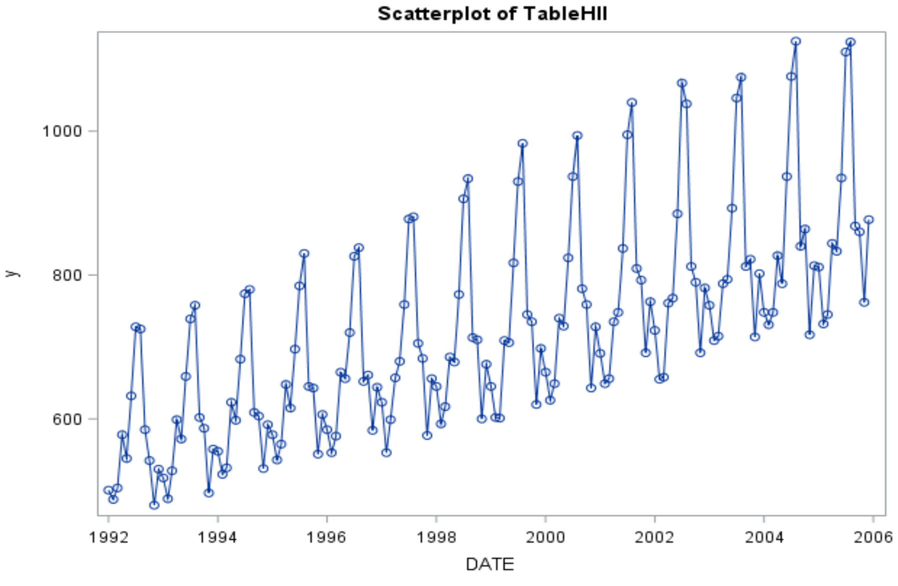
Stockholm University
Martin Ottosson 840413-0455, Herman Niclazon 951027-9434

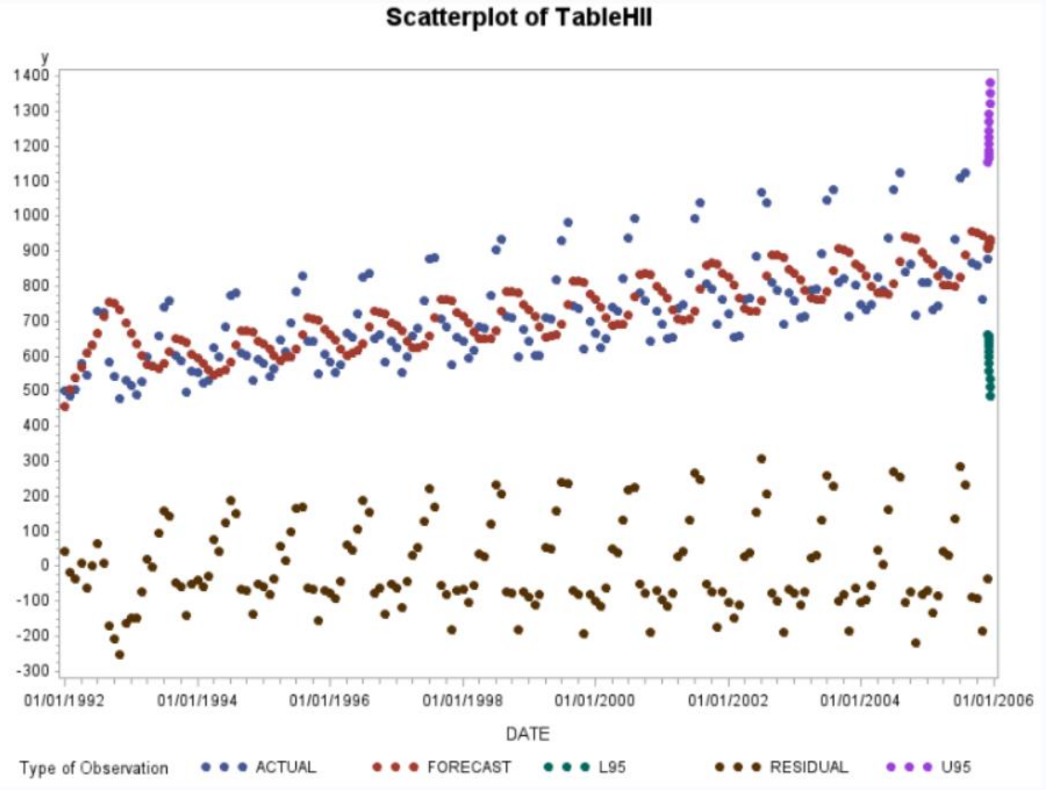
# Econometrics II

# Compulsory home assignment

**Problem 1.**




The process seems to be constant, but with a positive trend, so there is also a seasonal pattern present, and the variable $y_t$ seems to be peaking during the summer (July-August). When there is a seasonal pattern in the process, second-order exponential smoothing is not sufficient. Instead it is appropriate to use the smoothing technique introduced by Holt and Winters, where the linear trend model also has a seasonal adjustment. According to Montgomery, Jennings, Kulachi (2015), the recommendation is to choose 0.2 for all our smoothing constants. The seasonally adjusted smoothed values of $y_t$ can be seen below as the red dots.



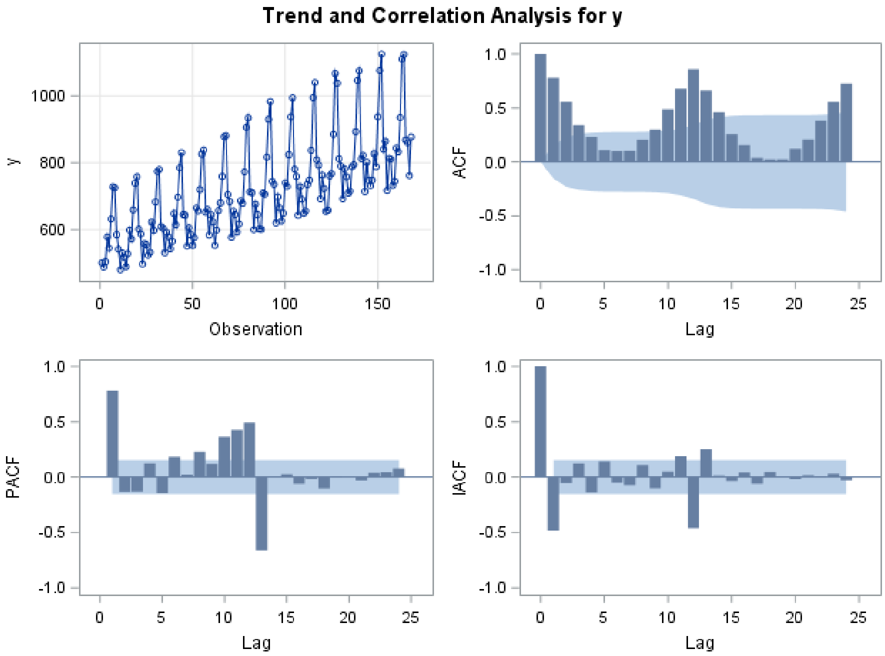
Since the amplitude of the process increases with time, giving a bigger range during the year, we have chosen the multiplicative seasonal model:

$$y_t = L_t S_t + \varepsilon_t$$

where the linear trend part is $L_t = \beta_0 + \beta_1 t$ and $S_t$ is the seasonal component.

Now we have an estimate of the signal, which is the intrinsic pattern of the real process, and the part of the process being noise has thus been separated. The forecast for this process is a continuing increasing trend with seasonal variation. Since exponential smoothing is a "rule of thumb"-technique, the accuracy may not be as high as with for example an ARIMA-model, but it is easier to use. Also the value of our smoothing constants is also not decided by using optimization techniques, but rather chosen by a common standard.

**Problem 2-3.**




Looking at the ACF, it seems as if lag nr 1-4, 10-15 and 24-25 are significant. This indicates that the process has 12 MA-terms. Similarly looking at the PACF, the significant lags seem to be number 1, 6, 8 and 10-12, that is 6 AR-terms. So far, indicating an ARMA(6,12)-process.

**Augmented Dickey-Fuller Unit Root Tests**

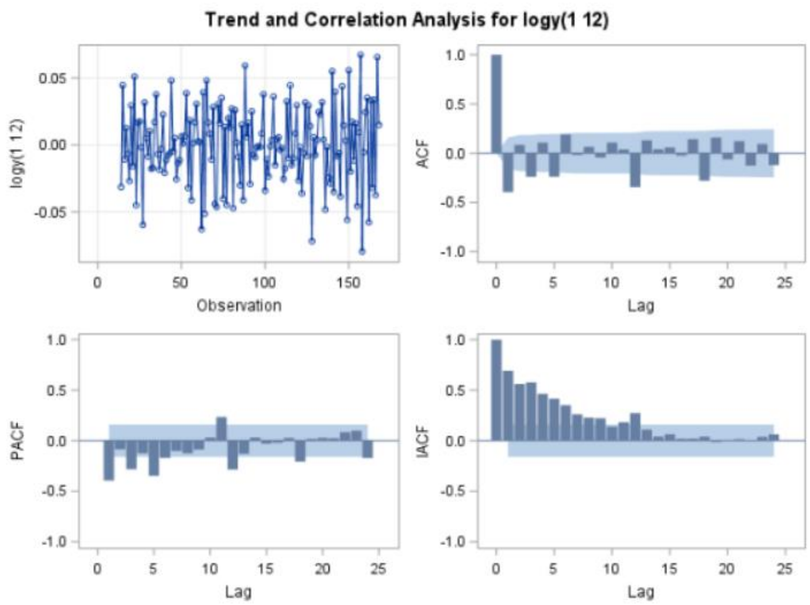
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| Zero Mean | 0 | -0.8276 | 0.5027 | -0.51 | 0.4931 | | |
| | 1 | -0.8692 | 0.4951 | -0.52 | 0.4885 | | |
| | 2 | -0.8371 | 0.5010 | -0.51 | 0.4940 | | |
| Single Mean | 0 | -35.5646 | 0.0013 | -4.48 | 0.0004 | 10.08 | 0.0010 |
| | 1 | -47.1766 | 0.0013 | -4.88 | 0.0001 | 11.98 | 0.0010 |
| | 2 | -62.2685 | 0.0013 | -5.13 | <.0001 | 13.22 | 0.0010 |
| Trend | 0 | -64.1224 | 0.0005 | -6.25 | <.0001 | 19.54 | 0.0010 |
| | 1 | -109.834 | 0.0001 | -7.33 | <.0001 | 26.85 | 0.0010 |
| | 2 | -286.133 | 0.0001 | -8.62 | <.0001 | 37.20 | 0.0010 |

Here we test if the process is stationary, with the Dickey-Fuller test.

$H_0$: The process is not stationary

$H_1$: The process is stationary

The $\tau$-statistic is significant, with very low p-values under 0.0004, for the single-mean model and the trend model. The absolute values of $\tau$ in the single mean, and trend tests are between 4.48 and 6.25 for the respective model with no lag, whereas the critical absolute values are 2.89 and 3.45 respectively at alpha 0.05. However, for the zero mean model the observed absolute value of tau, 0.51, is below the critical value 1.95. Therefore, we conclude that the null hypothesis is rejected and the process $Y_t$ is stationary around a deterministic trend, since there is a clear trend to be seen in the plot of the data. In order to fit an ARIMA-model, transformation is necessary, and because of the seasonality, which seems to have an increasing range as time goes on, we already concluded that we have multiplicative seasonality. The best transformation would then be to first make $y_t$ logarithmic and then use seasonal differencing, in this way we can remove the trend and make the process seasonally stationary.



When taking the first-order difference of $Y_t$, which is period to period change, and then the difference 12 periods ago, we manage to control for the monthly variation over the year. Now there are 6 significant lags in the ACF and 8 significant lags in the PACF. The new DF-table shows that our model has been improved, with reported p-values as low as they can become.

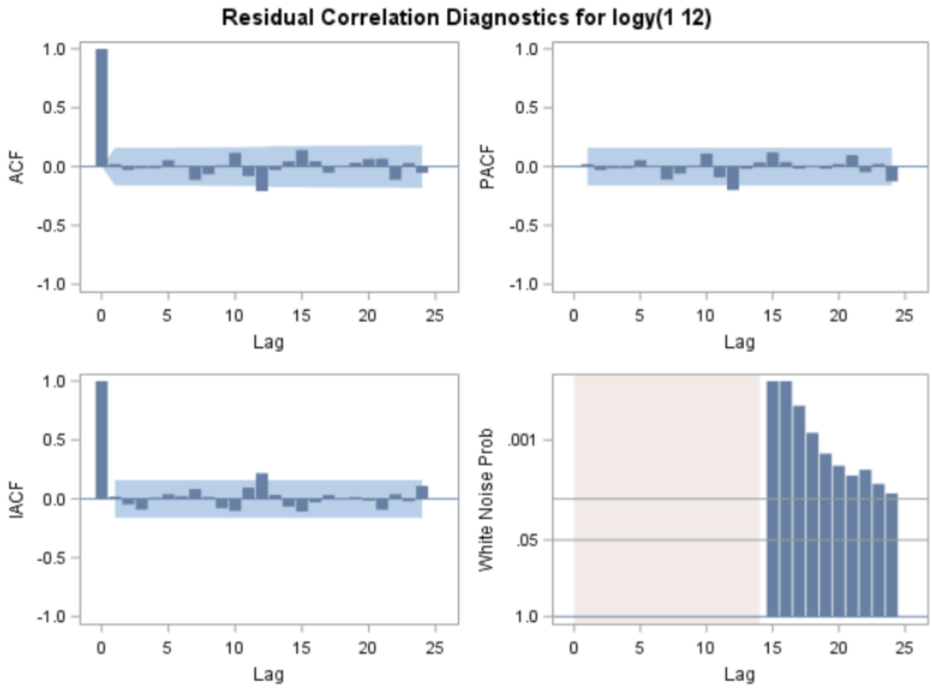
| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Type** | **Lags** | **Rho** | **Pr < Rho** | **Tau** | **Pr < Tau** | **F** | **Pr > F** |
| **Zero Mean** | 0 | -215.063 | 0.0001 | -18.88 | <.0001 | | |
| | 1 | -255.449 | 0.0001 | -11.03 | <.0001 | | |
| | 2 | -1618.81 | 0.0001 | -10.98 | <.0001 | | |
| **Single Mean** | 0 | -215.080 | 0.0001 | -18.82 | <.0001 | 177.18 | 0.0010 |
| | 1 | -255.451 | 0.0001 | -11.00 | <.0001 | 60.48 | 0.0010 |
| | 2 | -1618.67 | 0.0001 | -10.94 | <.0001 | 59.88 | 0.0010 |
| **Trend** | 0 | -215.095 | 0.0001 | -18.76 | <.0001 | 176.03 | 0.0010 |
| | 1 | -255.261 | 0.0001 | -10.96 | <.0001 | 60.14 | 0.0010 |
| | 2 | -1608.93 | 0.0001 | -10.90 | <.0001 | 59.53 | 0.0010 |

We decide to choose this model to represent our process, we call it an $ARIMA\ (8,1,6)x(8,1,6)_{12}$ model.

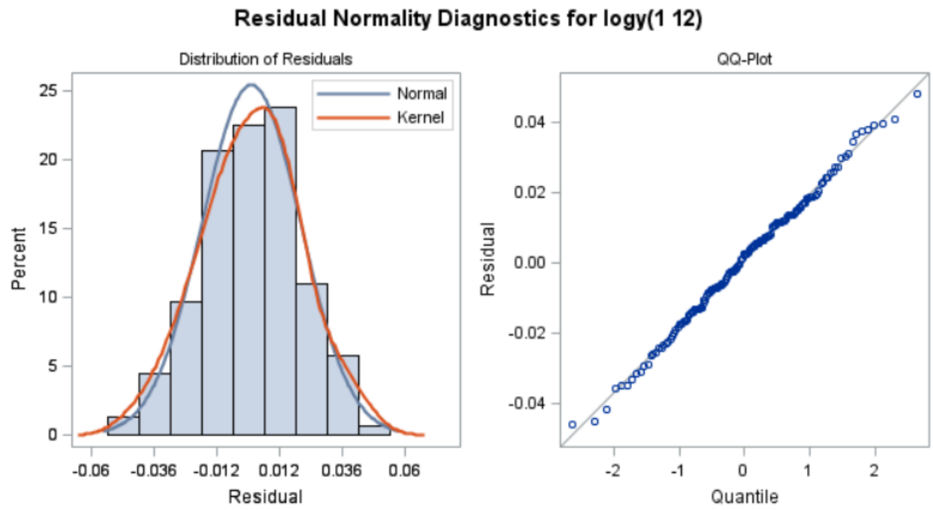
These factors tables show the estimated values of the significant parameters.

| Autoregressive Factors | |
|---|---|
| **Factor 1:** | 1 + 2.66887 B**(1) + 3.68507 B**(2) + 3.22009 B**(3) + 2.10397 B**(4) + 1.70902 B**(5) + 1.65363 B**(6) + 1.31169 B**(7) + 0.52096 B**(8) |

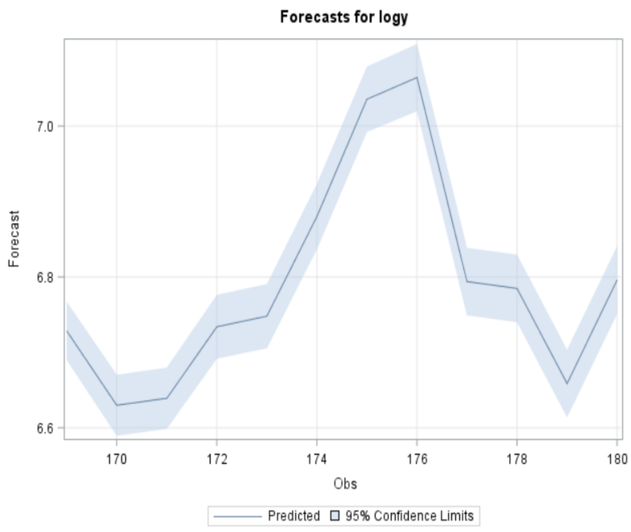
| Moving Average Factors | |
|---|---|
| **Factor 1:** | 1 + 1.97213 B**(1) + 1.62324 B**(2) - 0.28843 B**(3) - 1.79648 B**(4) - 1.53136 B**(5) - 0.66529 B**(6) |




The residual histogram and normal probability plot below indicate normality in our model, meaning that our forecast errors are Gaussian white noise, given uncorrelated observations and constant variance. The white noise probability diagram above also suggest that the estimated residuals are white noise up to the presented 25th lag. If this was not the case, we would need to start over, but since this assumption holds, we can continue and use our $ARIMA\ (8,1,6)x(8,1,6)_{12}$ model for forecasting.



Below is the forecast for 2006 from the ARIMA-model.

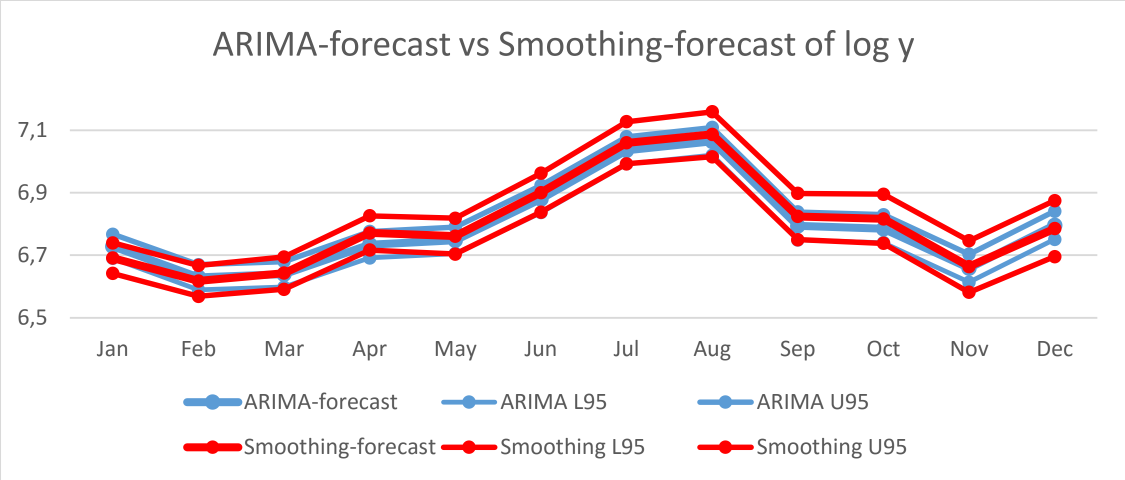



| Forecasts for variable logy | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 169 | 6.7285 | 0.0197 | 6.6898 | 6.7672 |
| 170 | 6.6298 | 0.0206 | 6.5894 | 6.6702 |
| 171 | 6.6391 | 0.0207 | 6.5985 | 6.6797 |
| 172 | 6.7338 | 0.0216 | 6.6916 | 6.7761 |
| 173 | 6.7479 | 0.0217 | 6.7053 | 6.7905 |
| 174 | 6.8800 | 0.0221 | 6.8366 | 6.9233 |
| 175 | 7.0355 | 0.0221 | 6.9922 | 7.0789 |
| 176 | 7.0643 | 0.0227 | 7.0199 | 7.1088 |
| 177 | 6.7938 | 0.0228 | 6.7491 | 6.8385 |
| 178 | 6.7847 | 0.0228 | 6.7400 | 6.8295 |
| 179 | 6.6586 | 0.0228 | 6.6138 | 6.7033 |
| 180 | 6.7959 | 0.0229 | 6.7510 | 6.8408 |

And in the table below are the forecast from the smoothing technique, in logarithmic form for comparative reasons. Later we will also present the results in the original scale.

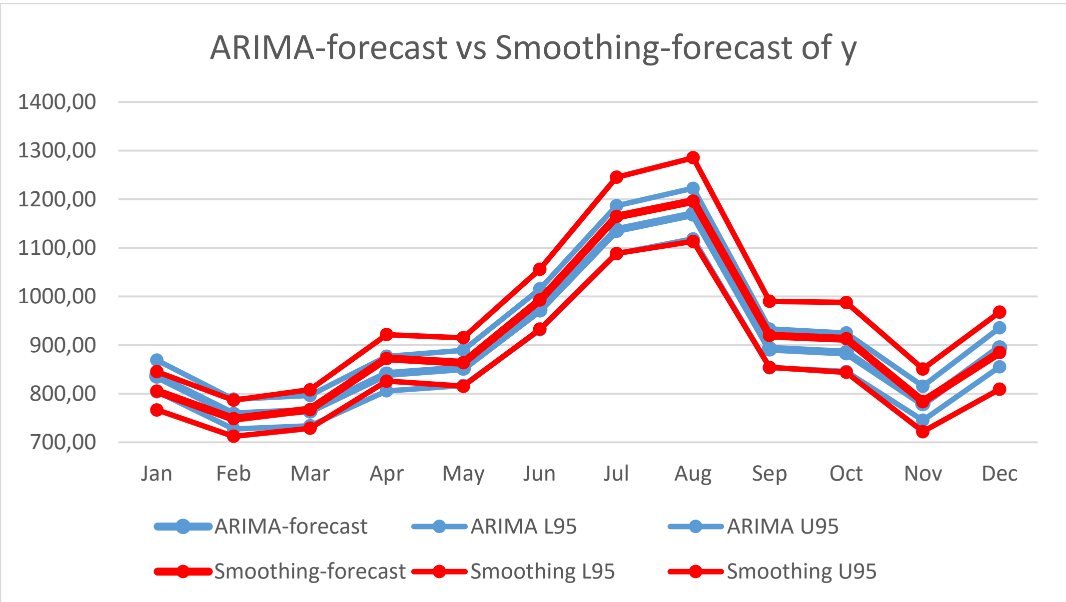
| | | | | |
|---|---|---|---|---|
| 505 | 01/01/2006 | FORECAST | 1 | 6.69087 |
| 506 | 01/01/2006 | L95 | 1 | 6.64208 |
| 507 | 01/01/2006 | U95 | 1 | 6.73965 |
| 508 | 02/01/2006 | FORECAST | 2 | 6.61808 |
| 509 | 02/01/2006 | L95 | 2 | 6.56848 |
| 510 | 02/01/2006 | U95 | 2 | 6.66767 |
| 511 | 03/01/2006 | FORECAST | 3 | 6.64284 |
| 512 | 03/01/2006 | L95 | 3 | 6.59127 |
| 513 | 03/01/2006 | U95 | 3 | 6.69440 |
| 514 | 04/01/2006 | FORECAST | 4 | 6.77130 |
| 515 | 04/01/2006 | L95 | 4 | 6.71645 |
| 516 | 04/01/2006 | U95 | 4 | 6.82615 |
| 517 | 05/01/2006 | FORECAST | 5 | 6.76114 |
| 518 | 05/01/2006 | L95 | 5 | 6.70360 |
| 519 | 05/01/2006 | U95 | 5 | 6.81867 |
| 520 | 06/01/2006 | FORECAST | 6 | 6.90015 |
| 521 | 06/01/2006 | L95 | 6 | 6.83813 |
| 522 | 06/01/2006 | U95 | 6 | 6.96217 |
| 523 | 07/01/2006 | FORECAST | 7 | 7.05977 |
| 524 | 07/01/2006 | L95 | 7 | 6.99245 |
| 525 | 07/01/2006 | U95 | 7 | 7.12709 |
| 526 | 08/01/2006 | FORECAST | 8 | 7.08691 |
| 527 | 08/01/2006 | L95 | 8 | 7.01499 |
| 528 | 08/01/2006 | U95 | 8 | 7.15883 |
| 529 | 09/01/2006 | FORECAST | 9 | 6.82389 |
| 530 | 09/01/2006 | L95 | 9 | 6.75004 |
| 531 | 09/01/2006 | U95 | 9 | 6.89774 |
| 532 | 10/01/2006 | FORECAST | 10 | 6.81665 |
| 533 | 10/01/2006 | L95 | 10 | 6.73788 |
| 534 | 10/01/2006 | U95 | 10 | 6.89541 |
| 535 | 11/01/2006 | FORECAST | 11 | 6.66385 |
| 536 | 11/01/2006 | L95 | 11 | 6.58161 |
| 537 | 11/01/2006 | U95 | 11 | 6.74608 |
| 538 | 12/01/2006 | FORECAST | 12 | 6.78520 |
| 539 | 12/01/2006 | L95 | 12 | 6.69579 |
| 540 | 12/01/2006 | U95 | 12 | 6.87461 |

However, for easier comparison it is easier to look at the graph below including both models below, where the ARIMA-model with 95% confidence limits are in blue and the seasonal smoothing in blue. As can be seen, the ARIMA model with its confidence limits fits within the 95% smoothing limits for most of the year, which is a sign of higher accuracy in the ARIMA-model. The smoothing technique is a less sophisticated model and therefore also has a wider interval with a higher degree of uncertainty. Thanks to the seasonal smoothing and seasonal differencing, both models show a similar pattern over the year, peaking during the summer. The models have therefore picked up the seasonal variation in the original data, which is what we wanted.



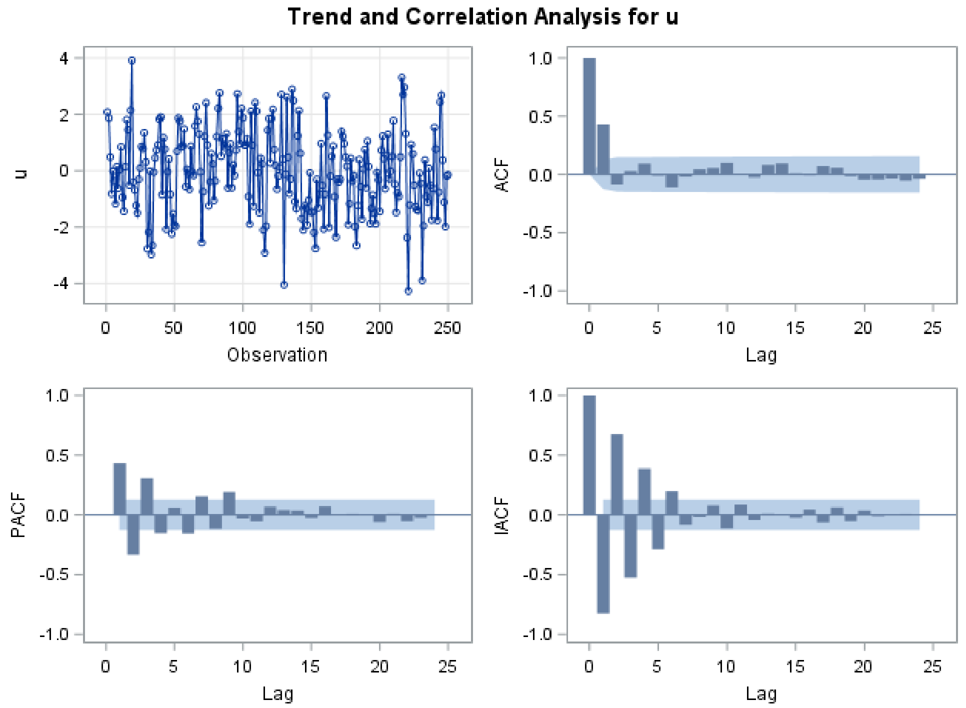

| | ARIMA-forecast | ARIMA L95 | ARIMA U95 | Smoothing-forecast | Smoothing L95 | Smoothing U95 |
|---|---|---|---|---|---|---|
| Jan | 6,73 | 6,69 | 6,77 | 6,69 | 6,64 | 6,74 |
| Feb | 6,63 | 6,59 | 6,67 | 6,62 | 6,57 | 6,67 |
| Mar | 6,64 | 6,60 | 6,68 | 6,64 | 6,59 | 6,69 |
| Apr | 6,73 | 6,69 | 6,78 | 6,77 | 6,72 | 6,83 |
| May | 6,75 | 6,71 | 6,79 | 6,76 | 6,70 | 6,82 |
| Jun | 6,88 | 6,84 | 6,92 | 6,90 | 6,84 | 6,96 |
| Jul | 7,04 | 6,99 | 7,08 | 7,06 | 6,99 | 7,13 |
| Aug | 7,06 | 7,02 | 7,11 | 7,09 | 7,01 | 7,16 |
| Sep | 6,79 | 6,75 | 6,84 | 6,82 | 6,75 | 6,90 |
| Oct | 6,78 | 6,74 | 6,83 | 6,82 | 6,74 | 6,90 |
| Nov | 6,66 | 6,61 | 6,70 | 6,66 | 6,58 | 6,75 |
| Dec | 6,80 | 6,75 | 6,84 | 6,79 | 6,70 | 6,87 |

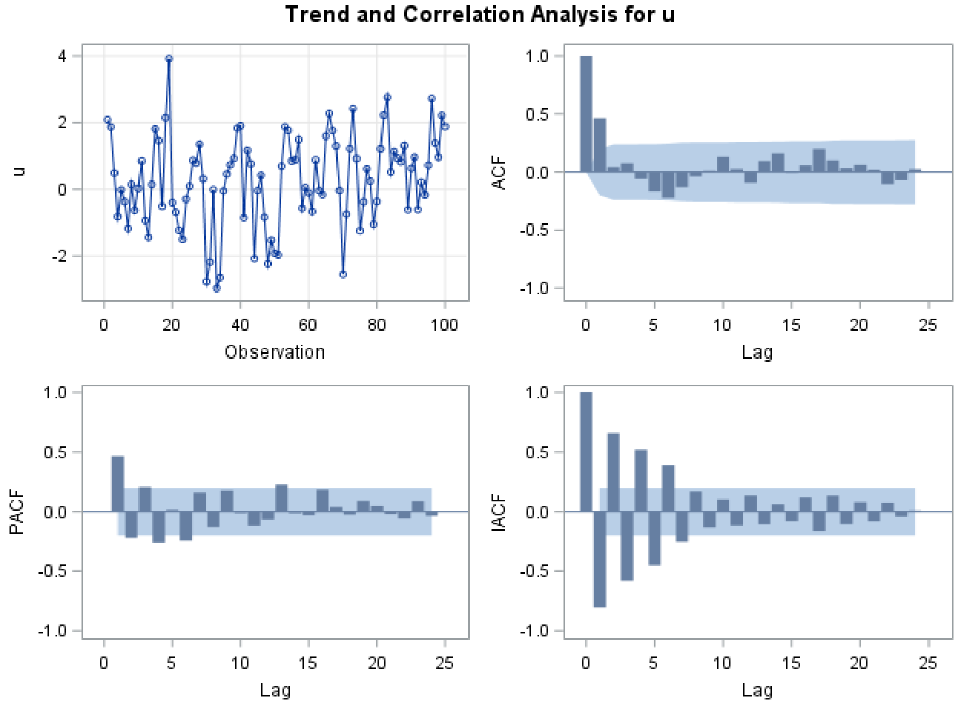
Below is the same graph when taking the antilog of log y, which is the original scale of $y_t$.

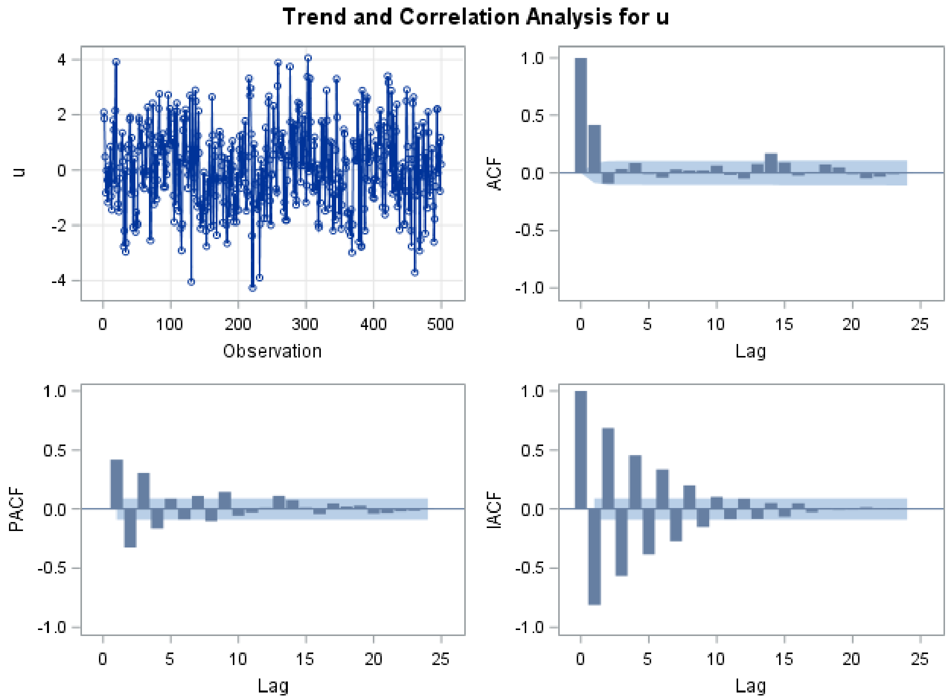


**Problem 4**

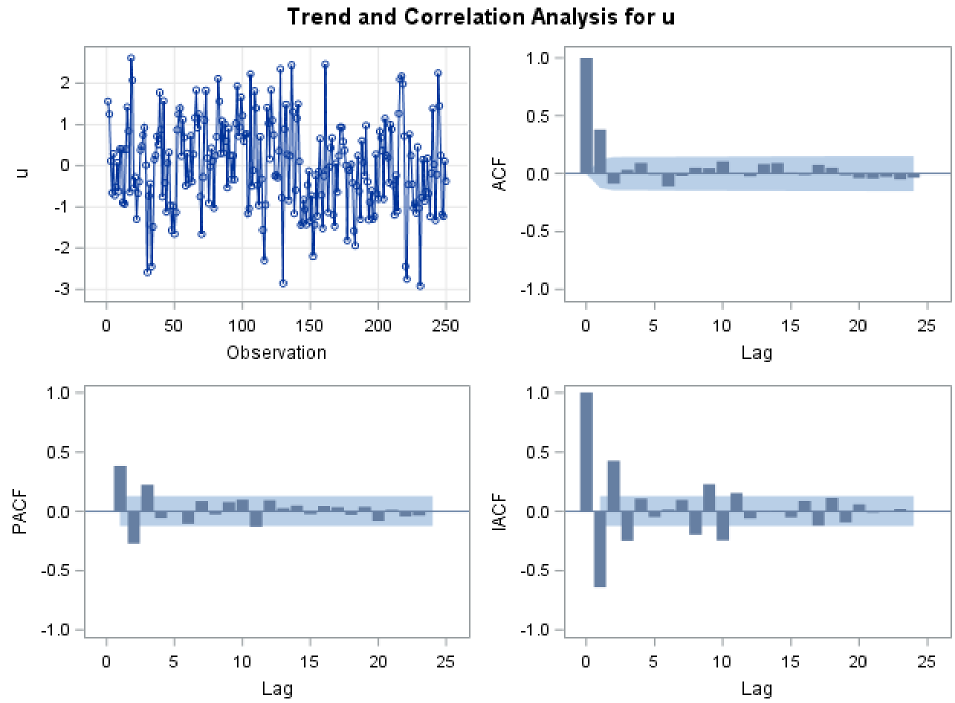
The model used in the first ARIMA graph is $y_t = 501 + 1{,}2\alpha$, which gave us an ARIMA (7,0,2).
We tried to make it somewhat close to our chosen model, to see how it behaved depending on the number of observations. By lowering them from 250 to 100 we still only had 2 MA terms, but we lost one AR term giving an ARIMA (6,0,2). And by increasing them to 500 we get one more AR term and one more MA-term, that is an ARIMA (8,0,3).

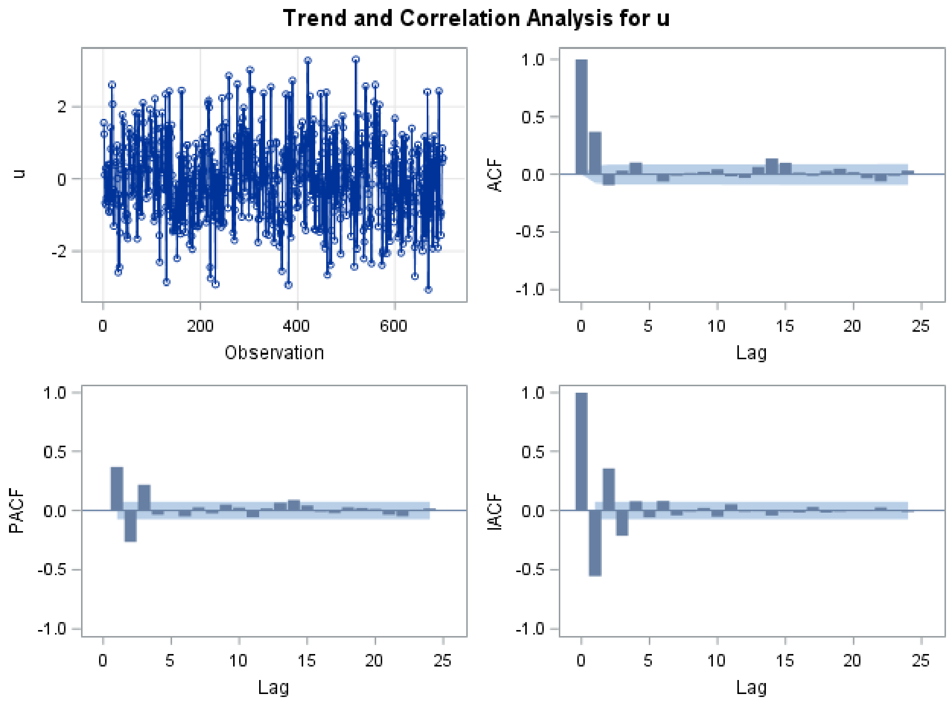





In the second ARIMA graph we used the following model $y_t = 250 + 0{,}6\alpha$, is an ARIMA (4,0,2). This time we just simulated and randomly selected the values, this is because we wan ted to see how much we could change the AR and MA terms by changing the number of observations. We started with n=250 as in our first simulated ARIMA model which gave us ARIMA (4,0,2). We tried to increase the number of observations to 700 to see if (p,d,q) changed, which it did. Now we got ARIMA (4,0,6). So by almost increasing the observations by 200% we got 4 more MA terms but still got the same amount of AR terms. And by lowering the number of observations to 50 we get ARIMA (1,0,2) which is also quite a big change from our original simulation.



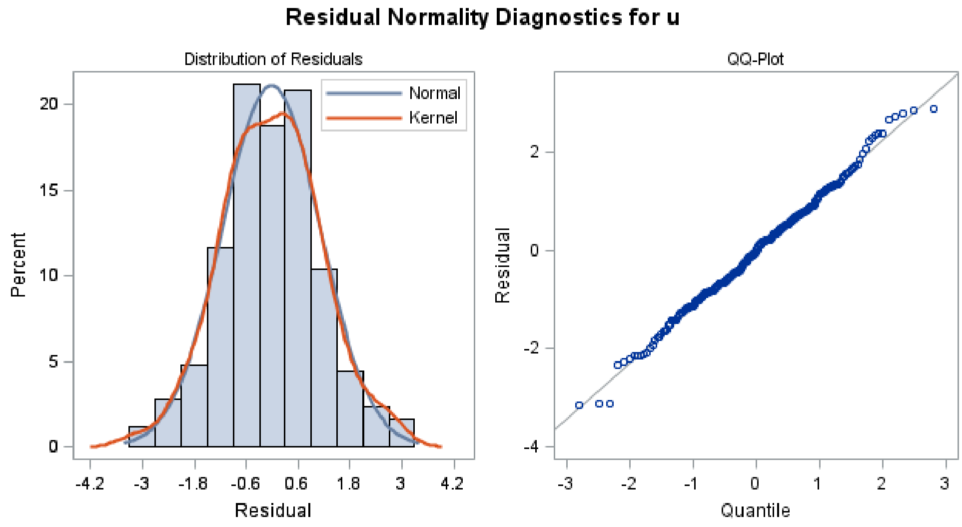


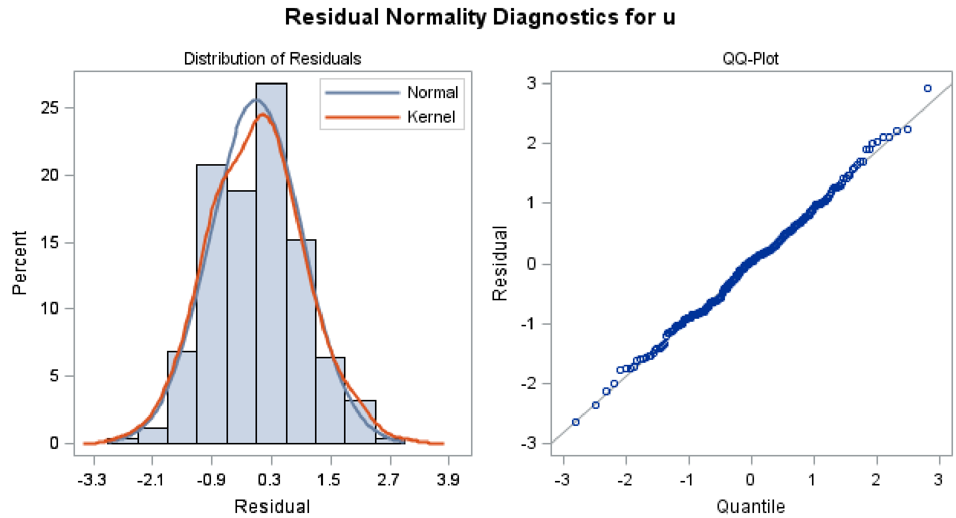


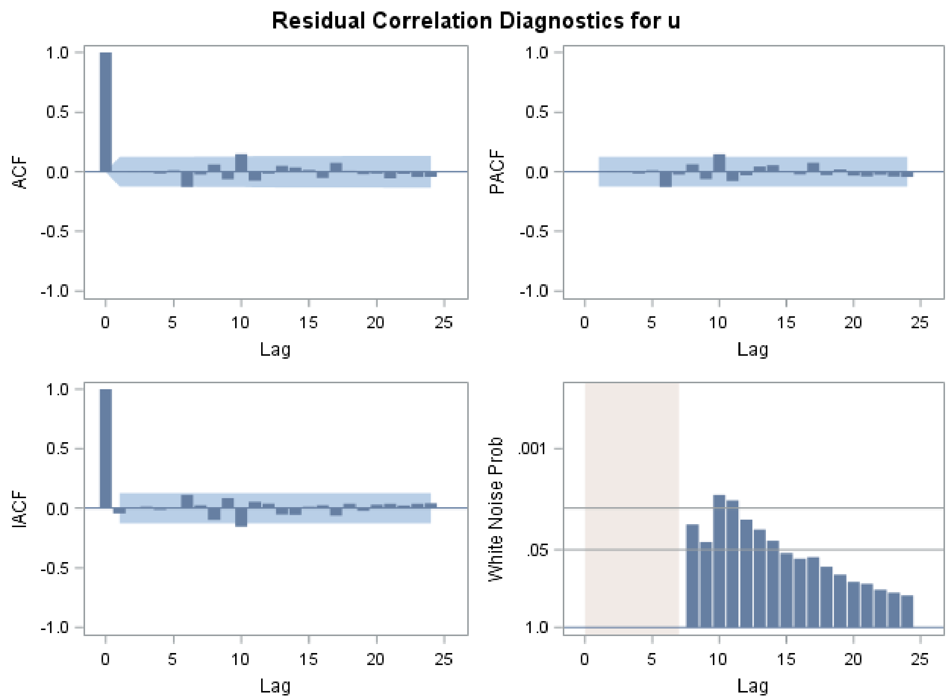
As we can see on the residual diagnostics for both the simulated ARIMA models, they both indicate normality. Which is positive, because it means that the forecast errors are Gaussian white noise. The white noise diagram also helps us to be even more sure that our assumption is correct.

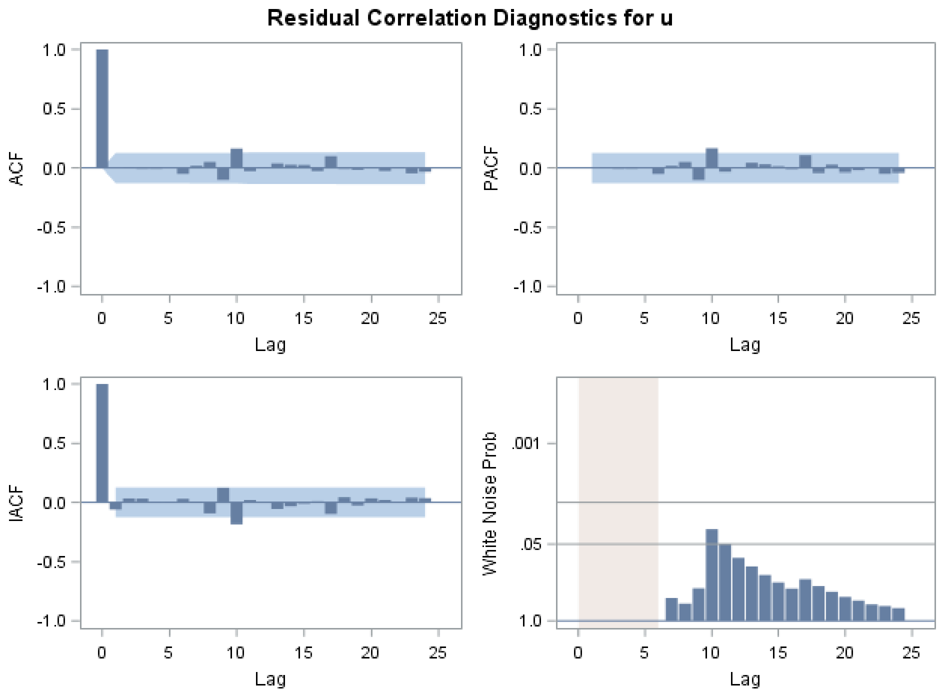
All three ARIMA models (simulated and theoretical) fulfill the necessary assumptions needed for a good model.

Stockholm University
Martin Ottosson 840413-0455, Herman Niclazon 951027-9434




Residual Correlation Diagnostics for u
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25