# Q5: Department Store Sales Time-Series Analysis

*Fall, 2022*

G Park

2022-12-04

## Question 5: Department Store Sales Time-Series (35 points)

The file `DepartmentStorSales.csv` contains data on the quarterly sales for a department store over a 6-year period from 2000 to 2005

| Quarter | Sales |
|---|---|
| 1 | 50147 |
| 2 | 49325 |
| 3 | 57048 |
| 4 | 76781 |
| 5 | 48617 |
| 6 | 50898 |

## We discuss four major components of time series (level, trend, seasonality, and noise). Discuss the meaning of these four major components. (3 points)

**Level: In time series analysis, level refers to the average value in the series. Usually, firms may work with the level value and then adjust their stock levels for seasonal and trend components.**

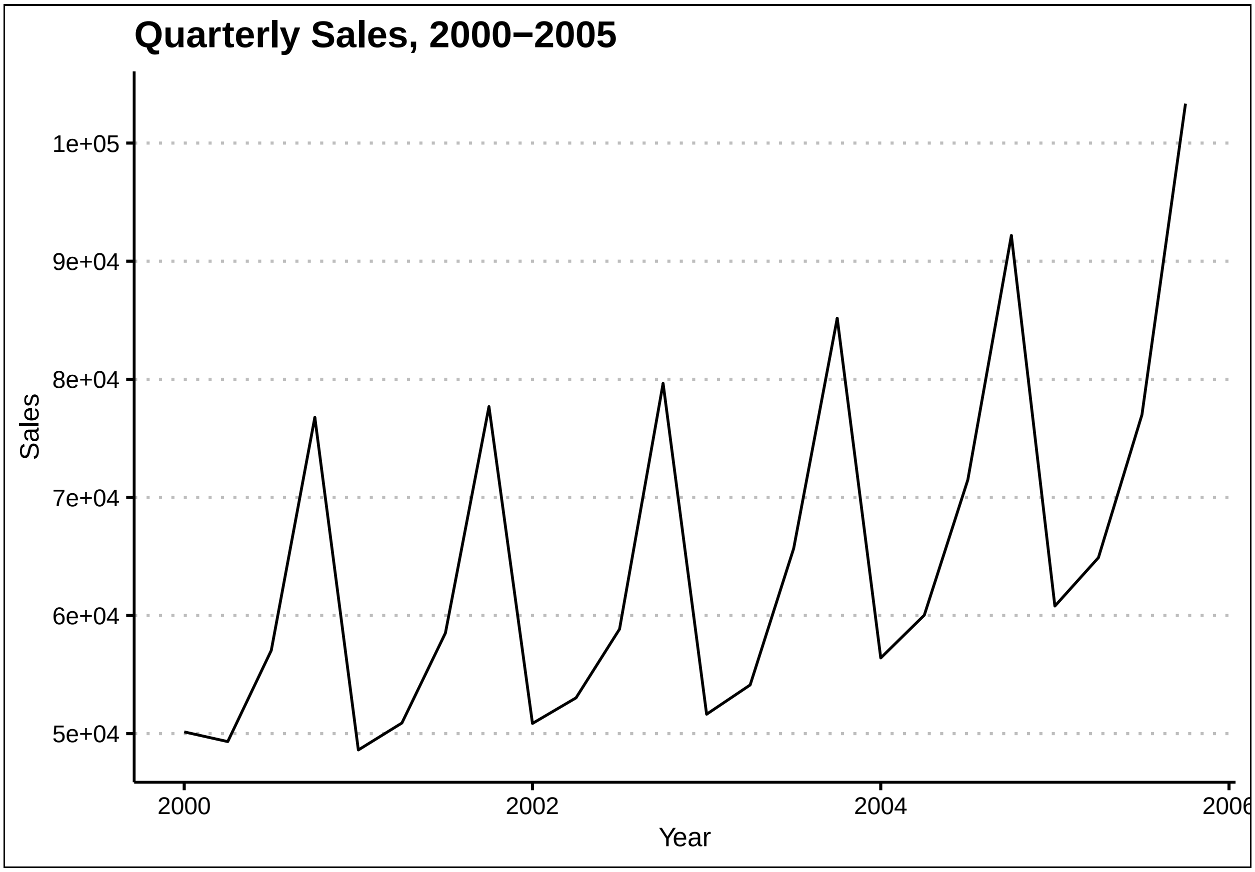
**Trend: Trend in time series analysis is the sustained increase or decrease observed in a series that spans accross any seasonal patterns. In the case of the departmental store, there is an overall upward trajectory in sales. This means that, on average, overall sales are increasing even before factoring in seasonality. This trajectory is also quite predictable. The store management can then plan to stock higher levels of stock in the current quarter compared to the corresponding stock levels in the previous quarter.**

**Seasonality: Seasonality refers to the short-term recurring cycles in the series. In other words, these are the periodic ups and downs observed in the data. In the case of our departmental store sales, we can observe a periodic rise and fall in demand that occurs every () quarters. The seasonal component is very predictable. In this case, the department should increase the stock during the peak periods and hold relatively lower stock during the off-peak periods.**

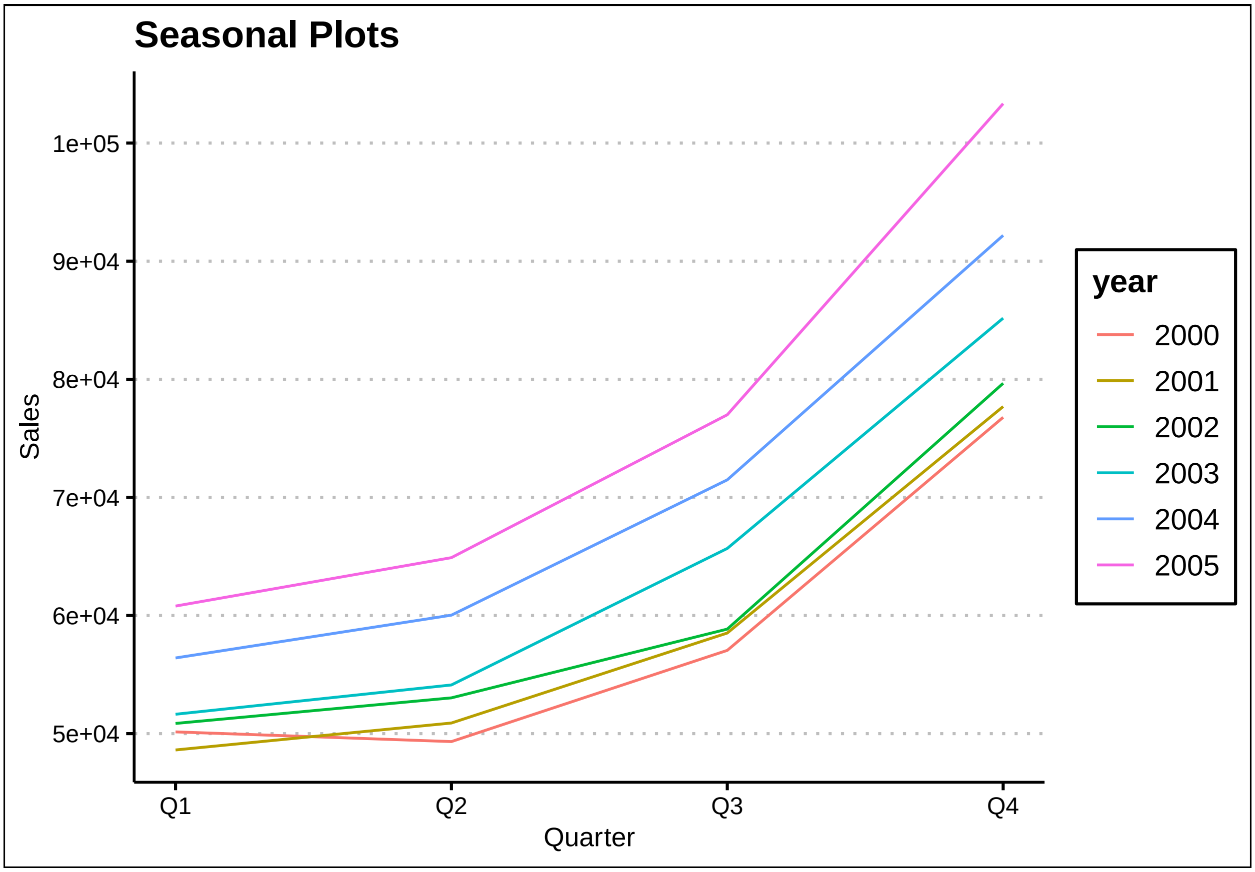
**Noise: Noise refers to the random variation in the series. This random variation is unpredictable. Noise arises from the unknown factors that affect demand or unmeasurable factors, the known-knowns and the unknown-unknowns. This is the reason many businesses hold working capital in the form of extra cash or inventory to take care of these unpredictable ups and downs in sales.**

## Show the data in a time-series format and create a well-performed time plot of the data. (2 points)

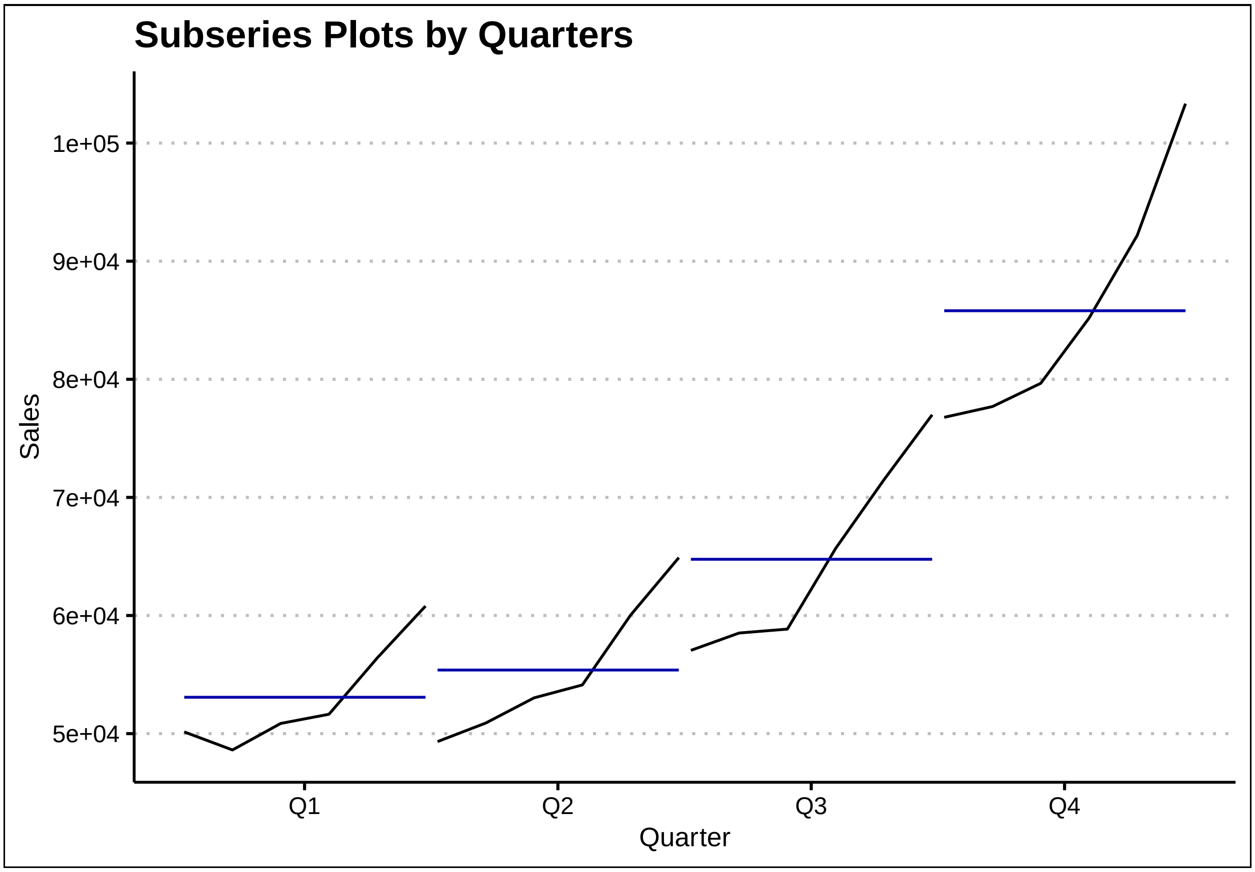
```
##         Quarter  Sales
## 2000 Q1       1  50147
## 2000 Q2       2  49325
## 2000 Q3       3  57048
## 2000 Q4       4  76781
## 2001 Q1       5  48617
## 2001 Q2       6  50898
## 2001 Q3       7  58517
## 2001 Q4       8  77691
## 2002 Q1       9  50862
## 2002 Q2      10  53028
## 2002 Q3      11  58849
## 2002 Q4      12  79660
## 2003 Q1      13  51640
## 2003 Q2      14  54119
## 2003 Q3      15  65681
## 2003 Q4      16  85175
## 2004 Q1      17  56405
## 2004 Q2      18  60031
## 2004 Q3      19  71486
## 2004 Q4      20  92183
## 2005 Q1      21  60800
## 2005 Q2      22  64900
## 2005 Q3      23  76997
## 2005 Q4      24 103337
```



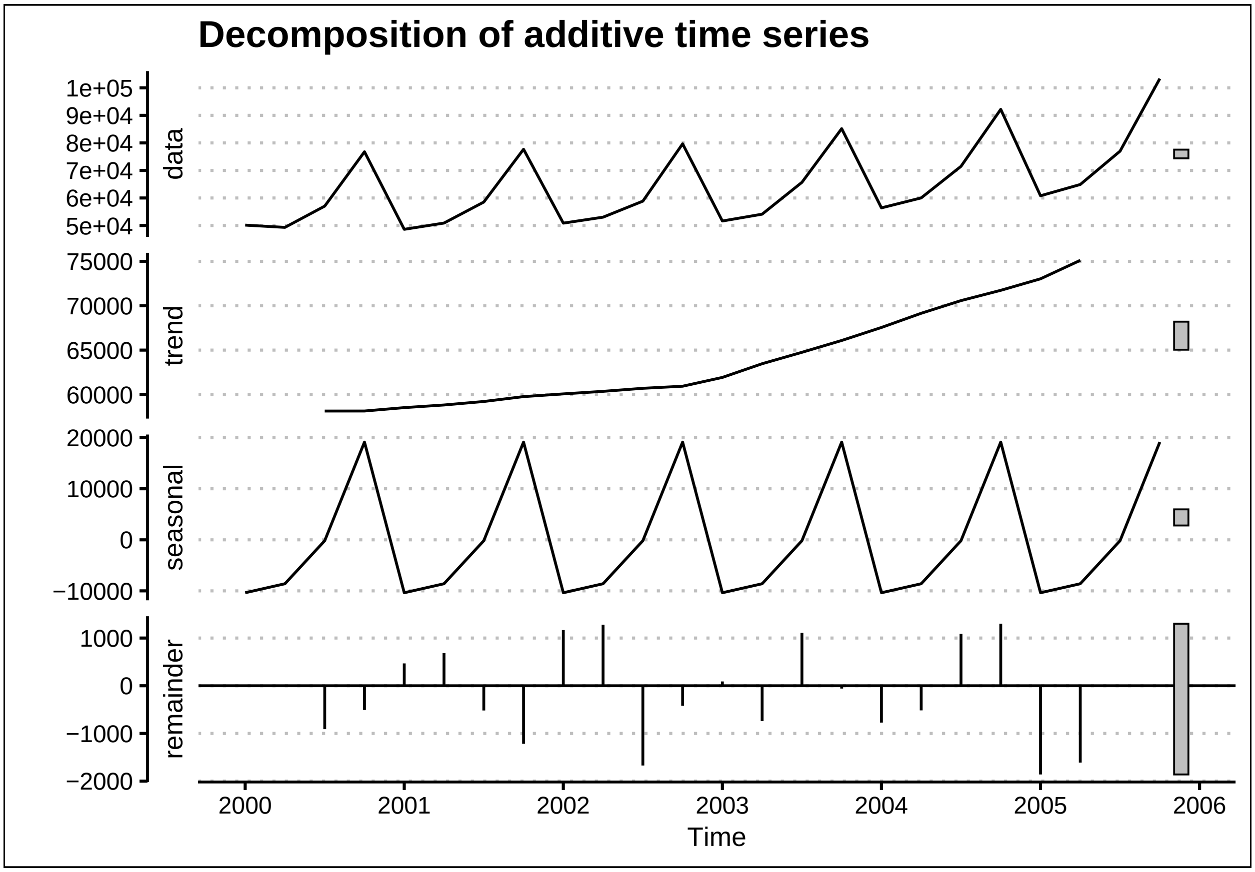

I also create a seasons plot that shows the sales for each year by season. The plot shows that in spite of the seasonality, sales in each subsequent year were higher than the last.




Likewise, the sub-series plot shows the trends in sales for each quarter. For instance, the first panel is a plot of sales in $2000Q1, 2001Q1 \cdots 2005Q1$. Again the plot shows a general upward trend.


Subseries Plots by Quarters
1e+05
9e+04
8e+04
7e+04
6e+04
5e+04
Sales
Q1
Q2
Q3
Q4
Quarter

## Decompose your time-series data into four major components (level, trend, seasonality, and noise) and plot these four components individually. Ensure to and explain your observations about different trends from your plots. (6 points)




The decomposition above consists of 4 plots. The first plot is the observed raw data. The second plot illustrates the general upward trend of sales over the period. The trend shows that quarter after quarter, the general trend is up. The management of the store could plan future inventory on this trend among other factors.

The third plot shows the seasonal component, the short swings in sales. when stocking, the management could consider these dynamics by stocking more when the demand is high compared the the off-peak periods.

The final plot shows the random component of the sales. Here, the data appears to follow a random walk with no apparent pattern. In planning for sales, it is hard to cater for this component except for keeping some buffer stock just in case of an upward but unexpected uptick in demand.

## We discuss different trend models in time series, including the linear trend regression model, exponential trend model, and polynomial trend model. Discuss and compare these three models. (4 points)

***Linear trend regression model***

The forecasting equation for the linear trend model is:

$y = a + \beta t$

Here, t is the time index. The $\alpha$ and $\beta$ (the "intercept" and "slope") are come from a simple regression in with Y as the dependent variable and the time t as the independent variable.

Although useful, these models perform poorly when the trend is not linear as is the case with our data.In this case a polynomial trend model does better. Even where the trend is linear, the linear model may fail to pick seasonal or cyclical fluctuations.

#### ***Exponential trend model***

Exponential smoothing methods use the weighted averages of past observations to make forecasts of the future. These weights reduce as the observations get older with more recent observations having a higher weight (Hyndman and Athanasopoulos 2018).

#### ***Polynomial trend model***

Polynomial trend models model the trend in data as a smooth variation in time. This model permits us to fit a either a line or a curve to the data as opposed to the linear model that assumes a purely linear relationship over time. The polynomial trend models provide a simple, flexible forms to describe the local trend components (West and Harrison 1989). These class of models are useful for modelling data that has an increasing or decreasing trend. More specifically, The polynomial trend models can be used in those situations where the relationship between study and explanatory variables is curvilinear.

Given a polynomial regression model in one variable as follows;

$y = \beta_0 + \beta_1 X + \beta_2 X^2 + e_2$

This is a second-order model or quadratic model. $\beta_1$ and $\beta_2$ are called the linear effect parameter and quadratic effect parameter, respectively.This class of models can extend to include several variables and their interactions and even higher orders like cubes.

## Partition the first 18 records as training sets and the rest of 6 records as validation sets. Fit a linear trend model on training sets. Based on this linear trend model, make a forecast for validation sets and show the results. (7 points)

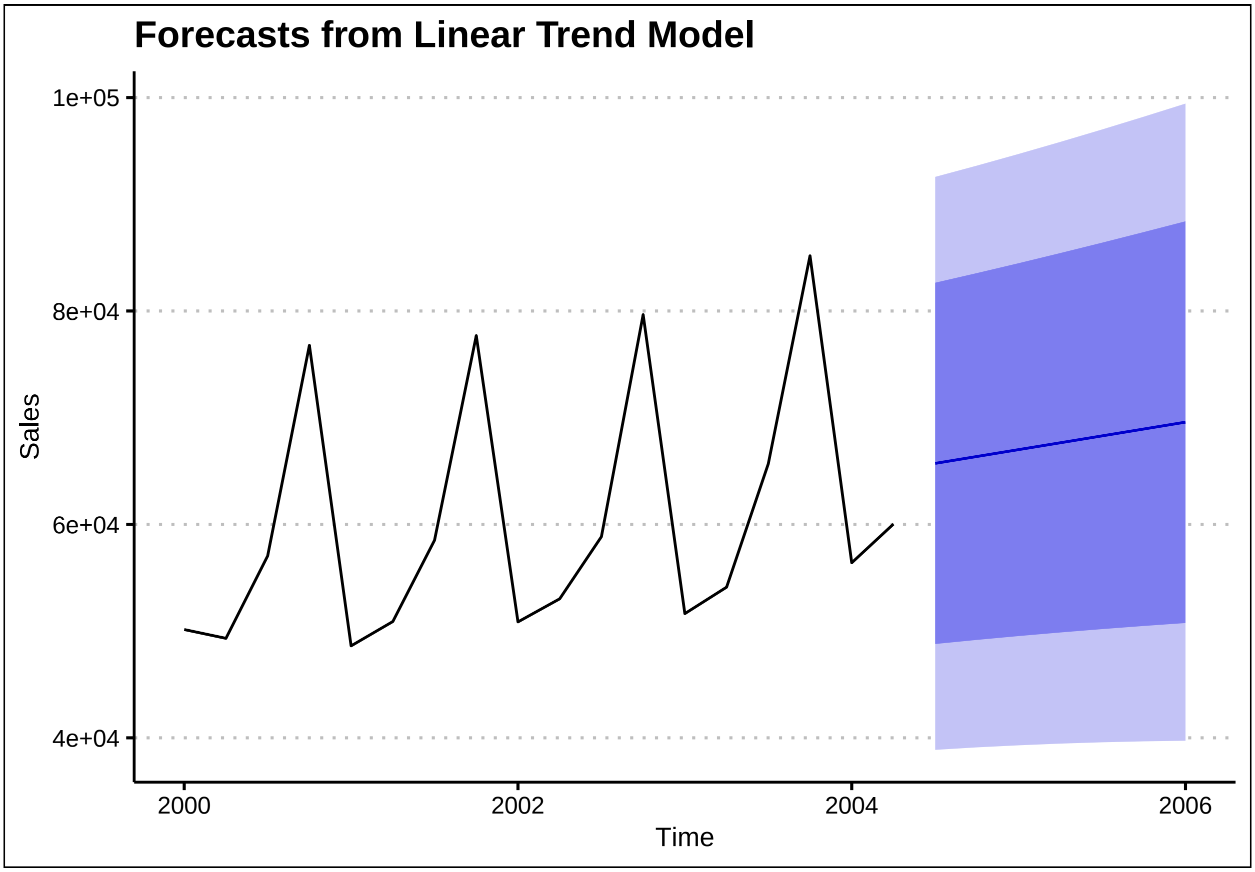
I then fit a linear trend model on the training data.

```
##
## Call:
## tslm(formula = Sales ~ Quarter, data = ts_training)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10862  -8390  -5160   1665  20742
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    54132       5674    9.54  5.3e-08 ***
## Quarter          644        524    1.23     0.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11500 on 16 degrees of freedom
## Multiple R-squared:  0.0861, Adjusted R-squared:  0.029
## F-statistic: 1.51 on 1 and 16 DF,  p-value: 0.237
```

```
##         Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
## 2004 Q3          65721 48791 82651 38872 92570
## 2004 Q4          66365 49175 83554 39105 93624
## 2005 Q1          67008 49536 84481 39299 94718
## 2005 Q2          67652 49873 85431 39457 95847
```

```
## 2005 Q3          68296 50189 86404 39580 97012
## 2005 Q4          68940 50483 87397 39670 98210
## 2006 Q1          69584 50758 88409 39729 99438
```

We then plot the linear model forecasts.

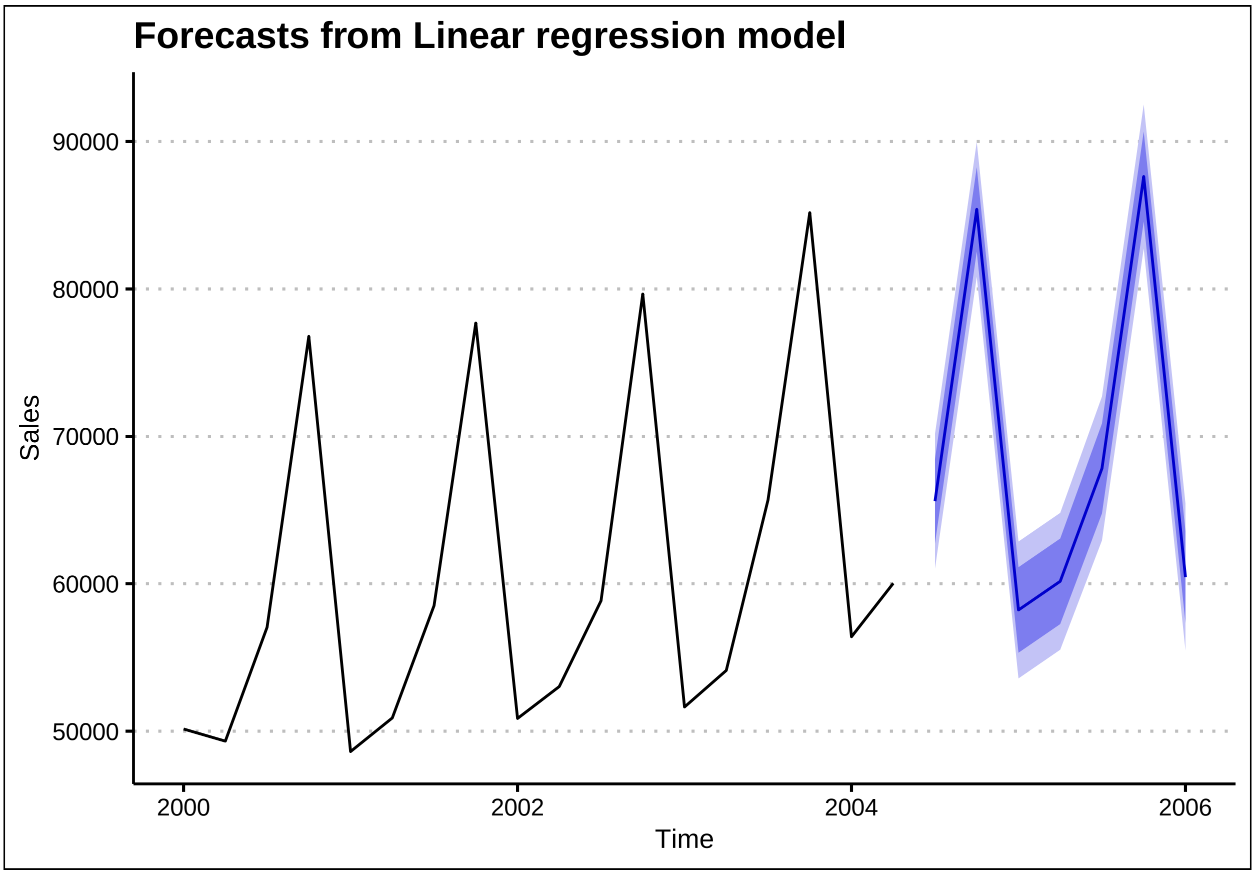



**Fit a polynomial trend model with seasonal effects on training sets. Based on this model, make a forecast for validation sets and show the results. Check accuracy measures of the forecasted model and show these accuracy measures. (8 points)**

```
##
## Call:
## forecast::tslm(formula = Sales ~ trend + season, data = ts_training)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2289   -938   -373    402   3070
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46519.5     1065.7   43.65  1.7e-15 ***
## trend          557.2       80.1    6.96  1.0e-05 ***
## season2       1388.8     1112.8    1.25     0.23
## season3       8489.5     1177.3    7.21  6.8e-06 ***
## season4      27735.4     1180.0   23.50  4.9e-12 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1750 on 13 degrees of freedom
## Multiple R-squared:  0.983,  Adjusted R-squared:  0.978
## F-statistic:  186 on 4 and 13 DF,  p-value: 2.49e-11
```

```
##         Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
## 2004 Q3          65596 62734 68457 61017 70174
## 2004 Q4          85399 82537 88260 80820 89977
## 2005 Q1          58221 55318 61123 53577 62864
## 2005 Q2          60166 57264 63069 55523 64810
## 2005 Q3          67824 64773 70876 62942 72707
## 2005 Q4          87627 84576 90679 82745 92510
## 2006 Q1          60449 57330 63569 55458 65441
```



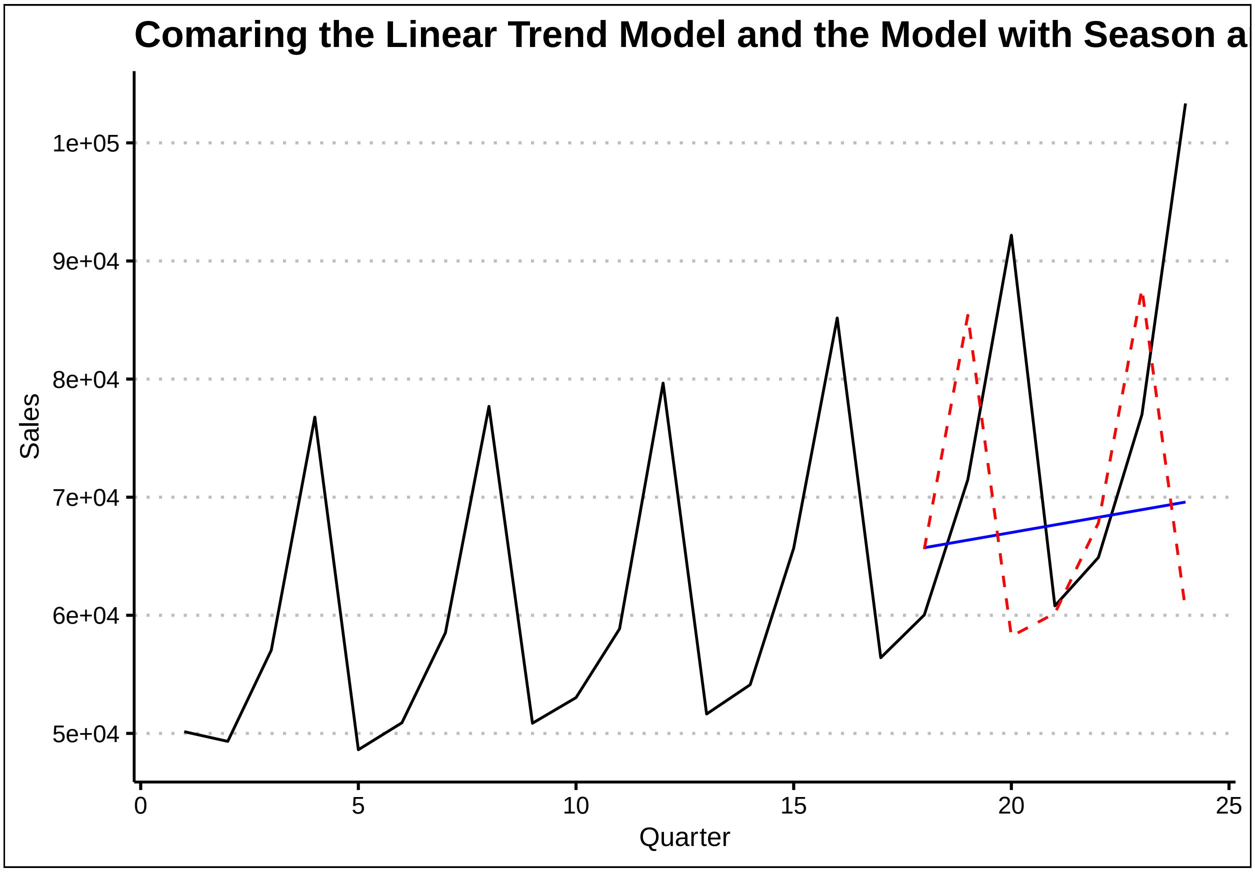

In this section I compute the error for both models. Note that the model with trend and seasons tends to do better than the linear trend model.

```
## [1] 15
```

```
## [1] 18
```

### Finally, based on your decomposition of four major components from 5.3, compare your 5.5 and 5.6 models and discuss which model is more appropriate in this particular setting and why. (5 points)

The plot below compares the comparison of both models. Note that the model with seasonality and trend (red broken line) captures the data much better than the linear model (blue line). I would choose the model with seasonality and trend given the nature of the data; It has both a trend component and a seasonal component. The linear trend model fares especially poorly in capturing seasons and cycles compared to the model with seasonality and trend.




# References


Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. Melbourne: OTexts.

West, Mike, and Jeff Harrison. 1989. "Polynomial Trend Models." In *Bayesian Forecasting and Dynamic Models*, 201–28. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-9365-9_7.