Lecture 6

Prediction and Forecasting with Stationary Time Series

Readings: Cryer & Chan Ch 9; Brockwell & Davis Ch 2.5 3.3; Shumway & Stoffer Ch 2.5

MATH 8090 Time Series Analysis
September 21 & September 23, 2021

Whitney Huang
Clemson University

Agenda

1. Linear Predictor
2. Prediction Equations
3. Examples
4. Case Study

Forecasting Stationary Time Series

Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Based on the observed data, $\boldsymbol{X}_n = (X_1, X_2, \cdots, X_n)^T$, we want to forecast X_{n+h} for some h, a positive integer

- Question: What is the best way to do so?
  $\Rightarrow$ Need to decide on what "best" means
- A commonly used metric for describing forecast performance is the mean square prediction error (MSPE):

$$\mathrm{MSPE} = \mathrm{E}\left[\left(X_{n+h} - m_n(\boldsymbol{X}_n)\right)^2\right].$$

  $\Rightarrow$ the best predictor (in terms of MSPE) is

$$m_n(\boldsymbol{X}_n) = \mathbb{E}\left[X_{n+h} | \boldsymbol{X}_n\right],$$

  the conditional expectation of X_{n+h} given $\boldsymbol{X}_n$

Linear Predictor

Calculating $\mathbb{E}\left[X_{n+h}|\boldsymbol{X}_n\right]$ can be difficult in general

- We will restrict to a linear combination of $X_1, X_2, \cdots, X_n$ and a constant ⇒ linear predictor:

$$\begin{aligned} P_n X_{n+h} &= c_0 + c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1 \\ &= c_0 + \sum_{j=1}^{n} c_j X_{n+1-j} \end{aligned}$$

- We select the coefficients that minimize the h-step-ahead mean squared prediction error:

$$\mathbb{E}\left(\left[X_{n+h} - P_n X_{n+h}\right]^2\right) = \mathbb{E}\left(X_{n+h} - c_0 - \sum_{j=1}^{n} c_j X_{n+1-j}\right)^2$$

- The best linear predictor is the best predictor if $\{X_t\}$ is Gaussian

How to Determine these Coefficients $\{c_j\}$?

The steps that we are about to follow to calculate the c_j values are the same as you would use for calculating ordinary least squares estimates

1. Take the derivative of the MSPE with respect to each coefficient c_j
2. Set each derivative equal to zero
3. Solve with respect to the coefficients

Forecasting Stationary Processes I

For simplicity, let's assume $\mu = 0$ (we can always achieve that by subtracting off μ) so that we don't need the constant term. We have

$$P_n X_{n+h} = c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1.$$

We want the MSPE

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$$

as small as possible.

From now on let's definite

$$\mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right] = S(c_1, \cdots, c_n)$$

We are going to take derivative of the $S(c_1, \cdots, c_n)$ with respect to each coefficient c_j

Forecasting Stationary Processes II

S is a quadratic function of $c_1, c_2, \cdots, c_n$, so any minimizing set of c_j's must satisfy these n equations:

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = 0, \quad j = 1, \cdots, n.$$

Since $S(c_1, \cdots, c_n) = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$, we have

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = -2\mathbb{E}\left[(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1})X_{n-j+1}\right] = 0$$

$$\Rightarrow \mathbb{Cov}(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}) = 0, \quad j = 1, \cdots, n$$

$\Rightarrow$ Prediction error is uncorrelated with all RVs used in corresponding predictor

Forecasting Stationary Processes III

Orthogonality principle:

$$\mathbb{Cov}\Big(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}\Big) = 0, \quad j = 1, \cdots, n.$$

We have

$$\mathbb{Cov}(X_{n+h}, X_{n-j+1}) - \sum_{i=1}^{n} c_i \mathbb{Cov}(X_{n-i+1}, X_{n-j+1}) = 0$$

We obtain $\{c_i; i = 1, \cdots, n\}$ by solving the system of linear equations:

$$\left\{\gamma(h+j-1) = \sum_{i=1}^{n} c_i \gamma(i-j) : j = 1, \cdots, n\right\},$$

to find n unknown c_i's

Computing $P_n X_{n+h}$ via Matrix Operations

We can rewrite the system of prediction equations as

$$\boldsymbol{\gamma}_n = \Sigma_n \boldsymbol{c}_n,$$

with $\boldsymbol{\gamma}_n = (\gamma(h), \gamma(h+1), \cdots \gamma(h+n-1))^T$, $\boldsymbol{c}_n = (c_1, c_2, \cdots, c_n)^T$ and

$$\Sigma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}$$

is the covariance matrix of $(X_1, X_2, \cdots, X_n)^T$.

Solving for $\boldsymbol{c}_n$ we have

$$\boldsymbol{c}_n = \Sigma_n^{-1} \boldsymbol{\gamma}_n$$

Properties of the Prediction Errors

The prediction errors are

$$\begin{aligned} U_{n+h} &= X_{n+h} - P_n X_{n+h} \\ &= (X_{n+h} - \mu) - \sum_{j=1}^{n} c_j (X_{n+1-j} - \mu). \end{aligned}$$

It then follows that

- The prediction error has mean zero

$$\mathbb{E}(U_{n+h}) = \mathbb{E}(X_{n+h} - P_n X_{n+h}) = 0$$

- The prediction error is uncorrelated with all RVs used in the predictor

$$\mathbb{Cov}(U_{n+h}, X_j) = \mathbb{Cov}(X_{n+h} - P_n X_{n+h}, X_j) = 0, \quad j = 1, \cdots, n$$

The Minimum Mean Squared Prediction Error

We obtain the minimum value of the MSPE by substituting the expression for $\boldsymbol{c}_n$ into $\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right]$:

$$
\begin{aligned}
\mathrm{MSPE} &= \mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] \\
&= \mathbb{E}\left[(X_{n+h} - \mu)^2\right] - 2\sum_{j=1}^{n} c_j \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+h} - \mu)\right] \\
&+ \mathbb{E}\left[\sum_{j=1}^{n} c_j (X_{n+1-j} - \mu)\right]^2 \\
&= \mathbb{E}\left[(X_{n+h} - \mu)^2\right] - 2\sum_{j=1}^{n} c_j \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+h} - \mu)\right] \\
&+ \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+1-k} - \mu)\right] \\
&= \gamma(0) - 2\sum_{j=1}^{n} c_j \gamma(h + j - 1) + \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k \gamma(k - j) \\
&= \gamma(0) - 2\boldsymbol{c}_n^T \boldsymbol{\gamma}_n + \boldsymbol{c}_n^T \Sigma_n \boldsymbol{c}_n.
\end{aligned}
$$

The Minimum Mean Squared Prediction Error (Cont'd)

From the previous slide we have

$$\mathrm{MSPE} = \gamma(0) - 2\boldsymbol{c}_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_n^T\Sigma_n\boldsymbol{c}_n$$

Recall that $\boldsymbol{c}_n = \Sigma_n^{-1}\boldsymbol{\gamma}_n$, therefore we have

$$\begin{aligned}\mathrm{MSPE} &= \gamma(0) - 2\boldsymbol{c}_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_n^T\Sigma_n\sigma_n^{-1}\boldsymbol{\gamma}_n \\ &= \gamma(0) - \boldsymbol{c}_n^T\boldsymbol{\gamma}_n \\ &= \gamma(0) - \sum_{j=1}^{n} c_j\gamma(h+j-1).\end{aligned}$$

If $\{X_t\}$ is a Gaussian process then an approximate $100(1-\alpha)$% prediction interval for X_{n+h} is given by

$$P_nX_{n+h} \pm z_{1-\alpha/2}\sqrt{\mathrm{MSPE}}.$$

One-Step Ahead Prediction of AR(1) Process

Consider AR(1) process $X_t = \phi X_{t-1} + Z_t$, where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, 1-\phi^2)$.

- Since $\mathbb{V}\mathrm{ar}(X_t) = 1$, $\gamma(h) = \rho(h) = \phi^{|h|}$
- To forecast X_{n+1} based upon $\boldsymbol{X}_n = (X_1, \cdots, X_n)^T$, using best linear predictor $P_n X_{n+1} = \boldsymbol{c}_n^T \boldsymbol{X}_n$, we need to solve $\Sigma_n \boldsymbol{c}_n = \boldsymbol{\gamma}_n$

$$\begin{bmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}$$

$\Rightarrow$ the solution is $\boldsymbol{c}_n = (\phi, 0, \cdots, 0)^T$, yielding

$$P_n X_{n+1} = \boldsymbol{c}_n^T \boldsymbol{X}_n = \phi X_n$$

One-Step Ahead Prediction of AR(1) Process (Cont'd)

- ϕX_n makes intuitive sense as a predictor since

$$X_{n+1} = \phi X_n + Z_{n+1}$$

- Prediction error is $X_{n+1} - \phi X_n = Z_{n+1}$ and

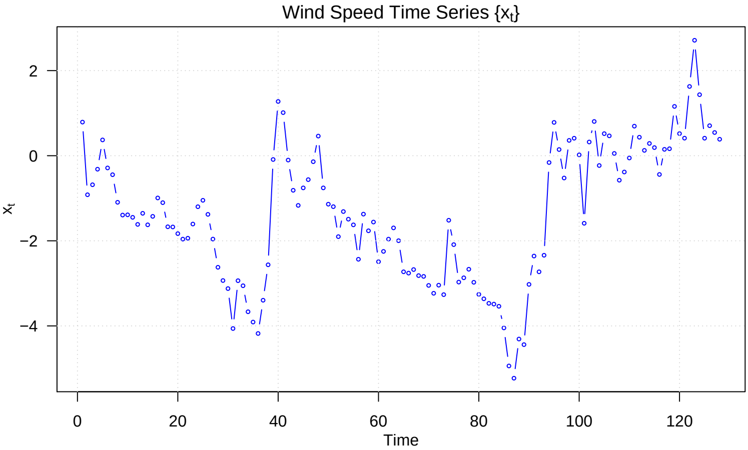
$$\mathbb{Cov}(Z_t, X_{n-j+1}) = 0,\ j = 1, \cdots, n$$

- MSPE is

$$\mathbb{Var}(X_{n+1} - \phi X_n) = \gamma(0) - \boldsymbol{c}_n^T \boldsymbol{\gamma}_n = 1 - \phi^2,$$

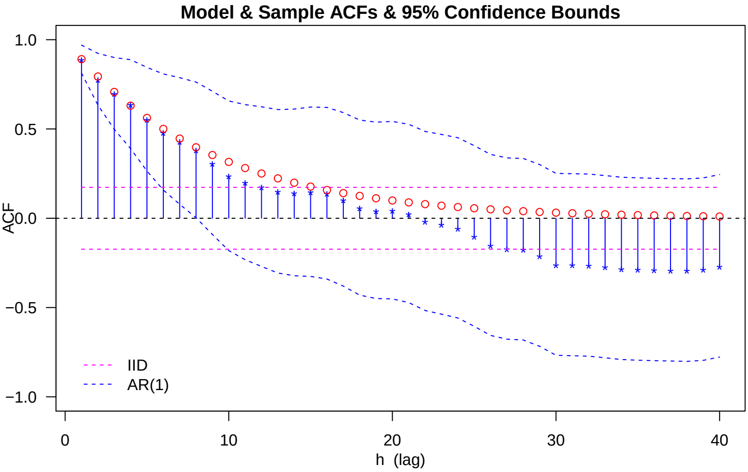
because $\boldsymbol{c}_n = (\phi, 0, \cdots, 0)^T$ and $\boldsymbol{\gamma}_n = (\phi, \phi^2, \cdots, \phi^n)^T$

Wind Speed Time Series Example [Source: UW stat 519 lecture notes by Donald Percival]



Let's use this series to illustrate forecasting one step ahead

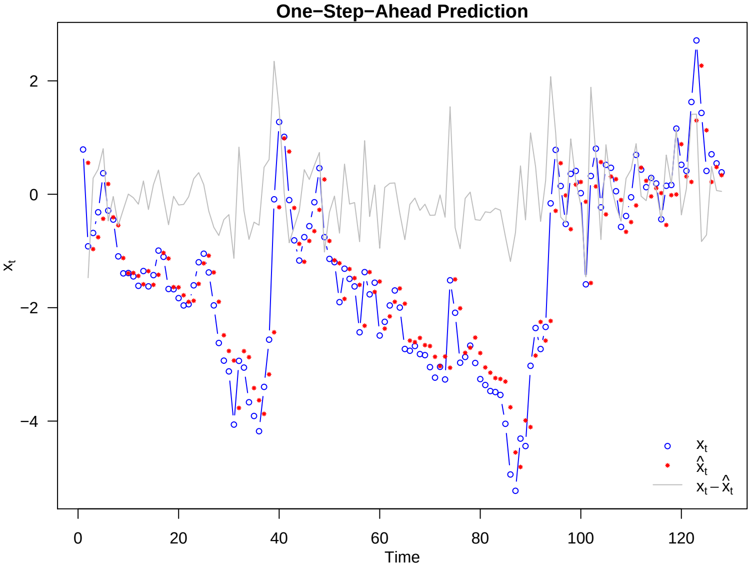
Model & Sample ACFs & 95% Confidence Bounds



The sample ACF indicates compatibility with AR(1) model

$\Rightarrow P_n X_{n+1} = \phi X_n$

One-Step-Ahead Prediction of Wind Speed Series



Predicting "Missing" Values

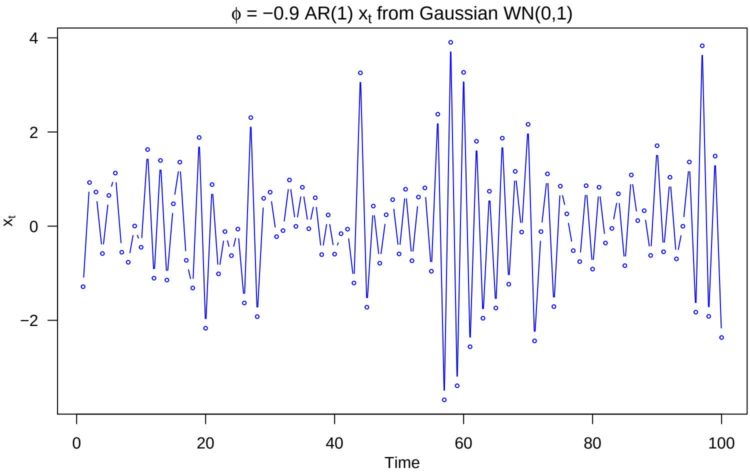
- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

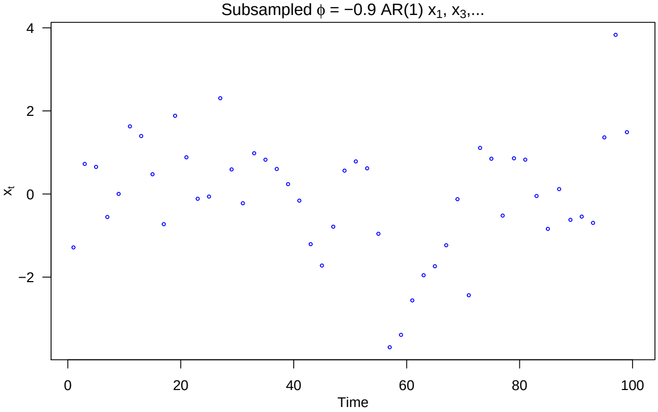
$$\begin{aligned}\mathrm{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3X_2})^2\right] \\ &= \mathbb{E}\left[(X_2 - c_0 - c_1 X_3 - c_2 X_1)^2\right]\end{aligned}$$

- Proceed as for the forecasting case to get the optimal coefficients:
  - Calculate derivatives
  - Set the derivatives equal to zero
  - Solve the linear system of equation

Another AR(1) Example with $\phi = -0.9$



Subsampled $X_1, X_3, \cdots$ and Removed $X_2, X_4, \cdots$



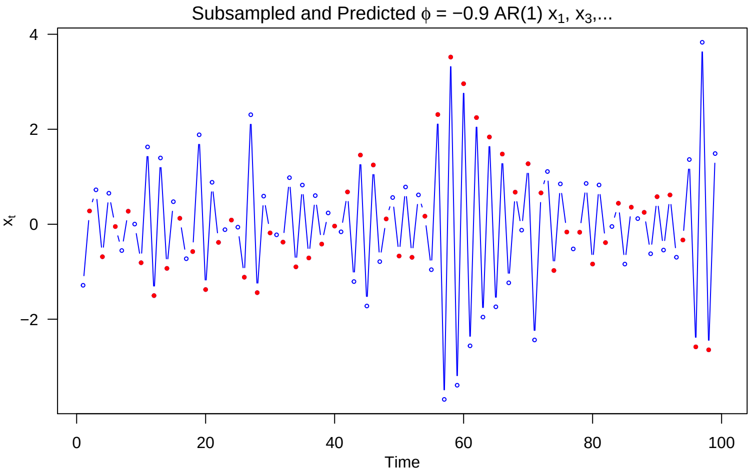
The best linear predictor of X_2 given X_1, X_3 is

$$\hat{X}_2 = \frac{\phi}{1+\phi^2}(X_1 + X_3),$$

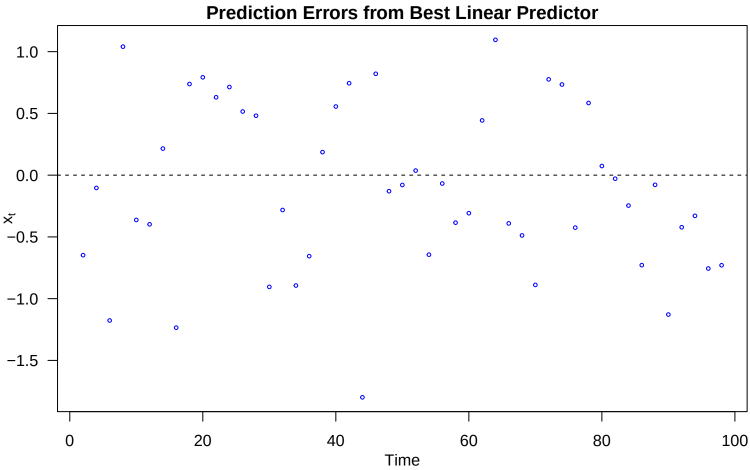
and the MSPE is

$$\frac{\sigma^2}{1+\phi^2}$$

Predict $X_2, X_4, \cdots$ Using Best Linear Predictor



Prediction Errors from Best Linear Predictor

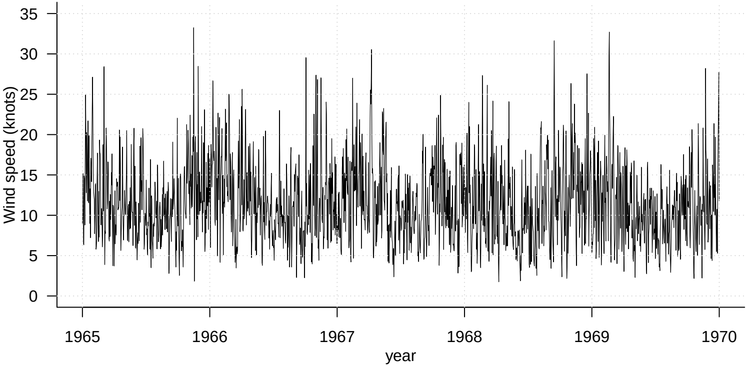


A Modeling Case Study of Ireland Wind Data

Data Description [Haslett & Raftery, 1989]

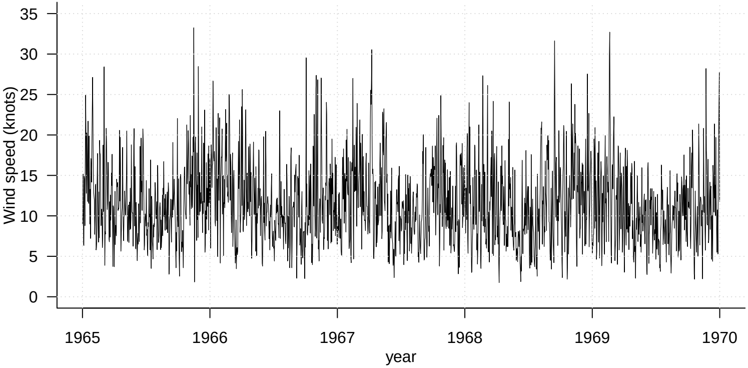
- 12 wind stations collected 6226 daily readings from 1/1/61 to 1/17/78. The wind speeds are measured in knots (1 knot = 0.5148 meters/second)
- We will focus on the wind data from 1965-1969 at the Rosslare station
- Modeling procedure:
  - Exploratory analysis
  - Model and remove the trend and seasonal components
  - Model identification, fitting, and selection
  - Perform forecast

Wind Speed Time Series at the Rosslare Station



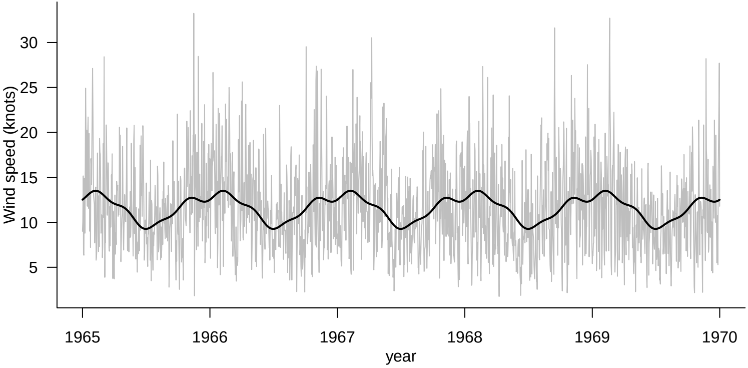
- No clear trend

Wind Speed Time Series at the Rosslare Station



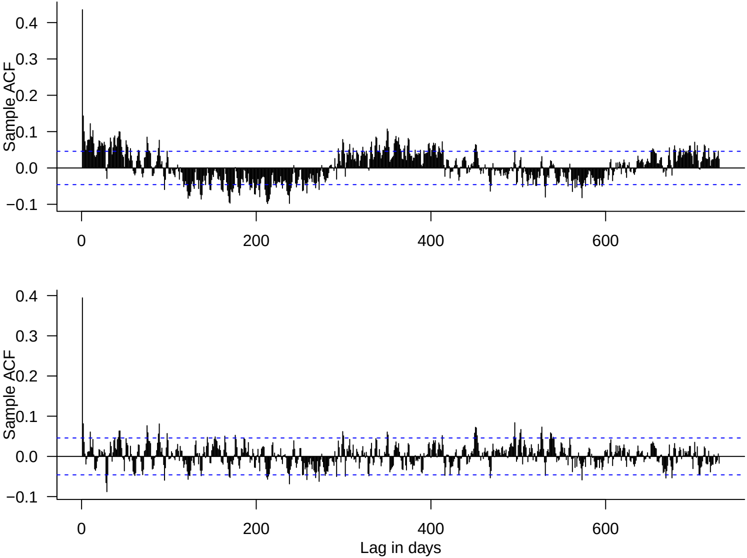
- No clear trend
- Seasonal Pattern

Estimating the Season Pattern

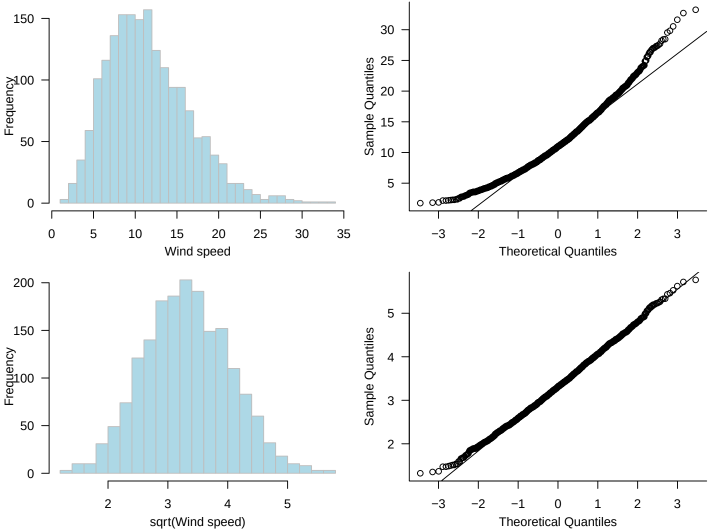


Here we fit a harmonic regression to account for the seasonal effects

ACF Plots: Original and Deseasonalized Series

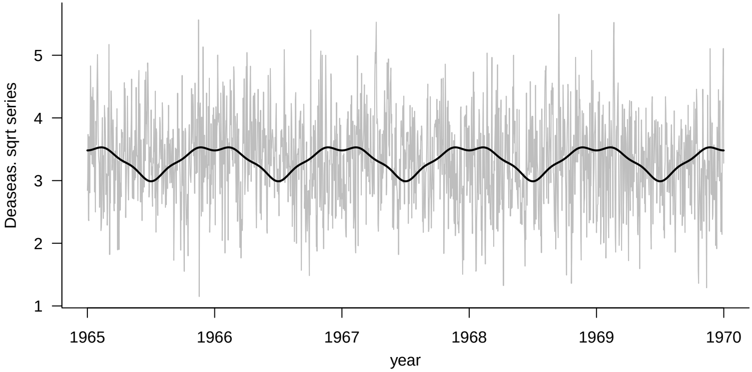


Apply Transformation to Make Wind Speeds More Gaussian Like



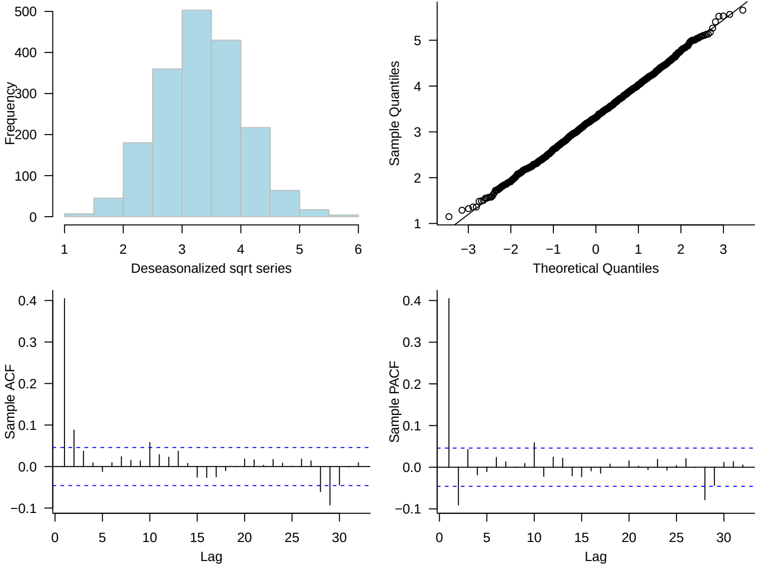
Now take square roots of the original data and deseasonalize again!

Estimating the Seasonal Component of the Transformed Series



Next, we need to check if the deseasonalized series Gaussian like

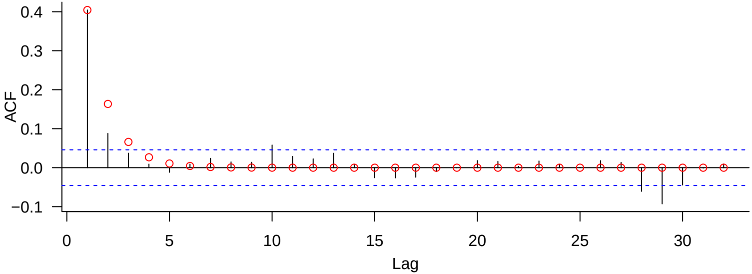
Marginal Distribution and ACF/PACF of the Deseasonalized Series



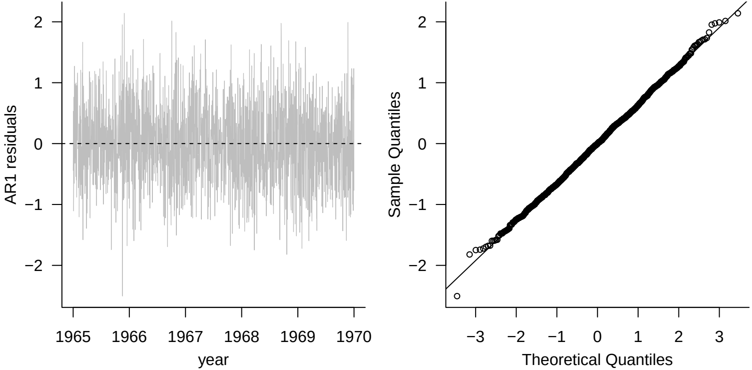
Based on ACF/PACF, which ARMA model would you choose?

Maximum Likelihood Estimation in R: AR(1)

```
> ## Fit an AR(1) model
> ar1.model <- arima(sqrt.rosslare.ds, order = c(1, 0, 0))
> ## summarize the model
> ar1.model

Call:
arima(x = sqrt.rosslare.ds, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.4044     3.3251
s.e.  0.0214     0.0253

sigma^2 estimated as 0.4149:  log likelihood = -1788.91,  aic = 3581.82
```



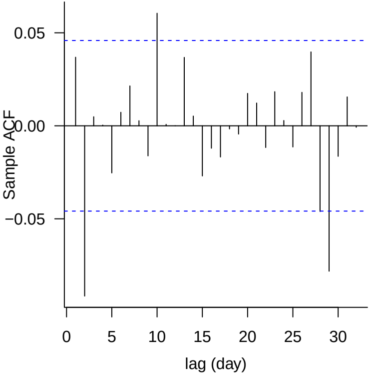
Residual Plots for the AR(1) Model

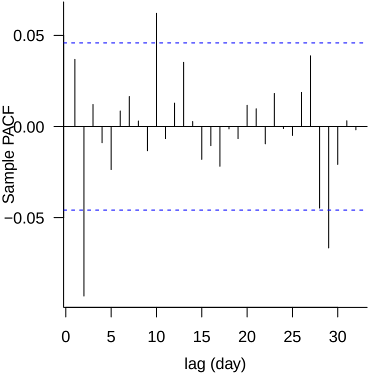


Normality assumption seems reasonable.

Next check the ACF/PACF and perform a Box test to assess if the AR(1) fit adequately account for temporal dependence strucuture

Diagnostic for the AR(1) Model

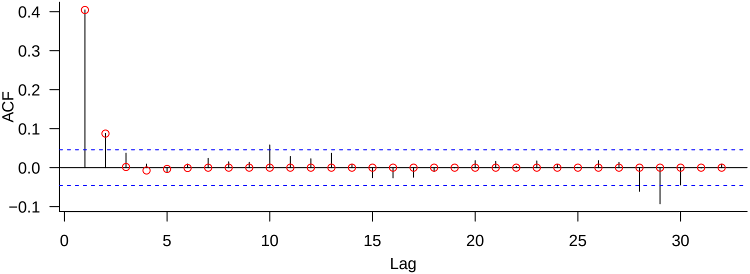




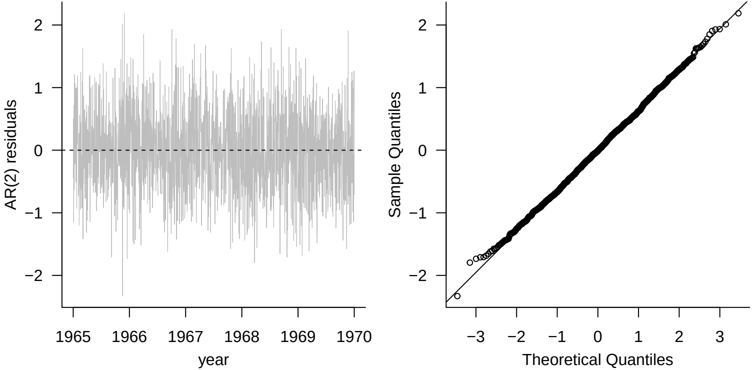

```
> Box.test(ar1.resids, lag = 32, type = "Ljung-Box")

        Box-Ljung test

data:  ar1.resids
X-squared = 53.656, df = 32, p-value = 0.009603
```

AR(2) Maximum Likelihood Estimation

```
> ## Fit an AR(2) model
> ar2.model <- arima(sqrt.rosslare.ds, order = c(2, 0, 0))
> ## summarize the model
> ar2.model

Call:
arima(x = sqrt.rosslare.ds, order = c(2, 0, 0))

Coefficients:
         ar1      ar2  intercept
      0.4413  -0.0911     3.3252
s.e.  0.0233   0.0233     0.0231

sigma^2 estimated as 0.4115:  log likelihood = -1781.32,  aic = 3568.65
```



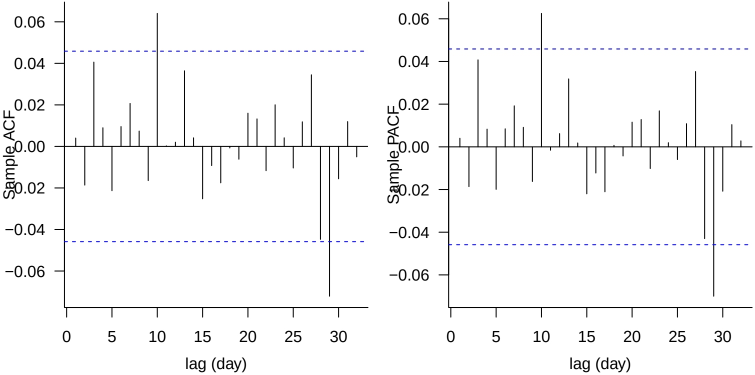
Residual Plots for the AR(2) Model



Normality assumption seems reasonable.

Next check the ACF/PACF and perform a Box test to assess if the AR(2) fit adequately account for temporal dependence strucuture

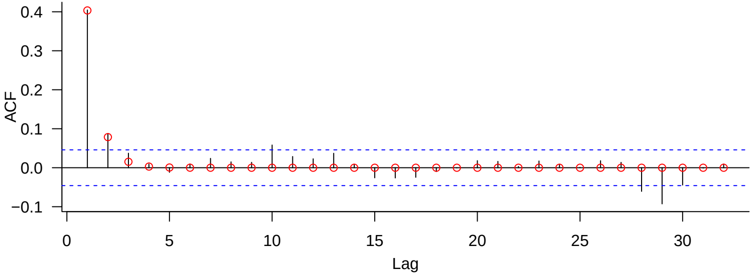
Diagnostic for the AR(2) Model



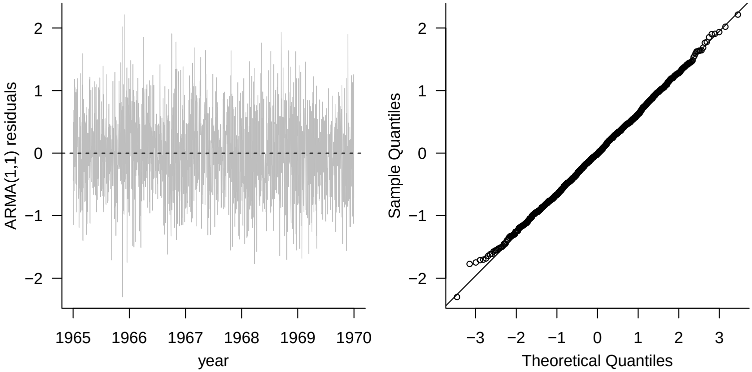

```
> Box.test(ar2.resids, lag = 32, type = "Ljung-Box")

        Box-Ljung test

data:  ar2.resids
X-squared = 36.852, df = 32, p-value = 0.2544
```

ARMA(1, 1) Maximum Likelihood Estimation

```
> ## Fit an ARMA(1,1) model
> arma11.model <- arima(sqrt.rosslare.ds, order = c(1, 0, 1))
> ## summarize the model
> arma11.model

Call:
arima(x = sqrt.rosslare.ds, order = c(1, 0, 1))

Coefficients:
         ar1     ma1  intercept
      0.1947  0.2521     3.3250
s.e.  0.0556  0.0553     0.0233

sigma^2 estimated as 0.4108:  log likelihood = -1779.92,  aic = 3565.83
```



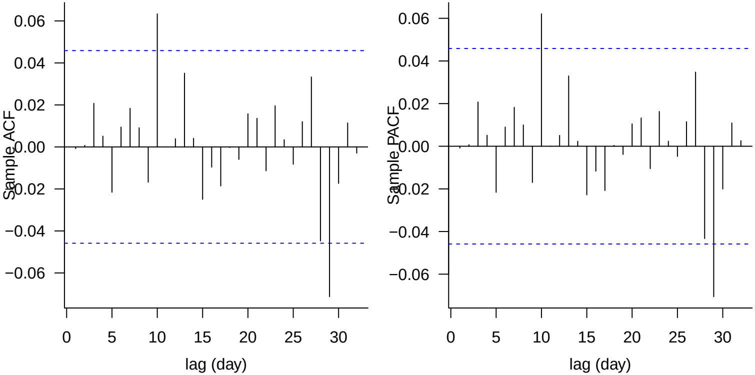
Residual Plots for the ARMA(1, 1) Model



Normality assumption seems reasonable.

Next check the ACF/PACF and perform a Box test to assess if the ARMA(1, 1) fit adequately account for temporal dependence strucuture

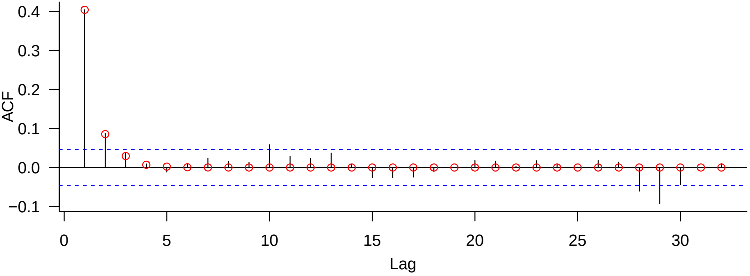
Diagnostic for the ARMA(1, 1) Model



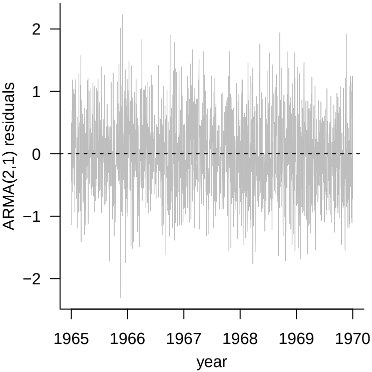

```
> Box.test(arma11.resids, lag = 32, type = "Ljung-Box")

        Box-Ljung test

data:  arma11.resids
X-squared = 33.09, df = 32, p-value = 0.4137
```

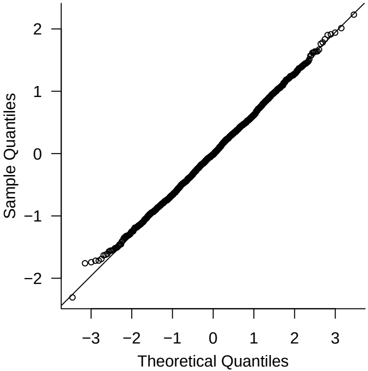
ARMA(2, 1) Maximum Likelihood Estimation

```
> ## Fit an ARMA(2,1) model
> arma21.model <- arima(sqrt.rosslare.ds, order = c(2, 0, 1))
> ## summarize the model
> arma21.model

Call:
arima(x = sqrt.rosslare.ds, order = c(2, 0, 1))

Coefficients:
         ar1     ar2     ma1  intercept
      0.0674  0.0584  0.3785     3.3247
s.e.  0.1693  0.0772  0.1665     0.0236

sigma^2 estimated as 0.4107:  log likelihood = -1779.66,  aic = 3567.32
```



Residual Plots for the ARMA(2, 1) Model

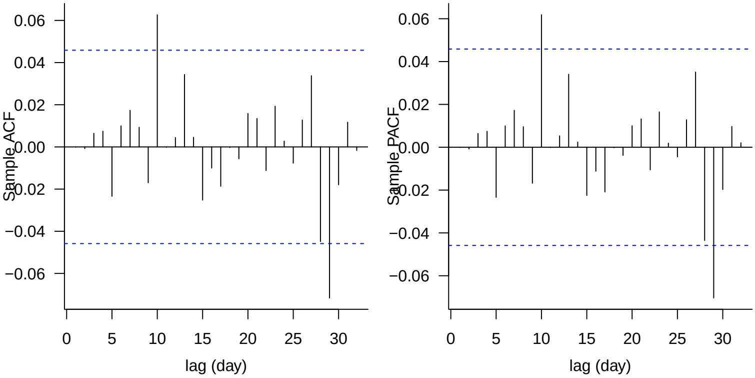




Normality assumption seems reasonable.

Next check the ACF/PACF and perform a Box test to assess if the ARMA(2, 1) fit adequately account for temporal dependence strucuture

Diagnostic for the ARMA(2, 1) Model




```
> Box.test(arma21.resids, lag = 32, type = "Ljung-Box")

        Box-Ljung test

data:  arma21.resids
X-squared = 32.537, df = 32, p-value = 0.4404
```

Comparing Models via Information Criteria

Model	AIC	AICC
AR(1)	3583.817	3583.824
AR(2)	3570.650	3570.663
ARMA(1, 1)	3567.833	3567.847
ARMA(2, 1)	3569.319	3569.341

Which model would you pick?

Forecasting future wind speeds

- Question: How do we predict wind speeds on the original scale, including the seasonality that was previously estimated?
- Suppose we want to predict the next month of wind speed values. We base our forecasts on the ARMA(1,1) model
- We need to reverse the order of our modeling

Forecasting future wind speeds, continued

- The forecasts for the next 31 days of deseasonalized square root values are:

```
> sqrt.rosslare.forecast <- predict(arma11.model, h)
> sqrt.rosslare.forecast$pred
 [1] 3.136357 3.288312 3.317896 3.323656 3.324778 3.324996 3.325039
 [8] 3.325047 3.325049 3.325049 3.325049 3.325049 3.325049 3.325049
[15] 3.325049 3.325049 3.325049 3.325049 3.325049 3.325049 3.325049
[22] 3.325049 3.325049 3.325049 3.325049 3.325049 3.325049 3.325049
[29] 3.325049 3.325049 3.325049
```

- The standard error for the forecasts are:

```
> round(sqrt.rosslare.forecast$se, 2)
 [1] 0.6409755 0.7020359 0.7042464 0.7043300 0.7043332 0.7043333
 [7] 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333
[13] 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333
[19] 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333
[25] 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333 0.7043333
[31] 0.7043333
```

Forecasting future wind speeds, continued

- Next, we add back in the seasonality to get:

```
> adj.forecast <- fitted(harm.model)[1:h] + sqrt.rosslare.forecast$pred
       1        2        3        4        5        6        7        8
3.292642 3.444667 3.474464 3.480576 3.482189 3.483033 3.483835 3.484730
       9       10       11       12       13       14       15       16
3.485742 3.486870 3.488110 3.489454 3.490896 3.492427 3.494039 3.495722
      17       18       19       20       21       22       23       24
3.497468 3.499267 3.501108 3.502981 3.504874 3.506778 3.508680 3.510569
      25       26       27       28       29       30       31
3.512434 3.514264 3.516047 3.517772 3.519428 3.521003 3.522487
```

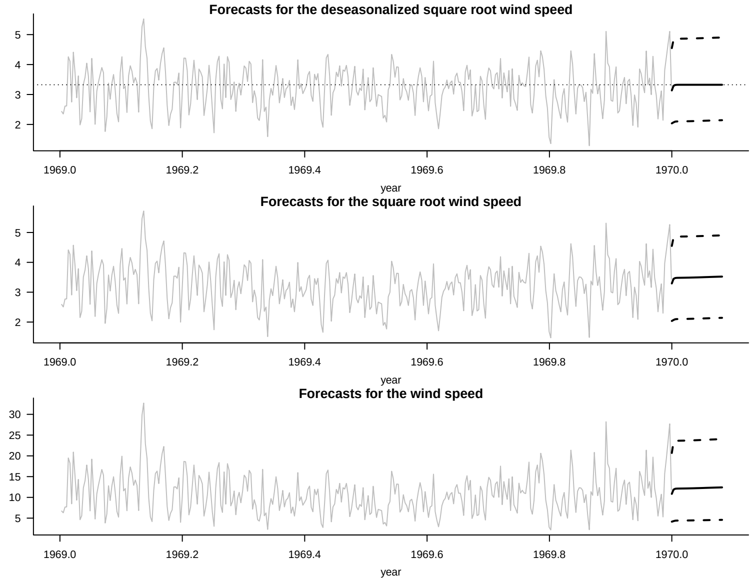
- Finally, we transform back to the original scale

```
       1        2        3        4        5        6        7        8
10.84149 11.86573 12.07190 12.11441 12.12564 12.13152 12.13710 12.14334
       9       10       11       12       13       14       15       16
12.15040 12.15826 12.16691 12.17629 12.18635 12.19704 12.20831 12.22007
      17       18       19       20       21       22       23       24
12.23229 12.24487 12.25776 12.27087 12.28414 12.29749 12.31083 12.32410
      25       26       27       28       29       30       31
12.33720 12.35005 12.36259 12.37472 12.38637 12.39746 12.40791
```

- To get the prediction limits, we need to transform the lower and upper prediction limits on the sqrt scale

```
> plus.or.minus <- qnorm(0.975) * sqrt.rosslare.forecast$se
> lower <- forecast - plus.or.minus
> upper <- forecast + plus.or.minus
```

Visualizing the Forecasts



Further Questions

- What is the full model for our time series data?
- Is there a better description for the trend rather than just a constant term?
- How well do we forecast?