Handout 6

Estimation and diagnostic Checking for ARMA and ARIMA Models

Class notes for Statistics 451: Applied Time Series
Iowa State University

February 14, 2007
14h 46min

ARMA Estimation Methods

For a specified ARMA(p,q) model and data, $(Z_1, Z_2, \ldots, Z_n)$, we "fit" the model by estimating the parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$.

- **Method of moments:** Compute $\widehat{\rho}_1, \widehat{\rho}_2, \ldots, \widehat{\rho}_{p+q}$, substitute into the "true" ACF equations $\rho_1 = \cdots, \rho_2 = \cdots, \ldots, \rho_{p+q} = \cdots$ and solve for $\widehat{\phi}_1, \ldots, \widehat{\phi}_p$, $\widehat{\theta}_1, \ldots, \widehat{\theta}_q$.
- **Nonlinear least squares:** Find $\widehat{\phi}_1, \ldots, \widehat{\phi}_p$, $\widehat{\theta}_1, \ldots, \widehat{\theta}_q$ to minimize
$$\text{SSQ} = \sum_{t=1}^{n} \widehat{a}_t^2 = \sum_{t=1}^{n} (Z_t - \widehat{Z}_t)^2$$
where $\widehat{Z}_t$ is the one-step-ahead forcast for Z_t.
- **Maximum Likelihood:** Find values of the parameters that maximize the probability of the data. Log(likelihood)$\approx -n\log(\text{SSQ})$.

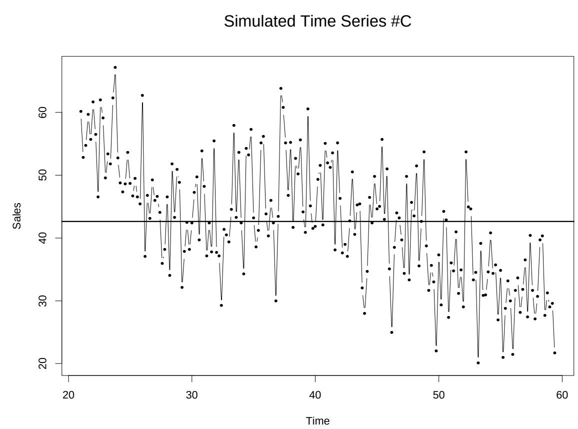
Example: Nonlinear Least Squares for IMA(1,1)

$$\begin{aligned}
\text{Minimize:}\quad \text{SSQ}(\theta) &= \sum_{t=1}^{n} \widehat{a}_t^2 = \sum_{t=1}^{n} (Z_t - \widehat{Z}_t)^2 \\
\text{IMA(1,1) forecast:}\quad \widehat{Z}_t &= Z_{t-1} - \theta_1 \widehat{a}_{t-1} \\
\text{SSQ}(\theta) &= \sum_{t=1}^{n} [Z_t - (Z_{t-1} - \theta_1 \widehat{a}_{t-1})]^2 \\
&= \sum_{t=1}^{n} (W_t + \theta_1 \widehat{a}_{t-1})^2
\end{aligned}$$

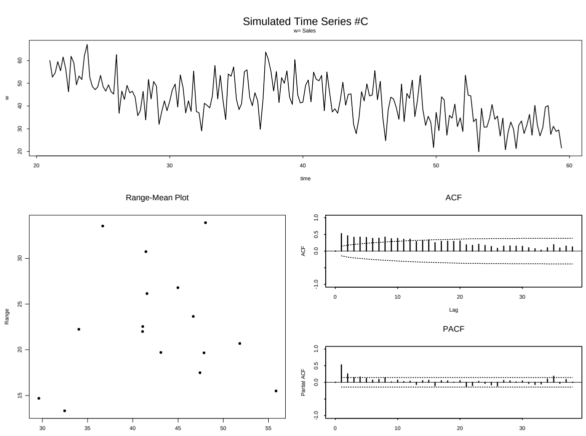
where $W_t = Z_t - Z_{t-1}$ and $\widehat{a}_{t-1} = Z_{t-1} - \widehat{Z}_{t-1} = W_{t-1} + \theta_1 \widehat{a}_{t-2}$.
$\widehat{\theta}$ is the value of θ that minimizes SSQ(θ).

- Need to compute $\widehat{a}$ recursively.
- Need some way to start the recursion (i.e., what is $\widehat{a}_0$?).
- Nonlinear optimization needed in general.

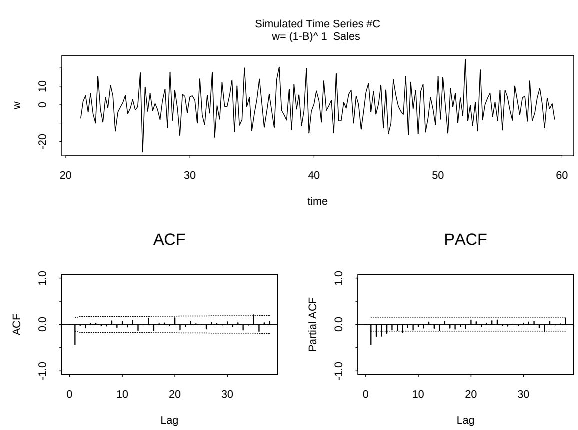
Time Series Plot of Simulated Series C



Function `iden` Output for Simulated Series C

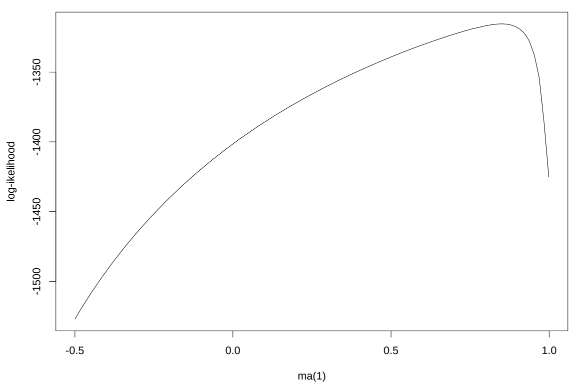


Function `iden` Output for the First Differences of Simulated Series C



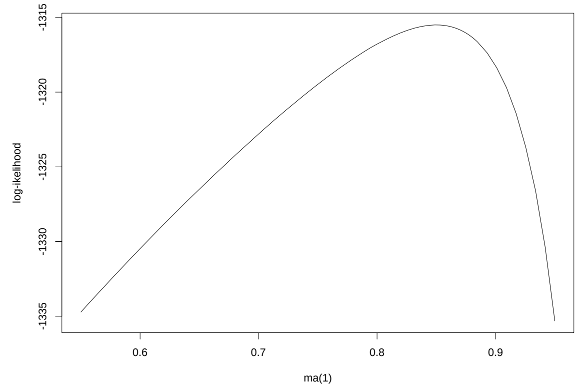
Plot of IMA(1,1) Model Log-likelihood for Simulated Series C

$$Z_t = Z_{t-1} - \theta_1 a_{t-1} + a_t$$

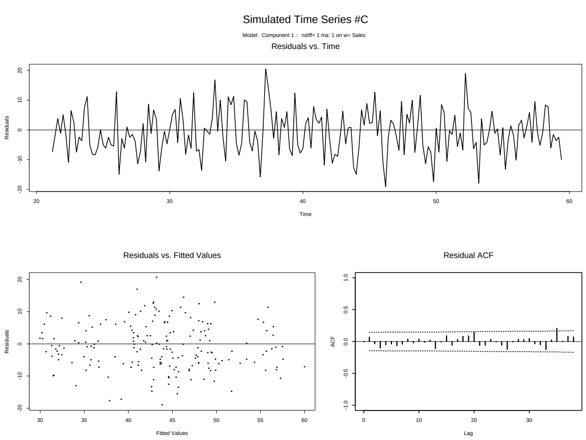


Plot of IMA(1,1) Model Log-likelihood for Simulated Series C

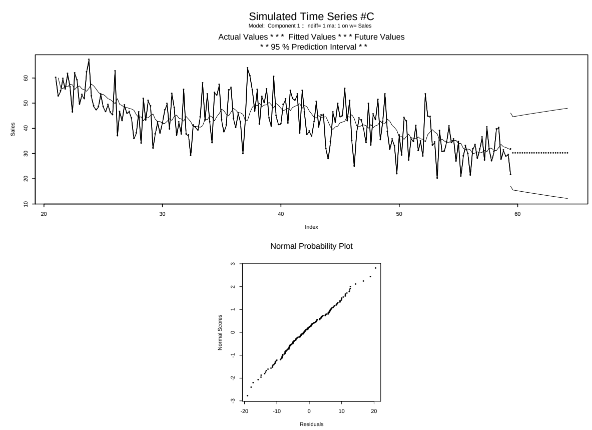
$$Z_t = Z_{t-1} - \theta_1 a_{t-1} + a_t$$



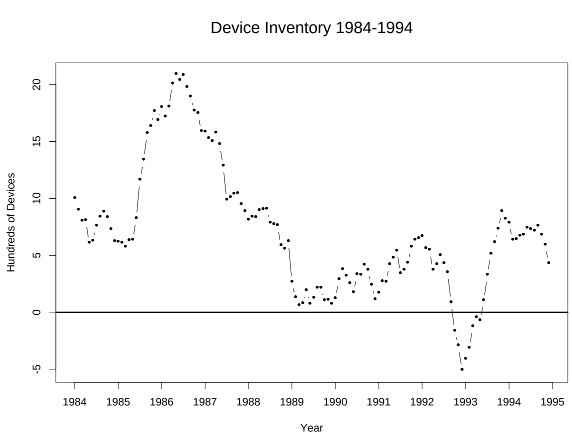
Function `esti` Output for Simulated Series C IMA(1,1) Model—Part 1



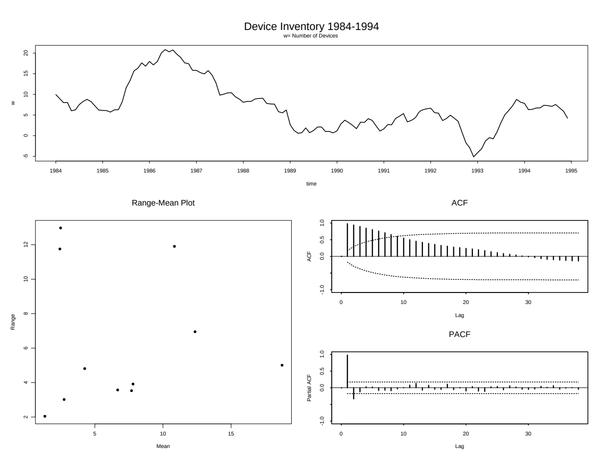
Function `esti` Output for for Simulated Series C IMA(1,1) Model—Part 2



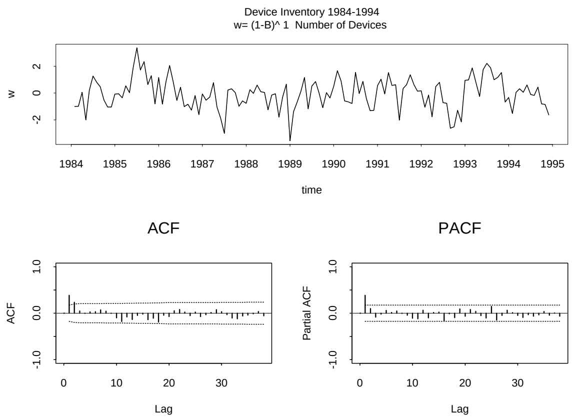
Time Series Plot of the Device Inventory Data (negative inventory indicates number of units that have been back-ordered)



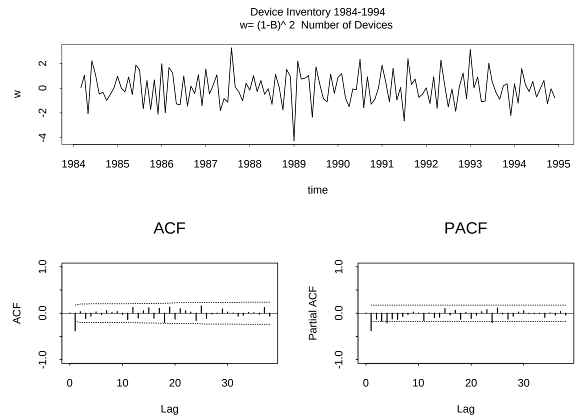
Function `iden` Output for The Device Inventory Data



Function `iden` Output for the First Differences of the Device Inventory Data



Function `iden` Output for the Second Differences of the Device Inventory Data

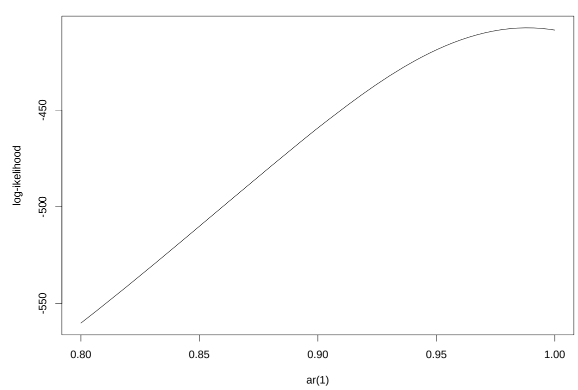


Graphical Outputs From the `esti` command

- Plot of residuals versus time
- Plot of residuals versus fitted values
- Plot of ACF of the residuals
- Plot showing original data, fitted values, forecasts, and forecast intervals
- Normal probability plot of the residuals

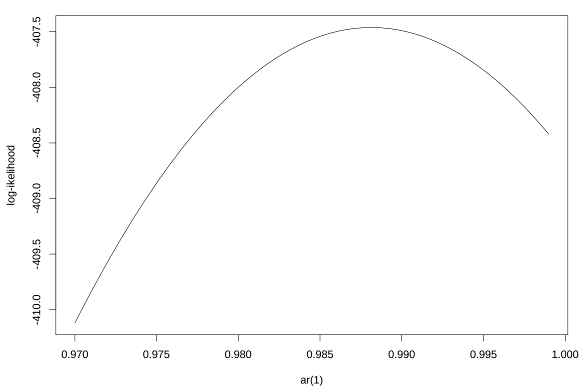
Plot of AR(1) Model Log-likelihood for the Device Inventory Data

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + a_t$$



Plot of AR(1) Model Log-likelihood for the Device Inventory Data

$$Z_t = \theta_0 + \phi_1 z_{t-1} + a_t$$

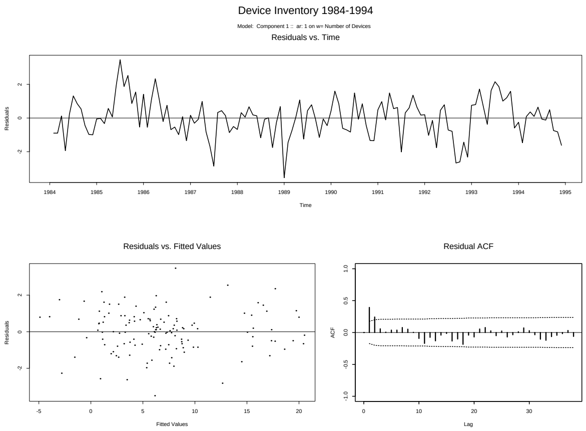


Commands for Making Plots of ARMA Model Log-likelihoods

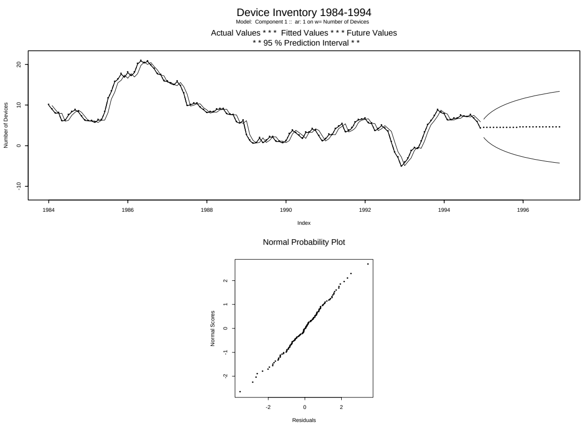
```
arima.likelihood.plot(device.inventory.d,
model = model.pdq(p = 1), list(1, seq(.8, .999, length = 50)))

#this is the zoom-in of the AR(1);
arima.likelihood.plot(device.inventory.d,
model = model.pdq(p = 1), list(1, seq(.97, .999, length =
50)))
```

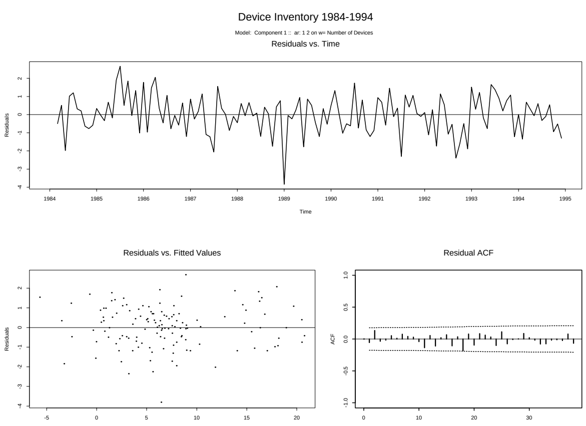
Function esti Output for the Device Inventory Data AR(1) Model—Part 1



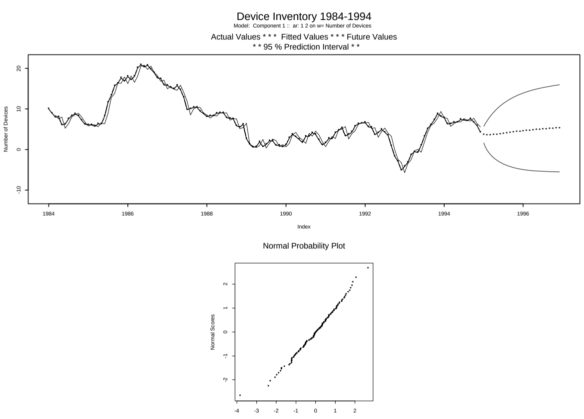
Function esti Output for the Device Inventory Data AR(1) Model—Part 2



Function esti Output for the Device Inventory Data AR(2) Model—Part 1



Function esti Output for the Device Inventory Data AR(2) Model—Part 2



Tabular Outputs From the esti Command

- Log likelihood, $S = \hat{\sigma}_a$, and the Akaike Information Criterion $\text{AIC}_c = -2\log(\text{Likelihood}) + 2M$ where M is the number of estimated parameters in the model (usually $M = p + q$). AIC_c is one, of many, "model-choice" criteria (see Section 7.7 of Wei for discussion of others).
- Table giving MLEs, SE of MLEs, t-ratio, and approximate confidence limits for each estimated parameter.
- Constant term estimate, SE, and t-ratio
- Variance-covariance and correlation matrices of the estimated partameters.
- Table of Ljung-Box statistics
- Table of ACF of the residuals, SE's and t-ratios
- Table of forecasts and prediction intervals

Tabular esti Output for the Device Inventory Data

```
>  esti(device.inventory.d,
      model=model.pdq(p=2),y.range =c(-12,22))
+ + Estimation/Forecasting Output for Device Inventory 1984-1994
ARIMA estimation results:
Model:  Component 1 ::  ar: 1 2
AICc: 385.1348
-2(Log Likelihood): 381.1348
S: 1.048085
Parameter Estimation Results
            MLE         se    t.ratio  95% lower  95% upper
ar(1)  1.3780133 0.08015244  17.192406  1.2209145  1.5351121
ar(2) -0.4059878 0.08015244  -5.065196 -0.5630866 -0.2488891
Constant term: 6.193324
Standard error: 0.09192312
t-ratio: 67.37504
```

Tabular esti Output for the Device Inventory Data

```
Correlation matrix
           ar(1)      ar(2)
ar(1)  1.0000000 -0.9801033
ar(2) -0.9801033  1.0000000
Ljung-Box Statistics
 dof Ljung-Box   p-value
   4  3.360982 0.4993286
   5  4.143612 0.5289303
   6  4.358709 0.6282547
   .
   .
   .
  36 31.849559 0.6663509
```


Tabular esti Output for the Device Inventory Data

```
Residual ACF
   Lag          ACF         se     t-ratio
 1   1 -0.054279383 0.08770580 -0.61888019
 2   2  0.132883340 0.08796383  1.51065894
 3   3 -0.035326410 0.08949467 -0.39473200
 .
 .
 .
38  38 -0.065383330 0.10392757 -0.62912398
Forecasts
         Lower Forecast     Upper
 1  1.59811508 3.652324  5.706532
 2 -0.02084244 3.476698  6.974238
 .
 .
 .
```


Diagnostic Checks Based on Residuals

- Need to check the assumption $a_t \sim \text{nid}(0, \sigma^2)$
- Plot of the residuals "time series" $\widehat{a}_t$ versus time
- ACF function $\widehat{\rho}_k(\widehat{a})$ of the residuals
  To judge $\text{H}_0 : \rho_k(a) = 0$, be suspicious of values of the associated t-like-ratio $t_k = \widehat{\rho}_k(\widehat{a})/\text{S}_{\widehat{\rho}_k(\widehat{a})}$ that are outside ± 1.5.
- Ljung-Box test for $\text{H}_0 : \rho_1(a) = \rho_2(a) = \ldots = \rho_K(a) = 0$

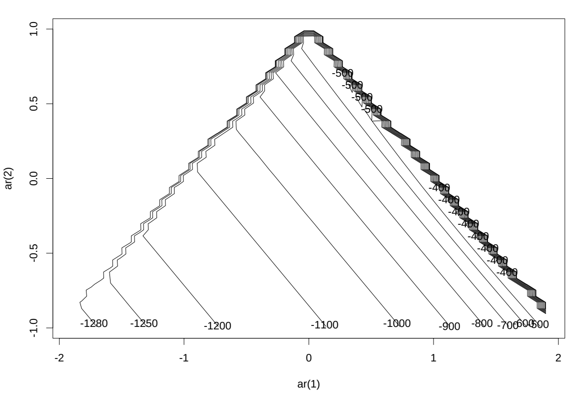
$$Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\widehat{\rho}_k^2(\widehat{a})$$

  Be suspicious if $Q > \chi^2_{(1-\alpha;K-p-q)}$
- Plot of residuals versus fitted values
- Normal probability plot of residuals to check normailty assumption

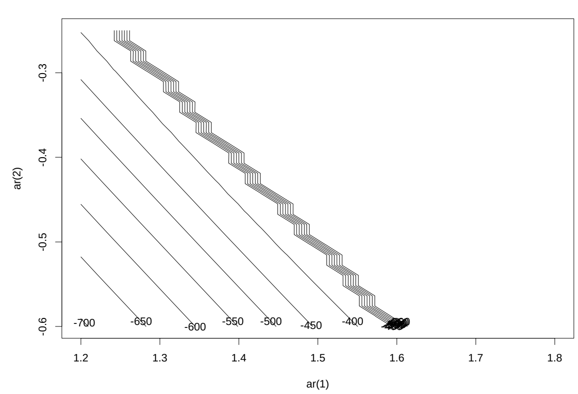
Other Diagnostic Checks

- Overfitting (add parameters to model looking for improvement)
- Make sure that the forecasts are reasonable; compare forecasts from different candadate models.
- Split the realization to see if the model and parameters agree on both sides (check for a change in the process).
- If there are outliers, do sensitivity analysis by moving outliers into the data
- If there is an indication of estimation problems, look at the likelihood (or sum of squares) surface

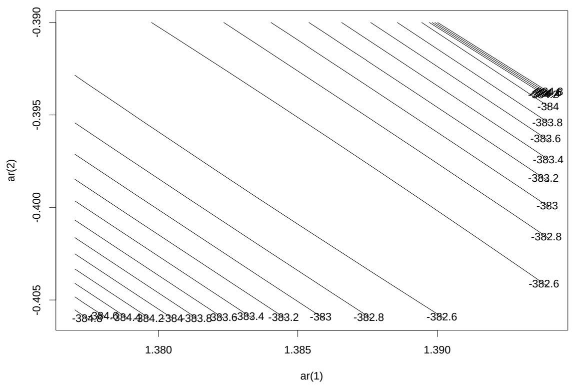
Plot of AR(2) Model Log-likelihood Surface for the Device Inventory Data



Plot of AR(2) Model Log-likelihood Surface for the Device Inventory Data



Plot of AR(2) Model Log-likelihood Surface for the Device Inventory Data Close-up View

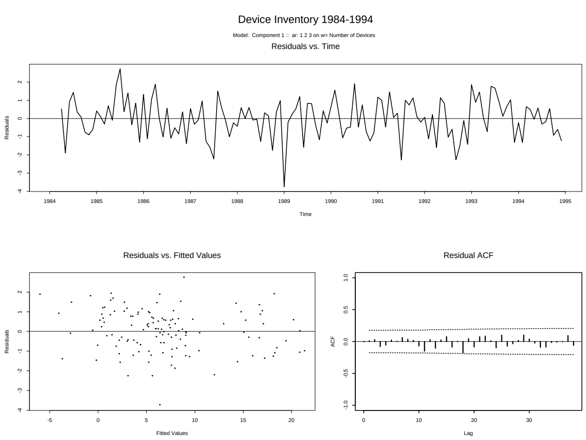


Commands for making Log-likelihood Surface Contour Plots

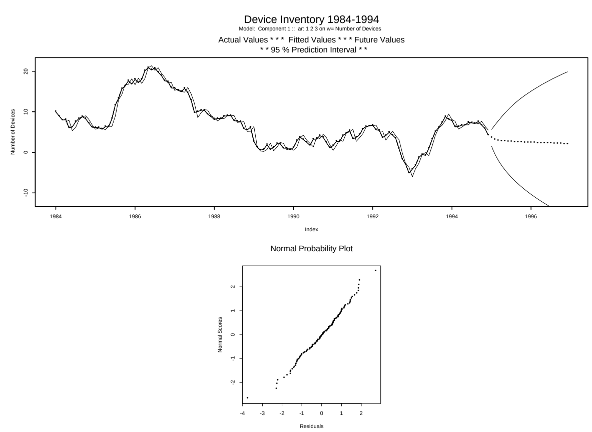
```
arima.contour(device.inventory.d,
model = model.pdq(p = 2),
list(1, seq(-1.9, 1.9, length = 50)),
list(2, seq(-.99, .99, length = 50)))

arima.contour(device.inventory.d,
model = model.pdq(p = 2),
list(1, seq(1.377, 1.394, length = 40)),
list(2, seq(-0.406, -0.390, length = 40)))
```

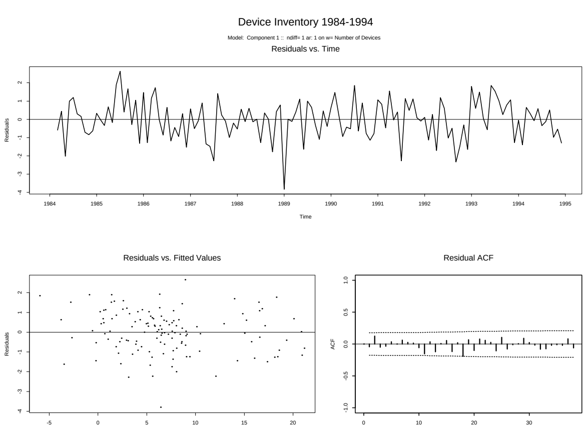

Function `esti` Output for the Device Inventory Data AR(3) Model—Part 1



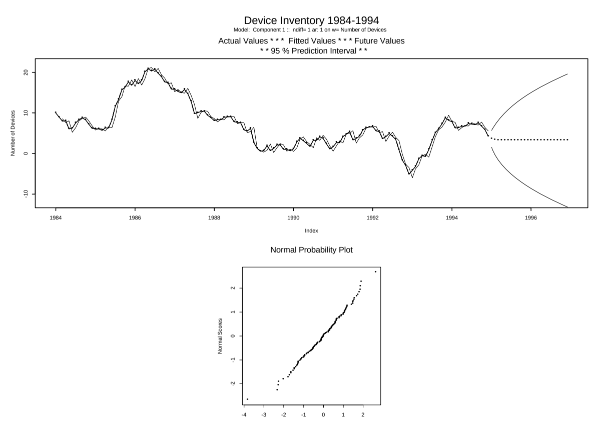
Function `esti` Output for the Device Inventory Data AR(3) Model—Part 2



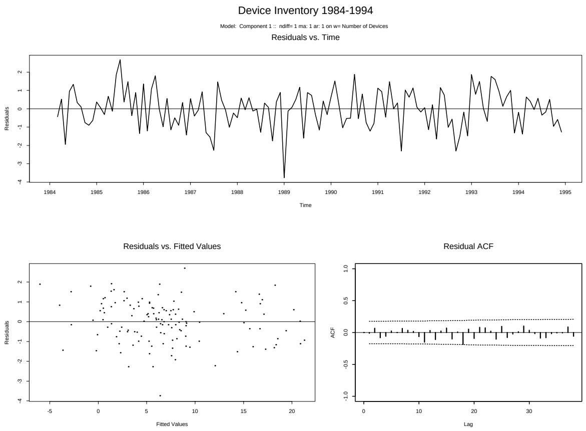
Function `esti` Output for the Device Inventory Data ARIMA(1,1,0) Model—Part 1



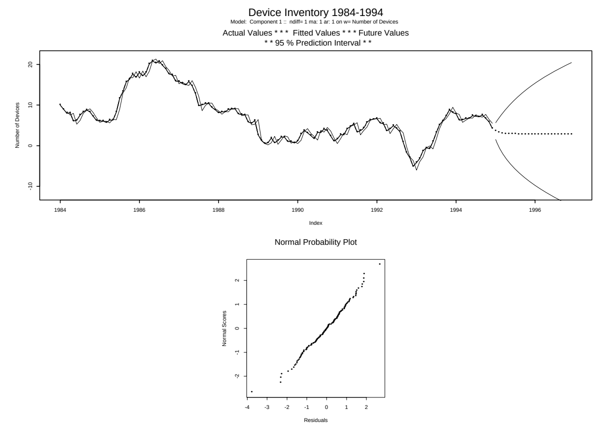
Function `esti` Output for the Device Inventory Data ARIMA(1,1,0) Model—Part 2



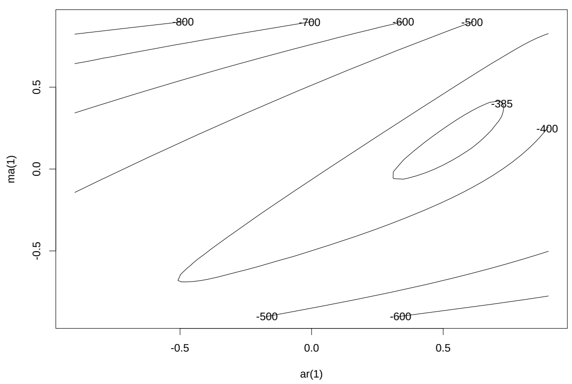
Function `esti` Output for the Device Inventory Data ARIMA(1,1,1) Model—Part 1



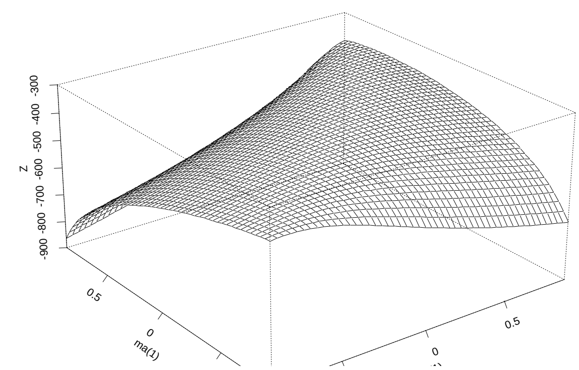
Function `esti` Output for the Device Inventory Data ARIMA(1,1,1) Model—Part 2



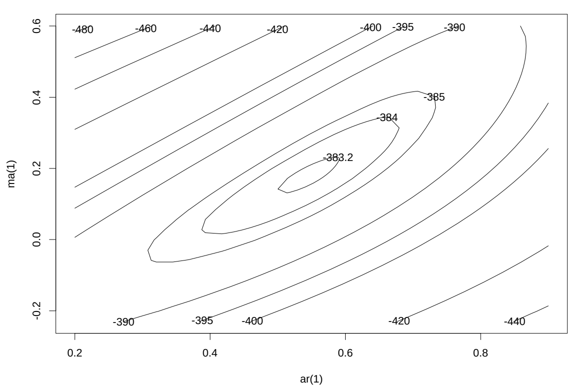
Plot of ARIMA(1,1,1) Model Log-likelihood Surface for the Device Inventory Data



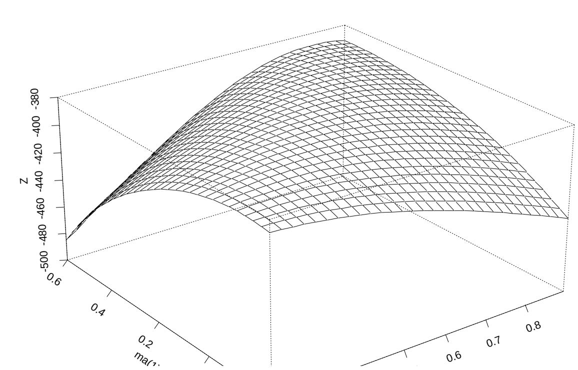
Plot of ARIMA(1,1,1) Model Log-likelihood Surface for the Device Inventory Data



Plot of ARIMA(1,1,1) Model Log-likelihood Surface for the Device Inventory Data



Plot of ARIMA(1,1,1) Model Log-likelihood Surface for the Device Inventory Data



Comparison of Modles for the Device Inventory Data

	ARIMA(p,d,q) Model				
	(1,0,0)	(2,0,0)	(3,0,0)	(1,1,0)	(1,1,1)
d (# differences)	0	0	0	1	1
ϕ_1	.98	1.37	1.35	.39	.55
	(55.4)	(17.2)	(15.4)	(4.9)	(3.0)
ϕ_2	—	-.4	-.24	—	—
		(-5.1)	(-1.7)		
ϕ_3	—	—	-.11	—	—
			(-1.2)		
θ_1	—	—	—	—	.18
					(.84)
Sig. $\widehat{\rho}_k(\widehat{a})$	1 , 2	2	—	—	—
S	1.14	1.04	1.06	1.06	1.06
AIC$_c$	409.05	385.13	386.35	386.13	387.08
$-2\log$(Likelihood)	407.05	381.13	380.35	384.13	383.08
Ljung-Box χ^2_6	30.26	4.35	2.17	3.49	2.73