Modeling and Forecasting Seasonality

Ka-fu WONG
University of Hong Kong

February 9, 2018

Contents

1. A recap of the unobserved components model

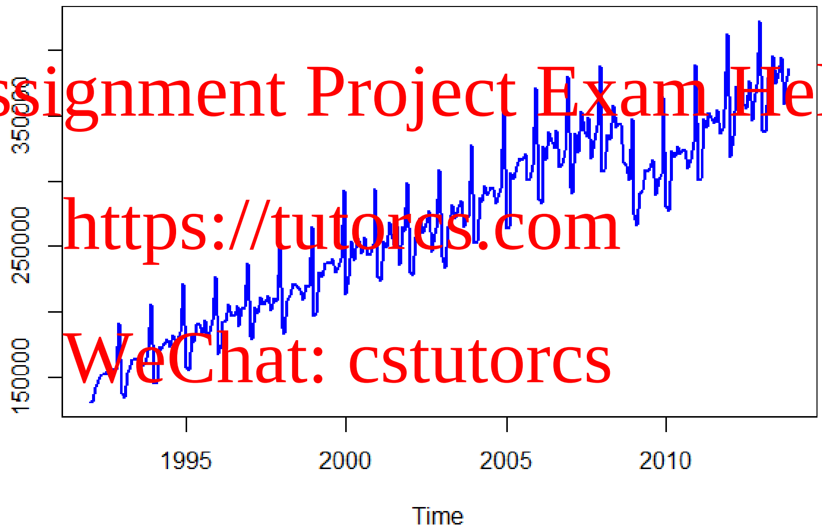
- According to the unobserved components model of a time series, the series y_t, is made up of the sum of three independent components
  - a time trend component
  - a seasonal component
  - an irregular or cyclical component.

$$y_t = \text{time trend} + \text{seasonal} + \text{cyclical} = T_t + S_t + C_t$$

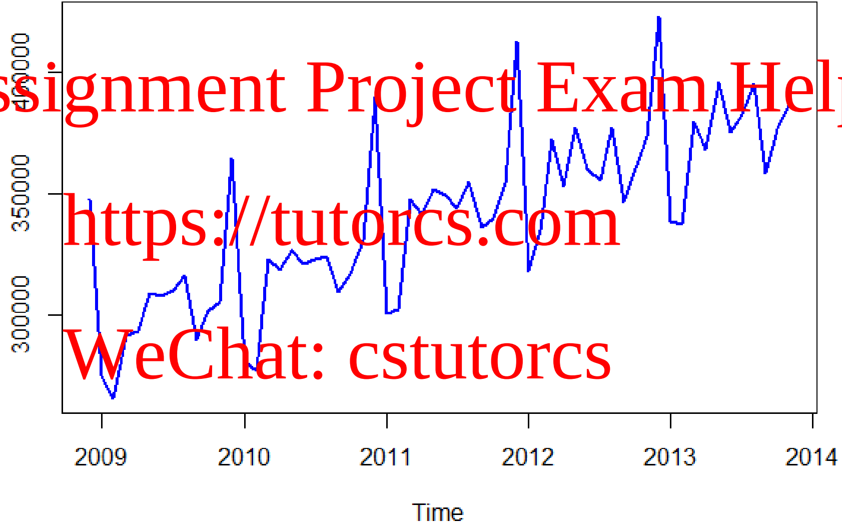
2. Seasonality

- Seasonality refers to the annual cyclical variation in a time series, which may be due to weather patterns, holiday patterns, school calendar patterns, etc.
  - Cyclical peaks in U.S. retail sales and employment during the last quarter of each calendar year due to the holiday shopping season
  - Cyclical troughs in U.S. housing starts during the winter months of each calendar year due to weather patterns.

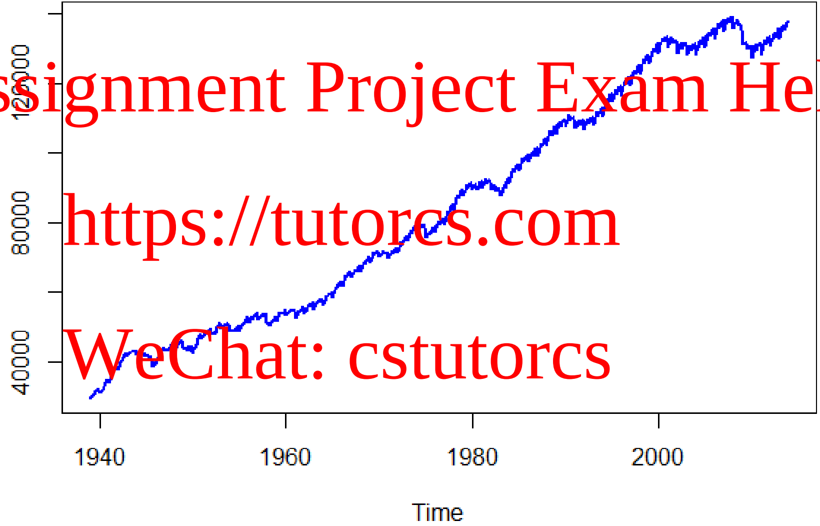
US Retailers Sales (1992:01 - 2013-11), Millions of Dollars



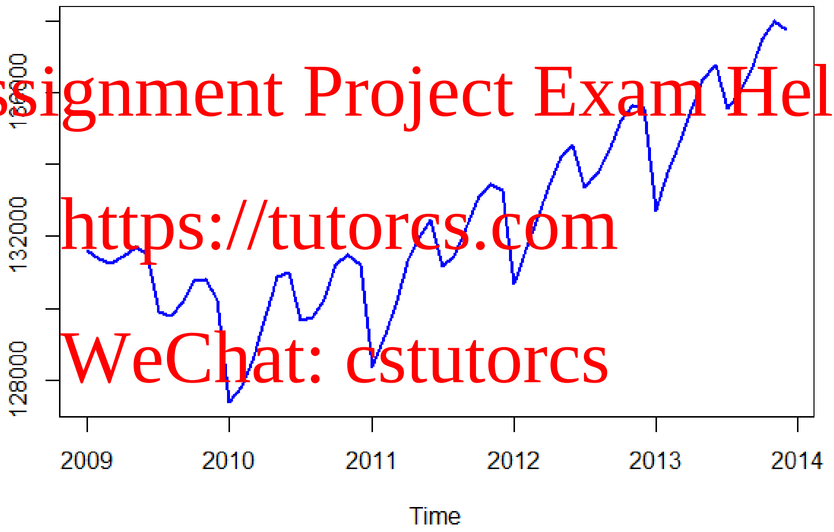
US Retailers Sales (2008:11 - 2013-11), Millions of Dollars



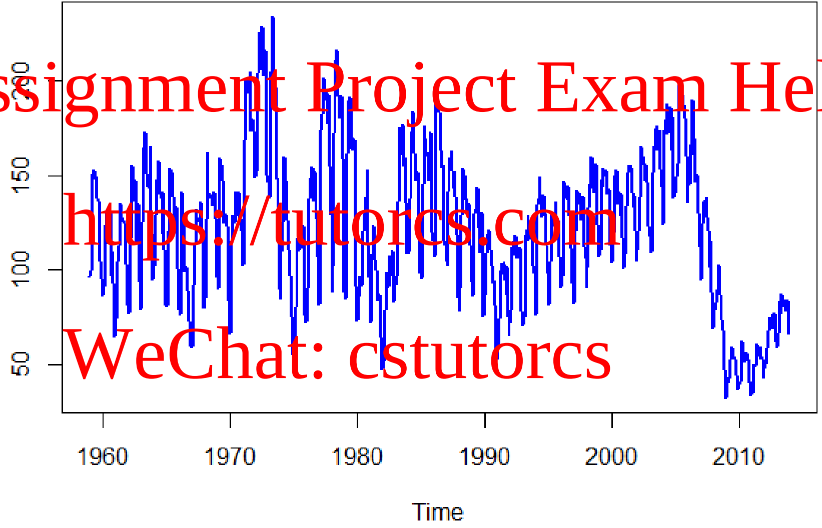
US Non-farm Employees (1939:01 - 2013-12), Thousands of Persons



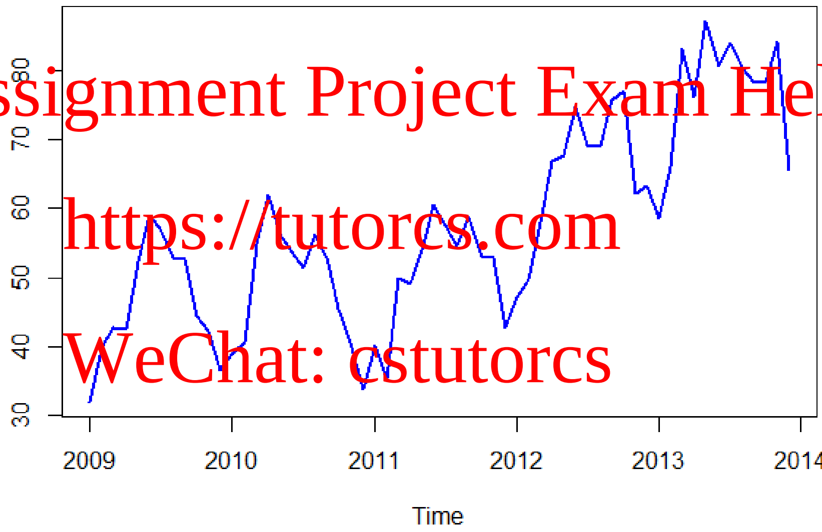
US Non-farm Employees (2009:01 - 2013-12), Thousands of Persons



US Housing Starts (1959:01 - 2013-12), Thousands of Units



US Housing Starts (2009:01 - 2013-12), Thousands of Units



3. Seasonal Adjustment

Most of the time series that we have ever used for economic analysis have been seasonally adjusted. That is,

$$y_{SA,t} = y_t - \hat{S}_t$$

where $\hat{S}_t$ is an estimate of the seasonal component of y_t.

3.1. Why use seasonally adjusted data.

- Typically, our interest in a macroeconomic time series is in the information it provides about the overall state of the economy and the direction the economy is heading.
- Suppose we did not seasonally adjust our data
  - we observe a huge increase in retail sales during the fourth quarter of 2005. Should we interpret this as a sign that the economy is suddenly booming?
    * No, unless retail sales are growing by more than normal for that part of the year.
  - we observe a huge increase in the unemployment rate during May/June of 2006 (after schools let out and there is a large temporary increase in the economy's labor force). Should I interpret this as a sign that the economy and its labor market are suddenly deteriorating?
    * No, unless the unemployment rate is increasing by more than normal for that part of the year.

- Seasonally adjusted data are meant to smooth out the data, i.e., to remove the regular ups and downs that are associated with the seasonal cycle.
- So, if seasonally adjusted retail sales increase during the fourth quarter or if the seasonally adjusted unemployment rate increases during May/June we can interpret these as movements beyond the movements that are part of the normal seasonal cycle.

3.2. Should use seasonally un-adjusted data?

- There are many business forecasting settings where the seasonal component of the series is fundamentally important and seasonally adjusted data would be inadequate and inappropriate.
- Some forecasters will be interested in predicting housing starts including the seasonal component.
  - A bank may be interested in forecasting housing starts in its area (or, if the bank is large enough, at a national level) in order to anticipate the demand for mortgage loans.
  - A business that provides building supplies to home-builders may need to forecast housing starts to anticipate the demand for its products and make current inventory decisions.
- In these cases, you would want to use seasonally unadjusted data, model the seasonality and forecast it, along with forecasts of the trend and cyclical components.

- Would anyone be interested in predicting the number of birth by Hong Kong residents?
  - Immigration Department
  - Hospital Authority
  - Private hospitals
  - Restaurant owners
  - Education Bureau

4. Modeling seasonal components as deterministic

- Two approaches to modeling seasonality –
  - deterministic seasonality and
  - stochastic seasonality
- The two approaches differ according to whether S_t is perfectly predictable or is subject to random disturbances.
- We will assume that the seasonal component is deterministic.

5. Modeling seasonality with dummies

- A straightforward and commonly used approach to modeling seasonality (which is, however, not the method government agencies typically use to seasonally adjust data) is to specify a seasonal dummy model.

5.1. **Modeling quarterly seasonality with dummies when we have quarterly data.** Suppose you are working with quarterly data and want to allow each quarter to have a distinct seasonal effect on the series.
The following two types of models will yield distinct quarterly seasonal effects.

(1) Model 1:

$$S_t = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{4t}$$

where

$$D_{it} = \begin{cases} 1 & \text{if } t = \text{quarter } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3, 4$. So,

- $S_t = \gamma_1$ if $t = \text{quarter } 1$,
- $S_t = \gamma_2$ if $t = \text{quarter } 2$,
- $S_t = \gamma_3$ if $t = \text{quarter } 3$,
- $S_t = \gamma_4$ if $t = \text{quarter } 4$.

(2) Model 2:

$$S_t = \delta_1 + \delta_2 D_{2t} + \delta_3 D_{3t} + \delta_4 D_{4t}$$

where

$$D_{it} = \begin{cases} 1 & \text{if } t = \text{quarter } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 2, 3, 4$. So,

- $S_t = \delta_1$ if $t = \text{quarter } 1$,
- $S_t = \delta_1 + \delta_2$ if $t = \text{quarter } 2$,
- $S_t = \delta_1 + \delta_3$ if $t = \text{quarter } 3$, and
- $S_t = \delta_1 + \delta_4$ if $t = \text{quarter } 4$.

	Seasonal Effects	
	Model 1	Model 2
Quarter 1	γ_1	δ_1
Quarter 2	γ_2	$\delta_1 + \delta_2$
Quarter 3	γ_3	$\delta_1 + \delta_3$
Quarter 4	γ_4	$\delta_1 + \delta_4$

5.2. **Modeling monthly seasonality with dummies when we have monthly data.** Suppose you are working with monthly data and want to allow each month to have a distinct seasonal effect on the series. The following two types of models will yield distinct monthly seasonal effects.

(1) Model 1:

$$S_t = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + ... + \gamma_{12} D_{12t}$$

where

$$D_{it} = \begin{cases} 1 & \text{if } t = \text{month } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3, ..., 12$. So,

- $S_t = \gamma_1$ if $t = \text{month } 1$,
- $S_t = \gamma_{12}$ if $t = \text{month } 2$,
- $\ldots$,
- $S_t = \gamma_{12}$ if $t = \text{month } 12$.

(2) Model 2:

$$S_t = \delta_1 + \delta_2 D_{2t} + \delta_3 D_{3t} + ... + \delta_{12} D_{12t}$$

where

$$D_{it} = \begin{cases} 1 & \text{if } t = \text{month } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 2, 3, ..., 12$. So,

- $S_t = \delta_1$ if $t = \text{month 1 (e.g., January)}$,
- $S_t = \delta_1 + \delta_2$ if $t = \text{month 2 (e.g., February)}$,
- $S_t = \delta_1 + \delta_3$ if $t = \text{month 3 (e.g., March)}$,
- ..., and
- $S_t = \delta_1 + \delta_{12}$ if $t = \text{month 12 (e.g., December)}$.

	Seasonal Effects	
	Model 1	Model 2
Month 1	γ_1	δ_1
Month 2	γ_2	$\delta_1 + \delta_2$
Month 3	γ_3	$\delta_1 + \delta_3$
...	...	...
Month 12	γ_{12}	$\delta_1 + \delta_{12}$

5.3. **Modeling quarterly seasonality with dummies when we have monthly data.** Suppose you are working with monthly data and want to allow each quarter to have a distinct seasonal effect on the series.

The following two types of models will yield distinct quarterly seasonal effects.

(1) Model 1:

$$S_t = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{4t}$$

where

$$D_{it} = \begin{cases} 1 & \text{if month } t \text{ is in quarter } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3, 4$. So,

- $S_t = \gamma_1$ if month t is in quarter 1,
- $S_t = \gamma_2$ if month t is in quarter 2,
- $S_t = \gamma_3$ if month t is in quarter 3,
- $S_t = \gamma_4$ if month t is in quarter 4.

(2) Model 2:

$$S_t = \delta_1 + \delta_2 D_{2t} + \delta_3 D_{3t} + \delta_4 D_{4t}$$

where

$$D_i = \begin{cases} 1 & \text{if month } t \text{ is in quarter } i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 2, 3, 4$. So,

- $S_t = \delta_1$ if $\text{month } t \text{ is in quarter } 1,$
- $S_t = \delta_1 + \delta_2$ if $\text{month } t \text{ is in quarter } 2,$
- $S_t = \delta_1 + \delta_3$ if $\text{month } t \text{ is in quarter } 3,$ and
- $S_t = \delta_1 + \delta_4$ if $\text{month } t \text{ is in quarter } 4.$

	Seasonal Effects	
	Model 1	Model 2
Month t in Quarter 1	γ_1	δ_1
Month t in Quarter 2	γ_2	$\delta_1 + \delta_2$
Month t in Quarter 3	γ_3	$\delta_1 + \delta_3$
Month t in Quarter 4	γ_4	$\delta_1 + \delta_4$

5.4. **Modeling irregular seasonality with dummies when we have monthly data.** Suppose you are working with monthly data and want to allow a distinct seasonal effect in December (month 12) and another for the remaining months (January to November) on the series.

The following two types of models will yield distinct December seasonal effects.

(1) Model 1:

$$S_t = \gamma_1 D_{1t} + \gamma_2 D_{2t}$$

where

$$D_{1t} = \begin{cases} 1 & \text{if } t \text{ is in month 1 to 11} \\ 0 & \text{otherwise} \end{cases}$$

$$D_{2t} = \begin{cases} 1 & \text{if } t \text{ is month 12} \\ 0 & \text{otherwise} \end{cases}$$

(2) Model 2:

$$S_t = \delta_1 + \delta_2 D_{2t}$$

where

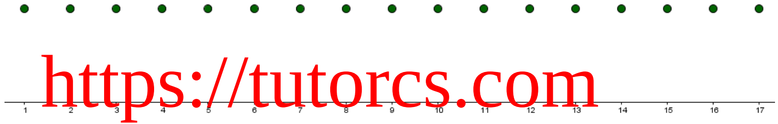
$$D_{2t} = \begin{cases} 1 & \text{if } t \text{ is month 12} \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we can model any distinct seasonal effects on data of any frequency. Of course, the number of distinct seasonal effects has to be less or equal to the frequency of data.

6. Estimating seasonal effects

For the sake of discussion, consider the following simple model

$$y_t = \beta + \epsilon_t, \quad \epsilon_t \sim (0, \sigma^2)$$



β can be estimated in two ways:

(1) Simple average of y_t.

$$\frac{1}{T}(y_1 + y_2 + \ldots + y_T) = \frac{1}{T}\sum_{t=1}^{T} y_t$$

(2) Regression:

$$y_t = \beta + \epsilon_t$$

Both methods yield identical results:

$$b = \frac{1}{T}\sum_{t=1}^{T} y_t$$

Under usual statistical conditions, the estimator

$$b \overset{A}{\sim} N(\beta, \sigma^2/T)$$

$$\frac{b-\beta}{\sigma/\sqrt{T}} \overset{A}{\sim} N(0,1)$$

Now, consider the following simple model with seasonal dummies

$$y_t = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \epsilon_t, \qquad \epsilon_t \sim (0, \sigma^2)$$

γ's can be estimated in two ways:

(1) Simple average of y_t:

(a) Use the observations that yields $D_{1t} = 1$, and compute simple average

$$\hat{\gamma}_1 = \frac{1}{\sum_{t=1}^{T} D_{1t}} \left(\sum_{t=1}^{T} D_{1t} y_t \right)$$

(b) Use the observations that yields $D_{2t} = 1$, and compute simple average

$$\hat{\gamma}_2 = \frac{1}{\sum_{t=1}^{T} D_{2t}} \left(\sum_{t=1}^{T} D_{2t} y_t \right)$$

(2) Regression:

$$y_t = \gamma_1 D_{1t} + \gamma_2 D_{1t} + \epsilon_t$$

Both methods yield identical results. Under usual statistical conditions, the estimator

$$\hat{\gamma}_1 \overset{A}{\sim} N\left(\gamma_1, \sigma^2 / \sum_{t=1}^{T} D_{1t}\right)$$

$$\hat{\gamma}_2 \overset{A}{\sim} N\left(\gamma_2, \sigma^2 / \sum_{t=1}^{T} D_{2t}\right)$$

where $\sigma^2 = Var(\epsilon_t)$.

7. Estimation of Trend and Seasonal components together

Suppose

$$y_t = T_t + S_t + \epsilon_t$$

where

- The trend is linear

$$T_t = \beta_0 + \beta_1 t$$

- The seasonal effects are modeled as quarterly

$$S_t = \delta_1 + \delta_2 D_{2t} + ... + \delta_4 D_{4t}$$

Estimation can take two steps.

(1) To estimate the trend model, ignoring seasonality, run a regression of

$$y_t = \beta_0 + \beta_1 t + u_t.$$

Let $\hat{u}_t$ be the estimated residuals.

(2) To estimate the seasonal model, run a regression of

$$\hat{u}_t = \delta_1 + \delta_2 D_{2t} + ... + \delta_4 D_{4t} + v_t$$

The alternative to estimate the trend and seasonality components in one step.

$$y_t = \beta_0 + \beta_1 t + \delta_1 + \delta_2 D_{2t} + ... + \delta_4 D_{4t} + \epsilon_t$$

Note that the model has a redundant parameter, since it has two constants, β_0 and δ_1. These two parameters cannot be estimated at the same time. In estimation, all statistical programs will return error (say, singular matrix encountered) because the variables included are perfectly co-linear, or, some set of variables are linearly dependent. Thus, we have to drop one of them.

Beware of perfect collinearity when including both trend and seasonal components!

- Model 1:

$$y_t = \beta_0 + \beta_1 t + \delta_2 D_{2t} + \ldots + \delta_4 D_{4t} + \epsilon_t$$

- Model 2:

$$y_t = \beta_1 t + \delta_1 D_{1t} + \delta_2 D_{2t} + \ldots + \delta_4 D_{4t} + \epsilon_t$$

According to this model of trend and seasonality, the nature of the seasonality is that the intercept of the trend line differs for each season.

- In Model 2, eliminating β_0 from the model has no consequence since the intercept of β_0 is accounted for by the seasonality jointly.
- In Model 1, eliminating δ_1 from the model has no consequence since the intercept of δ_1 is accounted for by β_0.

Remark: Although we focus on a linear trend model and quarterly seasonality, the discussion can be easily extended to models with more complicated trend and seasonality.

Usually we preferred the following model:

$$y_t = \beta_1 t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + ... + \gamma_4 D_{4t} + \epsilon_t$$

In the model, we can imagine the seasonal component

$$\gamma_1 D_{1t} + \gamma_2 D_{2t} + ... + \gamma_4 D_{4t}$$

is nothing more than allowing the constant term to vary with time. Thus, it is reasonable to use $\gamma_1 D_{1t} + \gamma_2 D_{2t} + ... + \gamma_4 D_{4t}$ to substitute out the constant term (β_0).
Using this interpretation, we will avoid the perfect collinear problem easily.

It is better to use $\gamma_1 D_{1t} + \gamma_2 D_{2t} + ... + \gamma_4 D_{4t}$ instead of $\delta_1 + \delta_2 D_{2t} + ... + \delta_4 D_{4t}$ because the estimated values of γ will immediately give us the pattern of seasonality, without the need to do additional calculations.

8. Forecasting y_{T+h}

The model

$$y_t = \beta_1 t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + ... + \gamma_4 D_{4t} + \epsilon_t$$

implies

$$y_{T+h} = \beta_1 (T+h) + \gamma_1 D_{1T+h} + \gamma_2 D_{2T+h} + ... + \gamma_4 D_{4T+h} + \epsilon_{T+h}$$

So, our forecast of y_{T+h} formed at time T will be:

$$\hat{y}_{T+h,T} = \hat{\beta}_1 (T+h) + \hat{\gamma}_1 D_{1T+h} + \hat{\gamma}_2 D_{2T+h} + ... + \hat{\gamma}_4 D_{4T+h} + \hat{\epsilon}_{T+h,T}$$

where $\hat{\epsilon}_{T+h,T}$ is the forecast of ϵ_{T+h} based on time T information.
If the ϵ's are i.i.d. with mean zero, then $\hat{\epsilon}_{T+h,T}=0$ and

$$\hat{y}_{T+h,T} = \hat{\beta}_1 (T+h) + \hat{\gamma}_1 D_{1T+h} + \hat{\gamma}_2 D_{2T+h} + ... + \hat{\gamma}_4 D_{4T+h}$$

If the ϵ's are i.i.d. $N(0, \sigma_\epsilon^2)$ then, ignoring parameter uncertainty, the forecast error $e_{T+h,T} = \epsilon_{T+h}$ and

$$\frac{y_{T+h} - y_{T+h,T}}{\hat{\sigma}_\epsilon} \approx N(0,1)$$

where

$$\hat{\sigma}_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2$$

When the parameter has to be estimated, we have the forecast error $e_{T+h,T} \neq \epsilon_{T+h}$ and

$$\frac{y_{T+h} - \hat{y}_{T+h,T}}{\hat{\sigma}_e} \approx N(0,1)$$

where

$$\hat{\sigma}_e^2 = \frac{1}{T-5} \sum_{t=1}^{T} \hat{e}_t^2$$

and 5 is the degree of freedom lost due to the estimation of 5 parameters (2 trend parameters + 3 seasonality parameters, or 1 trend parameters + 4 seasonality parameters, depending on the specification).

Hence, the 95% forecast interval for y_{T+h} is

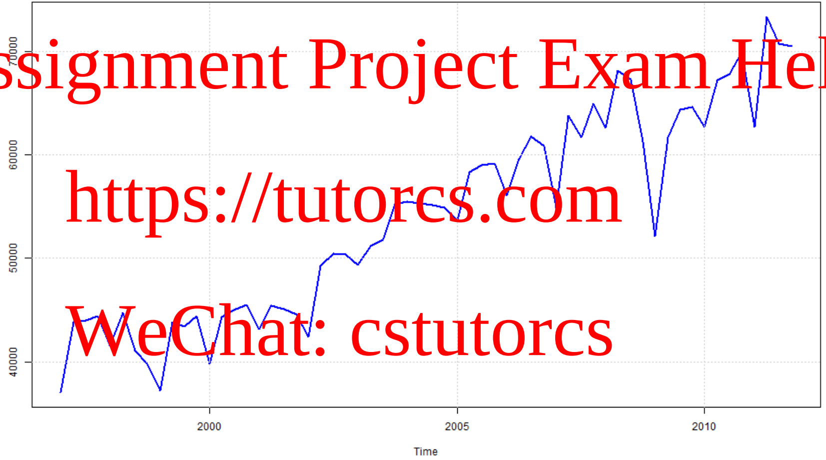
$$\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}_e$$

9. Application: Forecast of Hong Kong Port Cargo Throughput

- Quarterly Port Cargo Throughput data (in thousand tonnes) from 1997:01 to 2013:03, a total of 67 observations, were obtained from the Census and Statistics Department Website.[1]
- Port cargo comprises seaborne cargo and river cargo.
  - Seaborne cargo refers to cargo transported by vessels plying beyond the river trade limits.
  - River cargo refers to cargo transported by vessels plying exclusively within the river trade limits (i.e., the Pearl River, Mirs Bay and Macao, and other inland waterways in Guangdong and Guangxi).
- We save 7 observations (2012:01 to 2013:03) for checking the accuracy of our model out of sample. That is, estimation and model selection use only 60 observations.

[1] http://www.censtatd.gov.hk/hkstat/sub/sp130.jsp?tableID=085&ID=0&productType=8

HK Port Cargo Throughput (1997:01 - 2013:03), in thousands tonnes



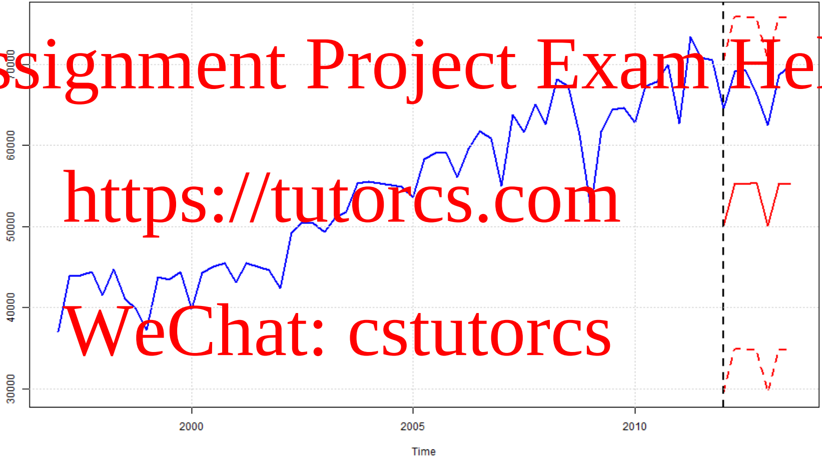
9.1. **Model with Seasonality Only.**

	Coefficient	Std. Error	t-Stat	p-value
d1	50052	2557	19.57	<2e-16
d2	55352	2557	21.65	<2e-16
d3	55249	2557	21.61	<2e-16
d4	55384	2557	21.66	<2e-16

Residual standard error: 9903 on 56 degrees of freedom
Multiple R-squared: 0.9696, Adjusted R-squared: 0.9675
F-statistic: 447 on 4 and 56 DF, p-value: < 2.2e-16
AIC=1280.204; SIC=1290.676

Residuals (Model with Seasonality Only)

Forecasts (Model with Seasonality Only)

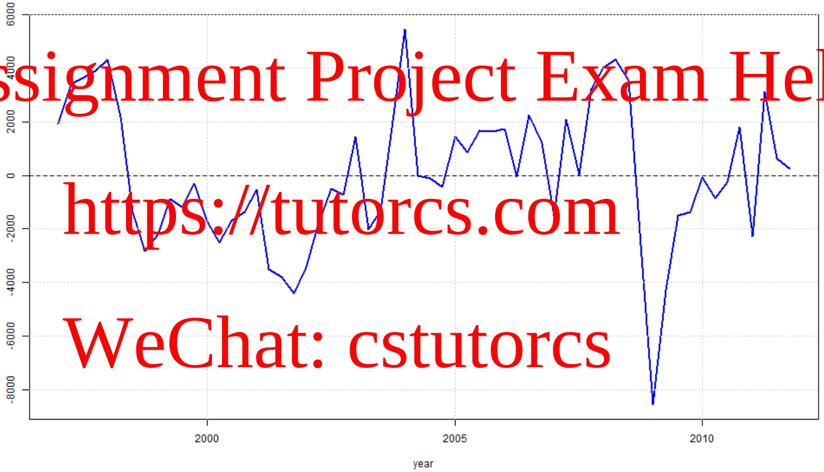


9.2. **Model with Seasonality and Linear Trend.**

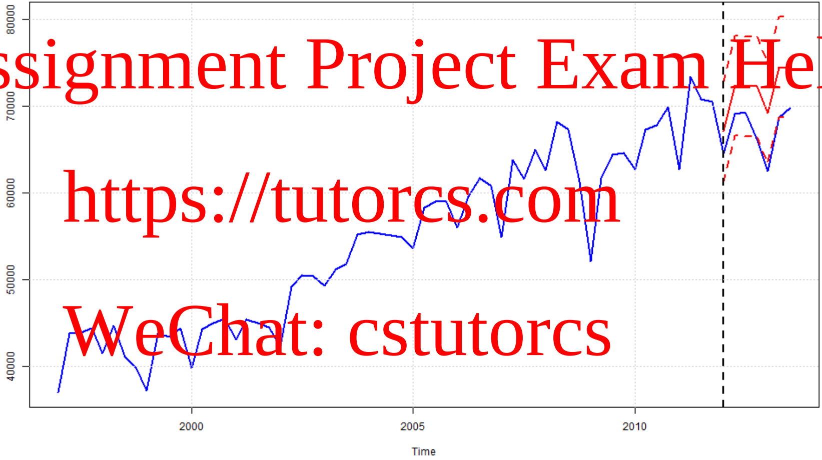
	Coefficient	Std. Error	t-Stat	p-value
d1	34601.73	915.46	37.8	<2e-16
d2	39369.25	928.63	42.4	<2e-16
d3	38733.36	942.06	41.12	<2e-16
d4	38335.55	955.73	40.11	<2e-16
TIME	532.75	20.29	26.26	<2e-16

Residual standard error: 2716 on 55 degrees of freedom
Multiple R-squared: 0.9978, Adjusted R-squared: 0.9976
F-statistic: 4891 on 5 and 55 DF, p-value: < 2.2e-16
AIC=1125.889; SIC=1138.455

Residuals (Model with Seasonality and Linear Trend)



Forecasts (Model with Seasonality and Linear Trend)



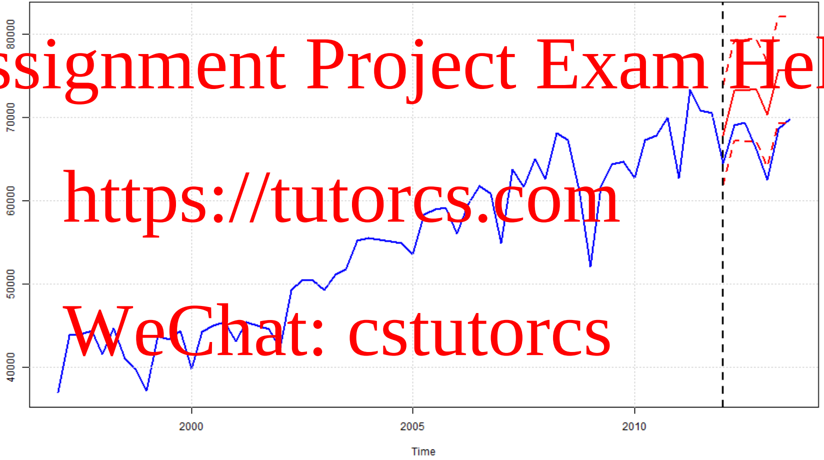
9.3. **Model with Seasonality and Quadratic Trend.**

	Coefficient	Std. Error	t-Stat	p-value
d1	35370.753	1232.462	28.699	<2e-16
d2	40140.714	1244.039	32.266	<2e-16
d3	39504.831	1254.115	31.5	<2e-16
d4	39104.571	1262.727	30.968	<2e-16
TIME	458.21	82.408	5.56	8.56E-07
TIME2	1.222	1.309	0.933	0.355

Residual standard error: 2719 on 54 degrees of freedom
Multiple R-squared: 0.9978, Adjusted R-squared: 0.9975
F-statistic: 4066 on 6 and 54 DF, p-value: $<$ 2.2e-16
AIC=1126.929; SIC=1141.589

Residuals (Model with Seasonality and Quadratic Trend)

Forecasts (Model with Seasonality and Quadratic Trend)

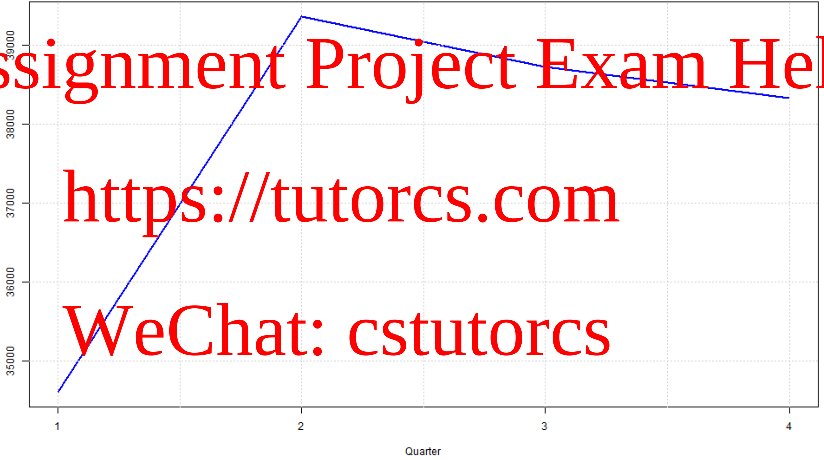


9.4. **Comparison.**

	In-sample criteria		Out-of-sample
	AIC	SIC	MSPE
Seasonality without Trend	1280.204	1290.676	180542811
Seasonality with Linear Trend	1125.889	1138.455	23403563
Seasonality with Quadratic Trend	1126.929	1141.589	34040818

- The best model chosen based on in-sample criteria (AIC and SIC) also yield the best out-of-sample performance (measured in terms of MSPE).
- The forecast with interval including the realized values need not be the best.

Seasonality Pattern (Model with Seasonality and Linear Trend)

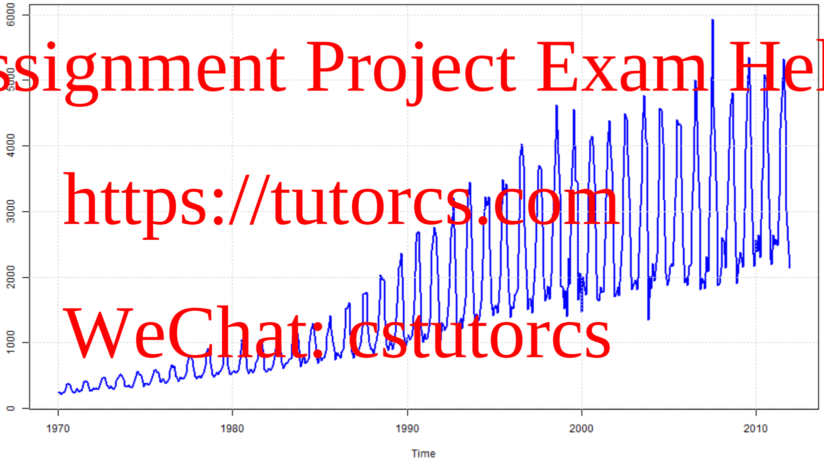


10. Application: Forecast of Hong Kong Electricity Consumption

- Monthly Domestic Electricity Consumption (in Terajoule) from 1970:01 to 2014:11, a total of 539 observations were obtained from the Census and Statistics Department Website.[2]
- We save 35 observations (2012:01 to 2014:11) for checking the accuracy of our model out of sample. That is, estimation and model selection use only 504 observations.

[2] http://www.censtatd.gov.hk/hkstat/sub/sp90.jsp?tableID=127&ID=0&productType=8

HK Domestic Electricity Consumption (1970:01 - 2011:12), in Terajoule



10.1. **Forecast models considered.**

- Model with seasonality only

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \epsilon_t$$

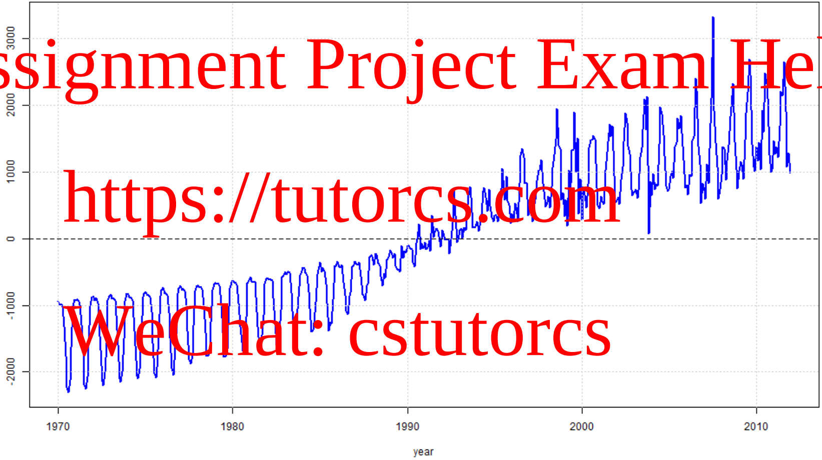
- Model with seasonality and linear trend

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \beta_1 TIME_t + \epsilon_t$$

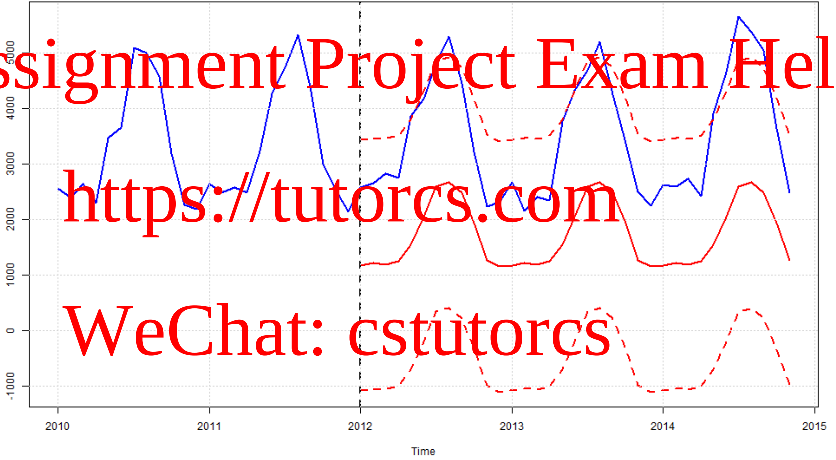
- Model with seasonality and quadratic trend

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \beta_1 TIME_t + \beta_2 TIME_t^2 + \epsilon_t$$

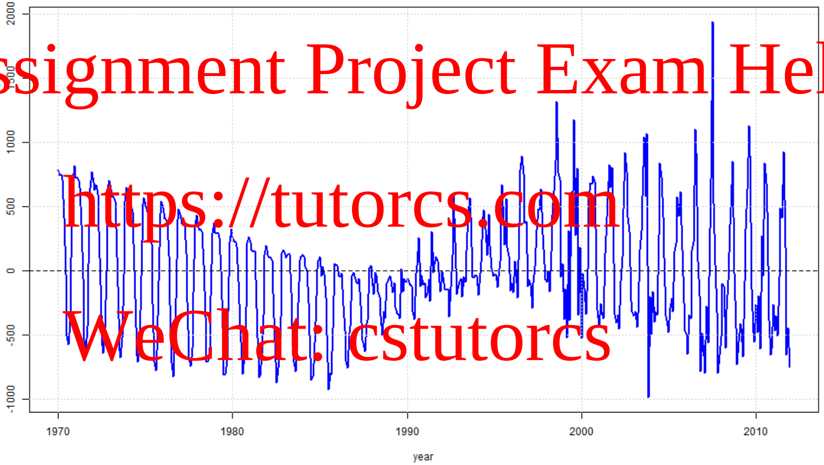
Residuals (Model with Seasonality Only)



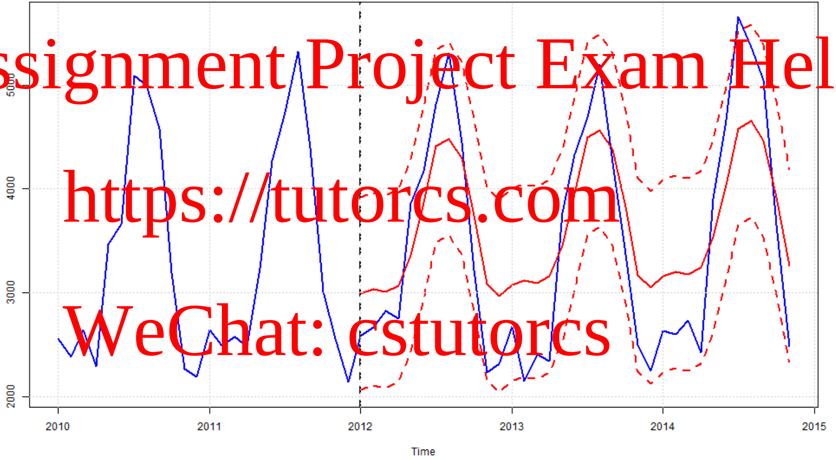
Forecasts (Model with Seasonality Only)



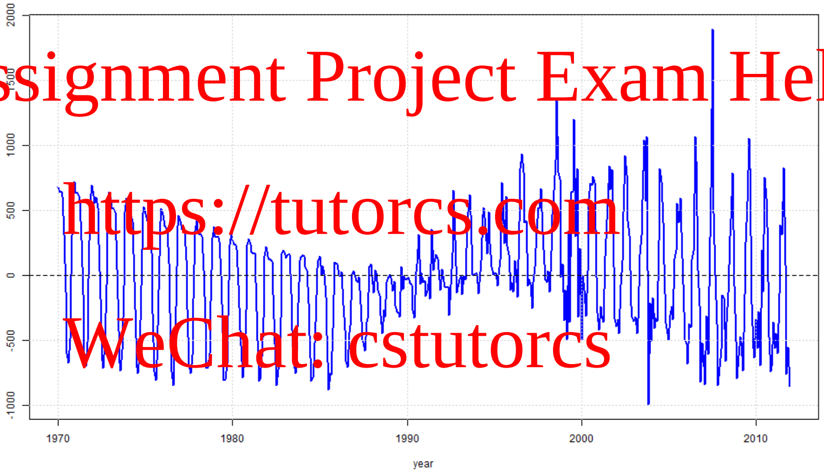
Residuals (Model with Seasonality and Linear Trend)



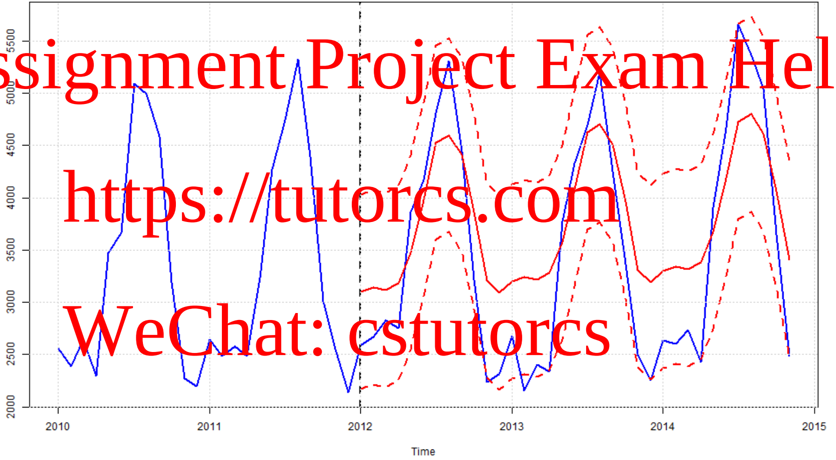
Forecasts (Model with Seasonality and Linear Trend)



Residuals (Model with Seasonality and Quadratic Trend)



Forecasts (Model with Seasonality and Quadratic Trend)

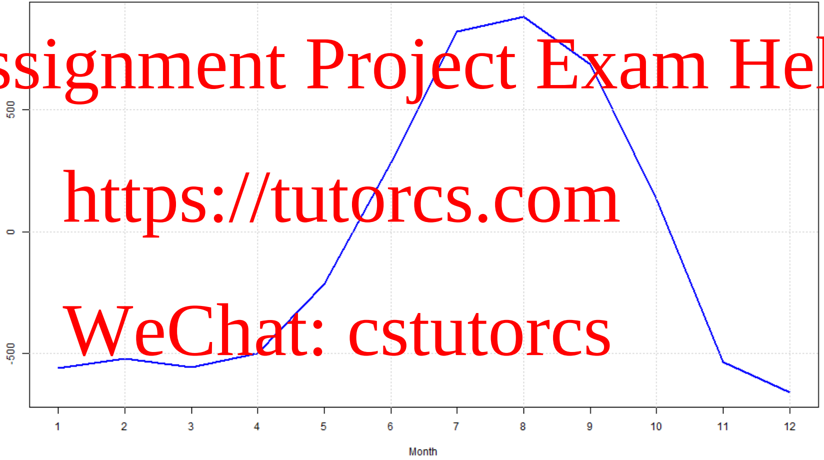


10.2. **Comparison.**

	In-sample criteria		Out-of-sample
	AIC	SIC	MSPE
Seasonality without Trend	8534.46	8589.35	3446619
Seasonality with Linear Trend	7635.65	7694.77	343331
Seasonality with Quadratic Trend	7632.32	7695.66	392512

- AIC indicates Seasonality with Quadratic Trend.
- SIC indicates Seasonality with Linear Trend.
- Seasonality with Linear Trend has the best out-of-sample performance (measured in terms of MSPE).

Seasonality Pattern (Model with Seasonality and Linear Trend)



10.3. **Log-Transformation.** From the data plot, we observe

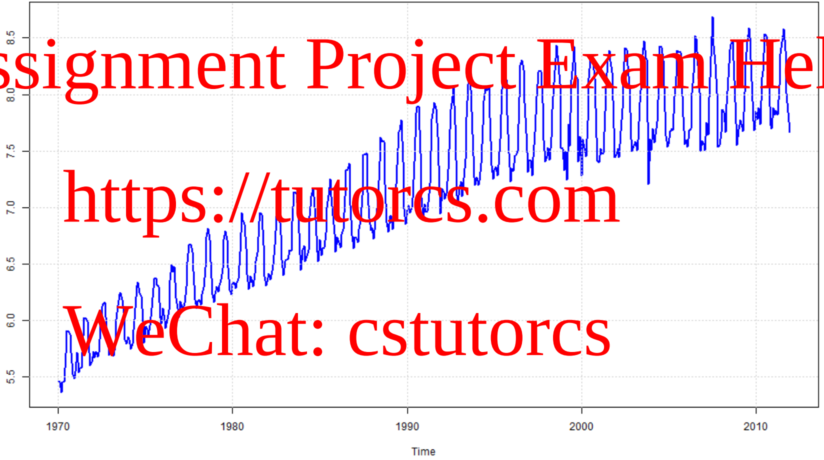
(1) Volatility increases with time.
(2) Volatility increases with the Electricity Consumption.

A log transformation will reduce the changing volatility with levels.
Regression based on data with similar volatility across observations tend to yield more precise estimates.

- y = log domestic electricity consumption.
- Once forecast of y is obtained, we take exponential transformation to obtain the forecast of domestic electricity consumption, i.e.,

$$\exp(\hat{y}_{T+h,T})$$

Log-transformed HK Domestic Electricity Consumption (1970:01 - 2011:12)



10.4. **Forecast models considered.**

- Model with seasonality only

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \epsilon_t$$

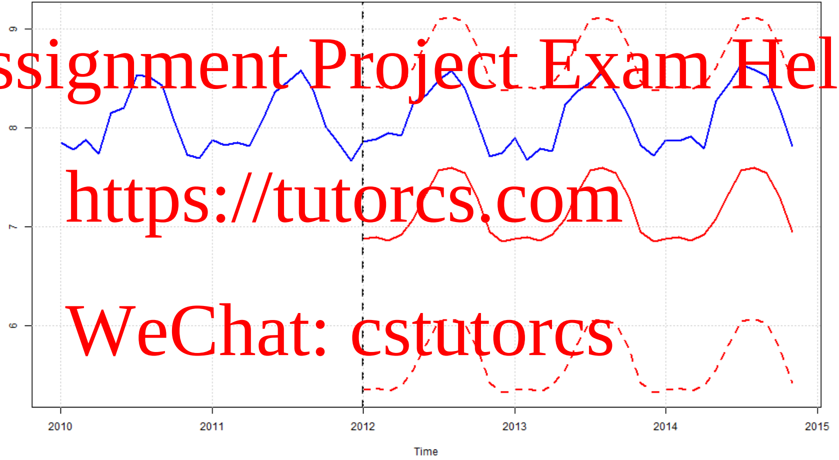
- Model with seasonality and linear trend

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \beta_1 TIME_t + \epsilon_t$$

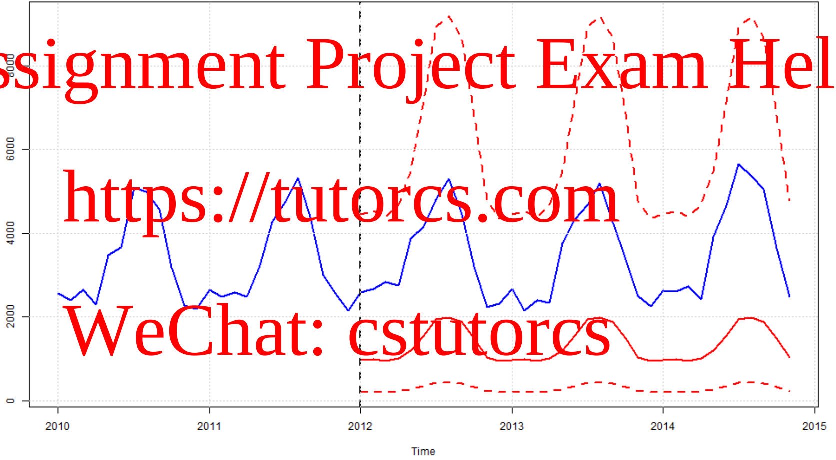
- Model with seasonality and quadratic trend

$$y_t = \gamma_1 M01_t + \gamma_2 M02_t + ... + \gamma_{12} M12_t + \beta_1 TIME_t + \beta_2 TIME_t^2 + \epsilon_t$$

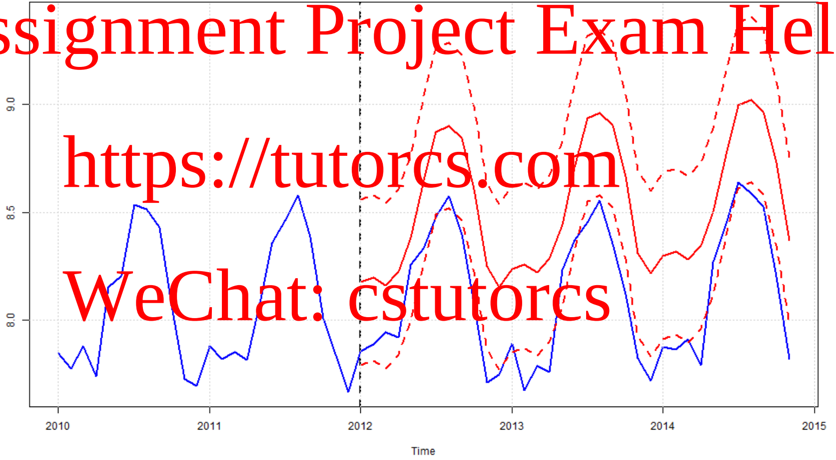
Forecast of Log Electricity Consumption (Model with Seasonality Only)



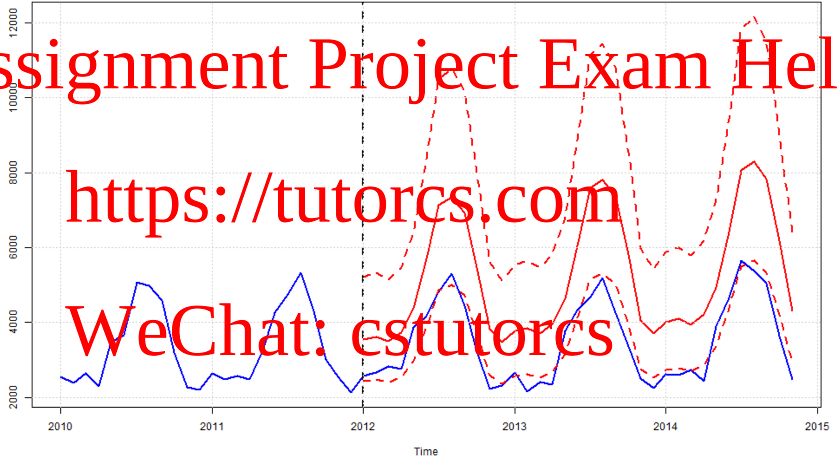
Forecast of Electricity Consumption (Model with Seasonality Only)



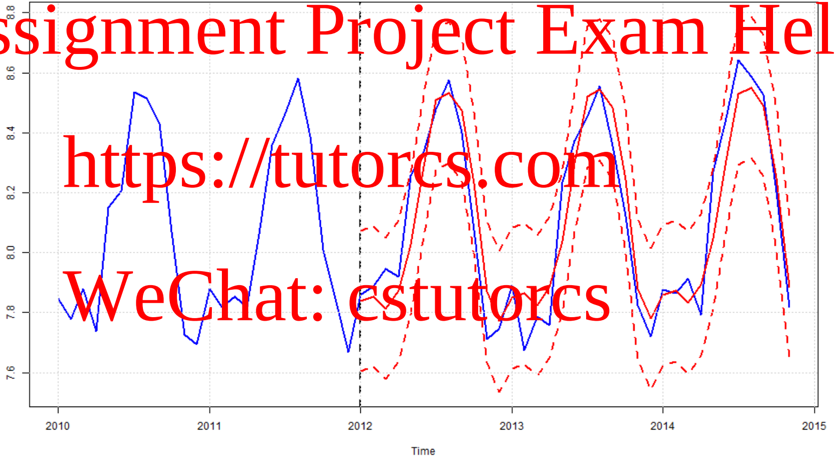
Forecast of Log Electricity Consumption (Model with Seasonality and Linear Trend)



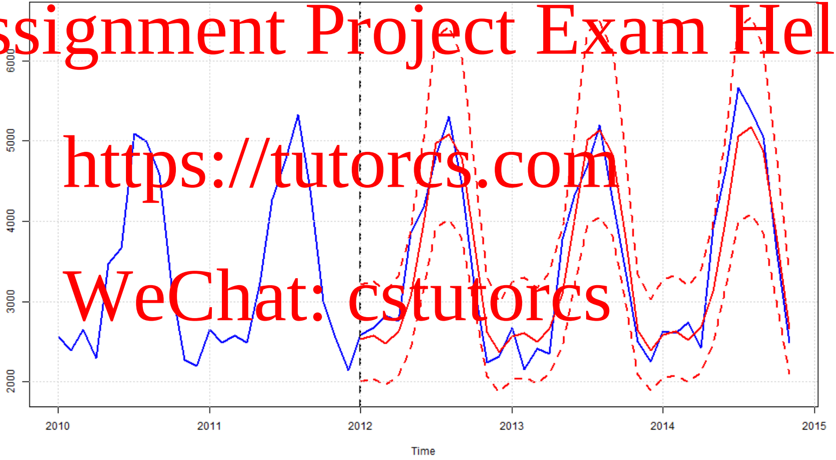
Forecast of Electricity Consumption (Model with Seasonality and Linear Trend)



Forecast of Log Electricity Consumption (Model with Seasonality and Quadratic Trend)



Forecast of Electricity Consumption (Model with Seasonality and Quadratic Trend)



10.5. **Comparison.**

	In-sample criteria		MSPE (Out-of-sample)
	AIC	SIC	consumption
Seasonality without Trend	1178.57	1233.464	5175881
Seasonality with Linear Trend	-215.887	-156.771	3524279
Seasonality with Quadratic Trend	-715.416	-652.078	131271

- Both AIC and SIC indicates Seasonality with Quadratic Trend.
- Seasonality with Quadratic Trend has the best out-of-sample performance (measured in terms of MSPE).

	MSPE (Out-of-sample)	
	Model based on	
	consumption	log consumption
Seasonality without Trend	3446619	5175881
Seasonality with Linear Trend	*343331*	3524279
Seasonality with Quadratic Trend	392512	*131271*

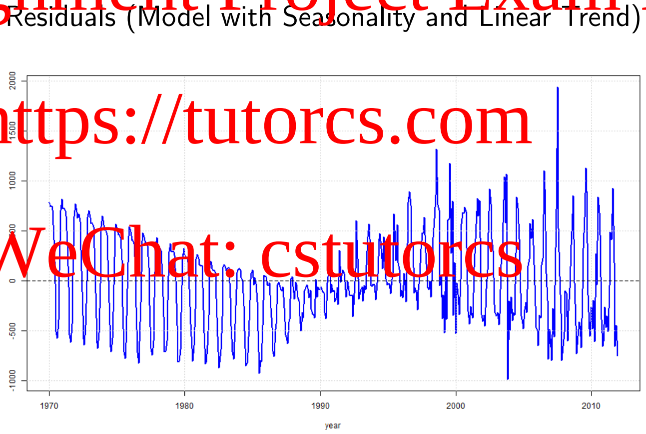
- The log-transformed model has a much better out-of-sample performance (measured in terms of MSPE)

Transformation that stabilizes volatility across observations tends to improve forecast performance.

- Gauss-Markov Theorem says OLS estimator is efficient when the observations are homoscedastic, not when observations are heteroscedastic.
- Efficiency means parameter uncertainty.
- Forecast uncertainty is a sum of the parameter uncertainty and fundamental uncertainty.
- If we can find a way to reduce parameter uncertainty, we will reduce the forecast uncertainty as well.

11. Additional discussions

Recall the best model on Electricity Consumption — seasonality with linear trend. The residuals still display certain kind of periodicity and serial correlation.



The presence of periodicity and serial correlation in the residuals suggest that there is much room to improve the model.

- It is possible the annual periodic pattern is partly driven by seasonal pattern of lunar calendar, in additional to the western calendar. Thus, adding an additional set of seasonal dummies to capture the lunar calendar may improve the fit, and hence the forecast.
- It is possible the serial correlation can be captured by the ARMA model of the cyclical components, to be discussed later.