Data 9. The total monthly expenditure on restaurants and takeaway food services in Australia, April 1980 - April 2015

Alejandro Astor Valentin   Ana Gil Molina

May 2024

Introduction

Review

- $\{X_t\}$ time series.
- We have observed $\{x_1, x_2, \ldots, x_T\}$ a realization of length T of $\{X_t\}$.
- We want to fit to the observed time series an ARIMA model,

$$\Phi_p(B^s)\phi_p(B)\nabla_s^D\nabla^d X_t = c + \theta_q(B)\Theta_Q(B^s)Y_t$$

- Box and Jenkins (1976) have proposed the following **three steps procedure**:

1. Identifying possible ARIMA models
2. Estimation and selection of ARIMA models
3. Diagnosis

Data overview

Time series of the total monthly expenditure on restaurants and takeaway food services in Australia, from April 1980 to April 2015.

- The length of the time series is 408 months ($T = 408$).

	x
	<dbl>
1	0.3424
2	0.3421
3	0.3287
4	0.3385
5	0.3315
6	0.3419

Figure: First values of the time series.

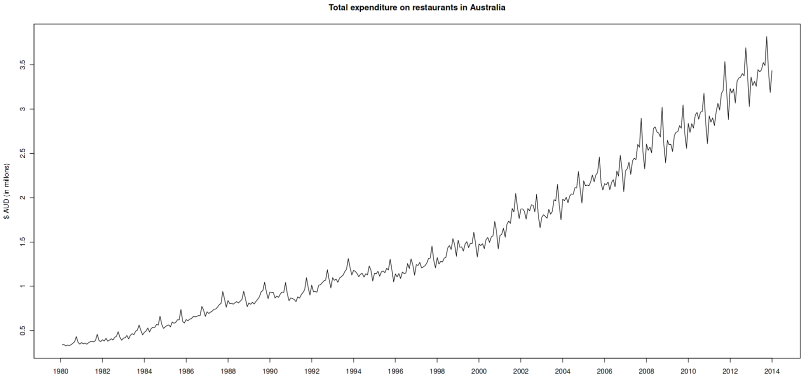


Figure: Data set representation.

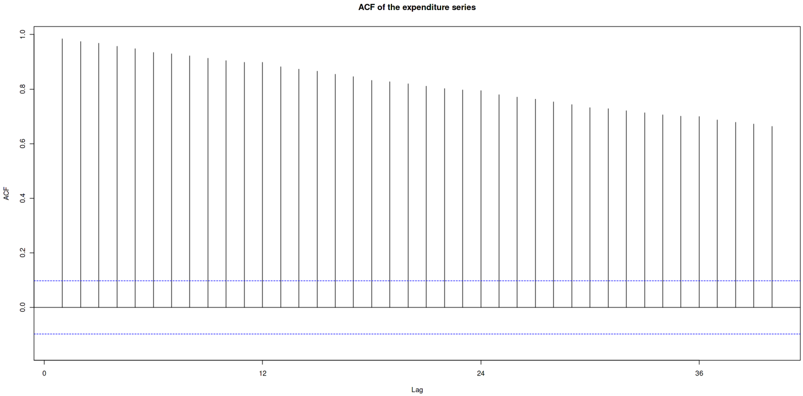


Figure: ACF from data set.

Is the time series stationary?

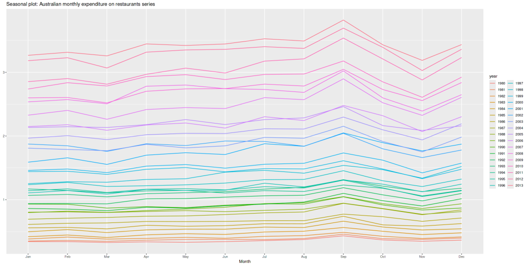


Figure: Seasonal plot from data set.

- The **variance increases** with the level of the series $\Rightarrow$ The **variance** is **not constant**.
- The presence of a **trend component** $\Rightarrow$ The **mean** is **not constant**.
- The presence of a **seasonal component** with period $s = 12 \Rightarrow$ The **mean** is **not constant**.

Remark

The presence of the seasonal component can actually be appreciated in the graph of the series.

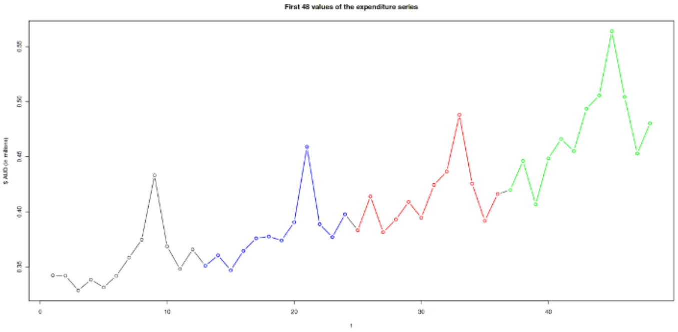


Figure: First 48 values of the Australian expenditure on restaurants series.

Variance stabilizing transformations

Box-Cox transformations

- Let $\{X_t\}$ be a time series.
- Let $\mu_t = E[X_t]$, $\sigma_t = \sqrt{Var[X_t]}$
- If $\sigma_t = k\mu_t^{1-\lambda}$ then the process $\{\tilde{X}_t\}$, with

$$\tilde{X}_t = \frac{X_t^{\lambda} - 1}{\lambda}$$

  has constant variance.
- For $\lambda \approx 0$, $\tilde{X}_t \approx \log(X_t)$

Remark

If the plot between variability and mean does not show a linear relationship, the Box-Cox transformations do not stabilize the variance.

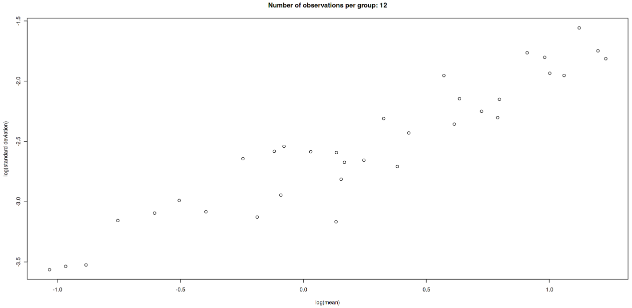


Figure: Relationship between **log(mean)** and **log(sd)**. Number of observations per group: 12.

Estimation of the parameter λ

- Form k consecutive groups, $I_1, I_2, \ldots, I_k$, each with approximately $\tilde{n}$ observations.
- Denote by $\bar{x}_i$ the mean and by s_i the standard deviation of the observations in $I_i, i = 1, 2, \ldots, k$. Fit the regression model:

$$\log s_i = \alpha + \beta \log(\bar{x}_i)$$

- $\hat{\lambda} = 1 - \hat{\beta}$

Remark

With seasonal time series of period s the groups size should be set equal to $s \Rightarrow \tilde{n} = 12$.

Plot of the transformed time series

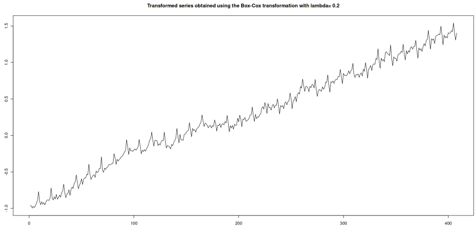


Figure: Transformed time series and estimated value of λ.

We represent the time series $\tilde{X}_t = \frac{X_t^\lambda - 1}{\lambda}$, where $\lambda = 0.2$.

Verification $\lambda \approx 0$

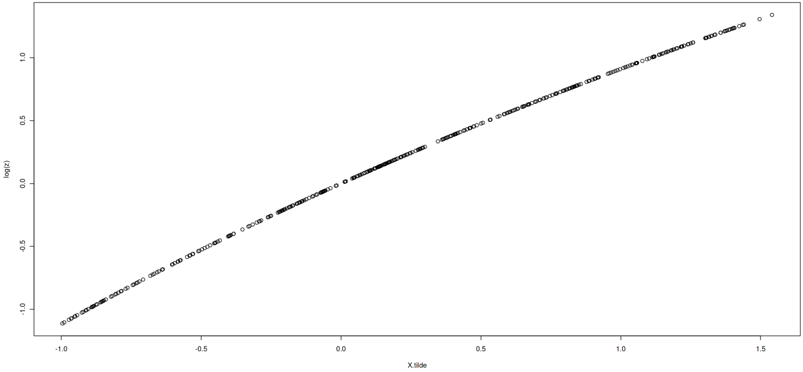


Figure: Representation of $\tilde{X}_t$ against $\log(X_t)$.

Mean stabilizing transformations

To stabilize the mean of the series we apply regular and seasonal differences.

1. **Determine the order of regular difference (d):** apply the regular difference operator $\nabla^d = (1-B)^d, d \geq 1$ when:
   - The graph of the series suggests the presence of a trend component.
   - The sample ACF does not fade for high values of the lag. In particular, if the ACF shows a slow and linear decay.
2. **Determine the order of seasonal differencing (s,D):** apply the seasonal differencing operator $(D = 1)$ of order s, $\nabla_s = (1 - B^s)$ when:
   - The graph of the series shows a repeating pattern of period s.
   - The ACF shows positive coefficients that slowly decrease in the lags $s, 2s, 3s, \ldots$

Removing trend component

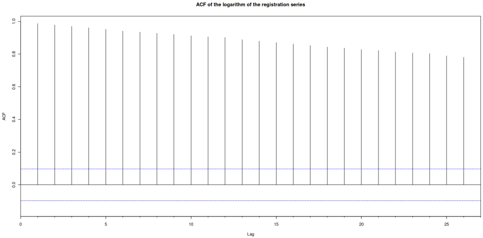


Figure: ACF from $\tilde{X}_t$.

ACF of **non**-stationary series **decays slowly and linearly**.

It suggests us to set $d = 1$.

Let $W_t = \nabla \tilde{X}_t = \tilde{X}_t - \tilde{X}_{t-1} = log(X_t) - log(X_{t-1})$.

Removing trend component

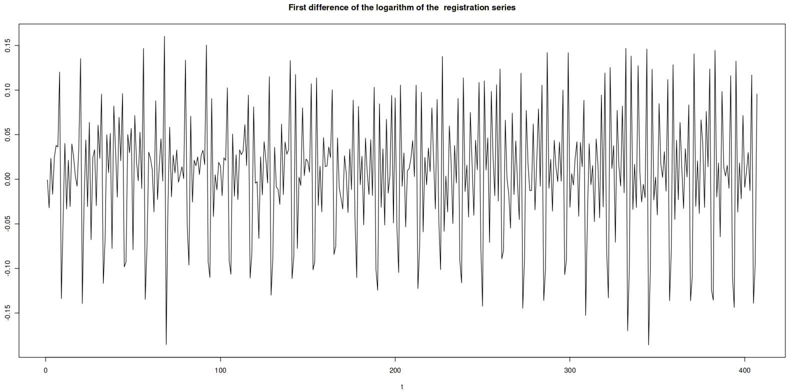


Figure: Representation of $W_t = \nabla \tilde{X}_t$.

Removing seasonal component

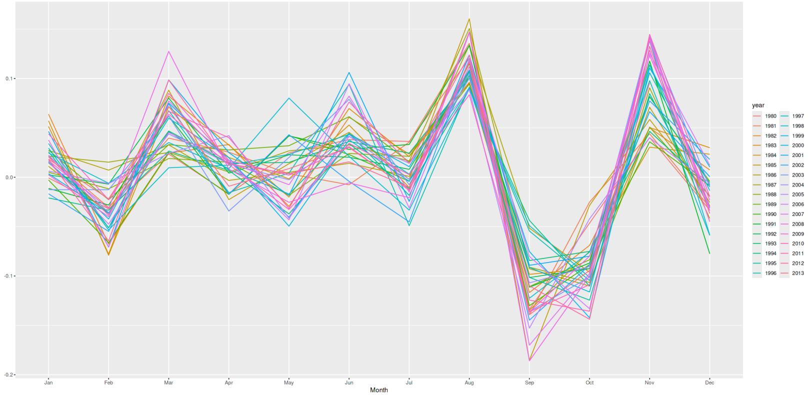


Figure: There is not trend component but there is a seasonal component.

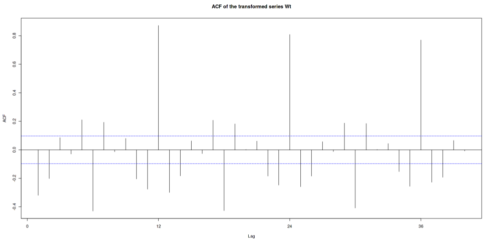


Figure: ACF of W_t.

The graph of the ACF shows positive coefficients that slowly decrease in the lags $12, 24, 36, \ldots$.
It suggests us to take $s = 12$.

Both the seasonal plot and the ACF suggest the presence of a seasonal component of period $s = 12 \Rightarrow$ We apply the seasonal differencing operator $\nabla_{12} = (1 - B^{12})$ to remove the seasonal component.

Remark

The presence of the seasonal component cannot be clearly appreciated in the graph of the transformed series $\{W_t\}$.

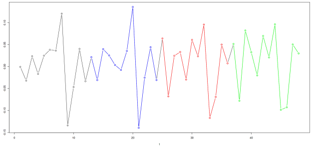


Figure: First 48 values of the transformed series $\{W_t\}$.

Removing seasonal component

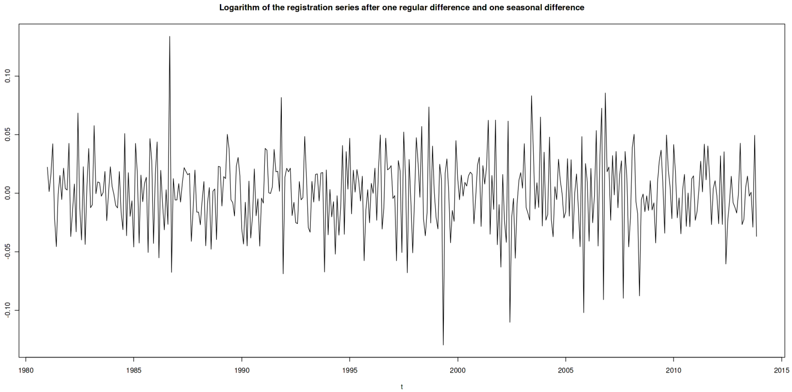


Figure: Apparently there is no seasonal component.

Removing seasonal component

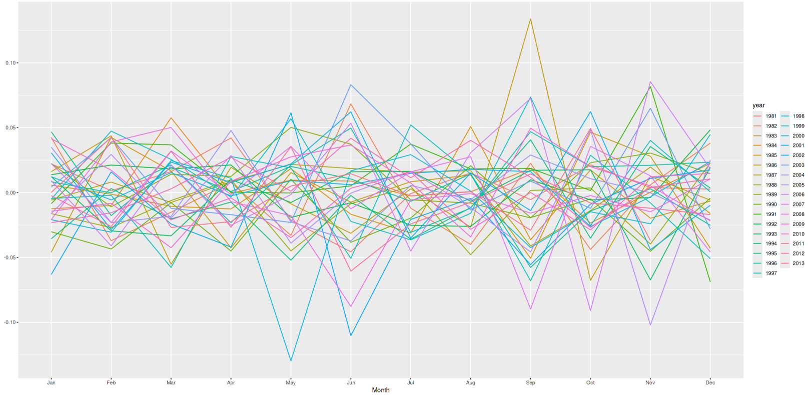


Figure: There is not seasonal component.

ACF and PACF

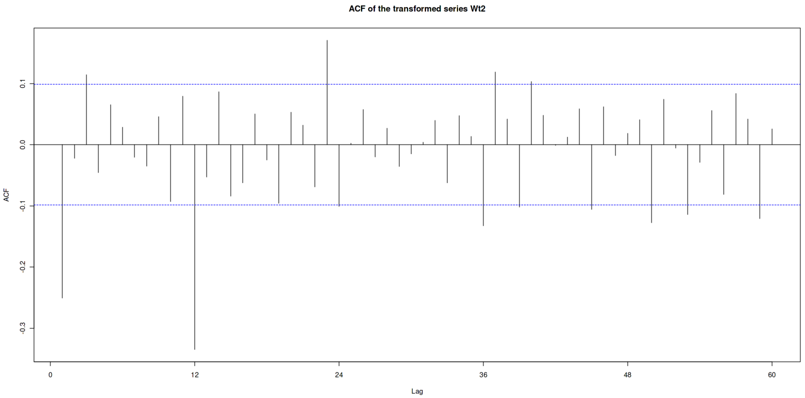


Figure: ACF of the stationary process.

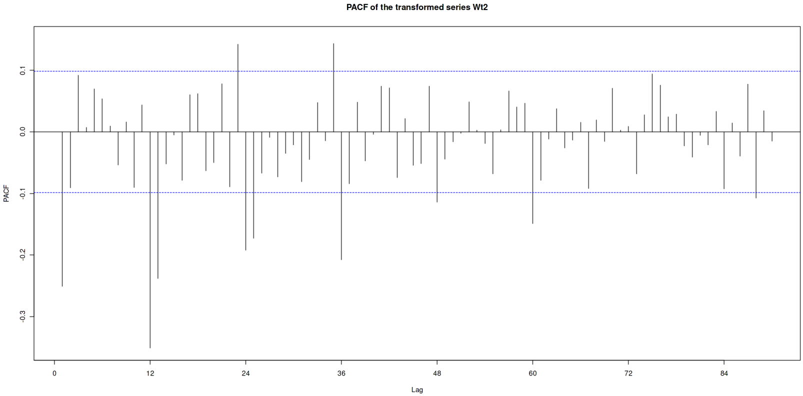


Figure: PACF of the stationary process.

Candidate model 1: AR(1)

$$X_t = \phi_1 X_{t-1} + Y_t,\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1$$

ACF

The ACF of an AR(1) process,

$$\rho_X(h) = \phi_1^h,\quad h \geq 1$$

has all the coefficients different from zero and decreases geometrically to zero.

Candidate model 1: AR(1)

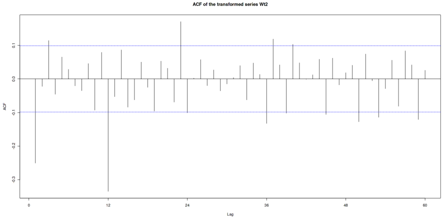


Figure: ACF of the stationary process.

- **First lags ACF (Medium evidence):** it approximately decreases geometrically to zero.
- **Interaction ACF (Medium evidence):** around each seasonal coefficient, it approximately decreases geometrically to zero.

Candidate model 1: AR(1)

$$X_t = \phi_1 X_{t-1} + Y_t,\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1$$

PACF

The PACF of an AR(1) process is given by

$$\alpha_X(h) = \begin{cases} \phi_1, & \text{if } h = 1 \\ 0, & \text{if } h \geq 2 \end{cases}$$

Only the first coefficient of the PACF is nonzero.

Candidate model 1: AR(1)

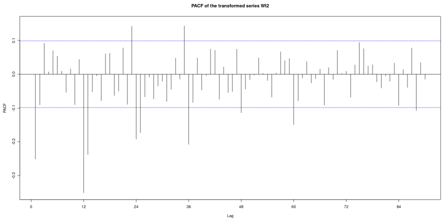


Figure: PACF of the stationary process.

- **First lags PACF (Strong evidence):** it could be considered that only the first coefficient is nonzero.
- **Interaction PACF (Strong evidence):** to the right of each seasonal coefficient, it could be considered that only the first coefficient is nonzero.
- **Interaction ACF (Medium evidence):** to the left of each seasonal coefficient, it approximately decreases geometrically to zero.

Candidate model 2: MA(1)

$$X_t = Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

The ACF of an MA(1) process is given by

$$\rho_X(h) = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & \text{if } h = 1 \\ 0, & \text{if } h \geq 2 \end{cases}$$

Only the first coefficient of the ACF is nonzero.

Candidate model 2: MA(1)

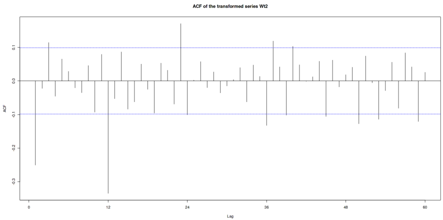


Figure: ACF of the stationary process.

- **First lags ACF (Strong evidence):** it could be considered that only the first coefficient is nonzero.
- **Interaction ACF (Medium evidence):** around some seasonal coefficients (not all of them), it could be considered that only the first coefficient is nonzero.

Candidate model 2: MA(1)

$$X_t = Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\theta_1| < 1$$

PACF

The coefficients of the PACF of an invertible MA(1) process are all different from zero and decay geometrically.

Candidate model 2: MA(1)

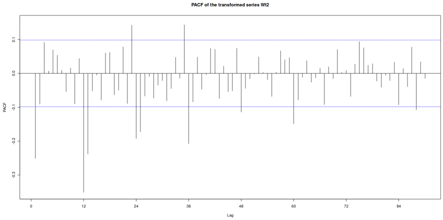


Figure: PACF of the stationary process.

- **First lags PACF (Medium evidence):** it approximately decreases geometrically to zero.
- **Interaction PACF (Weak/Medium evidence):** it approximately decreases geometrically to zero, although it is not really clear.
- **Interaction ACF (Medium evidence):** to the left of some seasonal coefficients, it could be considered that only the first coefficient is nonzero.

Candidate model 3: AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Y_t,\ \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

Let G_1 and G_2 be the factors of $\phi_2(B) = 1 - \phi_1 B - \phi_2 B^2$, $G_1 \neq G_2$, then the ACF of an AR(2) process is given by

$$\rho_X(h) = A_1 G_1^h + A_2 G_2^h, \quad h \geq 1$$

- If G_1 and G_2 are reals, the ACF is the sum of two exponentials and its shape will depend of whether G_1 and G_2 have equal signs or opposite signs.
- If G_1 and G_2 are complex conjugates, the ACF decreases sinusoidally.

In any case, all the coefficients are non-null.

Candidate model 3: AR(2)

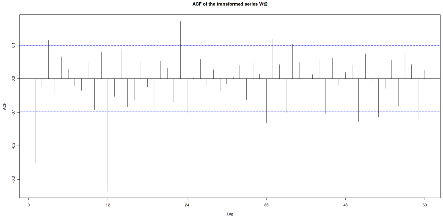


Figure: ACF of the stationary process.

- **First lags ACF (Medium evidence):** it approximately decreases sinusoidally.
- **Interaction ACF (Medium evidence):** around some seasonal coefficients, it approximately decreases sinusoidally.

Candidate model 3: AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Y_t,\ \{Y_t\} \sim WN(0, \sigma^2)$$

PACF

The PACF of an AR(2) process is given by

$$\alpha_X(h) = \begin{cases} \frac{\phi_1}{1-\phi_2}, & \text{if } h = 1 \\ \phi_2, & \text{if } h = 2 \\ 0, & \text{if } h \geq 3 \end{cases}$$

Only the first two coefficients of the PACF are nonzero.

Candidate model 3: AR(2)

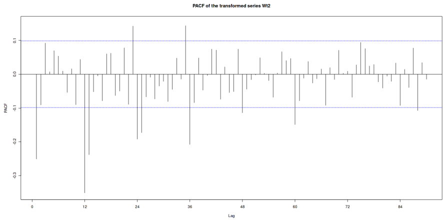


Figure: PACF of the stationary process.

- **First lags PACF (Medium evidence):** it could be considered that only the first (and maybe the second) coefficients are nonzero.
- **Interaction PACF (Weak/Medium evidence):** to the right of some seasonal coefficients, it could be considered that only the first (and maybe the second) coefficients are nonzero.
- **Interaction ACF (Medium evidence):** to the left of some seasonal coefficients, it approximately decreases sinusoidally.

Candidate model 4: AR(3)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Y_t,\ \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

Let G_1, G_2 and G_3 be the factors of $\phi_2(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3$, all different, then the ACF of an AR(3) process is given by

$$\rho_X(h) = A_1 G_1^h + A_2 G_2^h + A_3 G_3^h, \quad h \geq 1$$

The ACF of an AR(p) process is a mixture of exponentials, due to the terms with real roots, and sinusoidals, due to the complex conjugate roots. In any case, all the coefficients are non-null.

Candidate model 4: AR(3)

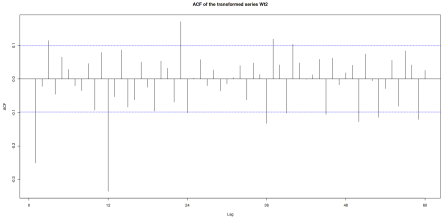


Figure: ACF of the stationary process.

- **First lags ACF (Medium evidence):** it approximately decreases sinusoidally.
- **Interaction ACF (Medium evidence):** around some seasonal coefficients, it approximately decreases sinusoidally.

Candidate model 4: AR(3)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Y_t, \ \{Y_t\} \sim WN(0, \sigma^2)$$

PACF

Only the first three coefficients of the PACF of an AR(3) process are nonzero.

Candidate model 4: AR(3)

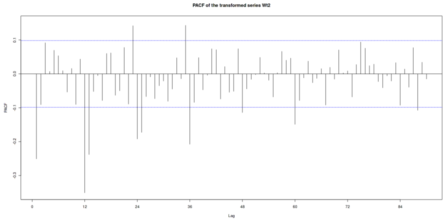


Figure: PACF of the stationary process.

- **First lags PACF (Strong evidence):** it could be considered that only the first three coefficients are nonzero.
- **Interaction PACF (Weak/Medium evidence):** to the right of some seasonal coefficients, it could be considered that only the first (and maybe the second, but not really the third) coefficients are nonzero.
- **Interaction ACF (Medium evidence):** to the left of some seasonal coefficients, it approximately decreases sinusoidally.

Candidate model 5: ARMA(1,1)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1,\ \phi_1 \neq \theta_1$$

ACF

The ACF of an ARMA(1,1) process,

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - 2\phi_1\theta_1 + \theta_1^2} \cdot \phi_1^{h-1}, & \text{if } h \geq 1 \end{cases}$$

has a first coefficient that depends on the parameters of the AR and MA operators, and from this coefficient on, it decays exponentially.

Candidate model 5: ARMA(1,1)

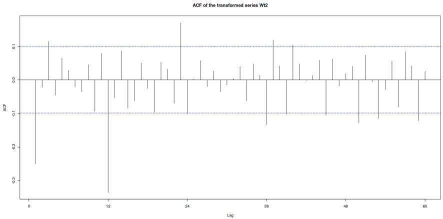


Figure: ACF of the stationary process.

- **First lags ACF (Medium evidence):** it approximately decreases exponentially from the first coefficient on.
- **Interaction ACF (Medium evidence):** around some seasonal coefficients, it approximately decreases exponentially from the first coefficient on.

Candidate model 5: ARMA(1,1)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1,\ |\theta_1| < 1$$

PACF

The PACF of an ARMA(1,1) has a first coefficient whose magnitude depends on $\phi_1 - \theta_1$, and from this coefficient on, it decays exponentially, with a rate of decay determined by θ_1.

Candidate model 5: ARMA(1,1)

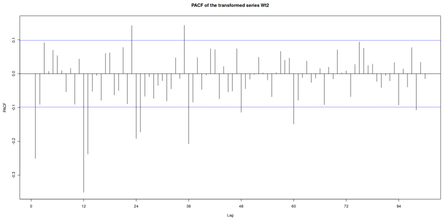


Figure: PACF of the stationary process.

- **First lags PACF (Medium / Strong evidence):** it approximately decreases exponentially from the first coefficient on.
- **Interaction PACF (Medium / Strong evidence):** to the right of each seasonal coefficient, it approximately decreases exponentially from the first coefficient on.
- **Interaction ACF (Medium evidence):** to the left of some seasonal coefficients, it approximately decreases exponentially from the first coefficient on.

Candidate model 1: MA(1)

$$X_t = Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

The ACF of an MA(1) process is given by

$$\rho_X(h) = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & \text{if } h = 1 \\ 0, & \text{if } h \geq 2 \end{cases}$$

Only the first coefficient of the ACF is nonzero.

Candidate model 1: MA(1)

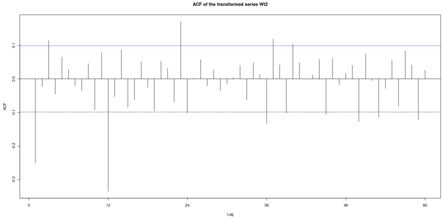


Figure: ACF of the stationary process.

- **ACF (Medium evidence):** it could be considered that only the first seasonal coefficient (at lag 12) is nonzero.

Candidate model 1: MA(1)

$$X_t = Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\theta_1| < 1$$

PACF

The coefficients of the PACF of an invertible MA(1) process are all different from zero and decay geometrically.

Candidate model 1: MA(1)

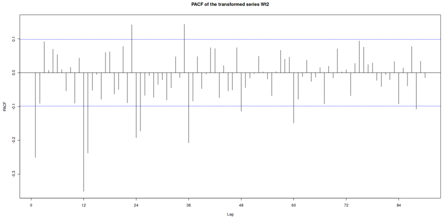


Figure: PACF of the stationary process.

- **PACF (Medium / Strong evidence):** at seasonal lags it approximately decreases geometrically.

Candidate model 2: MA(3)

$$X_t = Y_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \theta_3 Y_{t-3}, \ \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

The ACF of an MA(3) process is given by

$$\rho_X(h) = \begin{cases} \frac{\sum_{i=h}^{3} \theta_i \theta_{i-h}}{\sum_{i=0}^{3} \theta_i^2}, & \text{if } h = 1, 2, 3 \\ 0, & \text{if } h \geq 3 \end{cases}$$

Only the first three coefficients of the ACF are nonzero.

Candidate model 2: MA(3)

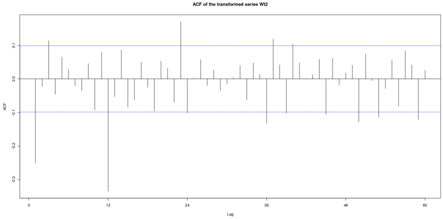


Figure: ACF of the stationary process.

- **ACF (Strong evidence):** it could be considered that only the first three seasonal coefficients (at lags 12, 24 and 36) are nonzero.

Candidate model 2: MA(3)

$$X_t = Y_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \theta_3 Y_{t-3}, \ \{Y_t\} \sim WN(0, \sigma^2)$$

PACF

The coefficients of the PACF of an invertible MA(3) process are all different from zero.

Candidate model 2: MA(3)

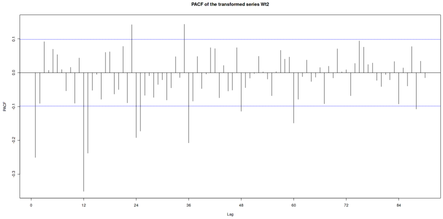


Figure: PACF of the stationary process.

- **PACF (Medium / Strong evidence):** it could be considered that almost all the seasonal coefficients are different from zero.

Candidate model 3: ARMA(1,1)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1}, \ \{Y_t\} \sim WN(0, \sigma^2), \ |\phi_1| < 1, \ \phi_1 \neq \theta_1$$

ACF

The ACF of an ARMA(1,1) process,

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - 2\phi_1\theta_1 + \theta_1^2} \cdot \phi_1^{h-1}, & \text{if } h \geq 1 \end{cases}$$

has a first coefficient that depends on the parameters of the AR and MA operators, and from this coefficient on, it decays exponentially.

Candidate model 3: ARMA(1,1)

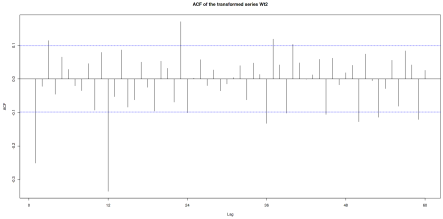


Figure: ACF of the stationary process.

- **ACF (Medium evidence):** at seasonal lags, it approximately decreases exponentially from the first coefficient (at lag 12) on.

Candidate model 3: ARMA(1,1)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1,\ |\theta_1| < 1$$

PACF

The PACF of an ARMA(1,1) has a first coefficient whose magnitude depends on $\phi_1 - \theta_1$, and from this coefficient on, it decays exponentially, with a rate of decay determined by θ_1.

Candidate model 3: ARMA(1,1)

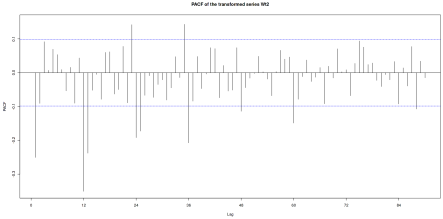


Figure: PACF of the stationary process.

- **PACF (Medium / Strong evidence):** at seasonal lags, it approximately decreases exponentially from the first coefficient (at lag 12) on.

Candidate model 4: ARMA(1,3)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \theta_3 Y_{t-3}, \; \{Y_t\} \sim WN(0, \sigma^2)$$

ACF

Since $1 \leq 3$, the ACF of an ARMA(1,3) process will have $3 + 1 - 1 = 3$ initial values with a structure that depends on the AR and MA parameters, and for $h \geq 3 + 1 - 1 = 3$, the coefficients $\rho_X(h)$ will decay as a mixture of exponentials and sinusoids, determined by the autoregressive part.

Candidate model 4: ARMA(1,3)

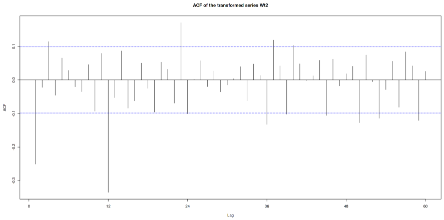


Figure: ACF of the stationary process.

- **ACF (Medium / Strong evidence):** at seasonal lags, it approximately decays from the third coefficient (at lag 36) on.

Candidate model 4: ARMA(1,3)

$$X_t = \phi_1 X_{t-1} + Y_t - \theta_1 Y_{t-1},\ \{Y_t\} \sim WN(0, \sigma^2),\ |\phi_1| < 1,\ |\theta_1| < 1$$

PACF

Since $1 \leq 3$, the ACF of an ARMA(1,3) process will have $3 + 1 - 1 = 3$ initial values with a structure that depends on the AR and MA parameters, and for $h \geq 3 + 1 - 1 = 3$, the coefficients $\rho_X(h)$ will decay as a mixture of exponentials and sinusoids, determined by the MA part.

Candidate model 4: ARMA(1,3)

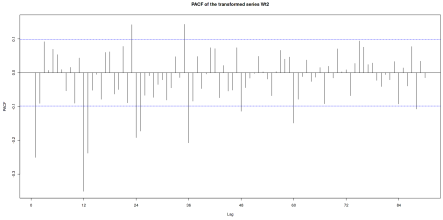


Figure: PACF of the stationary process.

- **PACF (Medium / Strong evidence):** at seasonal lags, it approximately decays from the third coefficient (at lag 36) on.

Summary

Candidate Model	First lags ACF	Interaction ACF	First lags PACF	Interaction PACF
AR(1)	Medium	Medium	Strong	Strong
MA(1)	Strong	Medium	Medium	Weak/Medium
AR(2)	Medium	Medium	Medium	Weak/Medium
AR(3)	Medium	Medium	Strong	Weak/Medium
ARMA(1,1)	Medium	Medium	Medium/Strong	Medium/Strong

Table: Candidate models for regular component.

Candidate Model	ACF	PACF
MA(1)	Medium	Medium/Strong
MA(3)	Strong	Medium/Strong
ARMA(1,1)	Medium	Medium/Strong
ARMA(1,3)	Medium/Strong	Medium/Strong

Table: Candidate model for seasonal component.

Summary

Model for $\{W_t\}$	Model for $\{X_t\}$	numer of parameters
$\text{ARMA}(0,3)_{12} \times (1,0)$	$\text{ARIMA}(0,1,3)_{12} \times (1,1,0)$	5
$\text{ARMA}(0,3)_{12} \times (3,0)$	$\text{ARIMA}(0,1,3)_{12} \times (3,1,0)$	7
$\text{ARMA}(0,3)_{12} \times (1,1)$	$\text{ARIMA}(0,1,3)_{12} \times (1,1,1)$	6
$\text{ARMA}(1,3)_{12} \times (1,0)$	$\text{ARIMA}(1,1,3)_{12} \times (1,1,0)$	6
$\text{ARMA}(1,3)_{12} \times (3,0)$	$\text{ARIMA}(1,1,3)_{12} \times (3,1,0)$	8
$\text{ARMA}(1,3)_{12} \times (1,1)$	$\text{ARIMA}(1,1,3)_{12} \times (1,1,1)$	7

Table: Candidate model $\{X_t\}$ and number of parameters.

Linear regression

Review

It is assumed that the observed data $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ arose from a statistical model:

- $Y_i = f^*(X_i) + \epsilon_i$, with $\epsilon_i \sim^{iid} N(0, \sigma^2)$
- $f^* \in \mathcal{F} = \{f_\beta(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \beta = (\beta_0, \ldots, \beta_p) \in \mathbb{R}^{p+1}\}$

Estimation procedure: $D \Rightarrow \hat{\beta} \Rightarrow \hat{f}^*(x) = f_{\hat{\beta}}(x)$

Linear regression

Estimation methods

- Least squares:
$\hat{\beta} = argmin_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} [y_i - \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2 = argmin_{\beta \in \mathbb{R}^{p+1}} \frac{\sum_{i=1}^{n} [y_i - f_\beta(x_i)]^2}{n}$
- Maximum likelihood: $(\hat{\beta}, \hat{\sigma}) = argmax_{(\beta,\sigma) \in \mathbb{R}^{p+1} \times \mathbb{R}^+} l(\beta, \sigma^2)$,
where: $l(\beta, \sigma^2) = \prod_{i=1}^{n} f_{Y|X=x_i}(y_i) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{\frac{-\sum_{i=1}^{n}[y_i - f_\beta(x_i)]^2}{2\sigma^2}}$

Linear regression

Consider m models, $M_1, \ldots, M_m$. Let $\hat{f}_i$ be the estimations of the regression function f^* obtained by fitting model M_i to the observed data.

$$MSE(\hat{f}) = E[(\hat{f}(X) - Y)^2] = \frac{\sum_{(x,y)\in P}[y - \hat{f}(x)]^2]}{|P|}$$

The **training error**:

$$\hat{MSE}[\hat{f}(\cdot)] = \frac{\sum_{(x_h,y_h)\in D}[\hat{f}(x_h) - y]^2]}{n}$$

always leads to select the most complex model $\Rightarrow$ **Overfitting**

Indirect methods: AIC and BIC

Let $\hat{f}$ be an estimated prediction function based on a linear regression model M with d parameters.

$$Score(\hat{f}) = g(\hat{MSE}[\hat{f}]) + h(d)$$

- $g(\hat{MSE}[\hat{f}])$: prediction error in the training set.
- $h(d)$: penalty term for the complexity of the model, taking into account the number of parameters d.

Let $L_M(\hat{f})$ be the maximum value of the likelihood function for model M, then we get the criterion:

- $AIC = -2\log L_M(\hat{f}) + 2d$
- $BIC = -2\log L_M(\hat{f}) + \log(n_{train})d$

AIC and BIC

Remark

1. AIC and BIC penalize the model complexity.

	AIC	AICc	BIC
Model 1	-1770.906	-1770.751	-1751.011
Model 2	-1778.172	-1777.883	-1750.320
Model 3	-1774.169	-1773.953	-1750.296
Model 4	-1774.180	-1773.963	-1750.306
Model 5	-1780.633	-1780.260	-1748.802
Model 6	-1777.034	-1776.745	-1749.182

- The best model in terms of **AIC** is **Model 5**.
- The best model in terms of **BIC** is **Model 1**.
- The second best model (both in terms of **AIC** and **BIC**) is **Model 2**.

Direct methods: Cross-Validation

- Choose $K \in \mathbb{N}$ and consider a random partition of D_{train} in K subsets with same number of observations. Let C_i be the set of indices of the observations in the i-th subset.
- Let $\hat{f}_{-k}$ be the estimator of the regression function using as training set the set obtained by removing from D_{train} the observations in C_k, and for $k \in \{1, \ldots, K\}$ compute:

$$SSE_k = \sum_{h \in C_k} [\hat{f}_{-k}(x_h) - y_h]^2$$

- Estimator of $MSE(\hat{f})$:

$$CV_K = \frac{\sum_{k=1}^{K} SSE_k}{n}$$

MSE and MAE

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
MSE	0.0003497877	0.0003433387	0.0003339444	0.0003391916	0.0003337607	0.0003239775
MAE	0.0074344690	0.0074424819	0.0072785899	0.0072581143	0.0073414805	0.0071415472

The best model in terms of **MSE** and **MAE** is **Model 6**.

Interpretation of MSE

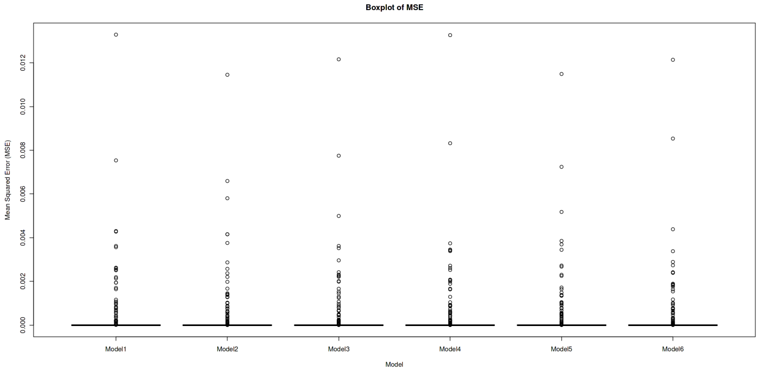


Figure: Boxplot of variability of MSE.

Interpretation of MAE

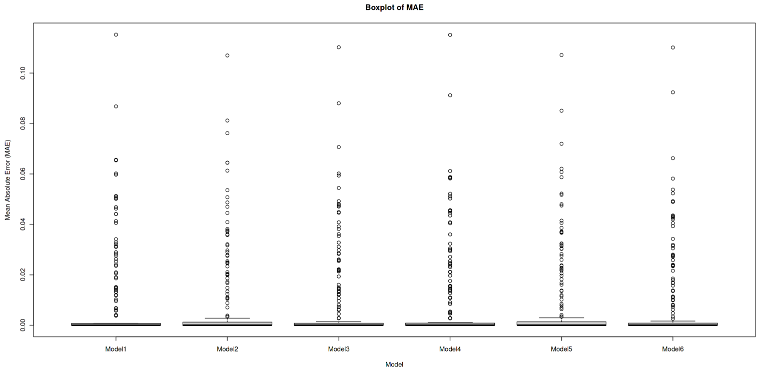


Figure: Boxplot of variability of MAE.

Diagnosis

Remark

The properties of the ML estimators as well as the value of the AIC, AICc, BIC criteria, the predictions and the properties of the confidence intervals for the predictions depend on the hypothesis $H_0 : Y_t \sim^{i.i.d.} N(0, \sigma^2)$

Diagnosis

The diagnosis of the model requires checking the basic hypotheses made with respect to the residuals:

$$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$$

$$H_{0b} : \{Y_t\} \sim N(0, \sigma^2)$$

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

Test based on the sample ACF

Under the null hypothesis H_{0a}, for large n, the coefficients of the sample ACF of the residuals:

$$\hat{\rho}_Y(h)|H_{0a} \sim^{i.i.d.} N(0, 1/n)$$

Under H_{0a} about 95% of the sample autocorrelation coefficients should fall between the bounds $\pm\frac{2}{\sqrt{n}}$.

Implementation

Reject the null hypothesis H_{0a} if more than 5% of the coefficients of the sample ACF fall outside the interval $\left(-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}}\right)$.

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

The Ljung-Box Test

For large n, $\hat{\rho}_Y(h)|H_{0a} \sim^{i.i.d.} N(0, \frac{1}{n})$. Thus, under H_{0a} the statistics $Q = n \cdot \sum_{h=1}^{k} \hat{\rho}_Y(h)^2$ is approximately distributed as a χ-squared with k degrees of freedom. Ljung and Box proposed to replace Q by:

$$Q_{LB} = n \cdot (n+2) \sum_{h=1}^{k} \frac{\hat{\rho}_Y(h)^2}{n-h}$$

whose distribution is better approximated by the χ-squared distribution.

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

Implementation

Reject H_{0a} at the level α, if the observed value of Q_{LB} is greater than the $(1-\alpha)$ quantile of the χ-squared distribution with $k - n_p$ degrees of freedom, where n_p is the number of estimated parameters.

For the Ljung-Box test:

- if the series is non-seasonal: $lag = \min(10, n/5)$
- if the series is seasonal with period s: $lag = \min(2s, n/5)$

In our case:
$lag = \min(2s, n/5) = \min(2 \cdot 12, 408/5) = \min(24, 81.6) = 24$

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

The turning point test

If $y_1, y_2, \ldots, y_n$ is a sequence of observations, we say that there is a turning point at $t = i$ if:

$$(y_{i-1} < y_i \text{ and } y_i > y_{i+1}) \text{ or if } (y_{i-1} > y_i \text{ and } y_i < y_{i+1}), i = 2, \ldots, n-1$$

Let T be the r.v. that represents the number of turning points of a realization of length n of $\{Y_t\}$. For large n:

$$T|H_{0a} \sim N(\mu_T, \sigma_T^2), \text{ where: } \mu_T = \frac{2(n-2)}{3}; \sigma_T^2 = \frac{16n-29}{90}$$

Implementation

Reject H_{0a} at level α if $\frac{|T-\mu_T|}{\sigma_T} > \phi_{1-\alpha/2}$ where $\phi_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

The Difference-sign test

Let S be the r.v. that represents the number of values of i such that $Y_i > Y_{i-1}$, $i = 2, \ldots, n$. For large n:

$$S|H_{0a} \sim N(\mu_S, \sigma_S^2), \text{ where: } \mu_S = \frac{n-1}{2}; \sigma_S^2 = \frac{n+1}{12}$$

Implementation

Reject H_{0a} at level α if $\frac{|S-\mu_S|}{\sigma_S} > \phi_{1-\alpha/2}$ where $\phi_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

$H_{0a} : \{Y_t\} \sim IID(0, \sigma^2)$

The rank test

Let P be the r.v. that represents the number of pairs (i, j) such that $Y_j > Y_i$, $i = 1, \ldots, n-1, j > i$. For large n:

$$P|H_{0a} \sim N(\mu_P, \sigma_P^2), \text{ where: } \mu_P = \frac{n(n-1)}{4}; \sigma_P^2 = \frac{n(n-1)(2n+5)}{72}$$

Implementation

Reject H_{0a} at level α if $\frac{|P-\mu_P|}{\sigma_P} > \phi_{1-\alpha/2}$ where $\phi_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

$H_{0b} : \{Y_t\} \sim N(0, \sigma^2)$

The null hypothesis H_{0b} can be checked using the following test:

- Shapiro-Wilk test (for small sample sizes, $n \leq 30$)
- Kolmogorov-Smirnov-Lilliefors (for larger sizes)
- QQ-plot, histrogram of residuals
- Test based on coefficients of asymmetry and kurtosis

Model 1: Diagnosis

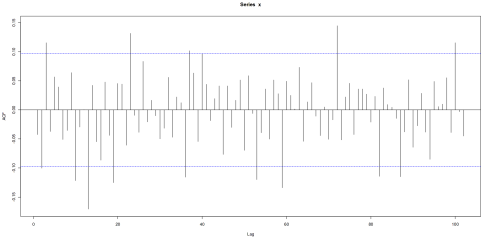


Figure: Residual ACF from model 1.

	Test Statistic	P-value
Turning Point test	1.3729186	0.1697777
Difference-sign Test	0.7707996	0.4408257
Rank Test	0.1264494	0.8993762

n	408.00
h	102.00
bound	0.10
# out of bounds	14.00
% out of bounds	13.73

Ljung-Box test:

k	Test Statistic	P-value
10	22.48815	4.227392e-04
15	37.20958	5.205800e-05
20	49.85999	1.269215e-05
24	59.88371	4.037194e-06
25	60.53381	5.886191e-06
30	65.04436	2.038370e-05

Model 1: Diagnosis (2)

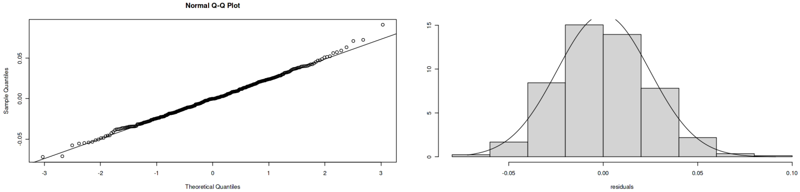


Figure: Checking residuals normality.

	Statistics	P-value
Shapiro-Wilxs	0.9970	0.6643
Lilliefors	0.0259	0.7256
Pearson Chi.square	17.6127	0.6129

Zero Mean Test:

$$\text{p-value} = 0.4696$$

Model 2: Diagnosis

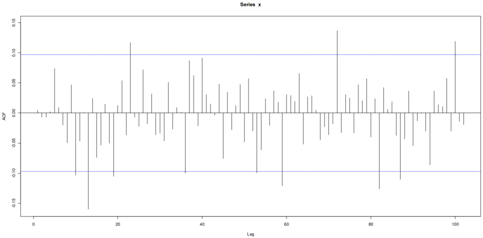


Figure: Residual ACF from model 2.

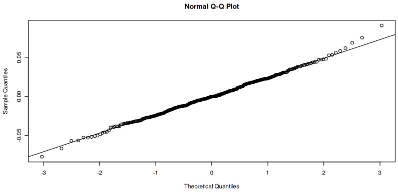
	Test Statistic	P-value
Turning Point test	0.4314887	0.6661131
Difference-sign Test	1.6272436	0.1036854
Rank Test	0.1744129	0.8615410

n	408.00
h	102.00
bound	0.10
# out of bounds	11.00
% out of bounds	10.78

Ljung-Box test:

k	Test Statistic	P-value
10	8.926261	0.030287525
15	23.165104	0.003158615
20	30.371287	0.004162430
24	38.161237	0.002335380
25	38.380354	0.003448410

Model 2: Diagnosis (2)



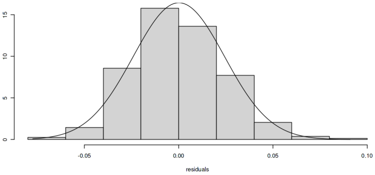


Figure: Checking residuals normality.

	Statistics	P-value
Shapiro-Wilxs	0.9968	0.6070
Lilliefors	0.0339	0.3074
Pearson Chi.square	22.6863	0.3045

Zero Mean Test:

$$\text{p-value} = 0.4661$$

Model 5: Diagnosis

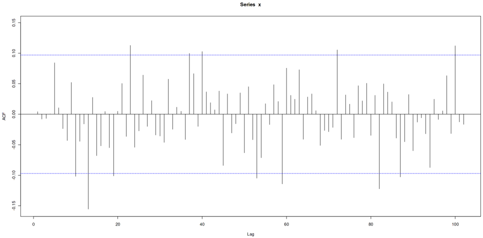


Figure: Residual ACF from model 5.

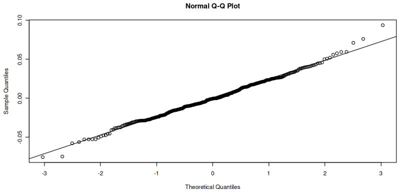
	Test Statistic	P-value
Turning Point test	0.4314887	0.6661131
Difference-sign Test	0.9420884	0.3461474
Rank Test	0.2805141	0.7790831

n	408.00
h	102.00
bound	0.10
# out of bounds	12.00
% out of bounds	11.76

Ljung-Box test:

k	Test Statistic	P-value
10	9.520098	0.008565189
15	22.946995	0.001741216
20	29.783357	0.003009864
24	38.250854	0.001393522
25	38.575006	0.002047423

Model 5: Diagnosis (2)



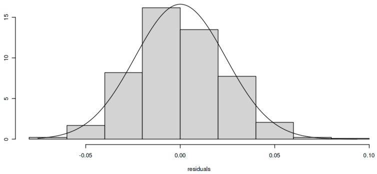


Figure: Checking residuals normality.

	Statistics	P-value
Shapiro-Wilxs	0.9945	0.1468
Lilliefors	0.0365	0.2076
Pearson Chi.square	34.2990	0.0242

Zero Mean Test:

$$\text{p-value} = 0.4840$$

Model 6: Diagnosis

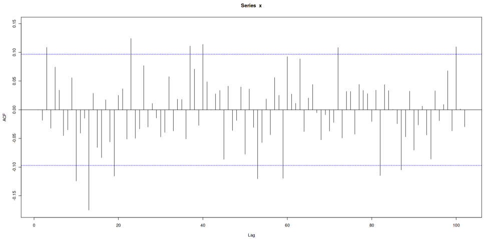


Figure: Residual ACF from model 6.

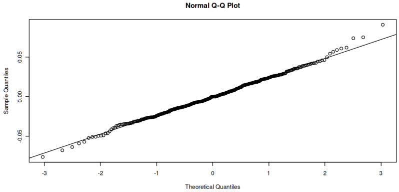
	Test Statistic	P-value
Turning Point test	1.3729186	0.1697777
Difference-sign Test	0.9420884	0.3461474
Rank Test	0.2528987	0.8003465

n	408.00
h	102.00
bound	0.10
# out of bounds	13.00
% out of bounds	12.75

Ljung-Box test:

k	Test Statistic	P-value
10	17.40700	5.827830e-04
15	33.36340	5.296066e-05
20	43.86600	3.229847e-05
24	53.38847	1.239618e-05
25	53.87686	1.917943e-05

Model 6: Diagnosis (2)



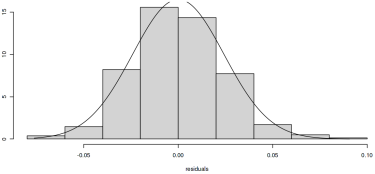


Figure: Checking residuals normality.

	Statistics	P-value
Shapiro-Wilxs	0.9956	0.3165
Lilliefors	0.0253	0.7591
Pearson Chi.square	11.2990	0.9381

Zero Mean Test:

$$\text{p-value} = 0.4898$$

Model fit and prediction

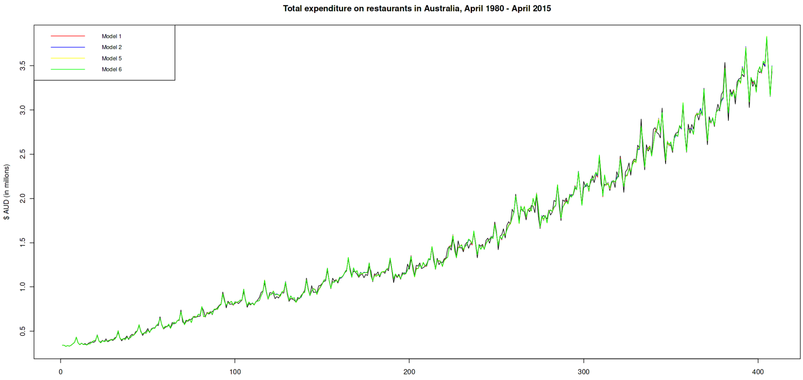


Figure: Every model fit with data representation.

Model 1

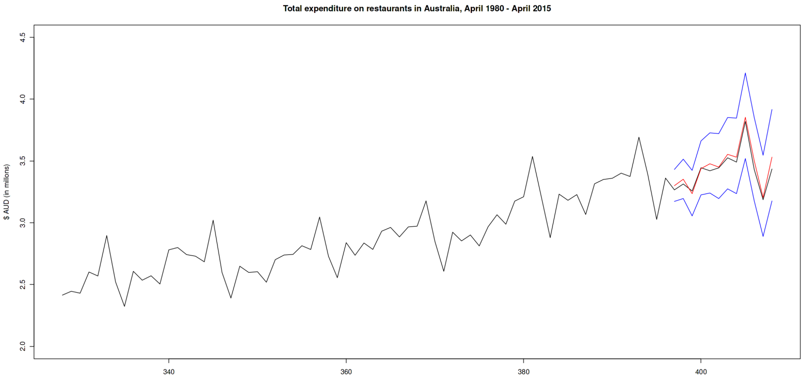


Figure: We fit the model using the firsts 398 observations and we forecast

Model 1

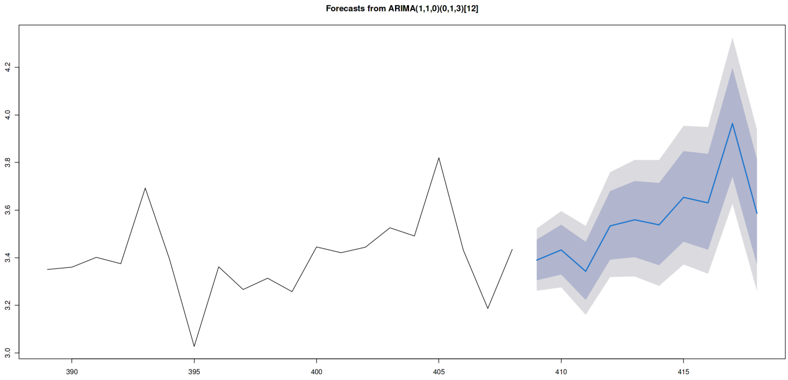


Figure: Confidence interval for the last 12 forecasting steps.

Model 2

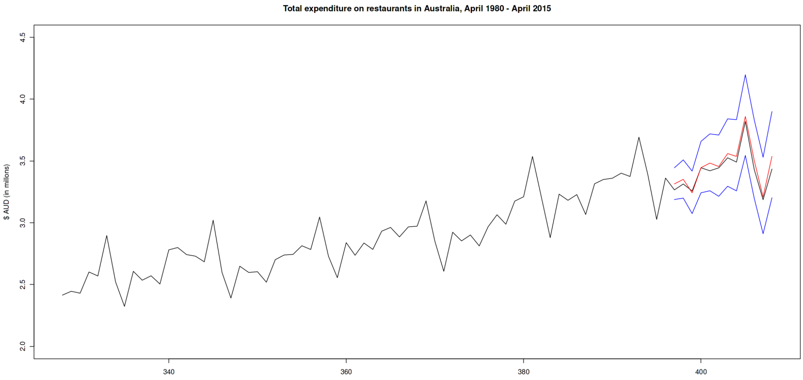


Figure: We fit the model using the firsts 398 observations and we forecast

Model 2

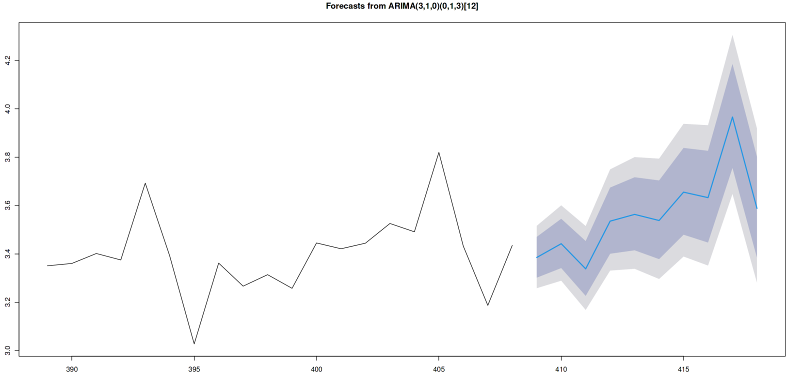


Figure: Confidence interval for the last 12 forecasting steps.

Model 5

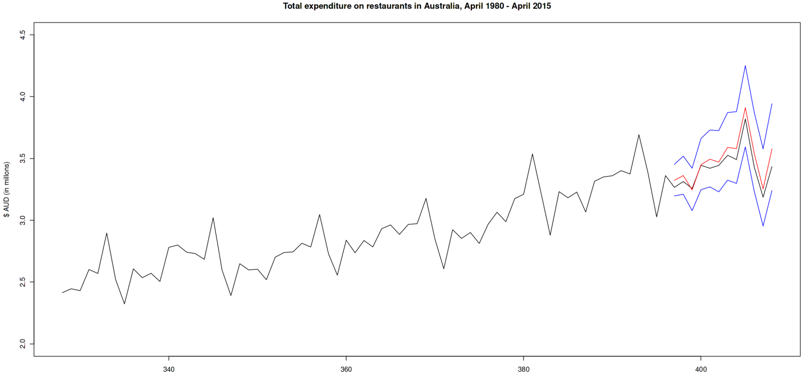


Figure: We fit the model using the firsts 398 observations and we forecast

Model 5

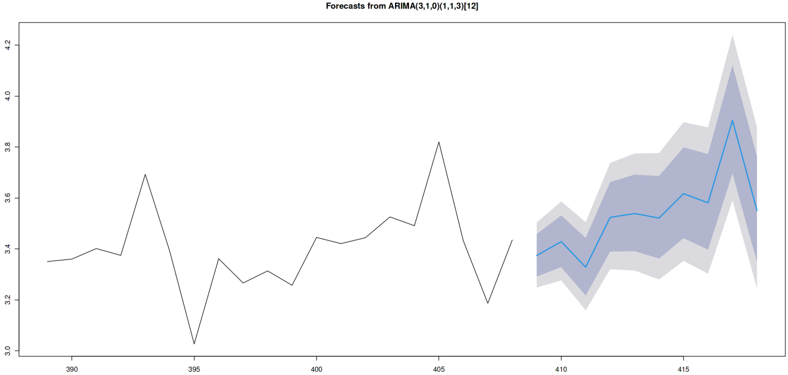


Figure: Confidence interval for the last 12 forecasting steps.

Model 6

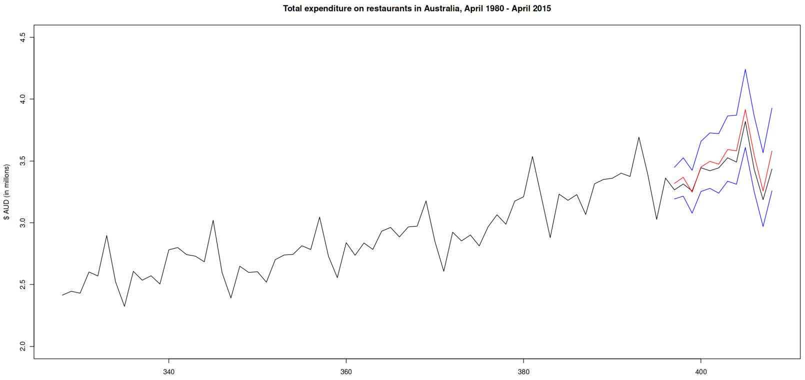


Figure: We fit the model using the firsts 398 observations and we forecast

Model 6

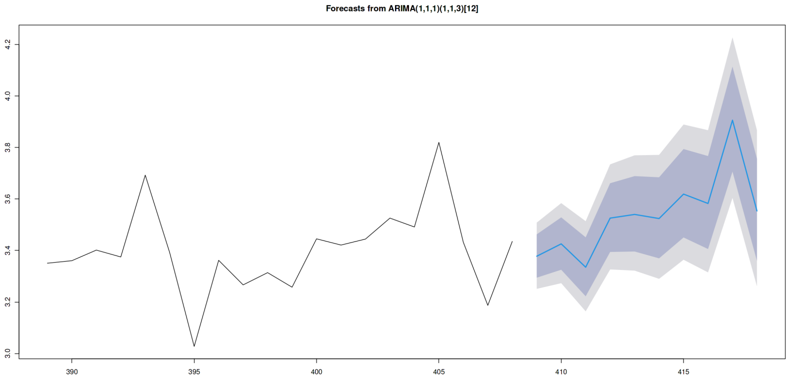


Figure: Confidence interval for the last 12 forecasting steps.

MSE and MAE for predicted values

	Model 1	Model 2	Model 5	Model 6
MSE	0.002032935	0.002431288	0.005781682	0.006133644
MAE	0.037469002	0.041012250	0.065481389	0.067663385

- The best model (both in terms of **MAE** and **MSE** of predicted values $\{X_{396+k}\}_{k=1}^{12}$) is **Model 1**.
- The second best model is **Model 2**.

Interpretation of MSE and MAE

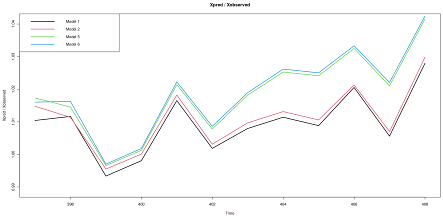


Figure: Plot of $\frac{X_{pred}}{X_{observed}}$

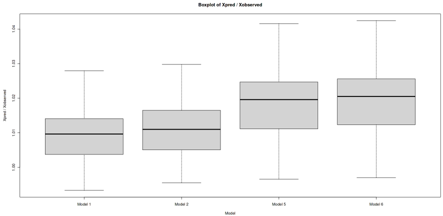


Figure: Boxplot of $\frac{X_{pred}}{X_{observed}}$

Interpretation of MSE and MAE

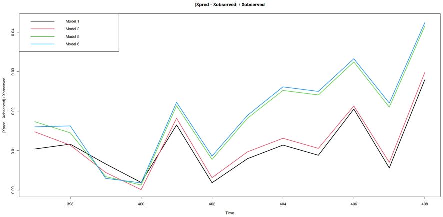


Figure: Plot of $\frac{|X_{pred} - X_{observed}|}{X_{observed}}$

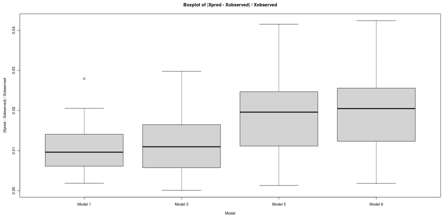


Figure: Boxplot of $\frac{|X_{pred} - X_{observed}|}{X_{observed}}$

Estimation of parameters

Let $\{X_t\} \sim \text{ARIMA}(0,1,3)_{12} \times (1,1,0)$ and let $W_t = \nabla_{12}\nabla X_t$.
Then $\{W_t\} \sim \text{ARMA}(0,3)_s \times (1,0)$ as it follows

$$(1 + 0.313B)W_t = (1 + 0.718B^{12} + 0.132B^{24} - 0.027B^{36})Y_t$$

where: $\{Y_t\} \sim WN(0, \sigma^2)$.

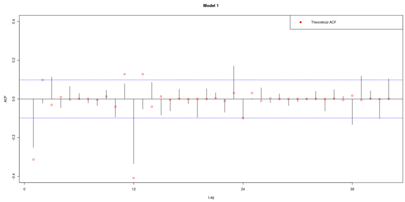


Figure: Comparison of the Sample ACF with Theoretical ACF.

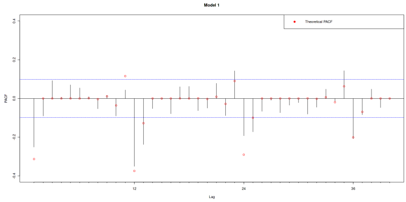


Figure: Comparison of the Sample PACF with Theoretical PACF.

Estimation of parameters

Let $\{X_t\} \sim \text{ARIMA}(0,1,3)_{12} \times (3,1,0)$ and let $W_t = \nabla_{12}\nabla X_t$.
Then $\{W_t\} \sim \text{ARMA}(0,3)_s \times (3,0)$ as it follows

$$(1 + 0.351B + 0.117B^2 - 0.084B^3)W_t = (1 + 0.74B^{12} + 0.117B^{24} - 0.037B^{36})Y_t$$

where: $\{Y_t\} \sim WN(0, \sigma^2)$.

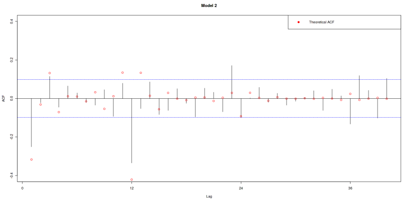


Figure: Comparison of the Sample ACF with Theoretical ACF.

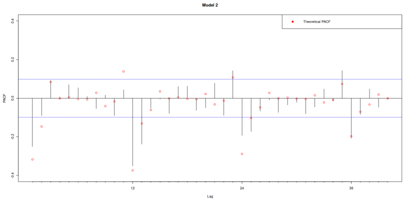


Figure: Comparison of the Sample PACF with Theoretical PACF.