Forecasting Trend, Seasonal and Cyclical Components

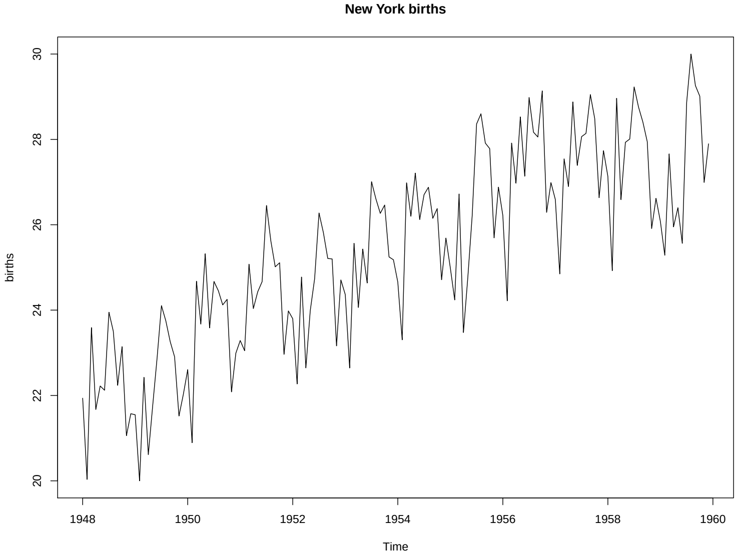
Dr. Natalia Khorunzhina

CBS

Time-Series Components

- Trend
  - Very long term (decades)
  - Smooth
- Seasonal
  - Patterns which repeat annually
  - May be constant or variable
- Cycle
  - Business cycle
  - Correlation over 2-7 years

New York births: additive components

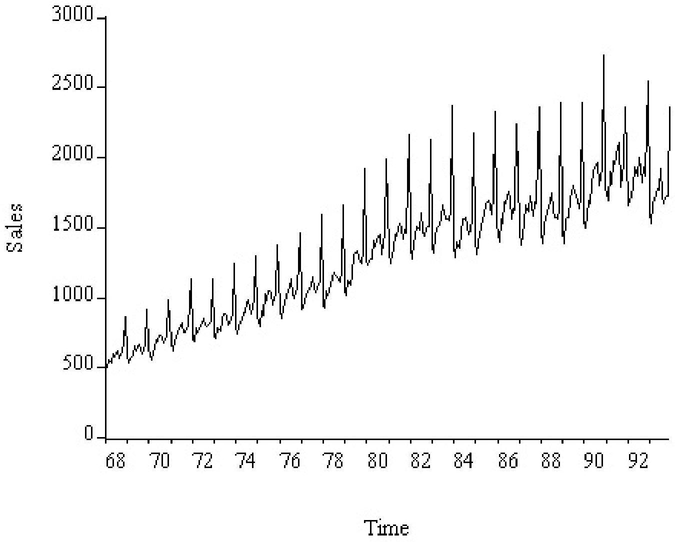


New York births is an example of time series that exhibits a **constant** seasonal variation. If parameters are not changing over time we can use the **additive model**:

$$y_t = T_t + S_t + C_t + \varepsilon_t$$

where
y_t is the observed value of the time series in period t
T_t the trend component in period t
S_t the seasonal component in period t
C_t the cyclical component in period t
ε_t the irregular component in period t

Liquor Sales: multiplicative components



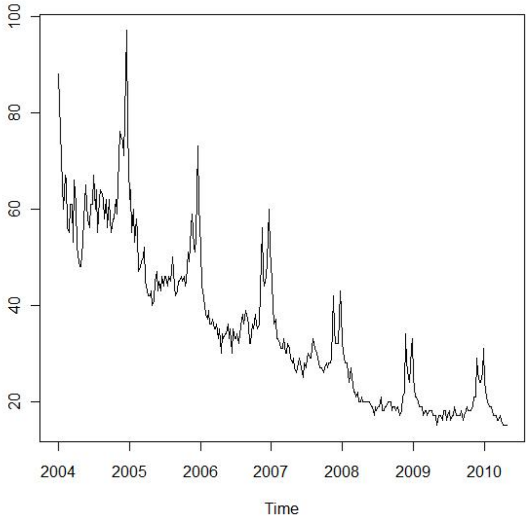
Liquor Sales, 1968.01 - 1993.12

Liquor sales is an example of time series that exhibits not a constant, but an **increasing** seasonal variation. Consider a time series that exhibits **increasing or decreasing** seasonal variation. If we assume that parameters are not changing over time we can use the **multiplicative model**:

$$y_t = Tr_t \times Sn_t \times Cl_t \times \epsilon_t$$

where
y_t is the observed value of the time series in period t
Tr_t the trend component in period t
Sn_t the seasonal component in period t
Cl_t the cyclical component in period t
ϵ_t the irregular component in period t

Decreasing seasonal variation: Google Query Index for Home Video Electronics



When the original model is multiplicative:

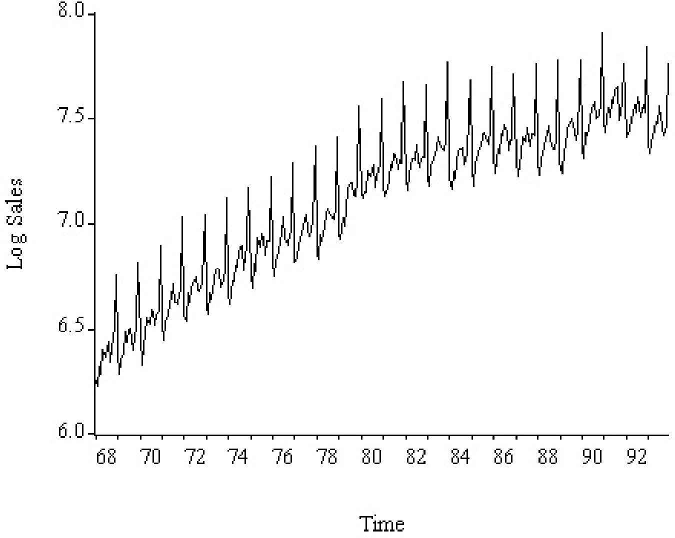
$$y_t = Tr_t \times Sn_t \times Cl_t \times \epsilon_t,$$

it can be made visually additive by taking a log-transformation

$$\begin{aligned} \ln y_t &= \ln Tr_t + \ln Sn_t + \ln Cl_t + \ln \epsilon_t \\ &= T_t \quad + \quad S_t \quad + C_t \quad + \varepsilon_t, \end{aligned}$$

where I simply redefined the components as $T_t = \ln Tr_t$, $S_t = \ln Sn_t$, $C_t = \ln Cl_t$, and $\varepsilon_t = \ln \epsilon_t$.

Log Liquor Sales, 1968.01 - 1993.12



Modeling trend

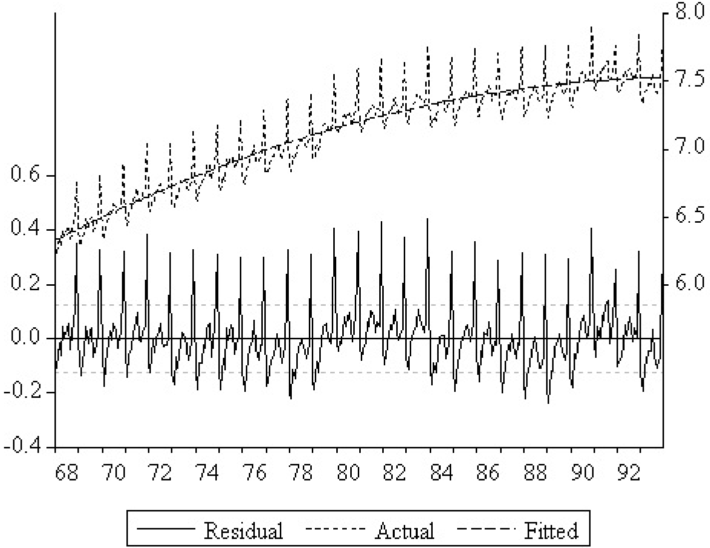
- We gradually add components to the model.
- We start with modeling trend.
- Liquor sales trend upward and the trend appears nonlinear.
  - We can model a simple linear trend and hope to pick up remaining non-linearity in it by the cyclical component.
  - Alternatively, to handle the nonlinear trend, we can adopt a quadratic trend model:

$$T_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2$$

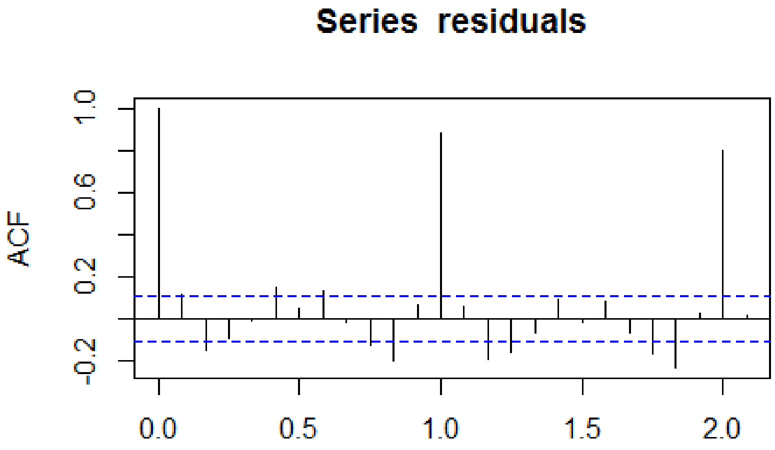
Log Liquor Sales, Quadratic Trend Regression

	Dependent variable:
	Log Liquor Sales
t	0.008^{***}
	(0.0003)
t2	-0.00001^{***}
	(0.00000)
Constant	6.231^{***}
	(0.021)
Observations	336
Adjusted R^2	0.903
Residual Std. Error	0.125 (df = 333)
F Statistic	$1{,}562.036^{***}$ (df = 2; 333)
Note:	*p<0.1; **p<0.05; ***p<0.01

Log Liquor Sales, Quadratic Trend Regression



Log Liquor Sales, Quadratic Trend Regression, Residual Sample Autocorrelation



Adding Seasonality

- We gradually add a seasonal component to the model.
- Recall that we either omit one dummy from the model or omit the intercept from the model.
- Liquor sales have a very strong seasonal pattern with the peak in December:

$$y_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2 + \sum_{j=1}^{11} \delta_j D_{jt} + \varepsilon_t$$

Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies: all seasonal dummies

LS // Dependent Variable is LSALES
Sample: 1968:01 1993:12
Included observations: 312

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TIME	0.007656	0.000123	62.35882	0.0000
TIME2	-1.14E-05	3.56E-07	-32.06823	0.0000
D1	6.147456	0.012340	498.1699	0.0000
D2	6.088653	0.012353	492.8890	0.0000
D3	6.174127	0.012366	499.3008	0.0000
D4	6.175220	0.012378	498.8970	0.0000
D5	6.246086	0.012390	504.1398	0.0000
D6	6.250387	0.012401	504.0194	0.0000
D7	6.295979	0.012412	507.2402	0.0000
D8	6.268043	0.012423	504.5509	0.0000
D9	6.203832	0.012433	498.9630	0.0000
D10	6.229197	0.012444	500.5968	0.0000
D11	6.259770	0.012453	502.6602	0.0000
D12	6.580068	0.012463	527.9819	0.0000

R-squared	0.986111	Mean dependent var	7.112383
Adjusted R-squared	0.985505	S.D. dependent var	0.379308
S.E. of regression	0.045666	Akaike info criterion	-6.128963
Sum squared resid	0.621448	Schwarz criterion	-5.961008
Log likelihood	527.4094	F-statistic	1627.567
Durbin-Watson stat	0.586187	Prob(F-statistic)	0.000000

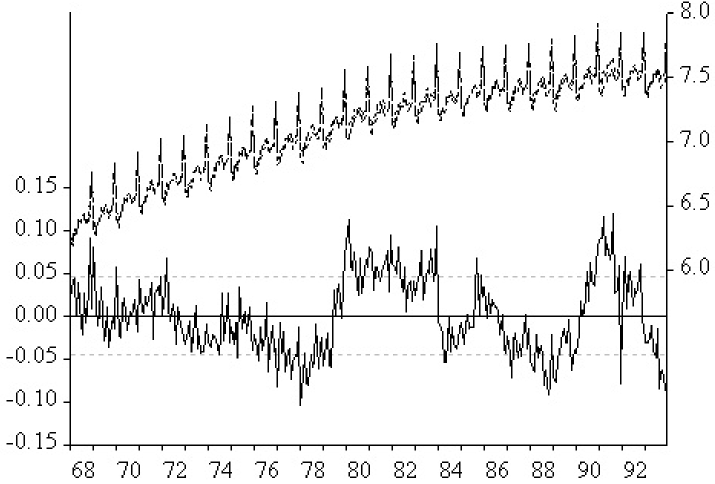
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies: January is an omitted season

```
tslm(formula = lny ~ t + t2 + season)

Residuals:
      Min        1Q    Median        3Q       Max
-0.105017 -0.034038 -0.004899  0.032406  0.132702

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.138e+00  1.121e-02 547.732  < 2e-16 ***
t            7.739e-03  1.039e-04  74.498  < 2e-16 ***
t2          -1.175e-05  2.985e-07 -39.368  < 2e-16 ***
season2     -5.694e-02  1.230e-02  -4.627 5.38e-06 ***
season3      3.021e-02  1.230e-02   2.455   0.0146 *
season4      3.122e-02  1.231e-02   2.537   0.0116 *
season5      1.002e-01  1.231e-02   8.143 8.54e-15 ***
season6      1.052e-01  1.231e-02   8.552 4.98e-16 ***
season7      1.492e-01  1.231e-02  12.125  < 2e-16 ***
season8      1.209e-01  1.231e-02   9.824  < 2e-16 ***
season9      6.104e-02  1.231e-02   4.960 1.15e-06 ***
season10     8.314e-02  1.231e-02   6.756 6.65e-11 ***
season11     1.152e-01  1.231e-02   9.356  < 2e-16 ***
season12     4.373e-01  1.231e-02  35.528  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04604 on 322 degrees of freedom
Multiple R-squared:  0.9875,    Adjusted R-squared:  0.9869
F-statistic:  1949 on 13 and 322 DF,  p-value: < 2.2e-16
```

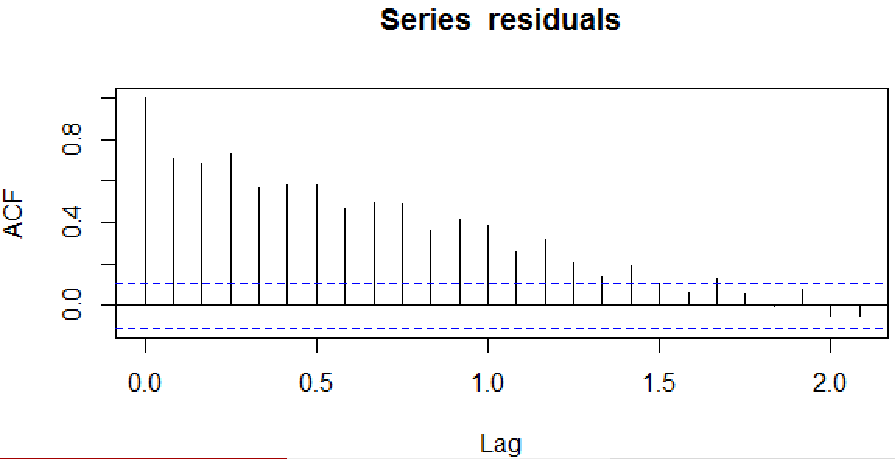
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies: December is an omitted season

	Log Liquor Sales	
	(1)	(2)
t	0.008*** (0.0003)	0.008*** (0.0001)
t2	−0.00001*** (0.00000)	−0.00001*** (0.00000)
MJan		−0.437*** (0.012)
MFeb		−0.494*** (0.012)
MMar		−0.407*** (0.012)
MApr		−0.406*** (0.012)
MMay		−0.337*** (0.012)
MJun		−0.332*** (0.012)
MJul		−0.288*** (0.012)
MAug		−0.316*** (0.012)
MSep		−0.376*** (0.012)
MOct		−0.354*** (0.012)
MNov		−0.322*** (0.012)
Constant	6.231*** (0.021)	6.576*** (0.011)
Observations	336	336
Adjusted R^2	0.903	0.987
DW test		0.58138
Residual Std. Error	0.125 (df = 333)	0.046 (df = 322)
F Statistic	1,562.036*** (df = 2; 333)	1,949.234*** (df = 13; 322)

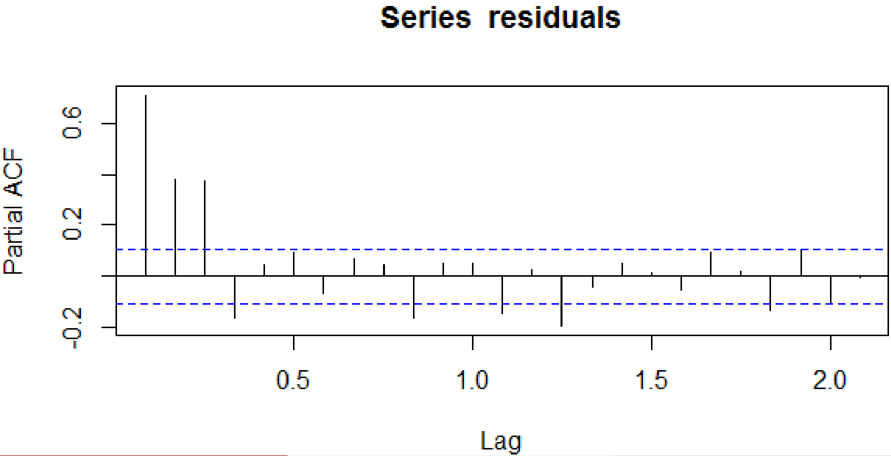
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies



Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies



Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies



Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances

General model with the time-series components can be formulated as:

$$\begin{aligned} y_t = \quad & \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2 + \sum_{j=1}^{11} \delta_j D_{jt} + \varepsilon_t + \\ & \rho_1 \tilde{y}_{t-1} + \rho_2 \tilde{y}_{t-2} + \ldots + \rho_p \tilde{y}_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \\ = \quad & \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2 + \sum_{j=1}^{11} \delta_j D_{jt} + \sum_{i=1}^{p} \rho_i \tilde{y}_{t-i} + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t, \end{aligned}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and
$\tilde{y}_t$ is detrended (if there is trend) and deseasonalized counterpart of y_t.
Quadratic trend regression model with seasonal dummies and AR(3):

$$y_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2 + \sum_{j=1}^{11} \delta_j D_{jt} + \rho_1 \tilde{y}_{t-1} + \rho_2 \tilde{y}_{t-2} + \rho_3 \tilde{y}_{t-3} + \varepsilon_t$$

Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances

LS // Dependent Variable is LSALES
Sample: 1968:01 1993:12
Included observations: 312
Convergence achieved after 4 iterations

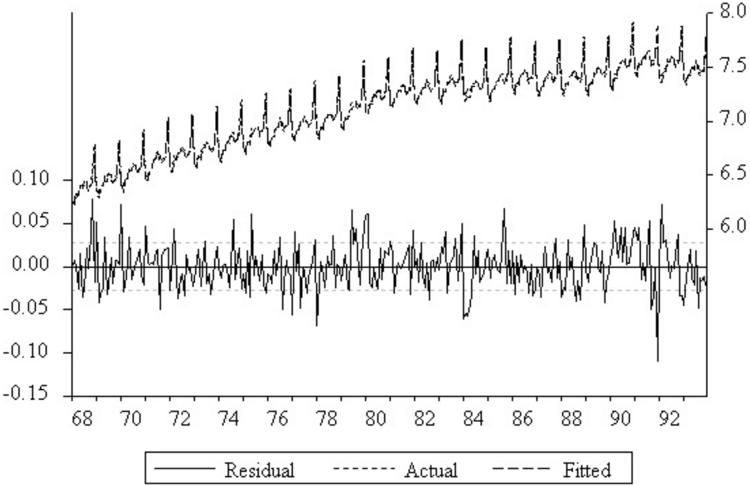
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TIME	0.008606	0.000981	8.768212	0.0000
TIME2	-1.41E-05	2.53E-06	-5.556103	0.0000
D1	6.073054	0.083922	72.36584	0.0000
D2	6.013822	0.083942	71.64254	0.0000
D3	6.099208	0.083947	72.65524	0.0000
D4	6.101522	0.083934	72.69393	0.0000
D5	6.172528	0.083946	73.52962	0.0000
D6	6.177129	0.083947	73.58364	0.0000
D7	6.223323	0.083939	74.14071	0.0000
D8	6.195681	0.083943	73.80857	0.0000
D9	6.131818	0.083940	73.04993	0.0000
D10	6.157592	0.083934	73.36197	0.0000
D11	6.188480	0.083932	73.73176	0.0000
D12	6.509106	0.083928	77.55624	0.0000
AR(1)	0.268805	0.052909	5.080488	0.0000
AR(2)	0.239688	0.053697	4.463723	0.0000
AR(3)	0.395880	0.053109	7.454150	0.0000

R-squared	0.995069	Mean dependent var	7.112383
Adjusted R-squared	0.994802	S.D. dependent var	0.379308
S.E. of regression	0.027347	Akaike info criterion	-7.145319
Sum squared resid	0.220625	Schwarz criterion	-6.941373
Log likelihood	688.9610	F-statistic	3720.875
Durbin-Watson stat	1.886119	Prob(F-statistic)	0.000000

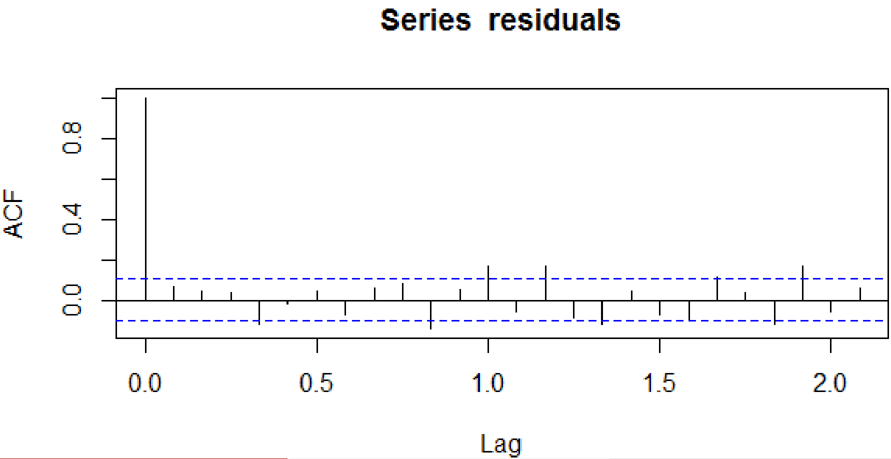
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances

	Log Liquor Sales		
	OLS		*ARIMA*
	(1)	(2)	(3)
ar1			0.288^{***} (0.050)
ar2			0.221^{***} (0.051)
ar3			0.381^{***} (0.049)
t	0.008^{***} (0.0003)	0.008^{***} (0.0001)	0.008^{***} (0.0003)
t2	-0.00001^{***} (0.00000)	-0.00001^{***} (0.00000)	-0.00001^{***} (0.00000)
MJan		-0.437^{***} (0.012)	-0.440^{***} (0.007)
MFeb		-0.494^{***} (0.012)	-0.497^{***} (0.007)
MMar		-0.407^{***} (0.012)	-0.410^{***} (0.006)
MApr		-0.406^{***} (0.012)	-0.408^{***} (0.007)
MMay		-0.337^{***} (0.012)	-0.339^{***} (0.007)
MJun		-0.332^{***} (0.012)	-0.334^{***} (0.007)
MJul		-0.288^{***} (0.012)	-0.289^{***} (0.007)
MAug		-0.316^{***} (0.012)	-0.317^{***} (0.007)
MSep		-0.376^{***} (0.012)	-0.377^{***} (0.006)
MOct		-0.354^{***} (0.012)	-0.355^{***} (0.007)
MNov		-0.322^{***} (0.012)	-0.322^{***} (0.006)
Constant	6.231^{***} (0.021)	6.576^{***} (0.011)	6.587^{***} (0.031)
Observations	336	336	336
Adjusted R^2	0.903	0.987	
DW test		0.58138	
Akaike Inf. Crit.			$-1,433.806$
Note:			*p<0.1; **p<0.05; ***p<0.01

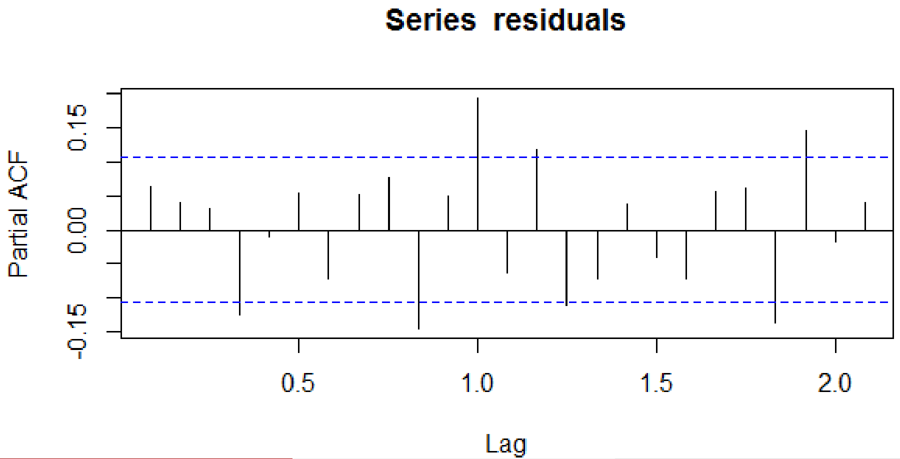
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances



Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances



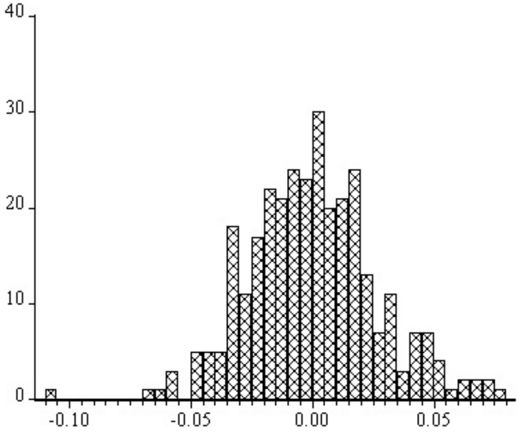
Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances



Log Liquor Sales, Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances

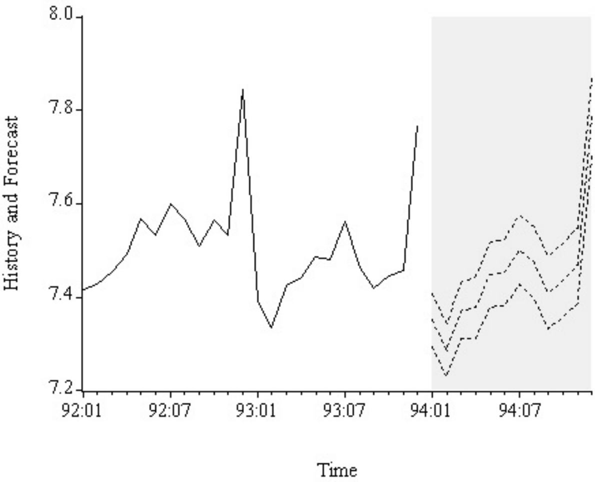
- All things considered, the quadratic trend, seasonal dummies and AR(3) specification for cycles seems adequate.
- We may perform a number of additional checks
- Recall that the forecast assumes that the model errors are normally distributed. We can check that the model residuals (our realistic counterparts to the theoretical model errors) are normally distributed.
- We construct a histogram and hope that it looks like a bell-shaped normal distribution.

Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances: Residual Histogram and Normality Test

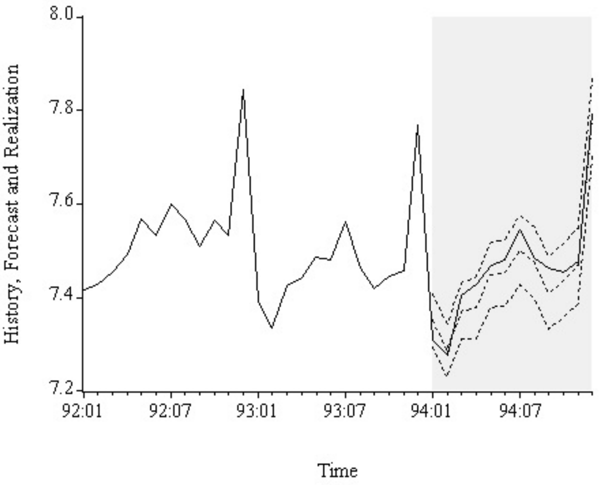


Series: Residuals	
Sample 1968:01 1993:12	
Observations 312	
Mean	3.77E-16
Median	-0.000160
Maximum	0.078468
Minimum	-0.109856
Std. Dev.	0.026635
Skewness	0.077911
Kurtosis	3.740378
Jarque-Bera	7.441714
Probability	0.024213

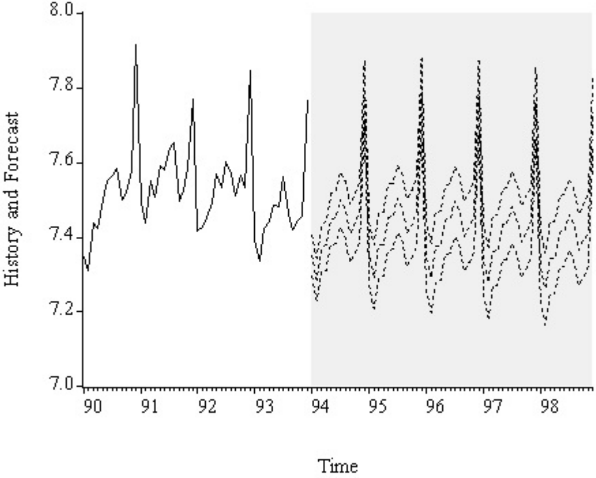
Log Liquor Sales. History and 12-Month-Ahead Forecast



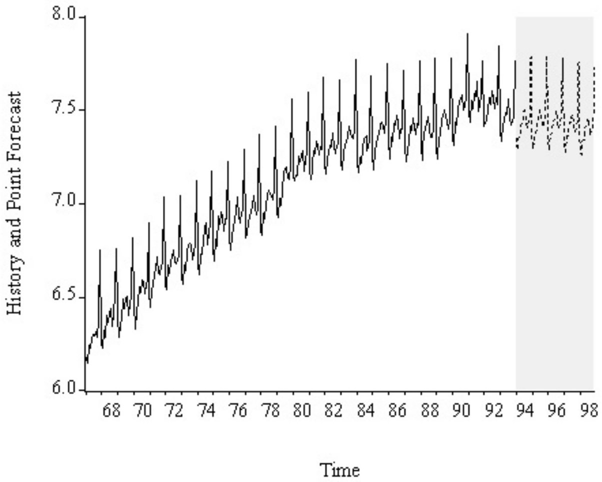
Log Liquor Sales. History, 12-Month-Ahead Forecast, and Realization



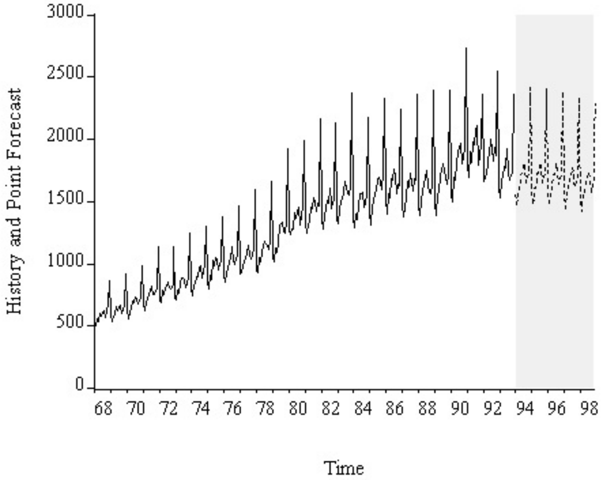
Log Liquor Sales. History and 60-Month-Ahead Forecast



Log Liquor Sales. Long History and 60-Month-Ahead Forecast



Liquor Sales. Long History and 60-Month-Ahead Forecast



Liquor Sales. Lab session

- As usual, in R we consider practical implementation of the application of **liquor sales**.
- For InClass application you can try:
  1. You can add cyclical component to the hotel room occupancy application
  2. Your application with seasonal data of your choice