# HW3 Econometrics 3

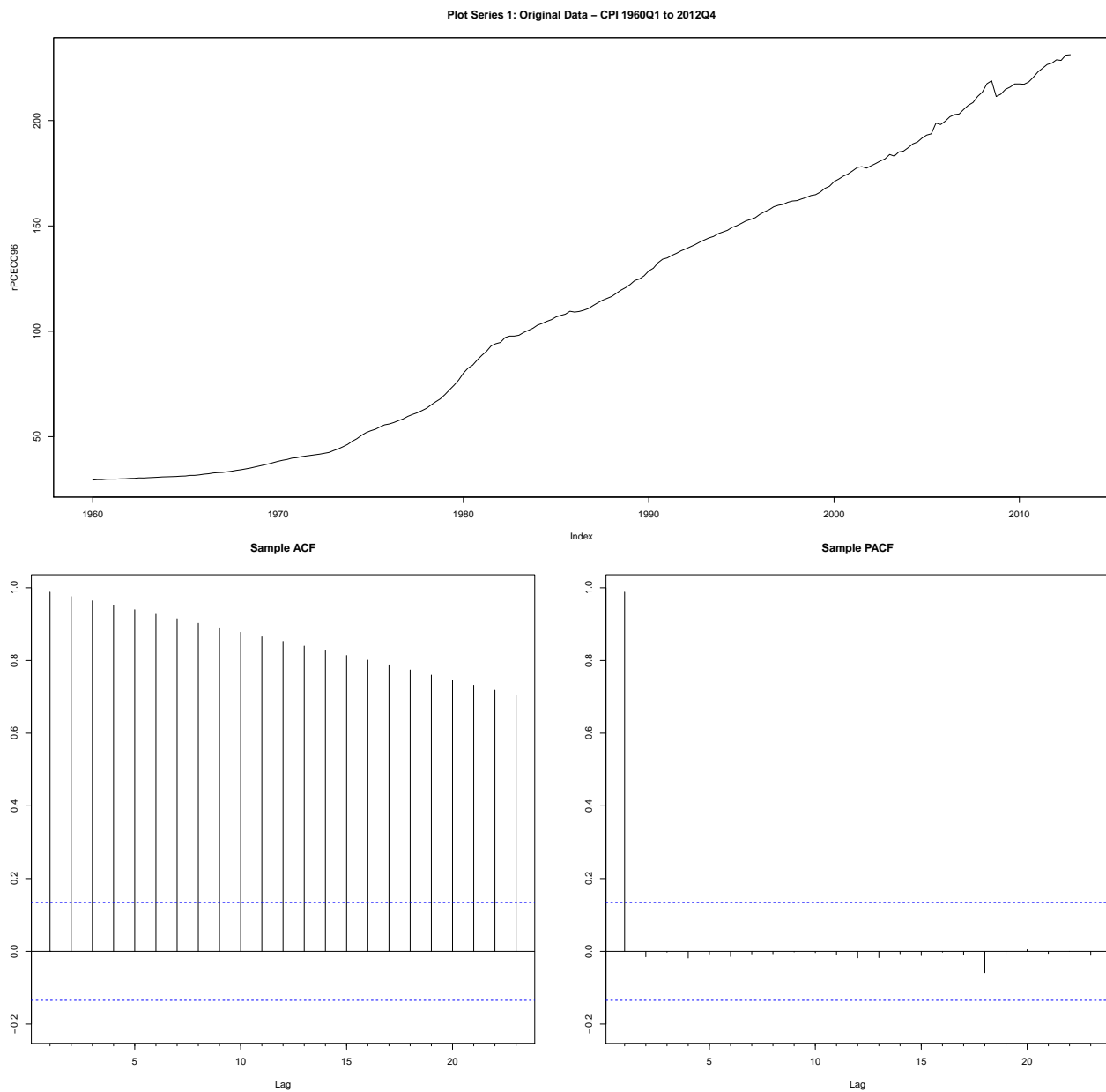*Matthew Aaron Looney*

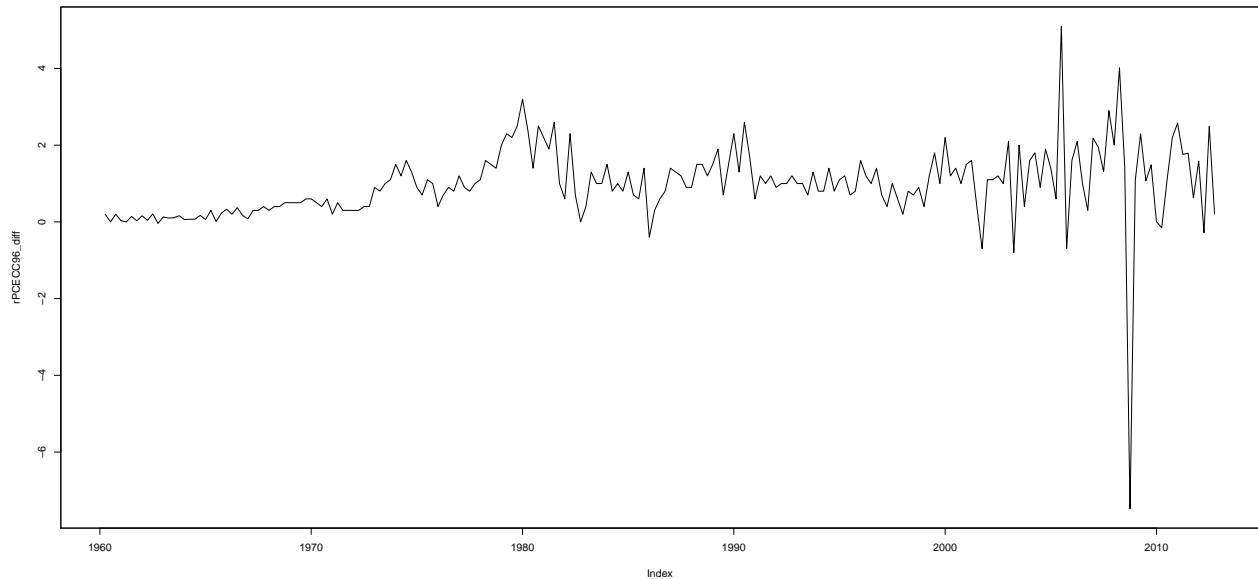*12/4/2017*

## Contents

# 1   1 The file CPIdata.xlsx contains quarterly data on U.S. CPI from 1960Q1 to 2012Q4.

**Plot Series 1: Original Data – CPI 1960Q1 to 2012Q4**



**Sample ACF**



**Sample PACF**

Plot Series 2: ∇Data − CPI 1960Q1 to 2012Q4



**Sample ACF**



**Sample PACF**

Plot Series 3: ∇(∇ Data) − CPI 1960Q1 to 2012Q4



**Sample ACF**
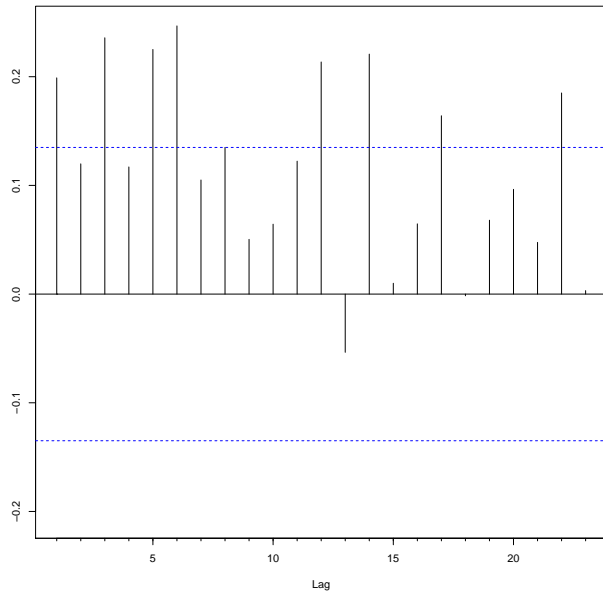
**Sample PACF**
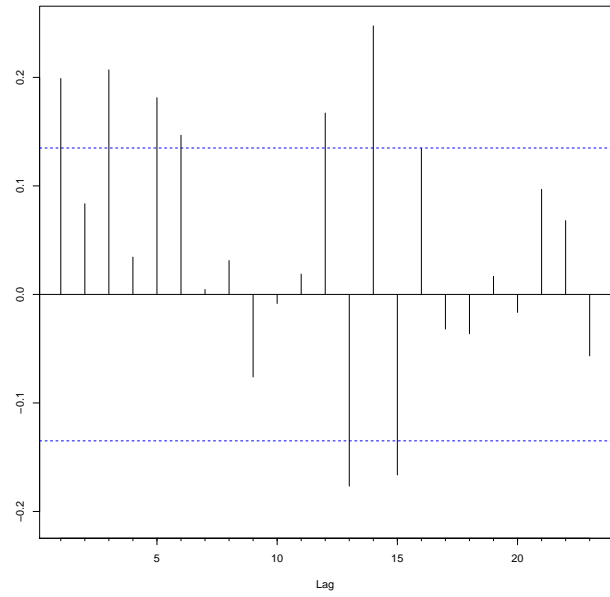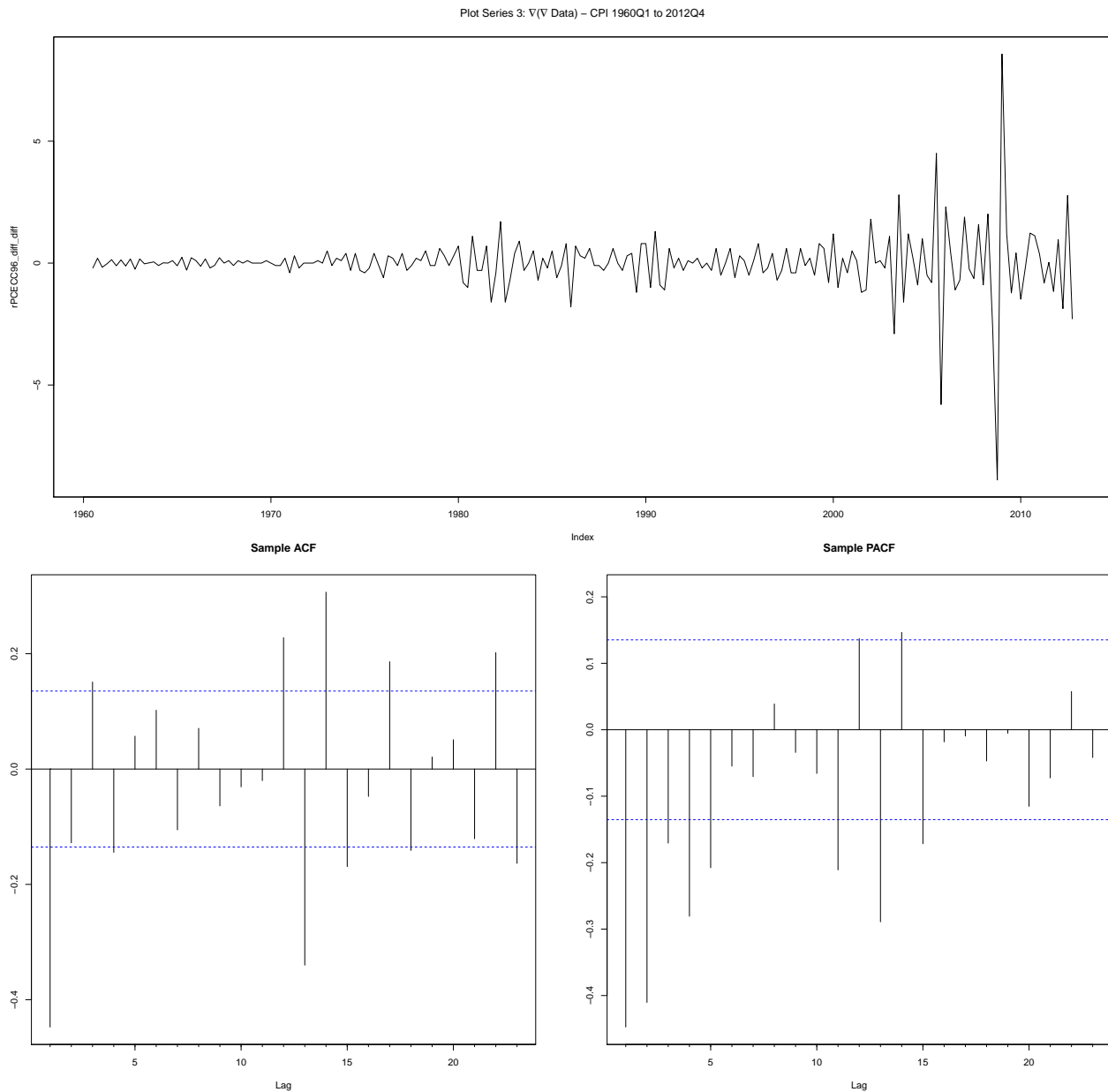


## 1.1 a) Plot the series against the time. Does the series appear to be stationary?

- According to the above Plot Series 1: Original Data, the series is clearly not stationary.

## 1.2 b) Plot and examine the autocorrelation and partial autocorrelation functions (ACF and PACF).

- Acording to the above ACF and PACF plots (Plot Series 1) the series in non-stationary. The ACF decays very slowly over time, indicating the presence of deterministic or stochastic trend.

## 1.3  c) Is the CPI series stationary? Formally test the stationarity of the series.

- When we preform the Augmented Dicky-Fuller test on the original data we obtain a p-value of 0.0635829 which indicates we FAIL to REJECT the null and determine the series is nonstationary.

- When we take the first difference of the series, (Plot Series 2) we see some indications that ther series has become stationary, but it is still possible to see a trend in the data. When we formally test the differenced data with the ADF test we obtain a p-value of 0.020394 which indicates the series is stationary.

- Just out of curiosity I take the second difference of the CPI data and plot the series (Plot Series 3). This time series now apeears to be completely stationary. It is interesting to see the non-constant variance though; which will be addressed in problem 2. When testing with ADF we obtain a p-value of 0.01. Since the p-value indicates stationarity and the plot of the second differenced data look better; I will assume the data is stationary with two differences.

## 1.4  d) Based on questions (b) and (c) propose at least three ARIMA models to study the CPI series and use all data in the sample to estimate them.

- According to the above ACF and PACF plots, (Plot Series 3) there is some ambiguity with identifying the proper ARIMA model to employ to model the time series. The three models I will use in this problem are: ARIMA(4, 2, 1); ARIMA(0, 2, 1); ARIMA(0,2,2).

## 1.5  e) Use AIC and BIC to pick the best of the three models.

|  | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| ar1 | 0.03 | | |
| | (0.10) | | |
| ar2 | −0.09 | | |
| | (0.09) | | |
| ar3 | 0.08 | | |
| | (0.09) | | |
| ar4 | −0.08 | | |
| | (0.08) | | |
| ma1 | −0.90 | −0.91 | −0.88 |
| | (0.07) | (0.03) | (0.07) |
| ma2 | | | −0.03 |
| | | | (0.08) |
| AIC | 580.40 | 576.12 | 577.93 |
| AICc | 580.81 | 576.17 | 578.05 |
| BIC | 600.48 | 582.81 | 587.98 |
| Log Likelihood | -284.20 | -286.06 | -285.97 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Statistical models

- The above summary, (Table 1) shows the estimation output of the three ARIMA models. Model 1= ARIMA(4, 2, 1); model 2= ARIMA(0, 2, 1); model 3= ARIMA(0,2,2).
  When we consider the AIC criterion we conclude that model 2 (ARIMA(0, 2, 1)) has the lowerest AIC and thus is the best model out of the three.

## 1.6 f) Compare the forecasting power of the three models by leaving out the last 50 observations. A graph showing the actual and the forecasted observed may help in this task.

**Model 1: Forecasts for ARIMA(4,2,1)**

**Model 1: Forecasts for ARIMA(0,2,1)**

**Model 1: Forecasts for ARIMA(0,2,2)**
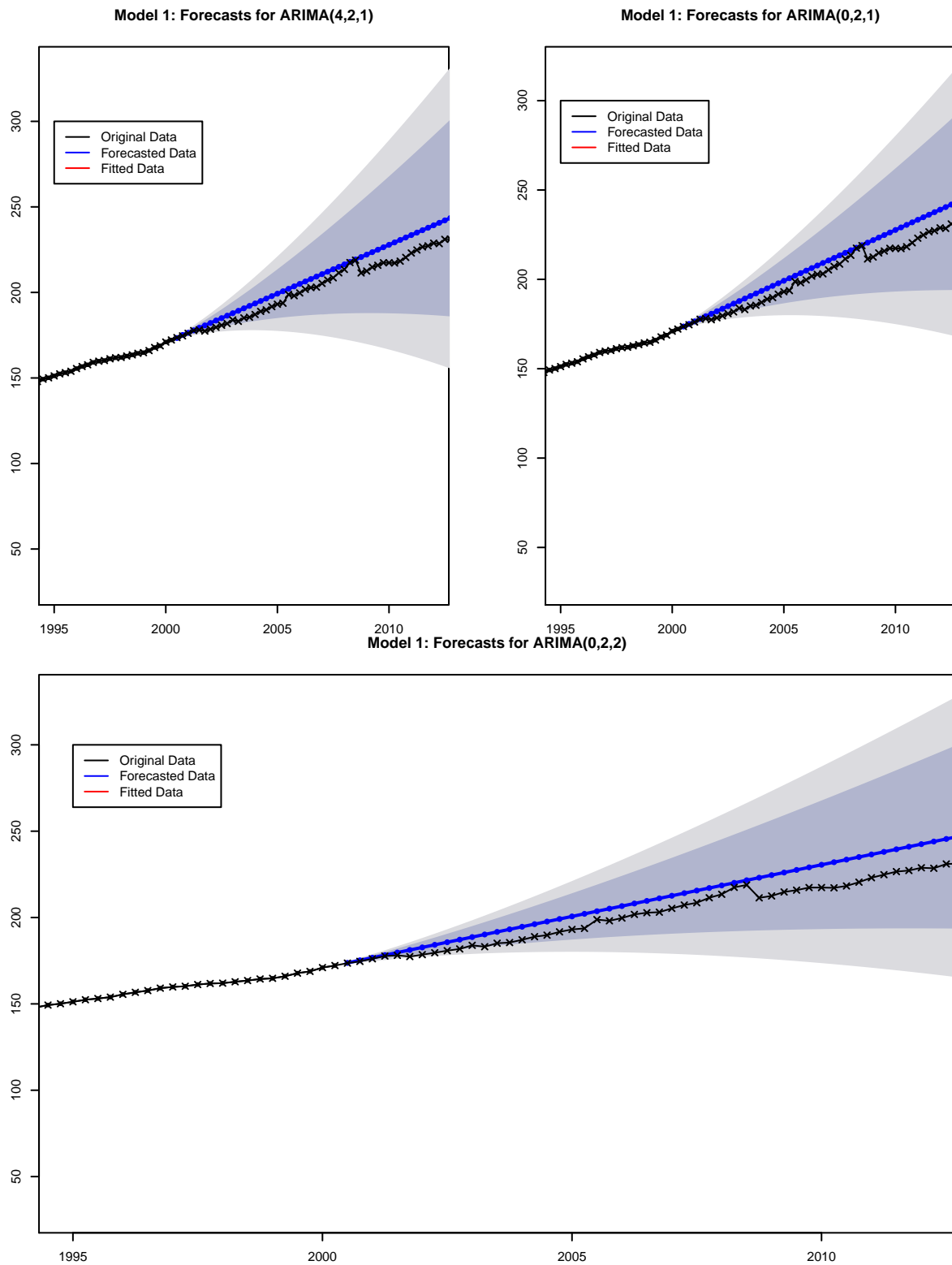
- It is not easy to assertain which model does a better job at forcating based on the above plots. There is a structural break in the data which is not modeled well by any of the models. However, aside from the structural break issue, the remainder of the data is resonably modeled by the forcast up to about four or five quarters. Beyond about five quarters the forcasting ability of the ARIMA(0,2,1) starts to fall apart.

# 2 In the previous problem, we assumed the CPI series has a constant variance. This assumption might not be appropriate.

## 2.1 a) Plot the squared residuals from your best model from problem 1. Can we assume the series is a constant variance one?
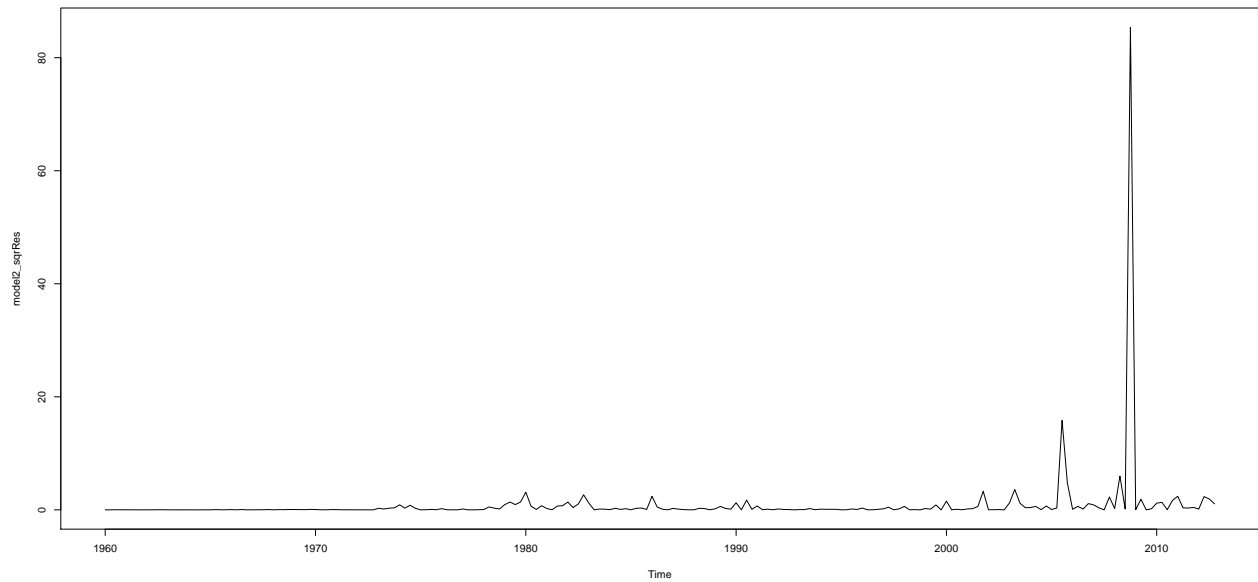


- It is clear by the above plot of the squared residuals of the ARIMA(0,2,1) model that the variance is not constant.

## 2.2 b) Formally test for ARCH or GARCH.

**Series model2$residuals** (ACF plot)

**Series model2$residuals** (Partial ACF plot)

**Series model2$residuals^2** (ACF plot)

**Series model2$residuals^2** (Partial ACF plot)
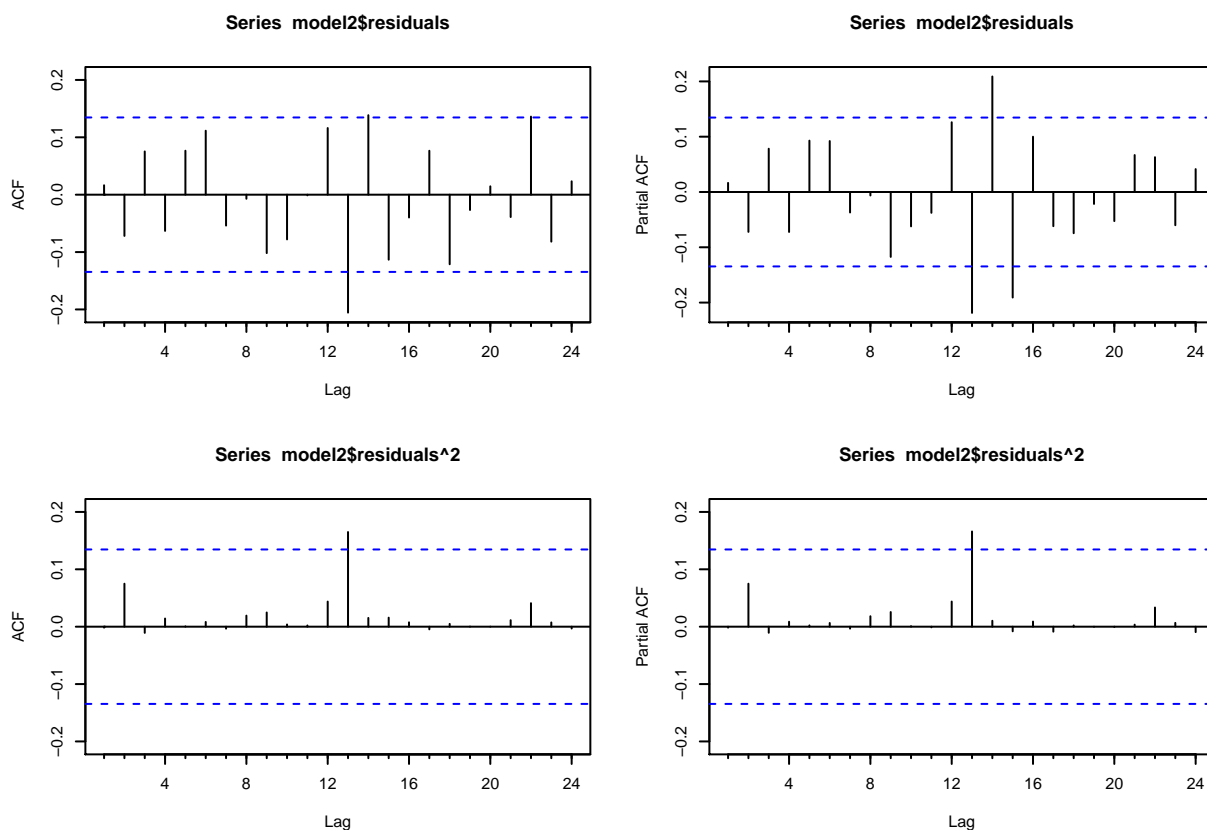
- When we look at the PACF of the Squared Residuals in the above plot we can see that the order of the ARCH model should be (G)ARCH(1).
  We can formally test the time series using the Ljung-Box Test. In the below results we see that the Ljung-Box Test shows we can reject the null hypothesis (independence) on variance, so it has significant serial correlation, in other words ARCH or GARCH effect.

```
##
##  Box-Ljung test
##
## data:  rPCECC96_diff_diff
## X-squared = 42.647, df = 1, p-value = 6.556e-11
```

## 2.3 c) Estimate and ARCH(4) for the CPI series and interpret the results.

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(0, 1) + garch(4, 0), data = coredata(rPCECC96_diff_diff),
##      trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(0, 1) + garch(4, 0)
## <environment: 0x7fbd3495a778>
```

```
## [data = coredata(rPCECC96_diff_diff)]
##
## Conditional Distribution:
## norm
##
## Coefficient(s):
##          mu          ma1        omega       alpha1       alpha2
## 0.00028571  -0.51604255   0.01269720   0.99999999   0.53309719
##      alpha3       alpha4
## 0.13879363   0.14912114
##
## Std. Errors:
## based on Hessian
##
## Error Analysis:
##          Estimate  Std. Error  t value Pr(>|t|)
## mu      0.0002857   0.0081042    0.035   0.9719
## ma1    -0.5160425   0.0663305   -7.780 7.33e-15 ***
## omega   0.0126972   0.0056166    2.261   0.0238 *
## alpha1  1.0000000   0.2124721    4.706 2.52e-06 ***
## alpha2  0.5330972   0.2052481    2.597   0.0094 **
## alpha3  0.1387936   0.0955394    1.453   0.1463
## alpha4  0.1491211   0.0971995    1.534   0.1250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
## -171.5236    normalized: -0.8167791
##
## Description:
## Thu Dec  7 14:47:24 2017 by user:
##
##
## Standardised Residuals Tests:
##                               Statistic p-Value
## Jarque-Bera Test   R   Chi^2  101.4903  0
## Shapiro-Wilk Test  R   W      0.9559571 4.490791e-06
## Ljung-Box Test     R   Q(10)  18.70095  0.04422963
## Ljung-Box Test     R   Q(15)  27.91281  0.02212123
## Ljung-Box Test     R   Q(20)  36.73851  0.01257579
## Ljung-Box Test     R^2 Q(10)  4.701659  0.9101962
## Ljung-Box Test     R^2 Q(15)  19.15944  0.2065997
## Ljung-Box Test     R^2 Q(20)  24.55866  0.2188397
## LM Arch Test       R   TR^2   5.757114  0.9278272
##
## Information Criterion Statistics:
##      AIC       BIC       SIC      HQIC
## 1.700225 1.811795 1.698097 1.745328
```

- When we estimate the ARIMA(0,2,1)-ARCH(4) model a few things stick out. First, the paramater on the MA(1) portion of the ARMIA(0,2,1) is significant. Second, the Alpha1 and Alpha2 paramater are significant, however, the Alpha3 and Alpha4 paramaters are not. This would indicate that a ARCH(1) or ARCH(2) may be a better model than the ARCH(4). I would say, overall, the ARCH(4) model is not the appropriate model to implement for this time series.

## 2.4 d) Estimate a GARCH(1,1) for the CPI series and interpret the results.

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(0, 1) + garch(1, 1), data = coredata(rPCECC96_diff_diff),
##      cond.dist = "std", trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(0, 1) + garch(1, 1)
## <environment: 0x7fbd34031040>
##  [data = coredata(rPCECC96_diff_diff)]
##
## Conditional Distribution:
##  std
##
## Coefficient(s):
##          mu          ma1        omega       alpha1        beta1
##  0.00028571  -0.63804007   0.00521365   0.87539897   0.55678576
##       shape
##  3.31516632
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##           Estimate  Std. Error  t value Pr(>|t|)
## mu       0.0002857   0.0067865    0.042 0.966419
## ma1     -0.6380401   0.0674053   -9.466  < 2e-16 ***
## omega    0.0052136   0.0042228    1.235 0.216962
## alpha1   0.8753990   0.3482521    2.514 0.011947 *
## beta1    0.5567858   0.0949074    5.867 4.45e-09 ***
## shape    3.3151663   0.9543647    3.474 0.000513 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  -153.5504    normalized:  -0.7311925
##
## Description:
##  Thu Dec  7 14:47:24 2017 by user:
##
##
## Standardised Residuals Tests:
##                                Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  267.9526  0
##  Shapiro-Wilk Test  R    W      0.9308093 2.100331e-08
##  Ljung-Box Test     R    Q(10)  16.90227  0.0765546
##  Ljung-Box Test     R    Q(15)  28.48165  0.01874135
##  Ljung-Box Test     R    Q(20)  35.54714  0.01737651
##  Ljung-Box Test     R^2  Q(10)  3.429973  0.9694227
##  Ljung-Box Test     R^2  Q(15)  22.8459   0.08746684
```

```
##  Ljung-Box Test      R^2  Q(20)   25.13694  0.1962421
##  LM Arch Test        R    TR^2    4.539569  0.9716204
##
## Information Criterion Statistics:
##      AIC      BIC      SIC     HQIC
## 1.519528 1.615159 1.517955 1.558188
```

- The ARIMA(0,2,1)-GARCH(1,1) seems to preform better than the other model variants. In the previous model ARMIA(0,2,1)-ARCH(4), when we examined the QQ-Plot it appeared the errors were not following a nomal distribution. In the ARIMA(0,2,1)-GARCH(1,1) we used a student t-innovation for the conditional distribution. By viewing the QQ-Plot we see that using the student t-innovation has helped with the normality of the errors but according to the Jarque-Bera Test statistic the models errors are still not strictly following a normal distribution. The papramaters on the MA(1), Alpha1 and Beta1 are significant and the Q(10) and Q(20) most all point to a good model fit. Overall, I would say the ARIMA(0,2,1)-GARCH(1,1) is a good model for this time series.

# 3 The number of customer arrivals per unit time at a supermarket checkout counter follows a Poisson distribution. Consider a random sample of the following 10 observations from this Poisson distribution $y_i$ =5,0,1,1,0,3,2,3,4,1. The density of each observation is given by:

$$f(y_i) = \frac{e^{-\theta}\theta^{y_i}}{y_i!}$$

## 3.1 a) What is the likelihood function of this process?

$$L(\theta|y_i) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!} \propto \theta^{\psi} e^{-n\theta}; \text{ where, } \psi = \sum_{i=1}^{n} y_i$$

## 3.2 b) Using the sample, find the maximum likelihood estimate of $\theta$.

- The MLE estimate of $\theta = 2.0999305$.

## 3.3 c) You consider to estimate $\theta$ using Bayesian procedure. To this end, you use Gamma distribution as a prior for $\theta$. Derive the posterior distribution for $\theta$.

- The conjugate prior in our problem is given by the *gamma* distribution which has a pdf proportional to:

$$\theta^{\alpha-1}e^{-\beta\theta}; \text{ for } 0 < \theta < \infty$$
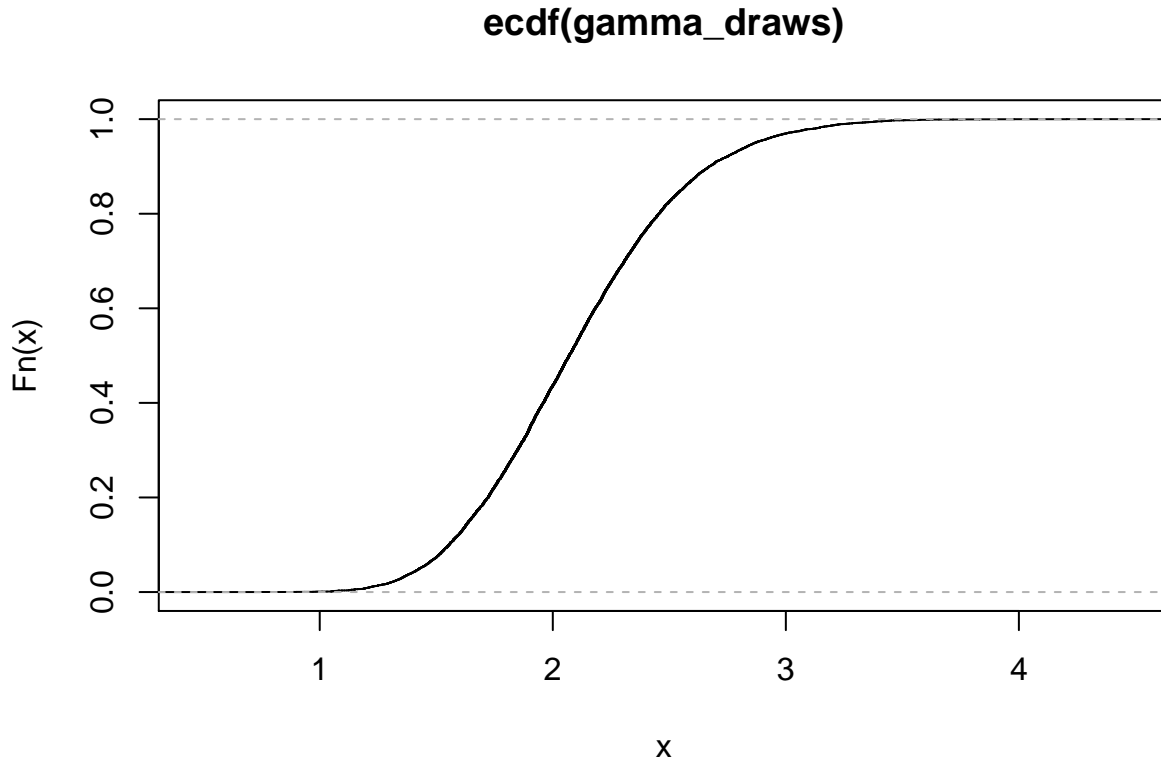
The posterior pdf is proportional to:

$$\theta^{\alpha-1}e^{-\beta\theta}\left(\theta^{\psi}e^{-n\theta}\right) = \theta^{\alpha+\psi-1}e^{-(\beta+n)\theta}$$

The above posterior pdf is proportional to the pdf of a $gamma\left(\alpha + \sum\limits_{i=1}^{n} x_i, \beta + n\right)$

## 3.4 d) Using this posterior distribution, generate 10,000 draws and do the following:

### 3.4.1 i) Plot the cumulative density function (CDF).

**ecdf(gamma_draws)**



### 3.4.2 ii) What is the mean of $\theta$?

- The posterior mean of $\theta = 2.0968314$.

### 3.4.3 iii) What is the variance of $\theta$?

- The posterior variance of $\theta = 0.1932839$.

### 3.4.4 iv) Construct a confidence interval for $\theta$.

- The 95% confidence interval for $\theta = ($ 1.2351515, 2.9585113 $)$.

### 3.4.5 v) What is the probability that we observe less than 4 customers per minute?

- The probability that we observe less than 4 customers per minute is 99.99.