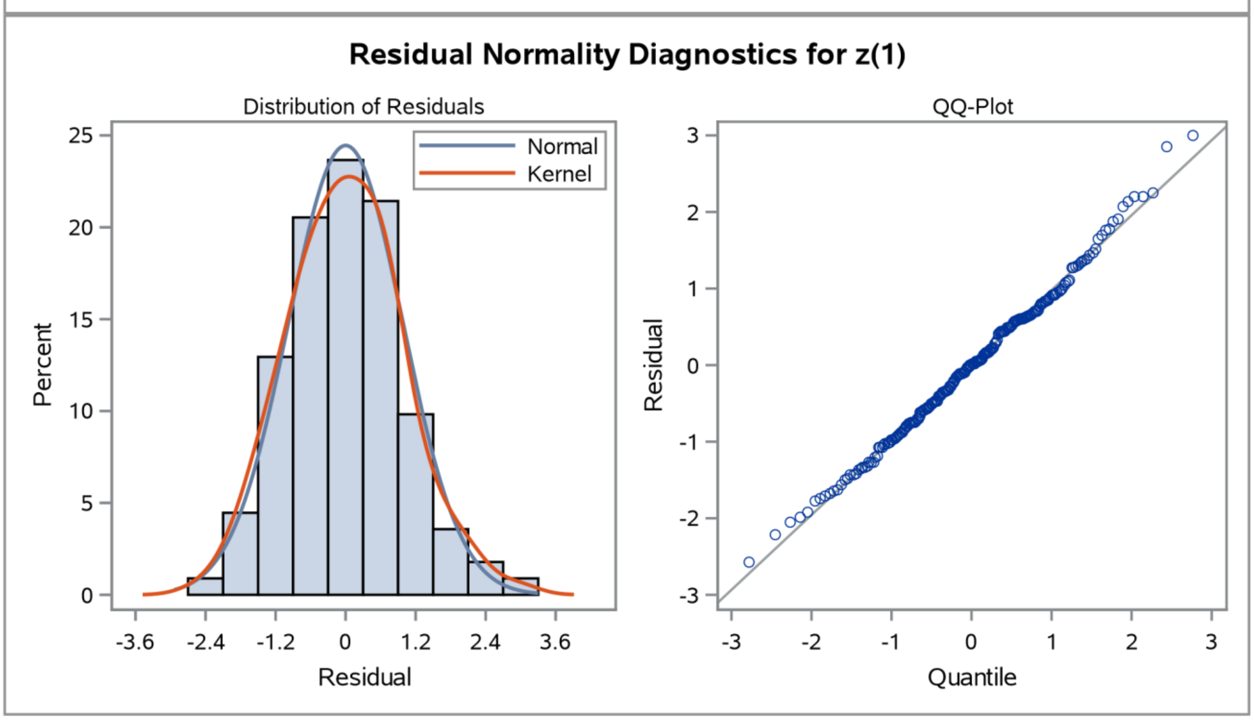
**Furthermore, the residuals of the model are normally distributed, and show all the signs of an excellent model.**

# ACF, PACF, IACF, Time Series Plot After Transformation and Differencing



# Residual Diagnostics



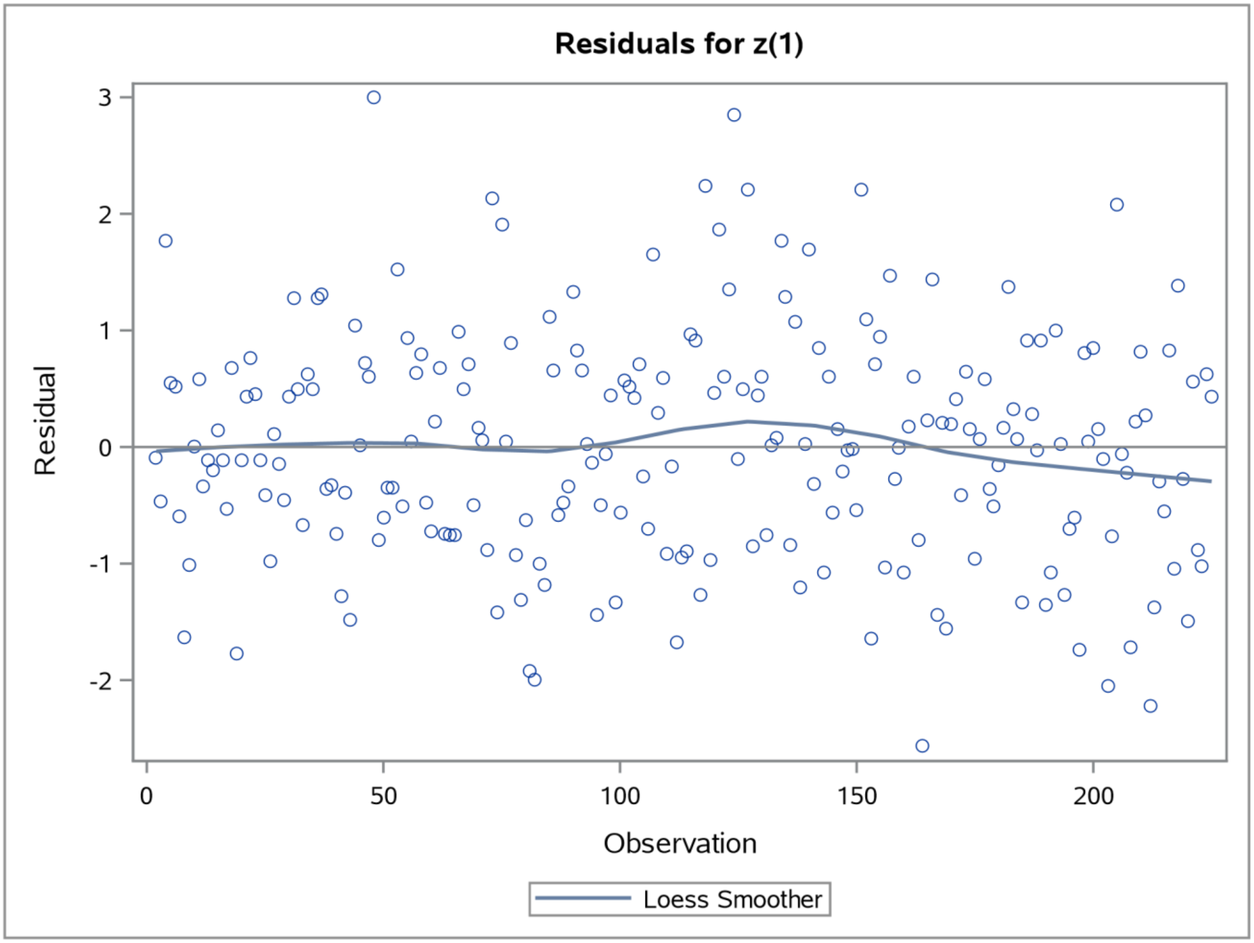
| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | -0.01262 | 0.24325 | -0.05 | 0.9586 | 0 |
| MA1,1 | 0.71394 | 0.07840 | 9.11 | <.0001 | 1 |
| MA1,2 | -0.27868 | 0.06943 | -4.01 | <.0001 | 2 |
| AR1,1 | 0.85089 | 0.05336 | 15.95 | <.0001 | 1 |

| | |
|---|---|
| Constant Estimate | -0.00188 |
| Variance Estimate | 0.97121 |
| Std Error Estimate | 0.9855 |
| AIC | 633.8866 |
| SBC | 647.5332 |
| Number of Residuals | 224 |

| Correlations of Parameter Estimates | | | | |
|---|---|---|---|---|
| Parameter | MU | MA1,1 | MA1,2 | AR1,1 |
| MU | 1.000 | -0.021 | -0.014 | -0.031 |
| MA1,1 | -0.021 | 1.000 | -0.236 | 0.562 |
| MA1,2 | -0.014 | -0.236 | 1.000 | 0.355 |
| AR1,1 | -0.031 | 0.562 | 0.355 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 2.76 | 3 | 0.4298 | -0.001 | 0.011 | 0.019 | -0.073 | 0.078 | 0.013 |
| 12 | 10.53 | 9 | 0.3092 | -0.010 | -0.102 | 0.024 | -0.046 | 0.121 | -0.071 |
| 18 | 18.81 | 15 | 0.2224 | -0.019 | 0.024 | -0.026 | 0.087 | 0.134 | 0.083 |
| 24 | 21.56 | 21 | 0.4250 | 0.050 | -0.021 | -0.059 | 0.020 | -0.064 | -0.012 |
| 30 | 24.07 | 27 | 0.6262 | 0.001 | -0.017 | -0.010 | -0.018 | -0.080 | 0.050 |
| 36 | 31.08 | 33 | 0.5629 | 0.103 | 0.077 | 0.071 | -0.049 | -0.050 | -0.006 |
| 42 | 39.17 | 39 | 0.4621 | -0.044 | -0.050 | -0.008 | -0.136 | -0.009 | -0.080 |

# Residual Time Series Plot




| Model for variable z | |
|---|---|
| **Estimated Mean** | -0.01262 |
| **Period(s) of Differencing** | 1 |

| Autoregressive Factors | |
|---|---|
| **Factor 1:** | 1 - 0.85089 B**(1) |

| Moving Average Factors | |
|---|---|
| **Factor 1:** | 1 - 0.71394 B**(1) + 0.27868 B**(2) |

# Residual vs Predicted Values