Dataset 1: Executive Report

The goal of this project is to fit a model to a simulated data set, and to use the model to make forecasts for the next 13 values; however, after reviewing the time plot of the data and the time plot of dataset 3 differenced once at lag 1, it became clear that an exact linear relationship existed. If the value of each point in dataset 1 is decreased by 75, then this shifted series is identical to the differenced dataset 3 from times 49 to 549. Figure 1.1 shows dataset 1 fitting perfectly over the differenced dataset 3.
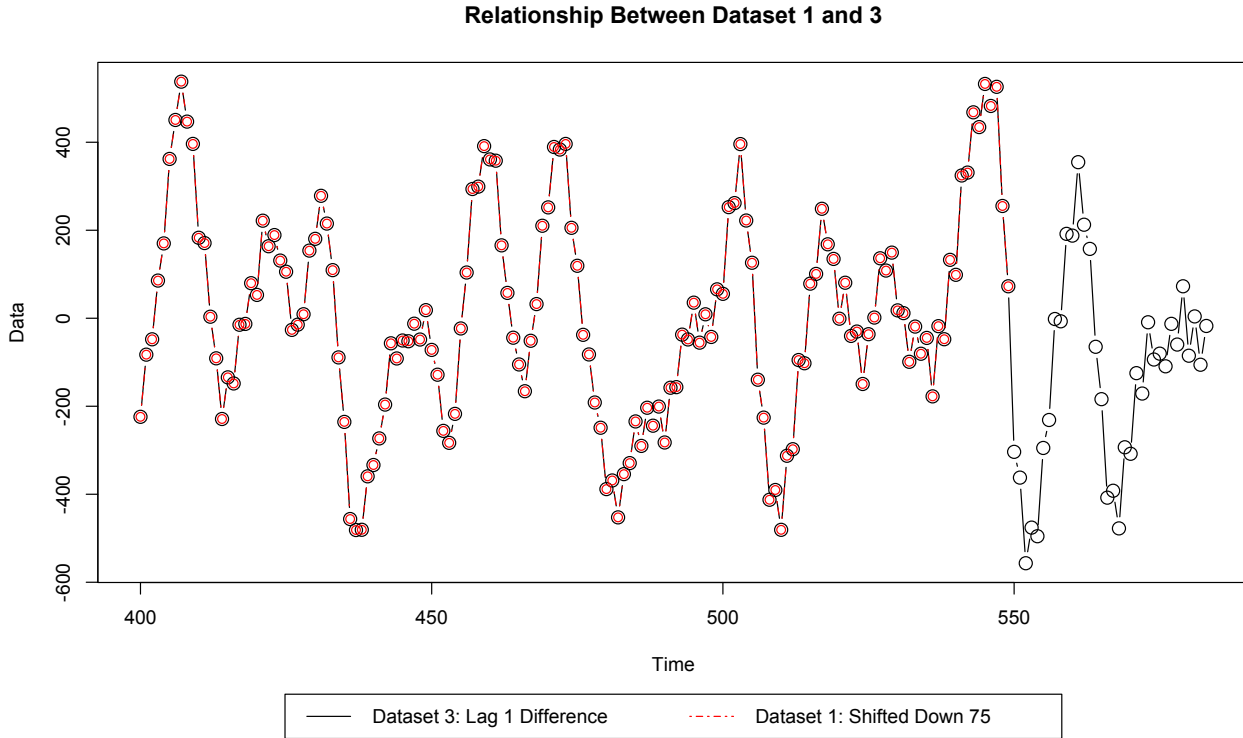

Relationship Between Dataset 1 and 3

Figure 1.1: Plot of the relationship between Dataset 1 and Dataset 3. For display purposes not all points are plotted.

Furthermore, dataset 3 differenced once at lag 1 has data from time 550 to 583; therefore, the next 34 observations of dataset 1 are known exactly. Only the first 13 are listed in the Table 1.1. It is worth noting that prediction intervals are not appropriate here because statistical forecasts are not being made.

Time	Prediction	Time	Prediction
502	-228.61953	509	73.26966
503	-287.16381	510	68.03457
504	-481.76411	511	266.79434
505	-400.79769	512	262.80876
506	-420.40200	513	429.74006
507	-219.78013	514	287.57687
508	-156.20658	-	-

Table 1.1: Predictions for the next 13 values of dataset one using the differenced dataset 3 plus 75.

This "method" is only useful up until time 535 for dataset 1. If true statistical forecasts are desired beyond that time, then the following model will provide accurate forecasts:

$$X_t = 18.4762 + 0.7045X_{t-1} + 0.8490X_{t-2} - 0.8502X_{t-3} + Z_t + 0.8953Z_{t-1}$$

Dataset 1: Technical Appendix

Since dataset 1 is essentially a subset of the differenced dataset 3, the ARMA components should be the same between the two models. See Dataset 3: Technical Appendix for more details regarding model selection and evaluation. The R output from fitting an ARMA(3,1) to the demeaned dataset 1 is below.

```
Call:
arima(x = p1.data.demean, order = c(3, 0, 1), include.mean = FALSE, method = "ML")

Coefficients:
         ar1      ar2      ar3      ma1
      0.7045   0.8490  -0.8502   0.8953
s.e.  0.0242   0.0101   0.0241   0.0211

sigma^2 estimated as 2258:  log likelihood = -2648.87,  aic = 5307.73
```

The intercept in the model specification is calculated using the following formula:

$$\bar{X}(1 - \hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3)$$

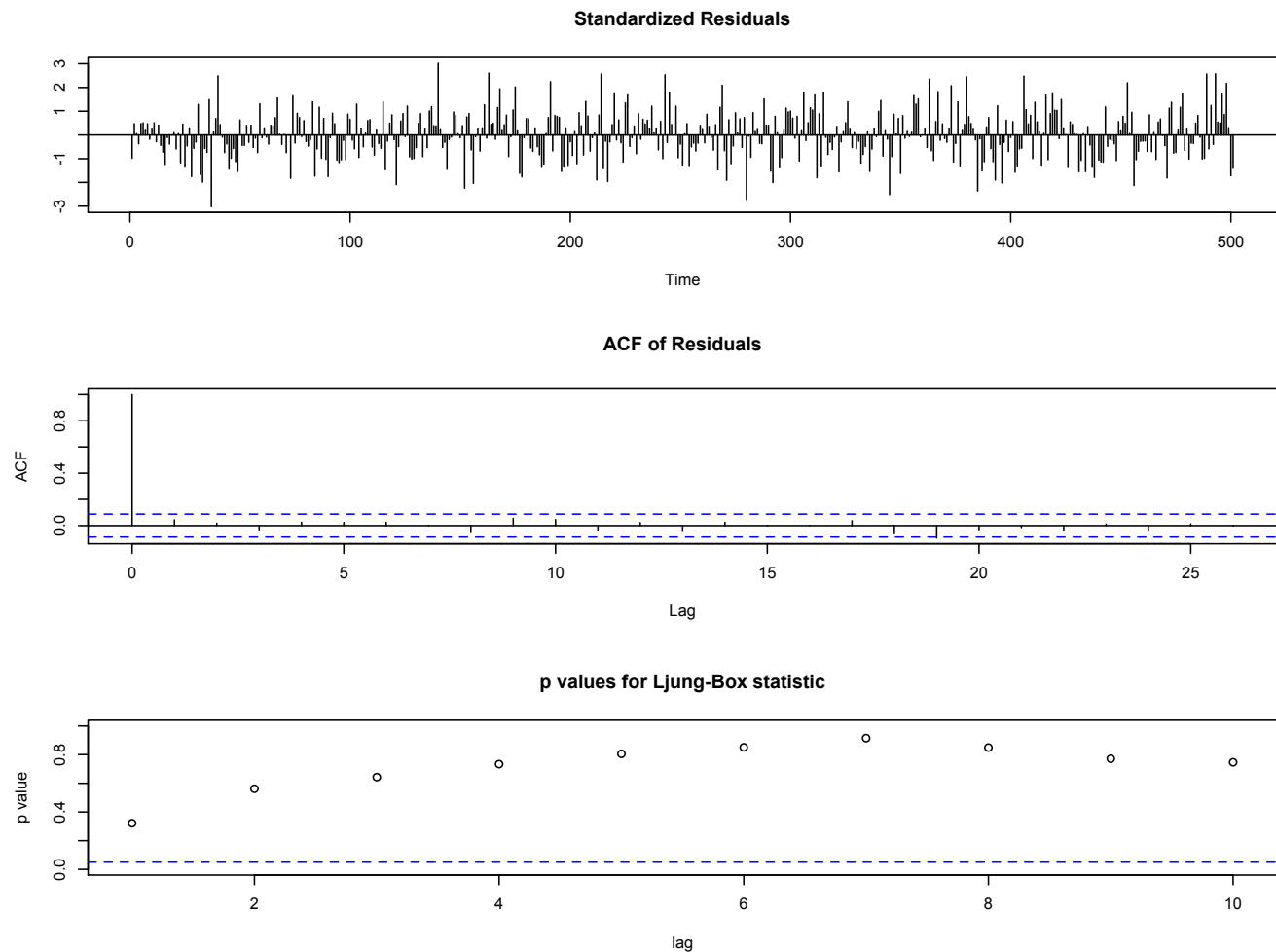The residual diagnostics of the ARMA(3,1) fit suggest that the model is adequate.

Figure 1.2: Residual diagnostics of dataset 1