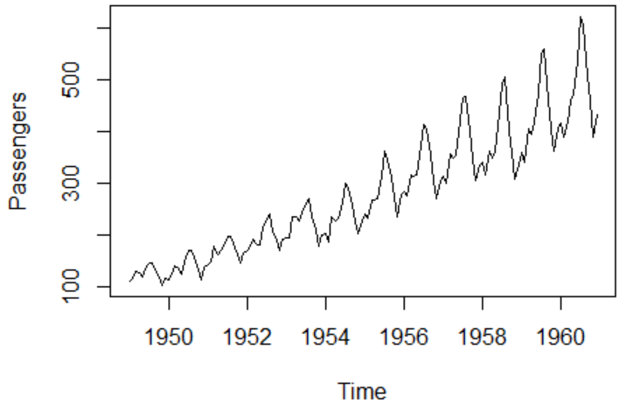
SERIE STORICA

IMPORTAZIONE DATI

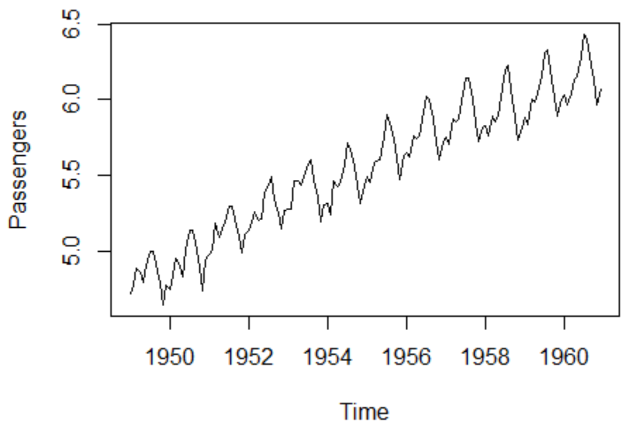
Abbiamo importato un dataset contenente informazioni relative al numero di passeggeri presenti sui voli aerei a partire da gennaio del 1949 fino a dicembre del 1961. L'obiettivo è quello di trovare il processo stocastico che ha generato la serie storica considerata e di farne una previsione.

GRAFICO DELLA SERIE



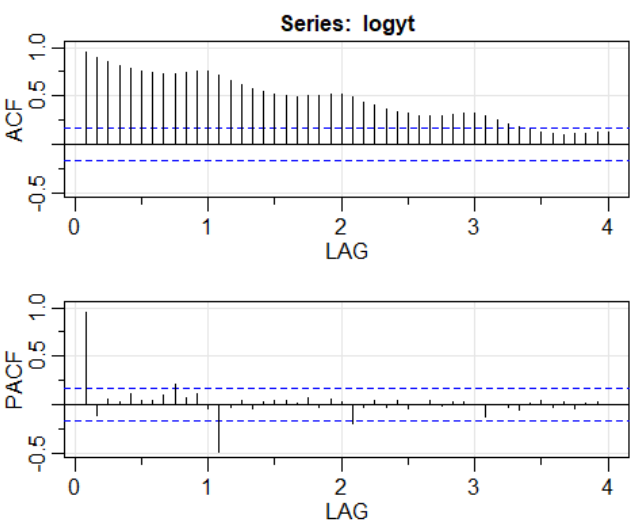
Osserviamo che la varianza non è costante e quindi applichiamo una trasformazione logaritmica per stabilizzarla.

TRASFORMAZIONE LOGARITMICA



La varianza ora è stabile.

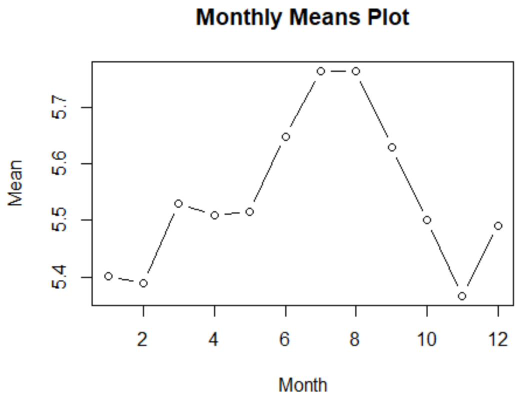
ACF E PACF



Componente stagionale: ipotizziamo AR(0) e MA(3).
Componente non stagionale: ipotizziamo AR(1) mentre, per quanto riguarda la componente a media mobile, vediamo che tutte le autocorrelazioni sono molto elevate.
Procediamo lavorando prima sulla componente stagionale.

COMPONENTE STAGIONALE

Controllo stagionalità

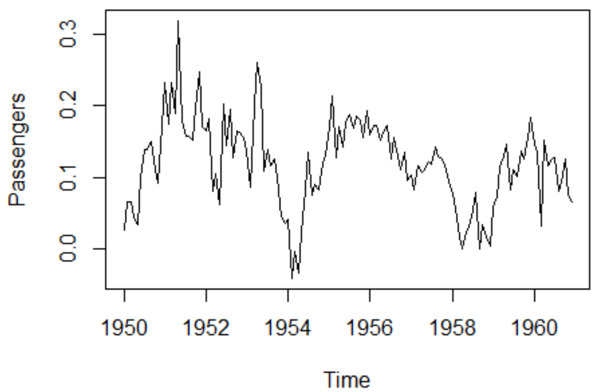


Poiché la media della risposta cambia nei diversi mesi, possiamo affermare che c'è chiara stagionalità.

Controllo differenziazioni

Vediamo che basta una sola differenziazione di ordine 12 per rendere la serie stazionaria nella sua componente stagionale.

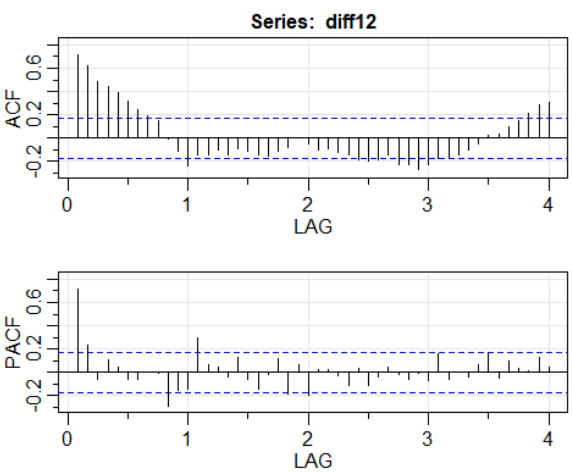
Differenziazione



Ora la nostra serie è stazionaria per quanto riguarda la sua parte stagionale.
Adesso possiamo procedere lavorando sulla componente non stagionale.

COMPONENTE NON STAGIONALE

ACF e PACF



ADF test

```
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
```

```
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.105416 -0.017336 -0.001595  0.018657  0.104638
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0548079  0.0200684   2.731  0.00742 **
## z.lag.1      -0.3469365  0.1105933  -3.137  0.00222 **
## tt           -0.0001897  0.0001212  -1.566  0.12045
## z.diff.lag1  -0.1773169  0.1102939  -1.608  0.11094
## z.diff.lag2   0.1327550  0.1055479   1.258  0.21129
## z.diff.lag3   0.0351851  0.1053101   0.334  0.73897
## z.diff.lag4   0.0410516  0.1054239   0.389  0.69778
## z.diff.lag5   0.1884087  0.1043706   1.805  0.07394 .
## z.diff.lag6   0.1939619  0.1042608   1.860  0.06566 .
## z.diff.lag7   0.0926435  0.1047215   0.885  0.37838
## z.diff.lag8   0.1842671  0.1023650   1.800  0.07474 .
## z.diff.lag9   0.3223770  0.1012341   3.184  0.00191 **
## z.diff.lag10  0.1736772  0.1035293   1.678  0.09643 .
## z.diff.lag11  0.0533492  0.1038694   0.514  0.60861
## z.diff.lag12 -0.2633223  0.0938585  -2.806  0.00600 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03745 on 104 degrees of freedom
## Multiple R-squared:  0.4236, Adjusted R-squared:  0.346
## F-statistic: 5.458 on 14 and 104 DF,  p-value: 9.959e-08
##
##
## Value of test-statistic is: -3.137 3.3411 4.9474
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

Osserviamo che value(-3.137) > tau3(-3.43) quindi non rifiutiamo H0: ro=0 e possiamo dire che la serie ha radici unitarie. Inoltre, value(4.9474) < phi3(6.49) quindi non rifiutiamo H0: beta2=0 e possiamo dire che la serie non ha un trend.

```
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.096405 -0.015333 -0.000812  0.018072  0.114315
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.029280   0.011782   2.485  0.01453 *
## z.lag.1      -0.249907   0.092231  -2.710  0.00787 **
## z.diff.lag1  -0.239858   0.103515  -2.317  0.02244 *
## z.diff.lag2   0.085232   0.101786   0.837  0.40429
## z.diff.lag3  -0.011181   0.101757  -0.110  0.91272
## z.diff.lag4  -0.006046   0.101738  -0.059  0.95273
## z.diff.lag5   0.144126   0.101158   1.425  0.15719
## z.diff.lag6   0.150986   0.101276   1.491  0.13900
## z.diff.lag7   0.049665   0.101756   0.488  0.62651
## z.diff.lag8   0.145999   0.100089   1.459  0.14764
## z.diff.lag9   0.290161   0.099804   2.907  0.00445 **
## z.diff.lag10  0.135007   0.101233   1.334  0.18521
## z.diff.lag11  0.012921   0.101302   0.128  0.89875
## z.diff.lag12 -0.293537   0.092486  -3.174  0.00197 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0377 on 105 degrees of freedom
## Multiple R-squared:   0.41,  Adjusted R-squared:  0.3369
## F-statistic: 5.612 on 13 and 105 DF,  p-value: 1.128e-07
##
##
## Value of test-statistic is: -2.7096 3.7342
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

Osserviamo che value(-2.7096) > tau2(-2.88) quindi non rifiutiamo H0: ro=0 e possiamo dire che la serie ha radici unitarie. Inoltre, value(3.7342) < phi1(4.63) quindi non rifiutiamo H0: beta1=0 e possiamo dire che la serie non ha un drift (infatti, mean=0.12 prossima a 0).

```
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.085309 -0.022677  0.001226  0.024508  0.116005
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.03082    0.02777  -1.110    0.270
## z.diff.lag1  -0.38153    0.08849  -4.312 3.64e-05 ***
## z.diff.lag2  -0.02665    0.09349  -0.285    0.776
## z.diff.lag3  -0.12269    0.09353  -1.312    0.192
## z.diff.lag4  -0.11825    0.09337  -1.267    0.208
## z.diff.lag5   0.03932    0.09417   0.418    0.677
## z.diff.lag6   0.04553    0.09418   0.483    0.630
## z.diff.lag7  -0.05936    0.09403  -0.631    0.529
## z.diff.lag8   0.04121    0.09296   0.443    0.658
## z.diff.lag9   0.18814    0.09317   2.019    0.046 *
## z.diff.lag10  0.03173    0.09454   0.336    0.738
## z.diff.lag11 -0.08796    0.09505  -0.925    0.357
## z.diff.lag12 -0.36951    0.08939  -4.133 7.16e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03861 on 106 degrees of freedom
## Multiple R-squared:  0.3758, Adjusted R-squared:  0.2993
## F-statistic:  4.91 on 13 and 106 DF,  p-value: 1.137e-06
##
##
## Value of test-statistic is: -1.1099
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

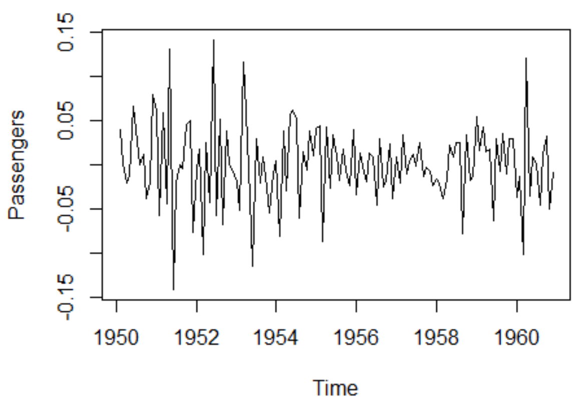
Osserviamo che value(-1.1099) > tau1(-1.95) quindi non rifiutiamo H0: ro=0 e possiamo dire che la serie ha radici unitarie.

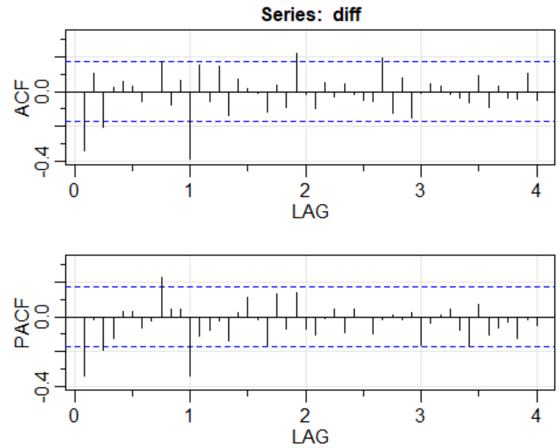
Controllo differenziazioni

Vediamo che è necessaria una differenziazione per rendere la nostra serie stazionaria.

Differenziazione

Grafici





Componente stagionale: ipotizziamo AR(1) e MA(1).
Componente non stagionale: ipotizziamo AR(1) e MA(1).

Funzione auto.arima

```
## Series: diff
## ARIMA(0,0,1)(0,0,1)[12] with zero mean
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 estimated as 0.001369:  log likelihood=244.7
## AIC=-483.39   AICc=-483.2   BIC=-474.77
```

La funzione auto.arima applicata alla serie resa stazionaria ci suggerisce un modello ARIMA(0,0,1)(0,0,1)[12] con media zero.

Proviamo a utilizzare la stessa funzione sul logaritmo della serie originale per vedere se vengono confermati i passaggi effettuati (differenziazione di ordine 12 prima e differenziazione di ordine 1 poi).

```
## Series: logyt
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 estimated as 0.001371:  log likelihood=244.7
## AIC=-483.4   AICc=-483.21   BIC=-474.77
```

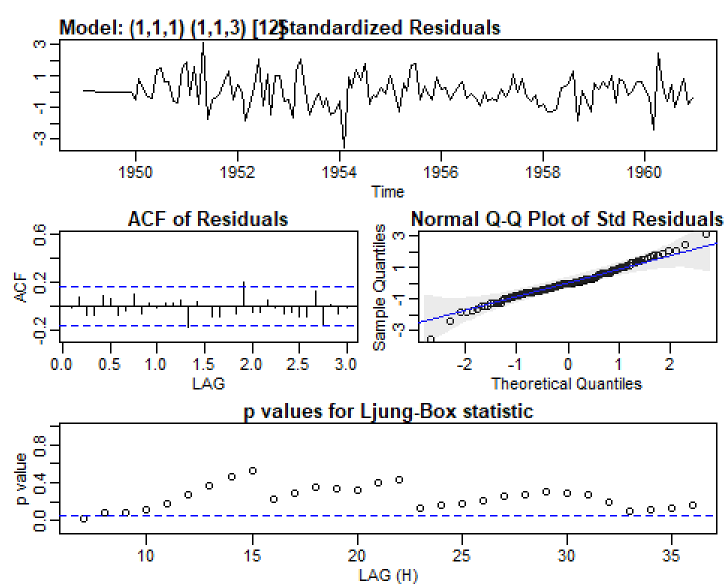
Ci viene suggerito un modello ARIMA(0,1,1)(0,1,1)[12].

MODELLO

Scelta del modello

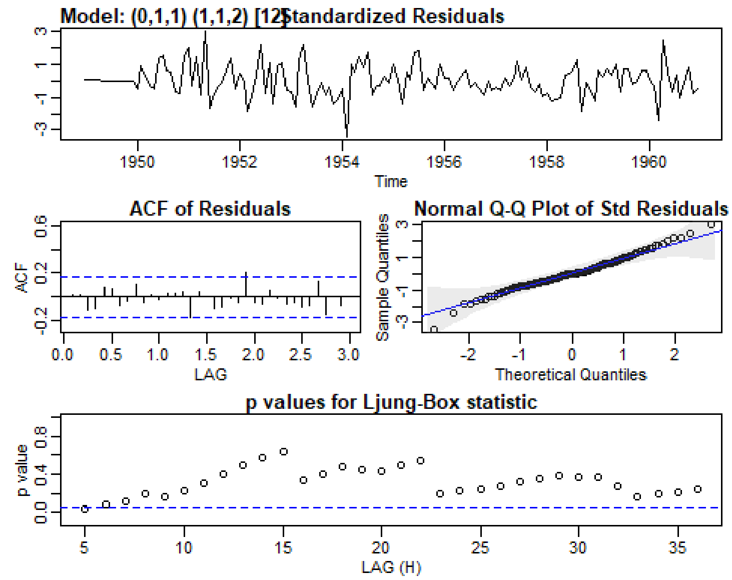
ARIMA (1,1,1) (1,1,3) [12]

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ar1      ma1     sar1    sma1     sma2    sma3
      0.1470  -0.5566  -0.9839  0.4109  -0.5091  0.0099
s.e.  0.2344   0.2012   0.1303  0.2151   0.1486  0.1057

sigma^2 estimated as 0.001308:  log likelihood = 245.64,  aic = -477.28

$degrees_of_freedom
[1] 125

$ttable
     Estimate     SE t.value p.value
ar1    0.1470 0.2344  0.6271  0.5318
ma1   -0.5566 0.2012 -2.7672  0.0065
sar1  -0.9839 0.1303 -7.5529  0.0000
sma1   0.4109 0.2151  1.9104  0.0584
sma2  -0.5091 0.1486 -3.4269  0.0008
sma3   0.0099 0.1057  0.0938  0.9254

$AIC
[1] -3.36111

$AICc
[1] -3.356729

$BIC
[1] -3.219375
```



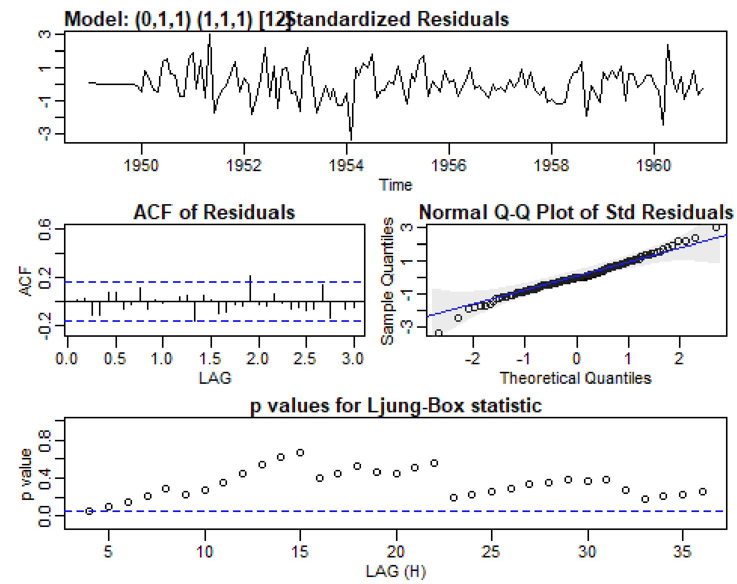
ARIMA (0,1,1) (1,1,2) [12]

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ma1     sar1    sma1     sma2
      -0.4288  -0.9736  0.4077  -0.4945
s.e.   0.0898   0.1736  0.2546   0.1729

sigma^2 estimated as 0.001315:  log likelihood = 245.45,  aic = -480.9

$degrees_of_freedom
[1] 127

$ttable
     Estimate     SE t.value p.value
ma1   -0.4288 0.0898 -4.7750  0.0000
sar1  -0.9736 0.1736 -5.6099  0.0000
sma1   0.4077 0.2546  1.6015  0.1118
sma2  -0.4945 0.1729 -2.8595  0.0050

$AIC
[1] -3.386609

$AICc
[1] -3.384552

$BIC
[1] -3.285369
```



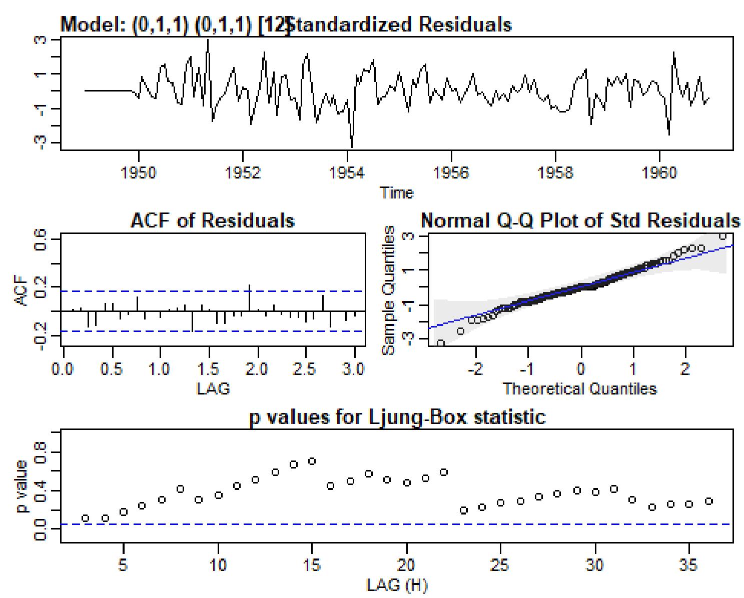
ARIMA (0,1,1) (1,1,1) [12]

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ma1     sar1     sma1
      -0.4143  -0.1116  -0.4817
s.e.   0.0899   0.1548   0.1363

sigma^2 estimated as 0.001341:  log likelihood = 244.96,  aic = -481.91

$degrees_of_freedom
[1] 128

$ttable
     Estimate     SE t.value p.value
ma1   -0.4143 0.0899 -4.6064  0.0000
sar1  -0.1116 0.1548 -0.7214  0.4720
sma1  -0.4817 0.1363 -3.5338  0.0006

$AIC
[1] -3.393754

$AICc
[1] -3.39253

$BIC
[1] -3.312763
```



ARIMA (0,1,1) (0,1,1) [12]

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4

$degrees_of_freedom
[1] 129

$ttable
     Estimate     SE t.value p.value
ma1   -0.4018 0.0896 -4.4825       0
sma1  -0.5569 0.0731 -7.6190       0

$AIC
[1] -3.404219

$AICc
[1] -3.403611

$BIC
[1] -3.343475
```



Partendo dal modello ARIMA(1,1,1,1,1,3)[12] e considerando la significatività dei parametri, otteniamo lo stesso modello suggeritoci dalla funzione auto.arima: ARIMA(0,1,1,0,1,1)[12]. I residui sembrano essere random, non autocorrelati (ACF=0 e p-value>0.05), distribuiti normalmente e l'AIC è pari a -483.4.

Fit

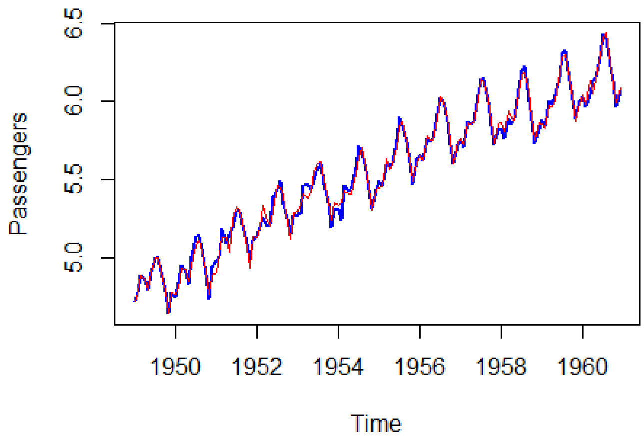
```
## Series: logyt
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 estimated as 0.001371:  log likelihood=244.7
## AIC=-483.4   AICc=-483.21   BIC=-474.77
```

Analisi residui

```
## Series: mod$residuals
## ARIMA(0,0,0) with zero mean
##
## sigma^2 estimated as 0.001228:  log likelihood=278.22
## AIC=-554.44   AICc=-554.41   BIC=-551.47
```

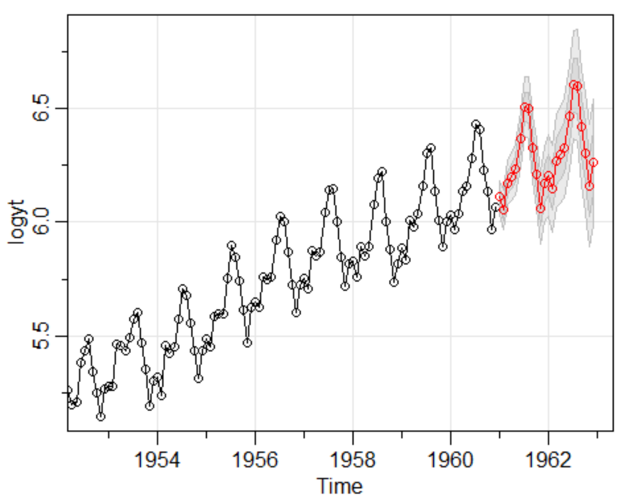
I residui del nostro modello sono white noise (infatti, da quanto visto prima, non risultavano problematici).

Plot (in blu logyt e in rosso i valori previsti)



Osserviamo che il modello fittato cattura molto bene l'andamento del logaritmo della nostra serie originale.

Forecast



Abbiamo fatto una previsione di due anni (24 mesi).

RIASSUNTO