3.8 Building ARIMA Models

- Plot data
- possibly transform the data
- identifying the dependence orders of the model
- parameter estimation
- diagnostics
- model selection

Growth Rate

Suppose a process evolves as a fairly small and stable percent change, such as an investment. We might have

$$x_t = (1 + p_t)x_{t-1}$$

where x_t is the value of the investment at time t and p_t is the percentage change from period $t-1$ to t, which may be negative.

$$\bigtriangledown[ln(x_t)] = ln(1 + p_t)$$

If the percent change p_t stays relatively small in magnitude, then $ln(1 + p_t) \approx p_t$, so $\bigtriangledown[ln(x_t)] \approx p_t$.

Selecting order of d

- A time plot of the data will typically suggest whether any differencing is needed

- If time plot is not stationary, differencing is called for, difference the data once, $d = 1$, and inspect the time plot of ∇x_t.

- Not to overdifference because this may introduce dependence where none exists. For example, $x_t = w_t$ is serially uncorrelated, but $\nabla x_t = w_t - w_{t-1}$ is $MA(1)$.

- ACF will not decay to zero fast as h increases. Thus, a slow decay in ACF is an indication that differencing may be needed.

Selecting order of p and q

- if $p = 0$ and $q > 0$, the ACF cuts off after lag q, and PACF tails off
- if $p > 0$ and $q = 0$, the PACF cuts off after lag p, and ACF tails off
- if $p > 0$ and $q > 0$, both the ACF and PACF will tail off

Model selection

most popular techniques, AIC, AICc and SIC

Comments

- Because we are dealing with the estimates, it will not always be clear whether the sample ACF or PACF is tailing off or cutting off

- Two models that are seemingly different can actually be very similar

- A few preliminary values of p, d and q should be at hand

- prevent overfitting, it is not always the case that more is better. Overfitting leads to less-precise estimators, and adding more parameters may fit the data better but may also lead to bad forecasts.

Example 3.35, 3.36 and 3.39 Analysis of GNP Data

We consider the analysis of quarterly U.S. GNP from 1747(1) to 2002(3), $n = 223$ observations. The data are real U.S. Gross National Product in billions of chained 1996 dollars and they have been seasonally adjusted. Data were obtained from the Federal Reserve Bank of St, Louis. Growth rate can be interpreted as the percentage quarterly growth of U.S. GNP.

- Plots

Figure 1: Plot, acf, first difference and growth rate of GNP

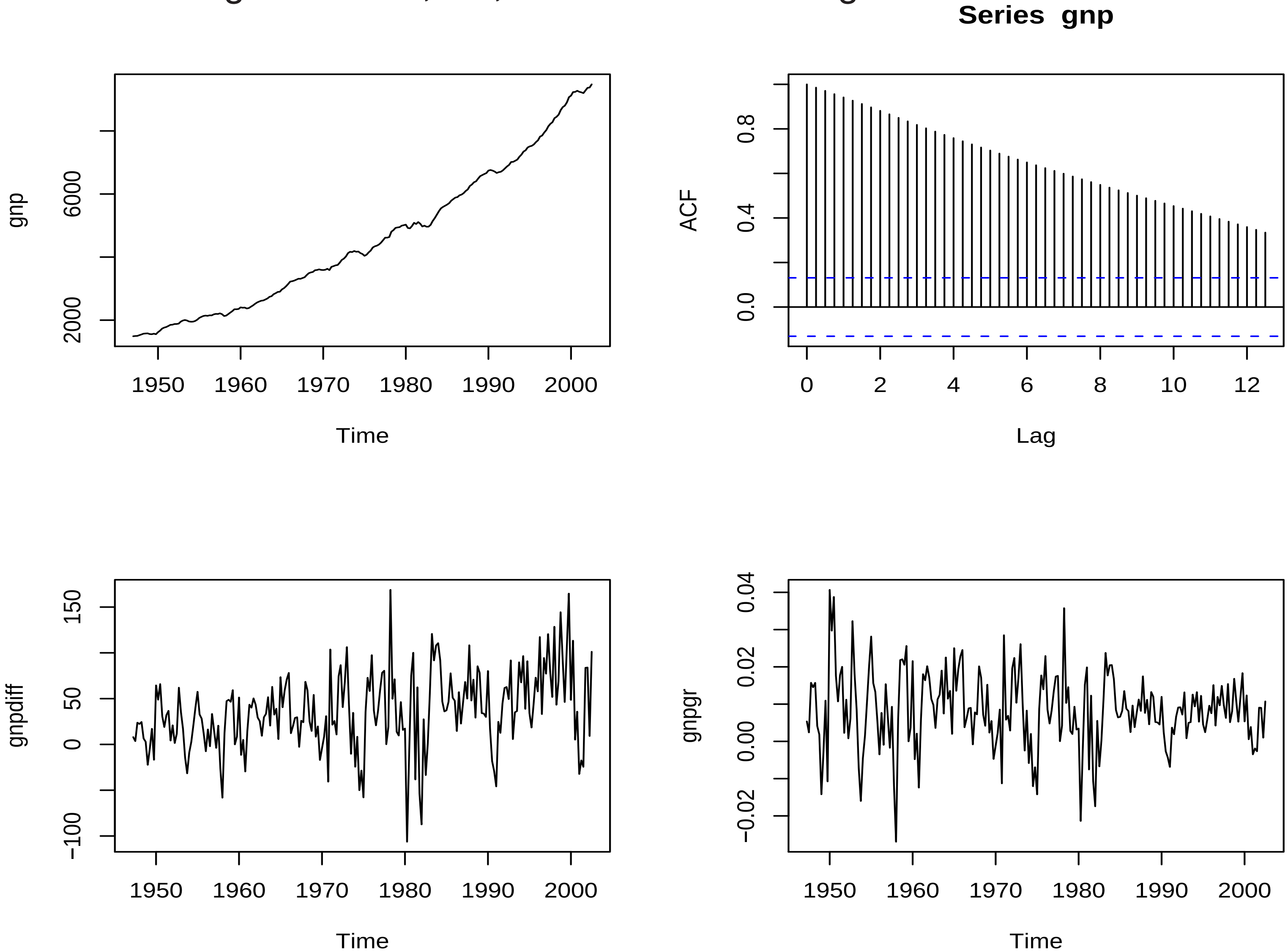
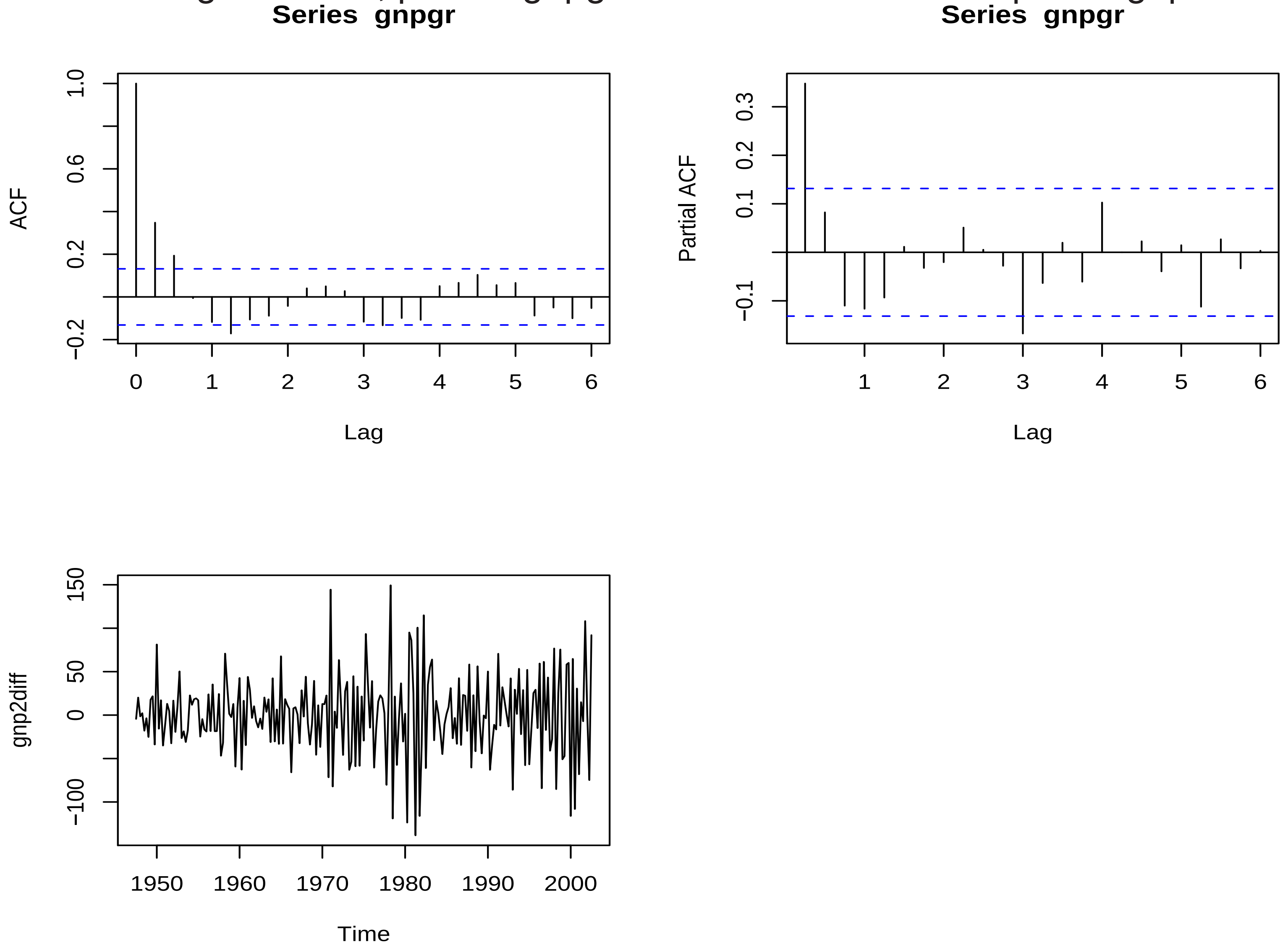


Figure 2: acf, pacf of gnpgr and second difference plot of gnp



- From Figure 2, we might feel that the ACF is cutting off at lag 2 and the PACF is tailing off. This suggests that GNP growth rate follows and $MA(2)$ process, or $log(GNP)$ follows an $ARIMA(0,1,2)$ process.

- From Figure 2, it also looks like that the ACF is tailing off and the PACF is cutting off after lag 1. This suggests and $AR(1)$ model for the growth rate, or $ARIMA(1,1,0)$ for $log(GNP)$.

- Fit both models

  Using MLE to fit the $MA(2)$ model for the growth rate, x_t, the estimated model is

$$x_t = .008_{(.001)} + .303_{(.065)}\hat{w}_{t-1} + .202_{(.064)}\hat{w}_{t-2} + \hat{w}_t,$$

  and $\hat{\sigma}_w = .0094$ (based on 219 degrees of freedom)

  The estimated $AR(1)$ model is

$$x_t = .008_{(.001)} + .347_{(.063)}x_{t-1} + \hat{w}_t,$$

  and $\hat{\sigma}_w = .0095$ (based on 220 degrees of freedom)

$$x_t = .35x_{t-1} + w_t,$$

write in its causal form, $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$, $\psi_0 = 1, \psi_1 = .35, \psi_2 = .123, \cdots, \psi_{10} = 0$, so

$$x_t \approx .35w_{t-1} + .12w_{t-2} + w_t$$

Diagnostics

- Time plot of the residuals, $x_t - \hat{x}_t^{t-1}$, or of the standardized residuals

$$e_t = (x_t - \hat{x}_t^{t-1}) / \sqrt{(\hat{P}_t^{t-1})},$$

where $\hat{x}_t^{t-1}$ is the one-step-ahead prediction of x_t based on the fitted model and $\hat{P}_t^{t-1}$ is the estimated one-step-ahead error variance. If the model fits well, the standardized residuals should behave as an iid sequence with mean zero and variance one.

- Histogram of the residuals, check for the normality
- Q-Q plot, can help in identifying departures from normality
- Plot of the sample autocorrelations of the residuals, $\hat{\rho}_e(h)$ v.s lag h. Check for obvious departures from the independence assumption. For a white noise sequence, the sample autocorrelations are approximately independently and normally distributed with zero means and variance $1/n$.
- Ljung-Box-Pierce Test

Q statistic

$$Q = n(n+2) \sum_{i=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h}$$

Typically, $H = 20$. Under the null hypothesis of model adequacy, asymptotically,

$$Q \sim \chi^2_{H-p-q}, (n \to \infty).$$

Reject the null hypothesis at level α if the value of Q exceeds the $(1-\alpha)$-quantile of the χ^2_{H-p-q}.

- Shapiro-Wilk test (Royston, 1982)

Test for normality.

- Diagnostics for $MA(2)$ fitting

Figure 3: Standardized Residuals Plot, ACF of Residuals and p-values for Ljung-Box Statistics

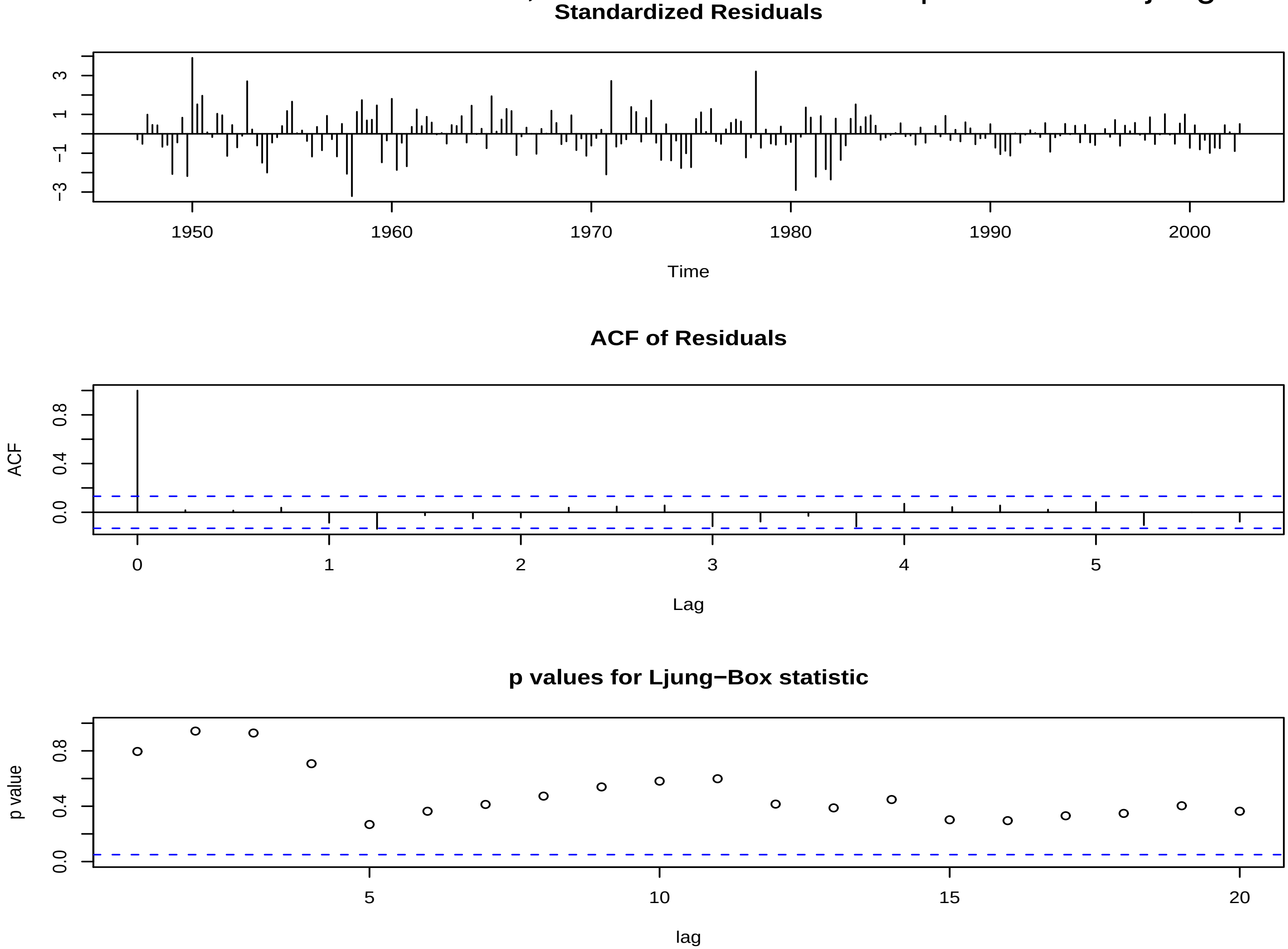
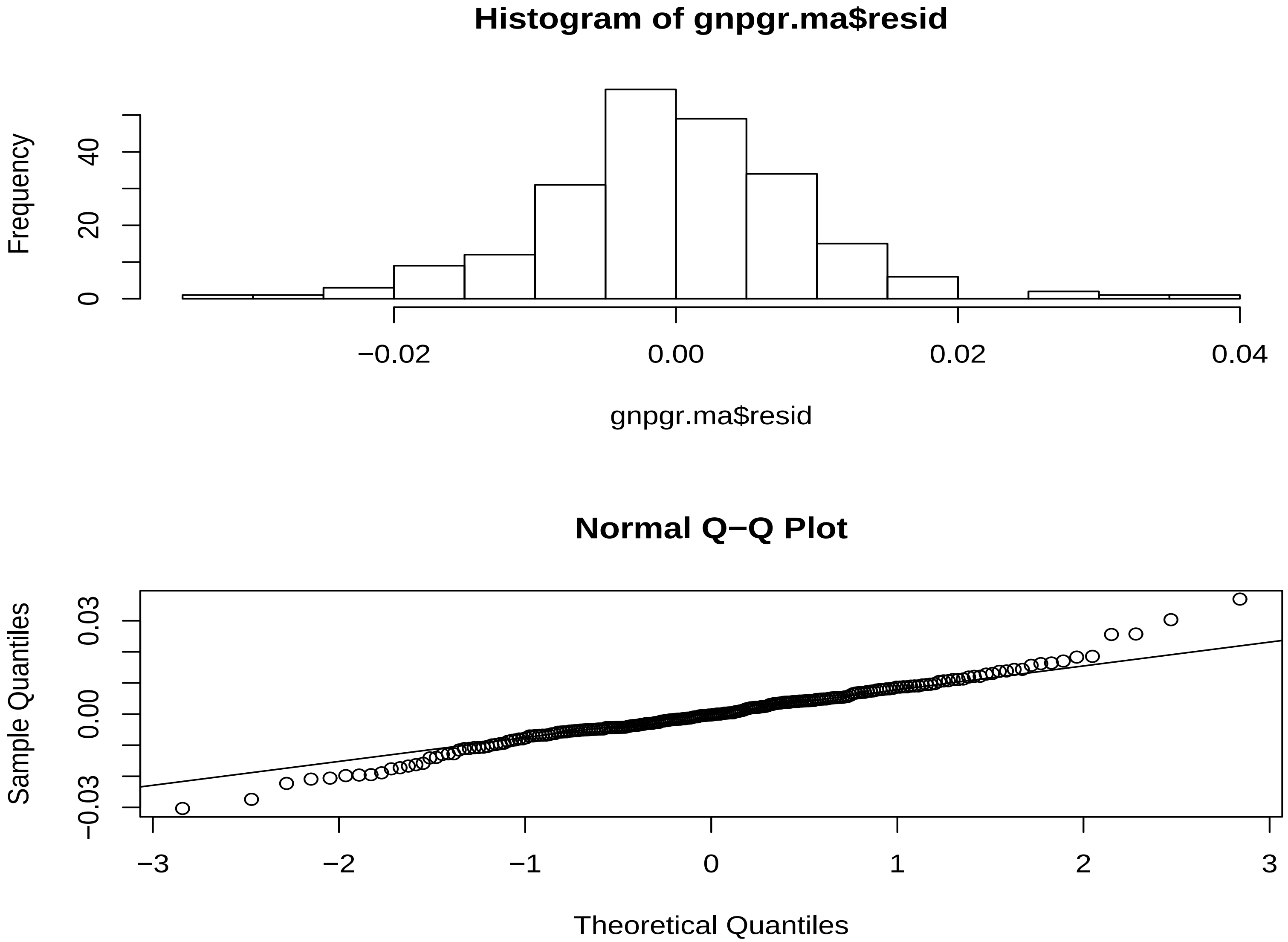


Figure 4: Histogram and Q-Q plot



Shapiro-Wilk test yields a p-value of $.003$, which indicates that the residuals are not normal. Hence, the model appears to fit well except for the fact that a distribution with heavier tails than the normal distribution should be employed.

- Diagnostics for $AR(1)$ fitting

Figure 5: Standardized Residuals Plot, ACF of Residuals and p-values for Ljung-Box Statistics

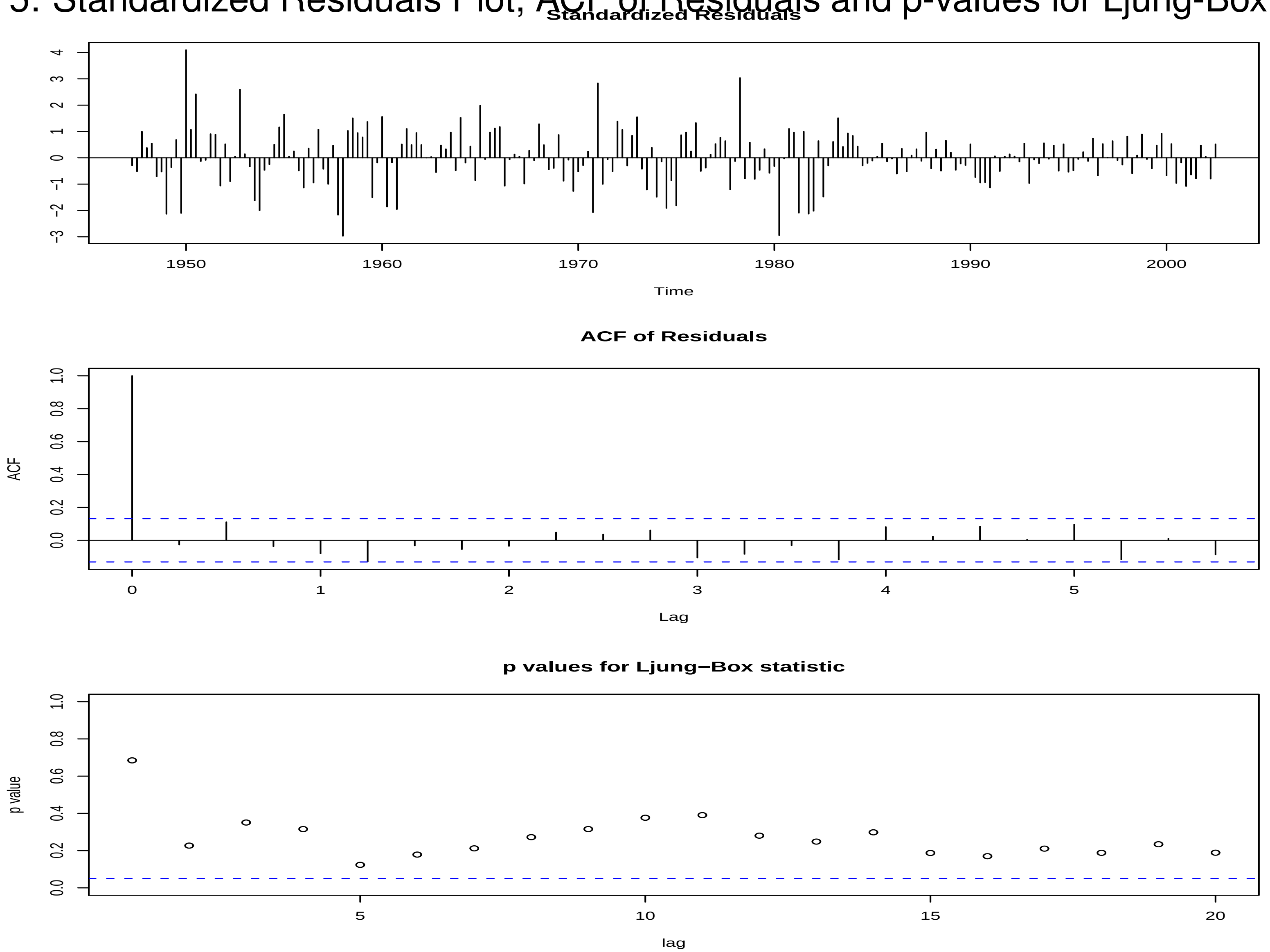
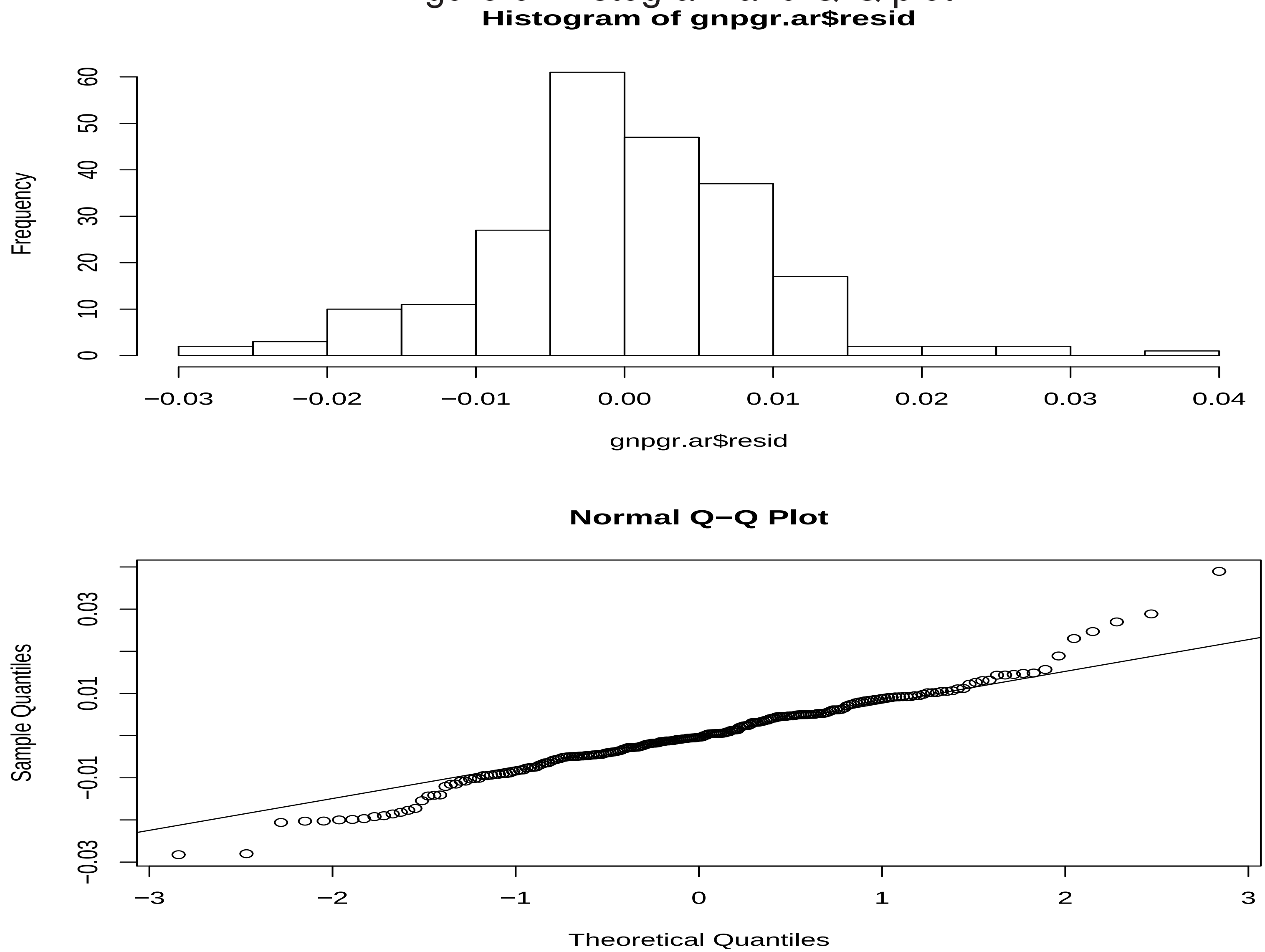


Figure 6: Histogram and Q-Q plot



Shapiro-Wilk test yields a p-value of 0.0006886, which indicates that the residuals are not normal. Hence, the model appears to fit well except for the fact that a distribution with heavier tails than the normal distribution should be employed.

- Choose the final model

	AIC	AICc	BIC
MA(2)	-1431.929	-8.297199	-9.276049
AR(1)	-1431.221	-8.294156	-9.288084

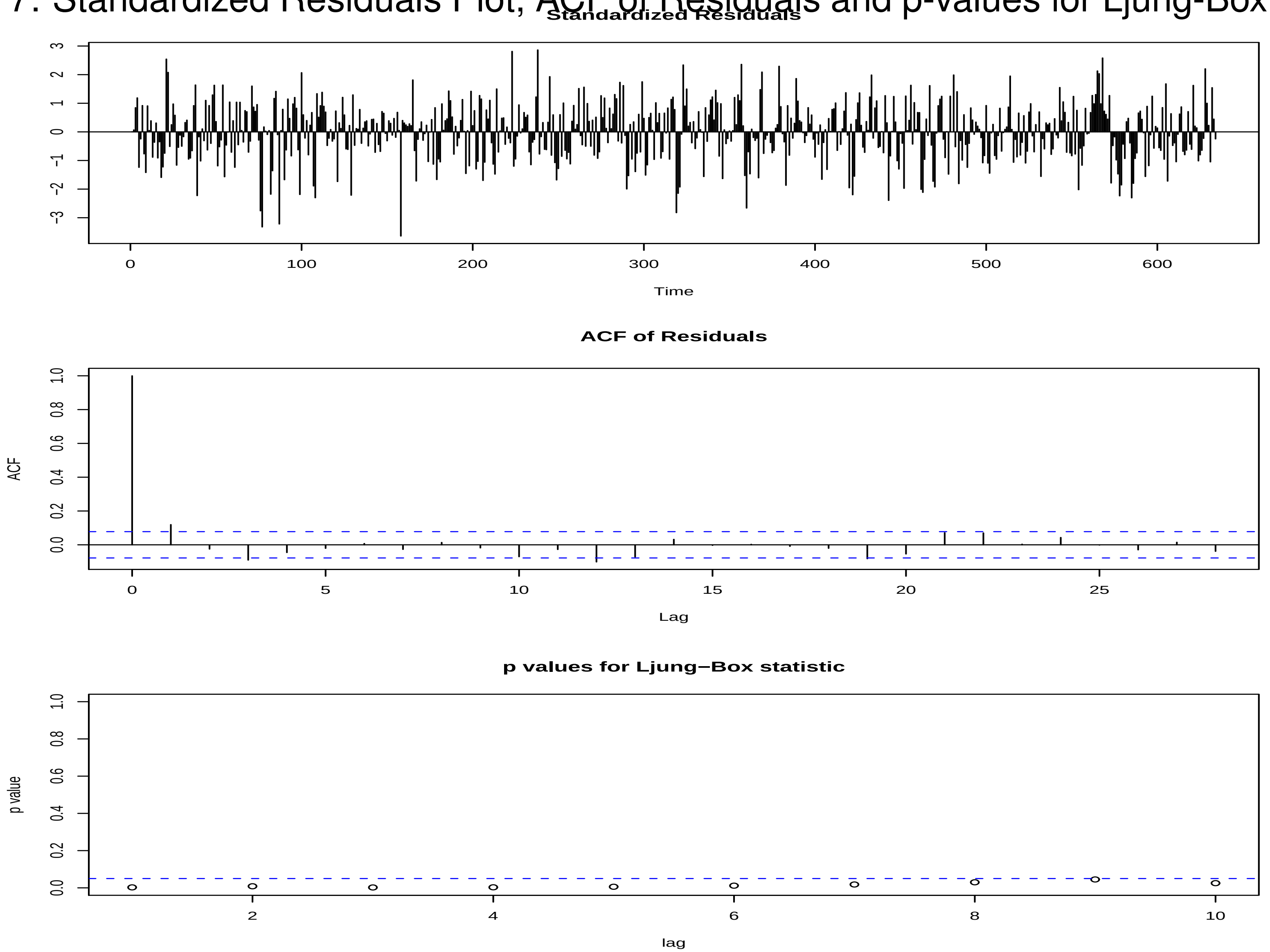
- Both AIC and AICc prefer the $MA(2)$ fit, while the BIC (or SIC) prefers the simpler $AR(1)$ model.

- It is often the case that the BIC will select a model of smaller order than the AIC or AICc
- Both of the model fit data well in this example
- Choose $AR(1)$ because pure autoregressive models are easier to work with

Example 3.37 Diagnostics for the Glacial Varve Series

- First fit $ARIMA(0,1,1)$ model to the logarithms of the glacial varve data.

Figure 7: Standardized Residuals Plot, ACF of Residuals and p-values for Ljung-Box Statistics

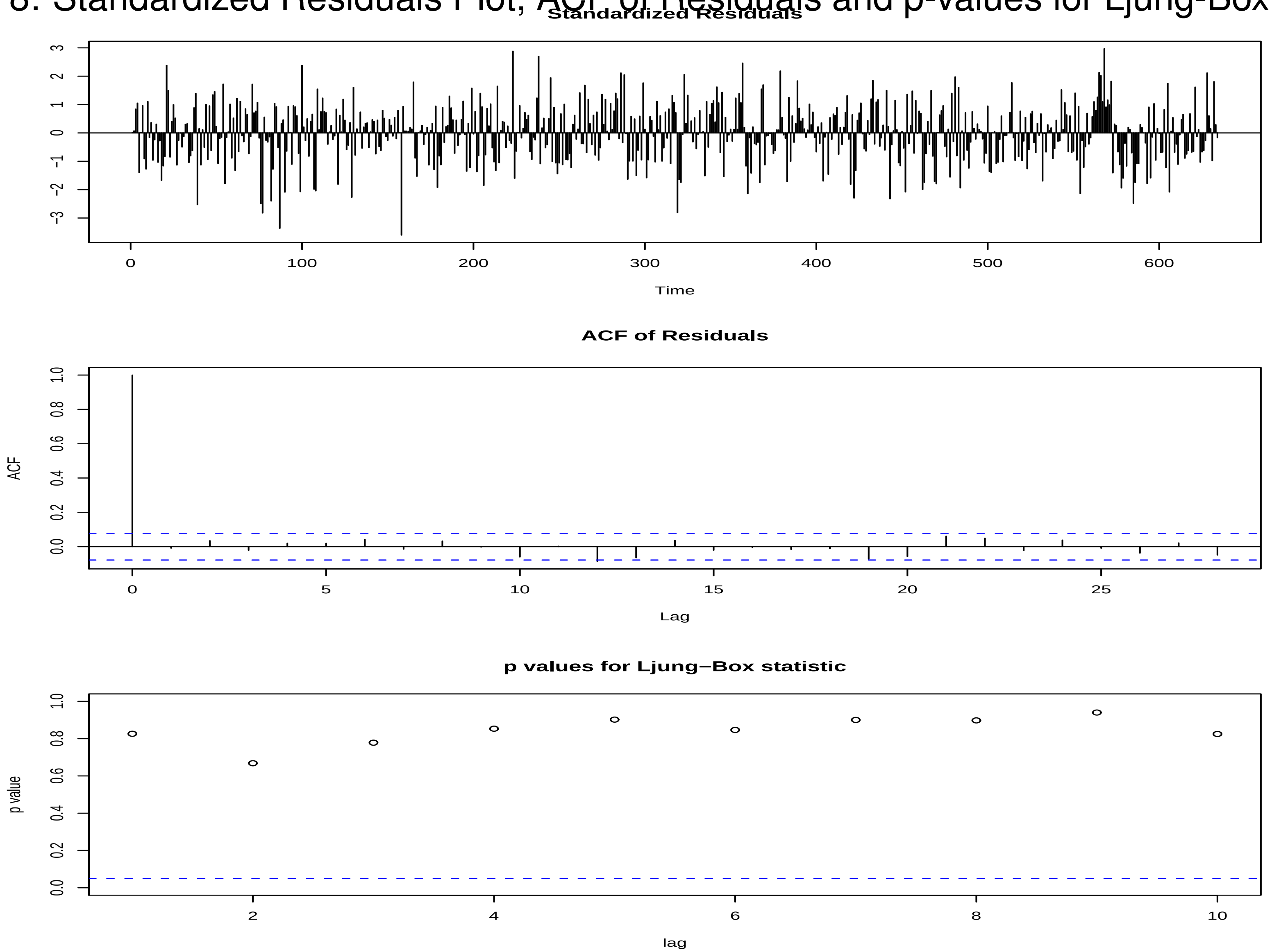


Notice a significant lag 1 displayed. ACF of the residuals appear to be tailing off, an AR term is suggested.

- Fit $ARIMA(1, 1, 1)$ to the logged varve data

$$x_t = .23_{(.05)}x_{t-1} - .89_{(.03)}\hat{w}_{t-1} + \hat{w}_t$$

Figure 8: Standardized Residuals Plot, ACF of Residuals and p-values for Ljung-Box Statistics



- Diagnostics for $ARMA(1,1)$ appear that this model fit the data well.
- AIC for $MA(1)$ is 885.435 and AIC for $ARMA(1,1)$ model is 868.875. AIC criterion also support $ARMA(1,1)$ model.

3.9 Multiplicative Seasonal ARIMA Models

- Seasonal and nonstationary behavior

- Dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag s.

  For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of $s = 12$, because of the strong connections of all activity to the calendar year.

- Natural phenomena such as temperature also have strong components corresponding to seasons.

- It is appropriate to introduce autoregressive and moving average polynomials that identify with the seasonal lags.

  $ARMA(P,Q)_s$ takes the form

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t,$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

  and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}.$$

  $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are the seasonal autoregressive operator and the seasonal moving average operator of orders P and Q, respectively, with seasonal period s.

Note: The pure seasonal $ARMA(P,Q)_s$ is causal only when the roots of $\Phi_P(z^s)$ lie outside the unit circle, and it is invertible only when the roots of $\Theta_Q(z^s)$ lie outside the unit circle.

Example 3.40 A seasonal ARMA series

$$(1-\Phi B^{12})x_t = (1+\Theta B^{12})w_t$$

or

$$x_t = \Phi x_{t-12} + w_t + \Theta w_{t-12}.$$

- First order seasonal autoregressive moving average series that could run over months

- Causal condition requires $|\Phi| < 1$ and invertible condition requires $|\Theta| < 1$.

  Consider first order seasonal $(s = 12)$ MA model,

$$x_t = w_t + \Theta w_{t-12}$$

- $\gamma_{(0)} = (1 + \Theta^2)\sigma^2$
- $\gamma(\pm 12) = \Theta\sigma^2$
- $\gamma(h) = 0, otherwise$
- $\rho(\pm 12) = \Theta/(1 + \Theta^2)$.

Consider first order seasonal $(s = 12)$ AR model,

$$x_t = w_t + \Phi x_{t-12}$$

- $\gamma(h) = \Phi\gamma(h-12), h \geq 1$
- $\gamma_{(0)} = \sigma^2/(1-\Phi^2)$
- $\gamma(\pm 12k) = \sigma^2\Phi^k/(1-\Phi^2), k = 1, 2, \cdots$
- $\gamma(h) = 0, otherwise$
- $\rho(\pm 12k) = \Phi^k, k = 0, 1, 2, \cdots$

Diagnostic Criterion

Behavior of the ACF and PACF for causal and invertible pure seasonal ARMA models

	$AR(P)_s$	$MA(Q)_s$	$ARMA(P,Q)_s$
ACF^*	Tails off at lags ks $k=1,2,\cdots$	Cuts off after lag Qs	Tails off at lag ks
$PACF^*$	cuts off after lag Ps	tails off at lags ks $k=1,2,\cdots$	Tails off at lags ks

The values at nonseasonal lags $h \neq ks$ for $k = 1, 2, \cdots$, are zero.

$ARMA(p,q) \times (P,Q)_s$

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(B)w_t$$

- Combination of the seasonal and nonseasonal operators into a multiplicative seasonal autoregressive moving average model
- Diagnostic criterion are not strictly true for the overall mixed model. We tend to see a mixture of the facts listed for ARMA models and seasonal ARMA models
- Focus on the seasonal autoregressive and moving average component first generally leads to more satisfactory results

Example 3.41 A mixed seasonal model

$ARMA(0,1) \times (1,0)_{12}$ model

$$\Phi_1(B^{12})x_t = \theta(B)w_t$$

$$(1 - \Phi B^{12})x_t = (1 + \theta B)w_t$$

$$x_t = \Phi x_{t-12} + w_t + \theta w_{t-1},$$

where $|\Phi| < 1$ and $|\theta| < 1$.

$$\gamma(0) = \Phi^2\gamma(0) + \sigma_w^2 + \theta^2\sigma_w^2,$$

or

$$\gamma(0) = \frac{1 + \theta^2}{1 - \Phi^2}\sigma_w^2.$$

Multiplying the model by $x_{t-h}, h > 0$, and taking expectations, we have

$$\gamma(1) = \Phi\gamma(11) + \theta\sigma_w^2$$

$$\gamma(h) = \Phi\gamma(h-12), h \geq 2$$

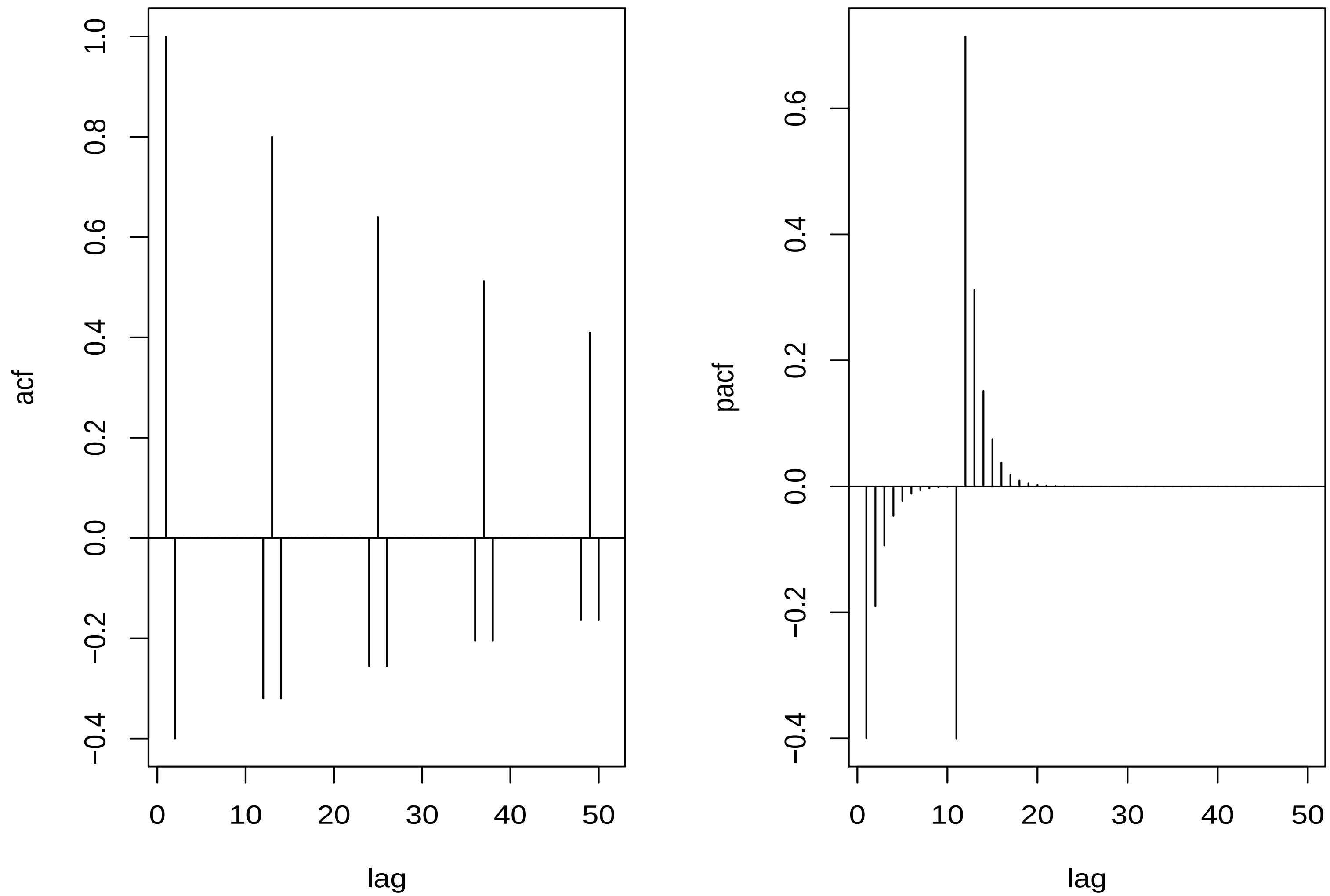
ACF of this model is

$$\rho(12h) = \Phi^h, h = 1, 2, \cdots$$

$$\rho(12h-1) = \rho(12h+1) = \frac{\theta}{1+\theta^2}\Phi^h, h = 0, 1, 2, \cdots$$

$$\rho(h) = 0, \text{otherwise}$$

Figure 9: ACF and PACF for the model $x_t = \phi x_{t-12} + w_t + \theta w_{t-1}$ with $\phi = .8$ and $\theta = -.5$



Difference of Seasonal Data

$$x_t = S_t + w_t,$$

- x_t is a temperature time series data for example
- S_t is a seasonal component that varies slowly from one year to the next, according to a random walk $S_t = S_{t-12} + v_t$
- w_t and v_t are uncorrelated white noise processes.
- $(1 - B^{12})x_t = x_t - x_{t-12} = v_t + w_t - w_{t-12}$ is $MA(1)_{12}$

Seasonal Difference of order D

$$\nabla_s^D x_t = (1 - B^s)^D x_t,$$

where $D = 1, 2, \cdots$. Typically, $D = 1$ is sufficient to obtain seasonal stationarity.

Definition 3.13 The multiplicative **seasonal autoregressive integrated moving average** model, or **SARIMA** model. of Box and Jenkins (1970) is given by

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \alpha + \Theta_Q(B^s)\theta(B)w_t,$$

where w_t is the usual Gaussian white noise process. The general model is denoted as $\mathbf{ARIMA(p,d,q)} \times \mathbf{(P,D,Q)_s}$. The ordinary autoregressive and moving average components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q, respectively, and the seasonal autoregressive and moving average components by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q and ordinary and seasonal difference components by $\nabla^d = (1-B)^d$ and $\nabla_s^D = (1-B^s)^D$.

Example 3.42 SARIMA Model

Model $ARIMA(0,1,1) \times (0,1,1)_{12}$ often provides a reasonable representation for seasonal, nonstationary, economic time series.

$$\nabla_{12}^1 \nabla^1 x_t = \Theta_1(B^{12})\theta(B) w_t$$

$$(1 - B^{12})(1 - B) x_t = (1 + \Theta B^{12})(1 + \theta B) w_t.$$

$$(1 - B - B^{12} + B^{13}) x_t = (1 + \theta B + \Theta B^{12} + \Theta\theta B^{13}) w_t,$$

or in difference equation form

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + w_t + \theta w_{t-1} + \Theta w_{t-12} + \Theta\theta w_{t-13}.$$

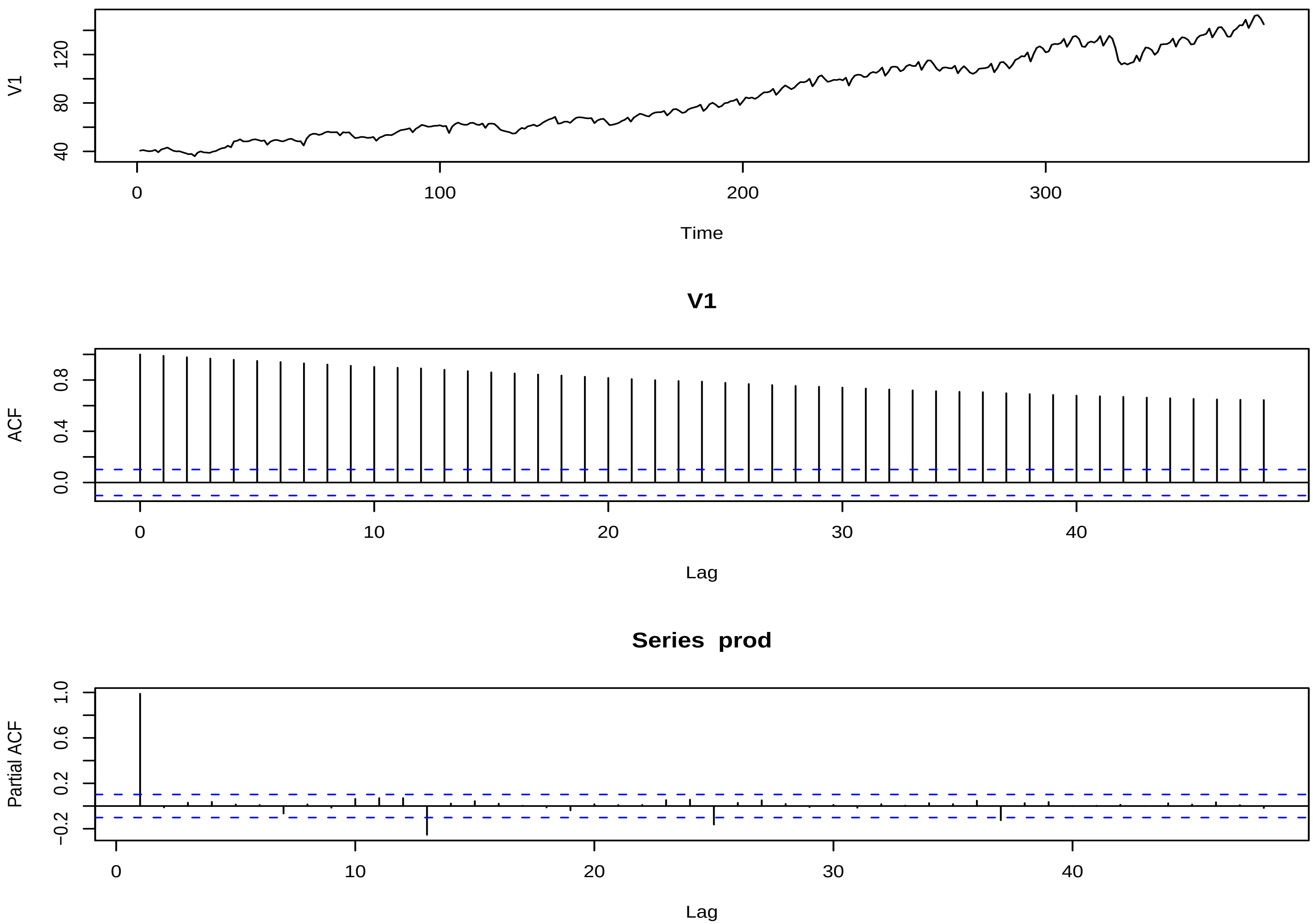
Selecting Appropriate Model

- Plot data
- possibly transform the data
- identifying the dependence orders of the model
- parameter estimation
- diagnostics
- model selection

Analysis of the Federal Reserve Board Production Index

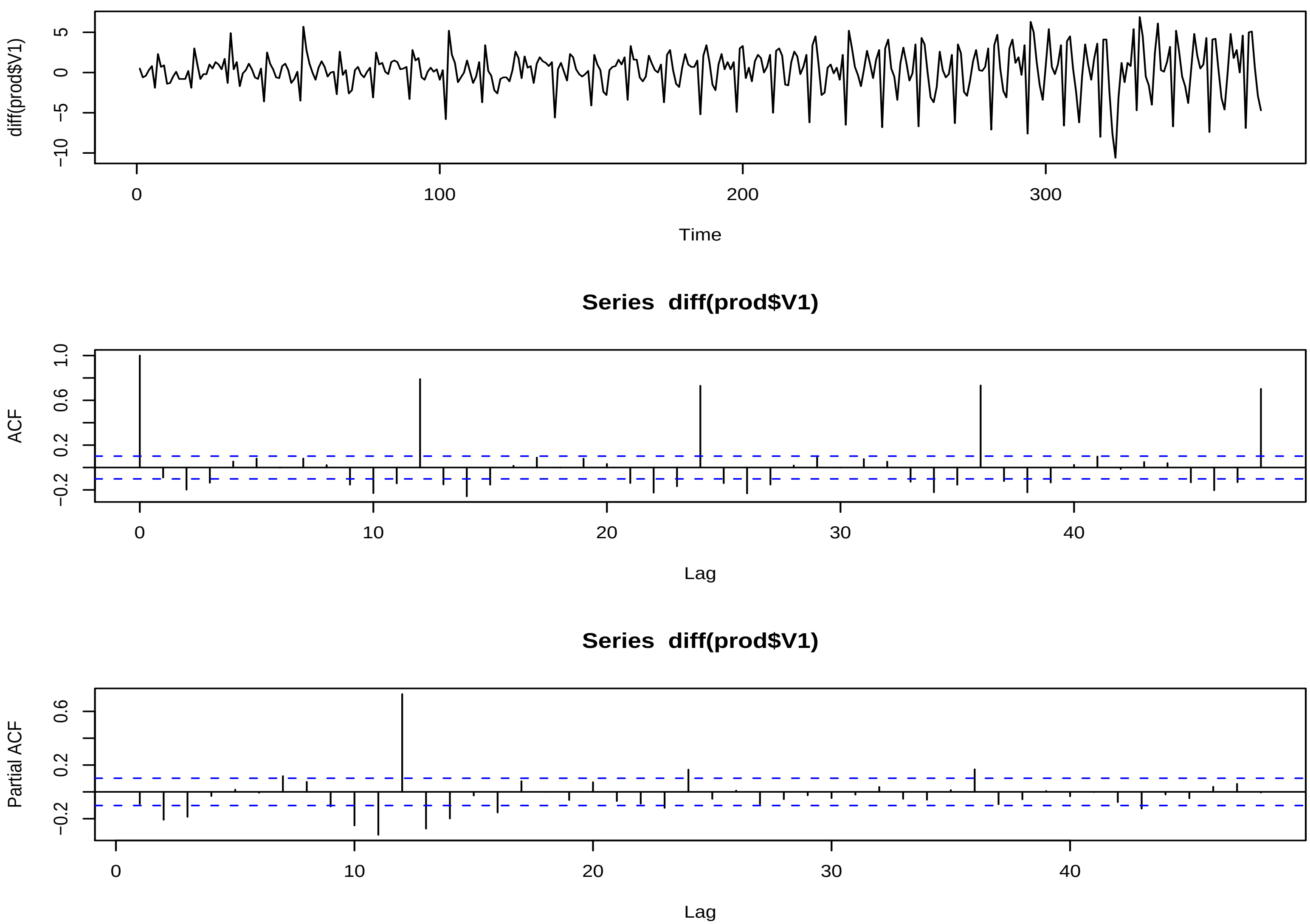
- identifying a model
- producing forecasts

Figure 10: Plot, ACF and PACF for Federal Reserve Board Production Index



- Slow decay in the ACF
- peak at lag $h = 1$ in the PACF
- both indicate nonstationary behavior

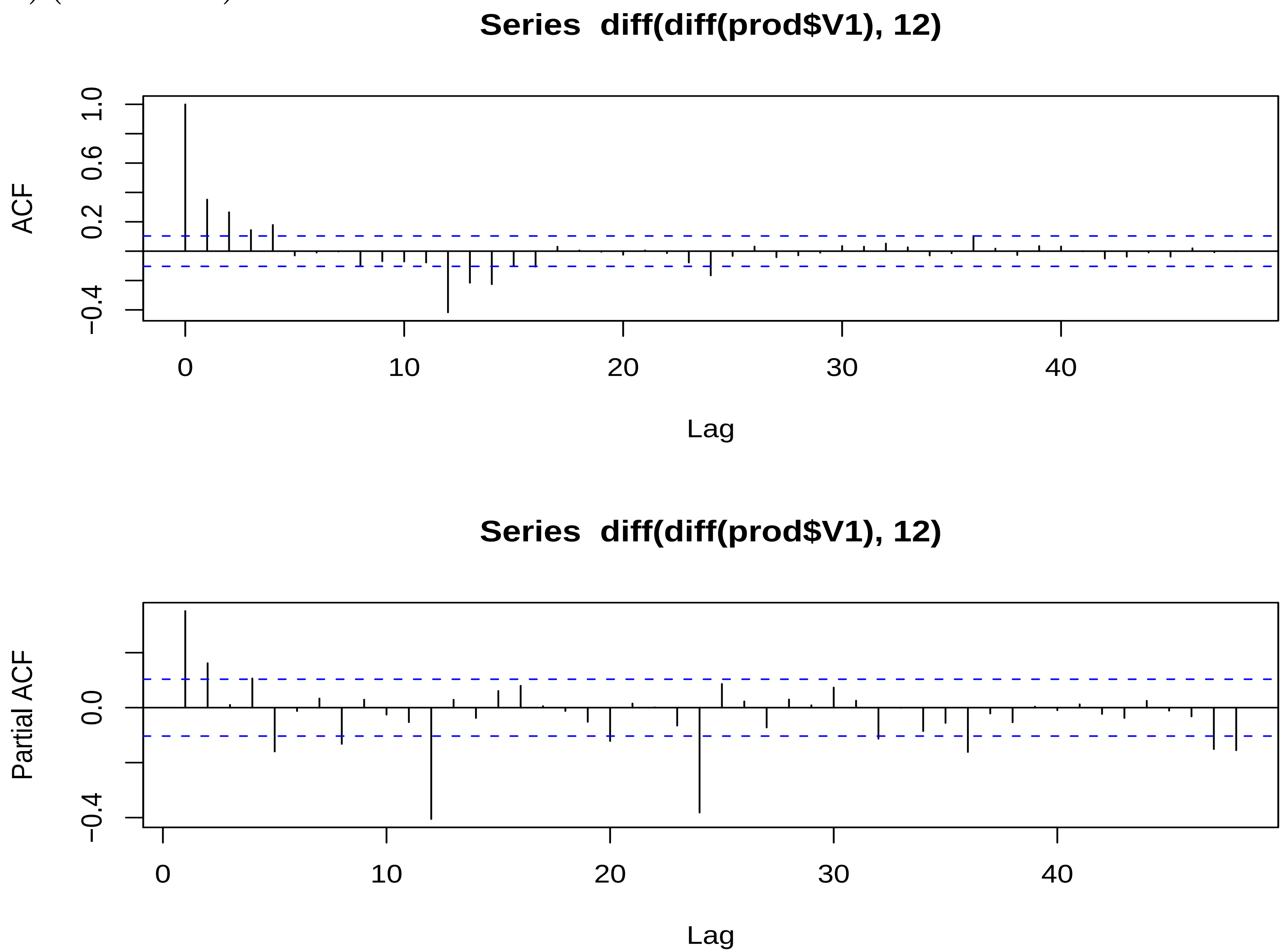
Figure 11: Plot, ACF and PACF of differenced production, $(1 - B)x_t$.



- Noting the peaks at $12, 24, 36,$ and 48 with relatively slow decay suggested a seasonal difference

$$\nabla_{12}\nabla x_t = (1 - B^{12})(1 - B)x_t$$

Figure 12: ACF and PACF of first differenced and then seasonally differenced production, $(1-B)(1-B^{12})x_t$.



- ACF and PACF tend to show a strong peak at $h = 12$ in the autocorrelation function, with smaller peaks appearing at $h = 24, 36$, combined with peaks at $h = 12, 24, 36, 48$, in the partial auttocorrelation function.

- Either a seasonal moving average of order $Q = 1$, a seasonal autoregression of possible order $P = 2$, or due to the fact that both the ACF and PACF may be tailing off at the seasonal lags, perhaps both components, $P = 2, Q = 1$ are needed

- Inspecting ACF and PACF at the within season lags, $h = 1, 2, \cdots, 11$, it appears that both the ACF and PACF are tailing off. We should consider fitting a model with both $p >$

$0, q > 0$ fir the nonseasonal components. Consider $p = 1, q = 1$.

- Fitting three models

$$ARIMA(1,1,1) \times (0,1,1)_{12}$$

$$ARIMA(1,1,1) \times (2,1,0)_{12}$$

$$ARIMA(1,1,1) \times (2,1,1)_{12}$$

Figure 13: Diagnostics for the $ARIMA(1,1,1) \times (0,1,1)_{12}$ fit on the production data.

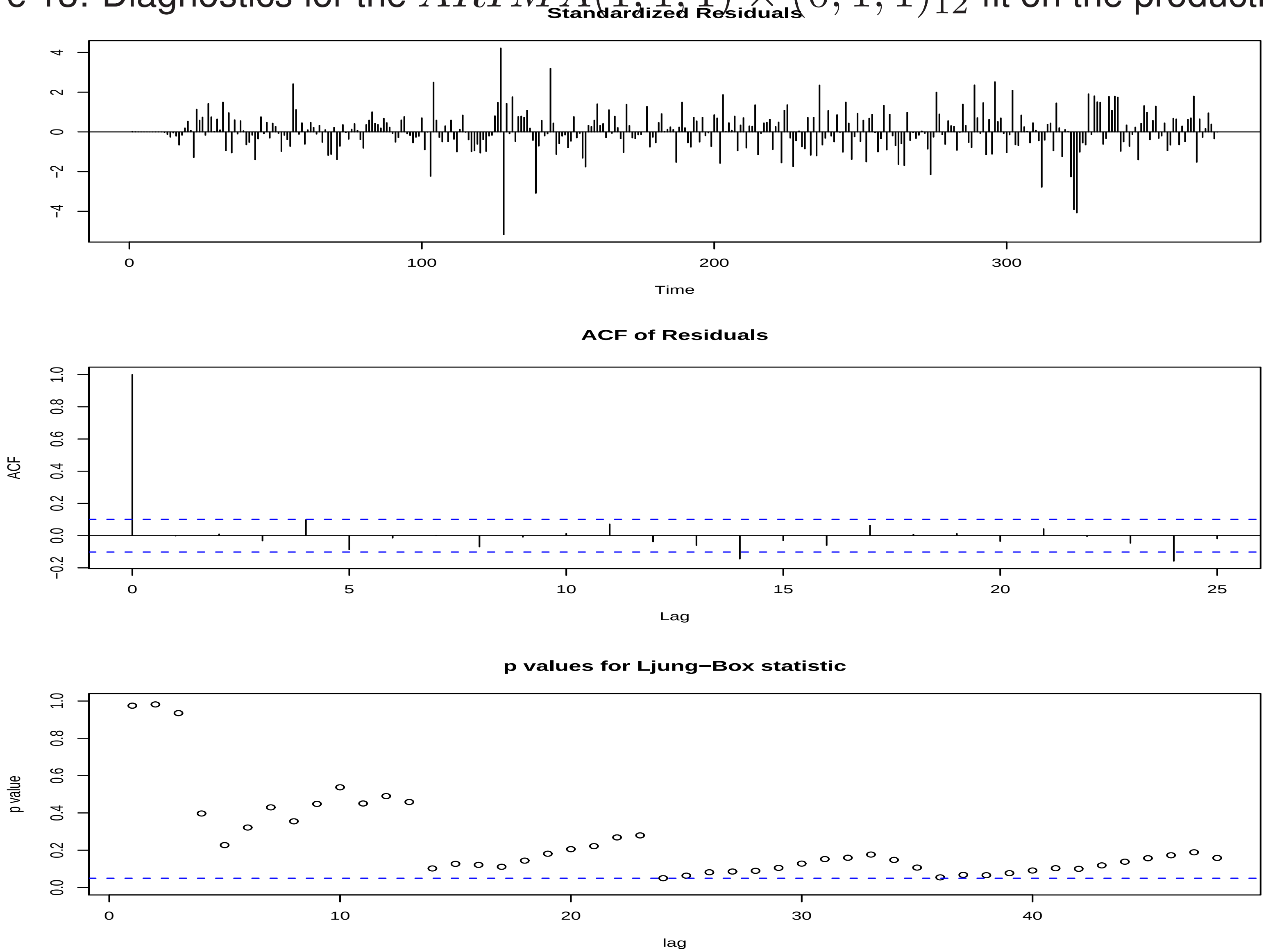


Figure 14: Diagnostics for the $ARIMA(1,1,1) \times (0,1,1)_{12}$ fit on the production data.

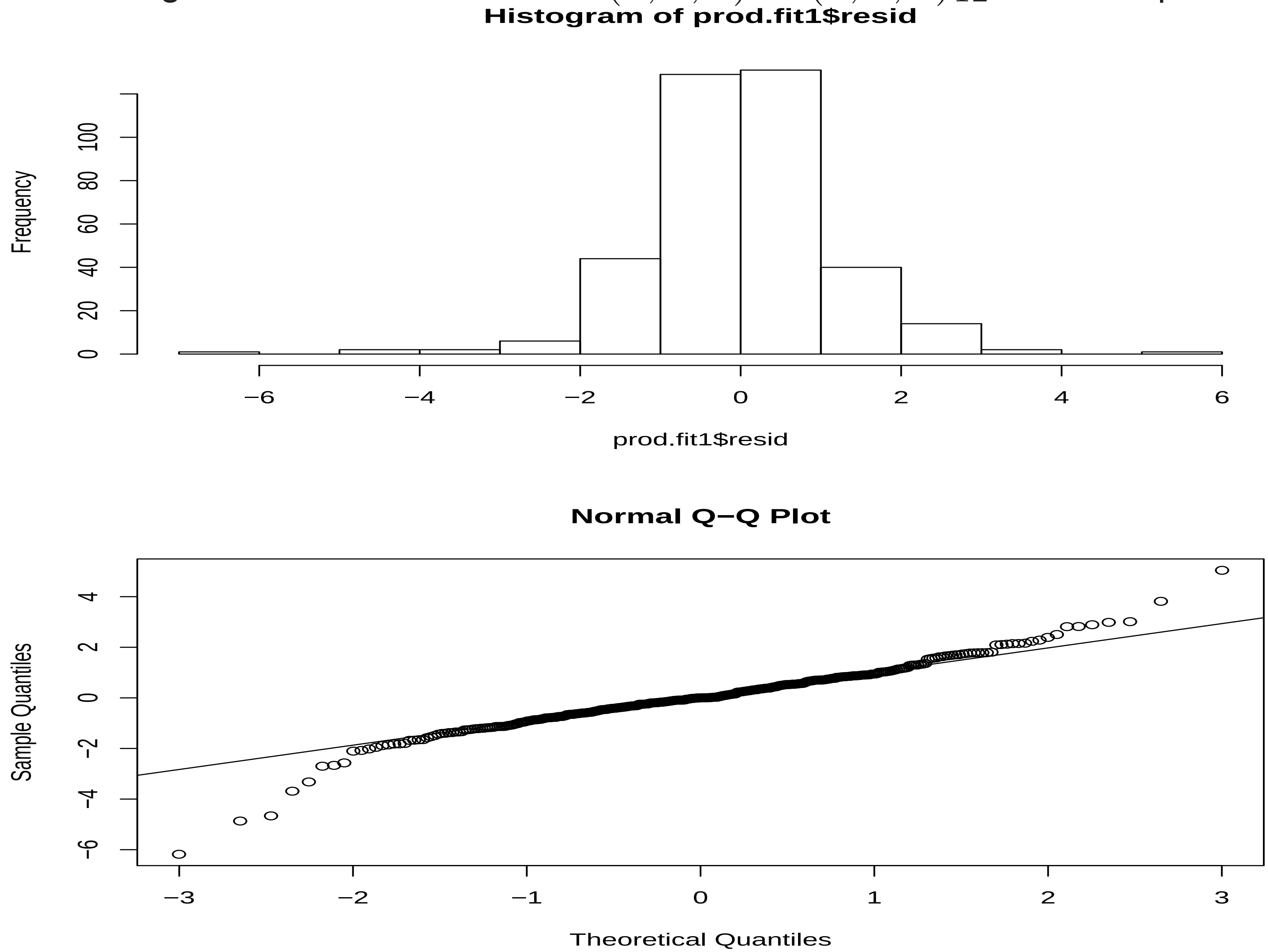


Figure 15: Diagnostics for the $ARIMA(1,1,1) \times (2,1,0)_{12}$ fit on the production data.

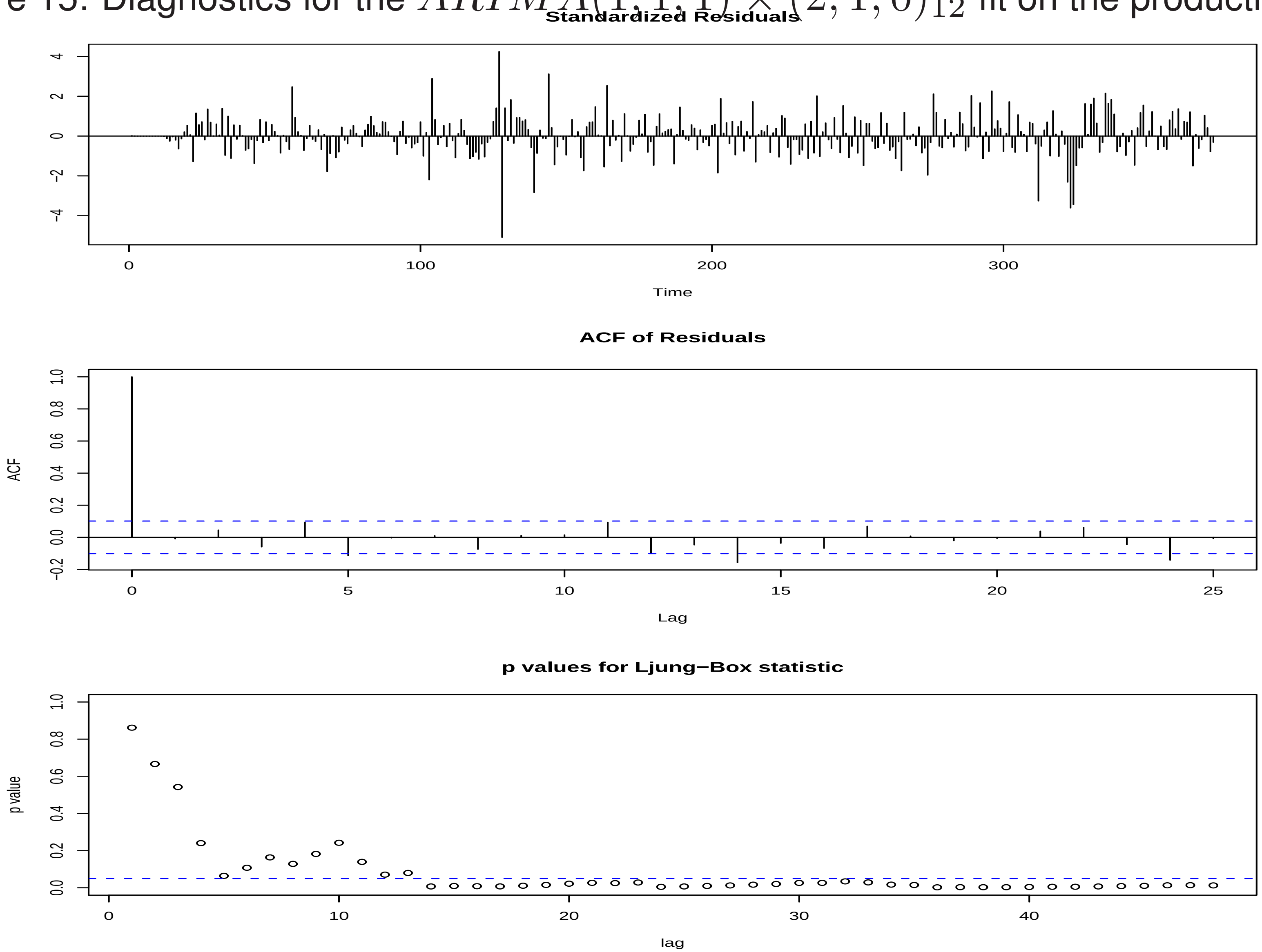


Figure 16: Diagnostics for the $ARIMA(1,1,1) \times (2,1,0)_{12}$ fit on the production data.

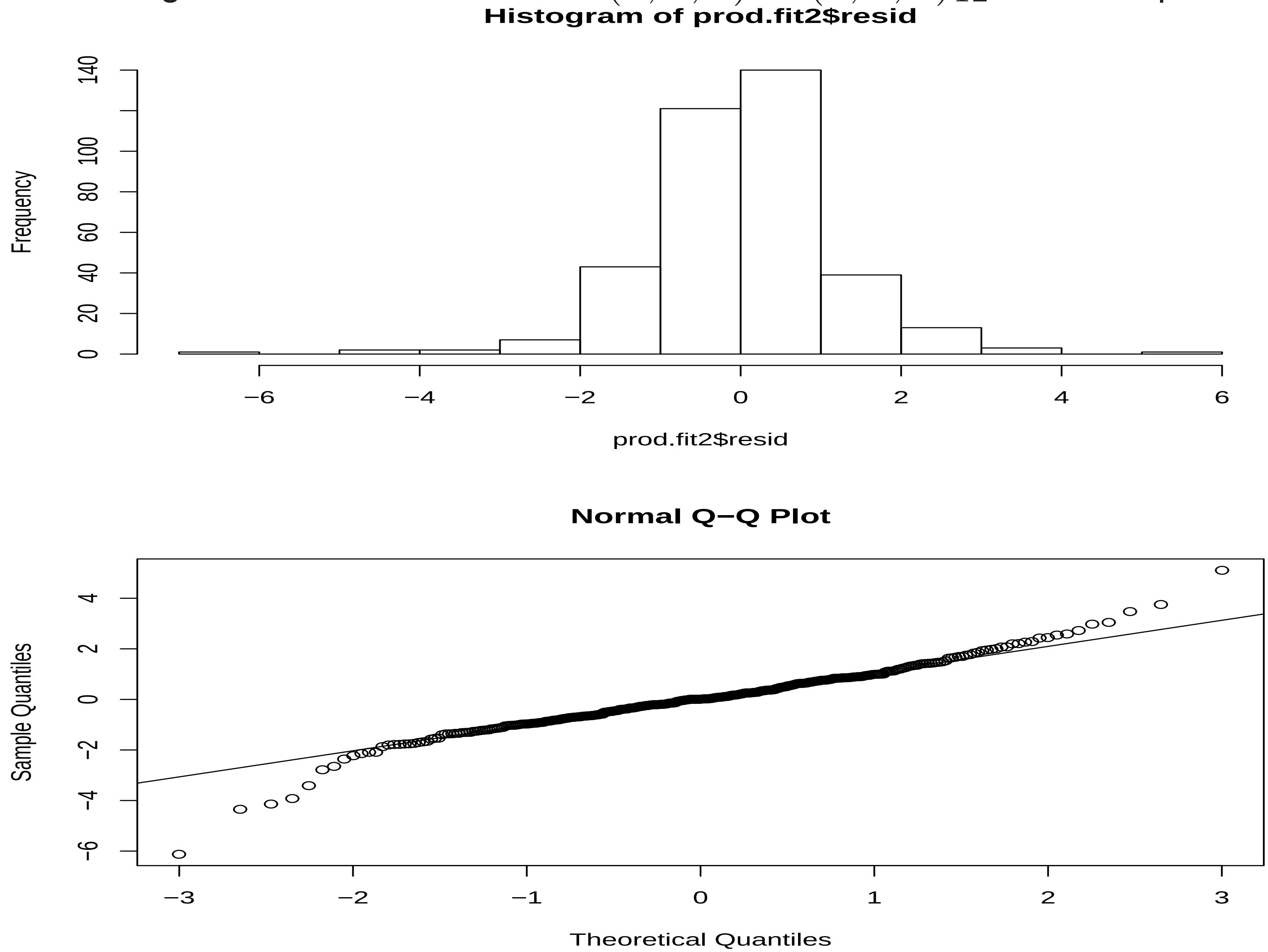


Figure 17: Diagnostics for the $ARIMA(1,1,1) \times (2,1,1)_{12}$ fit on the production data.

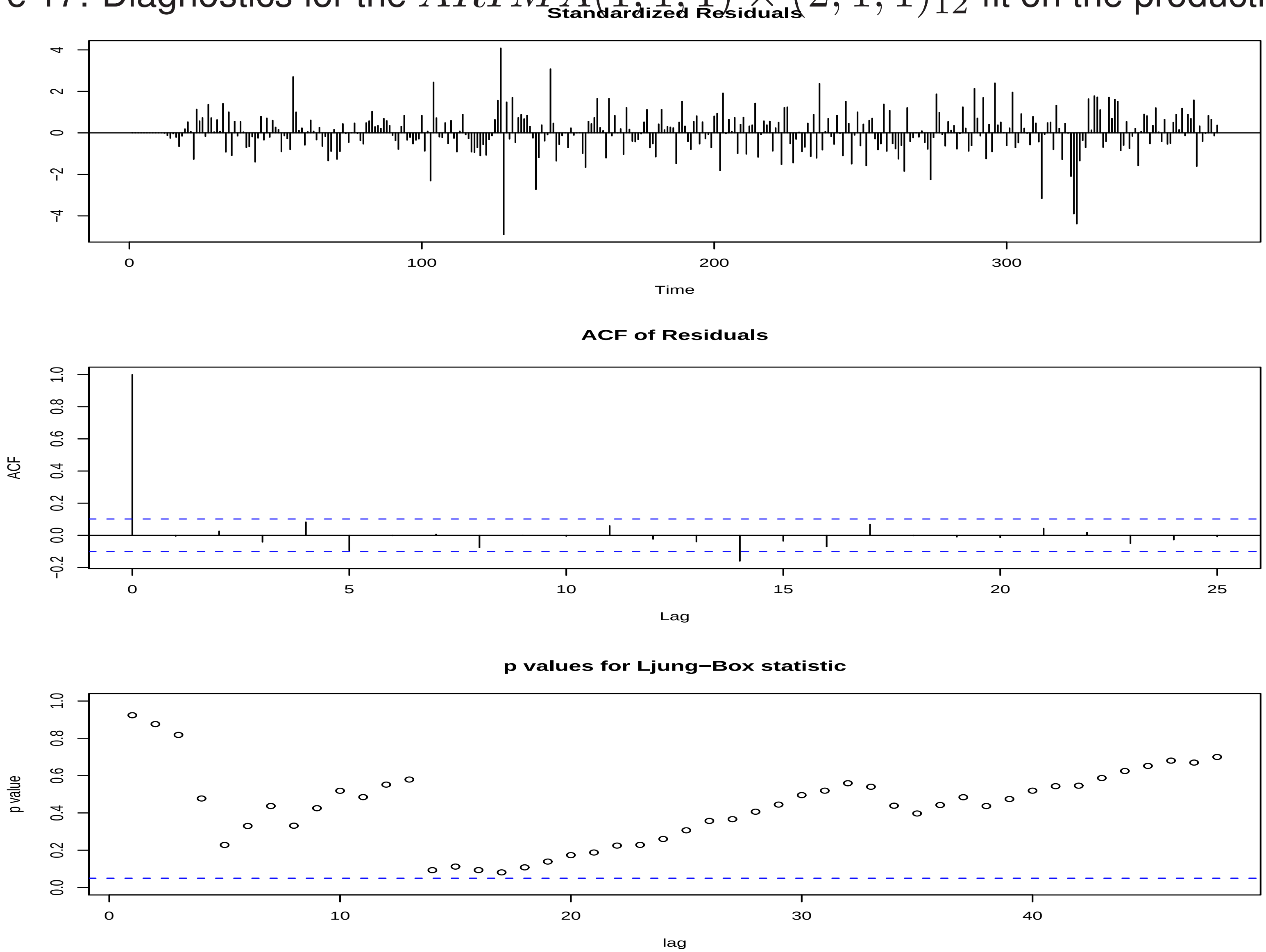
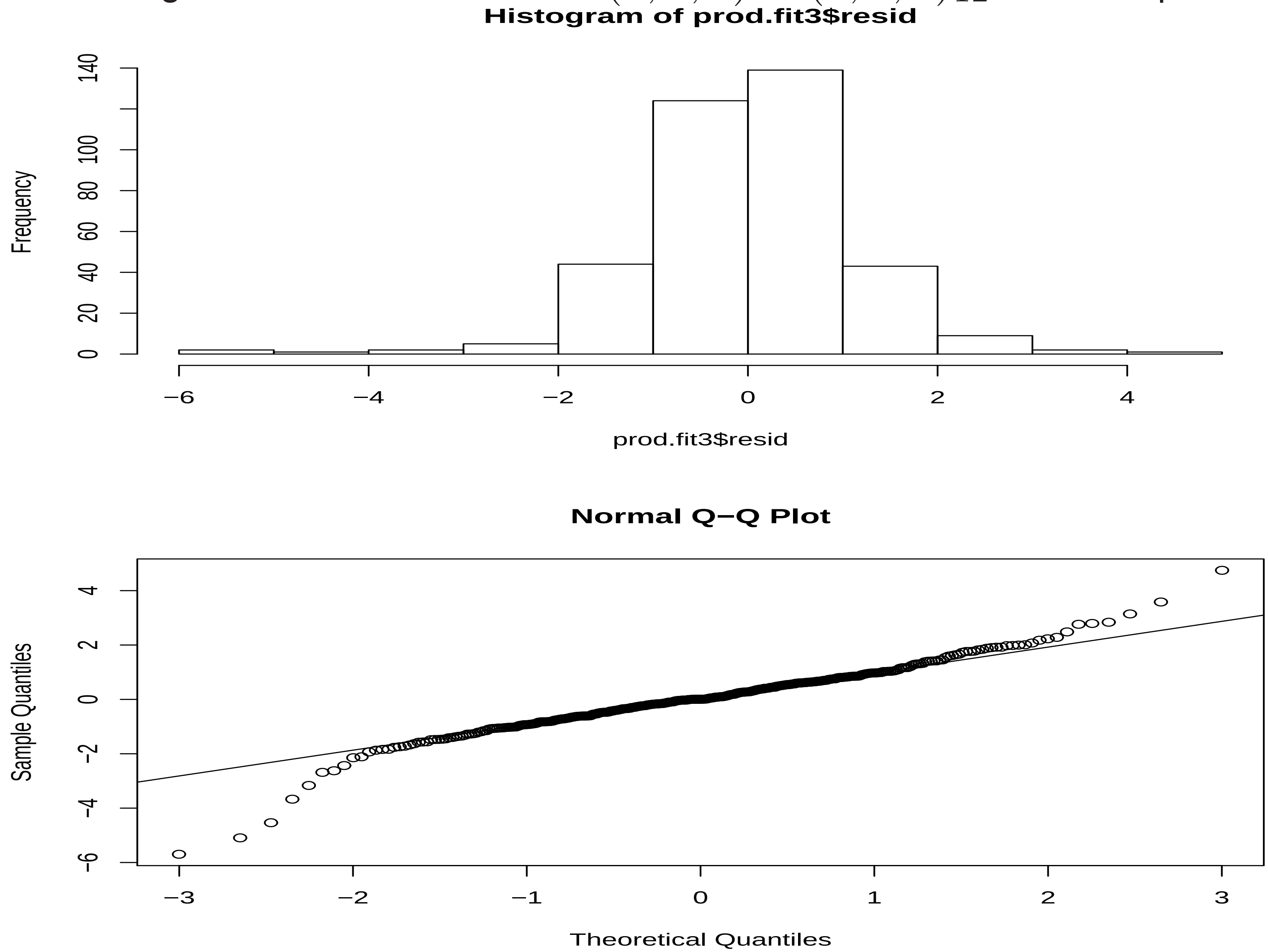


Figure 18: Diagnostics for the $ARIMA(1,1,1) \times (2,1,1)_{12}$ fit on the production data.



- Fitting three models

$$ARIMA(1,1,1) \times (0,1,1)_{12}, AIC = 1162.30$$

$$ARIMA(1,1,1) \times (2,1,0)_{12}, AIC = 1169.04$$

$$ARIMA(1,1,1) \times (2,1,1)_{12}, AIC = 1148.43$$

- On the basis of AICs, we prefer the

$$ARIMA(1,1,1) \times (2,1,1)_{12}$$

- $ARIMA(1,1,1) \times (2,1,1)_{12}$ fit is adequate

- A few outliers present

The fitted $ARIMA(1,1,1) \times (2,1,1)_{12}$

$$(1 + .22_{(.08)}B^{12} + .28_{(.06)}B^{24})(1 - .58_{(.11)}B)\nabla_{12}\nabla\hat{x}_t$$
$$= (1 - .50_{(.07)}B^{12})(1 - .27_{(.13)}B)\hat{w}_t$$

with $\hat{\sigma}_w^2 = 1.35$.

Figure 19: Forecasts and limits for production index. The vertical dotted line separates the data from the prediction.