THE UNIVERSITY OF MELBOURNE

FORECASTING IN ECONOMICS & BUSINESS ECOM90024

LECTURE 2: INTRODUCTION & REVIEW

LECTURE OUTLINE

- Brief review of multiple regression
- Time series as unobserved components
- Deterministic trends
- Forecasting and regression analysis
- Model selection
- Estimating and forecasting trends in ***R***

LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- The simple linear regression framework can be easily extended to the case of multiple regressors.

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

$$E\left[\varepsilon_i \middle| X_{1i}, X_{2i}, \ldots, X_{ki}\right] = 0$$

- This model permits estimating the effect on Y of changing one variable while holding the other regressors constant.
- The intercept α is the expected value of Y when all the $X's$ are equal to zero.

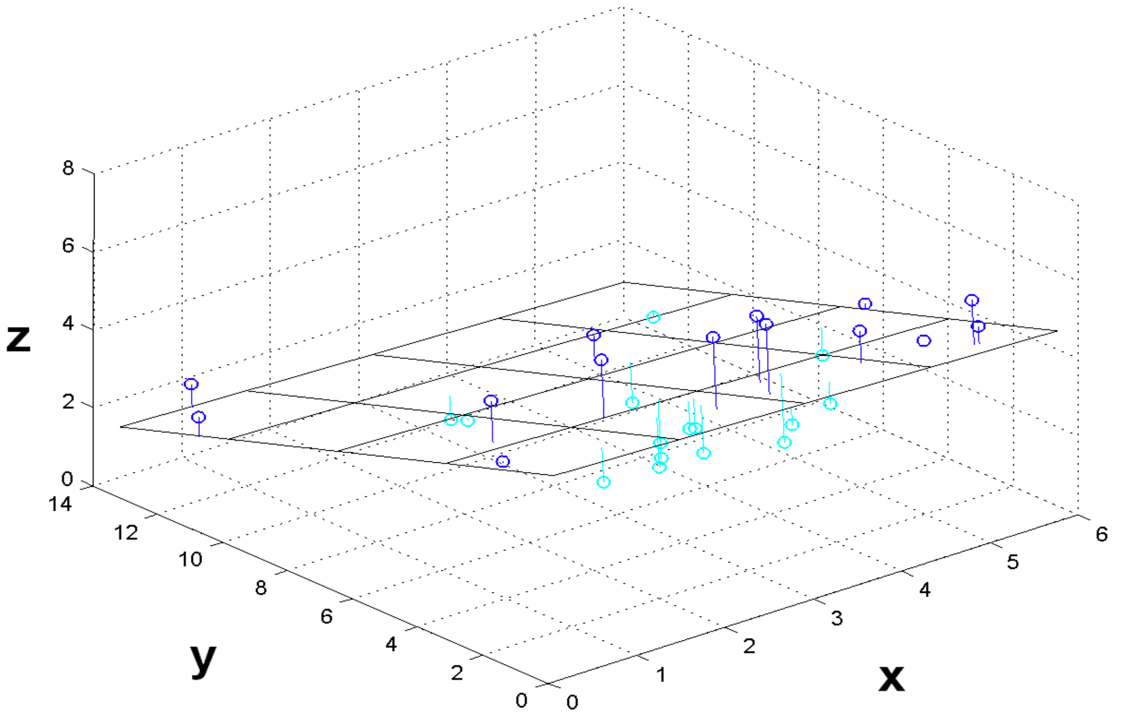
LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- As is the case with simple linear regression, the method of OLS can be used to estimate the coefficients $\alpha, \beta_1, \beta_2, \dots, \beta_k$.
- Suppose that we had n observations of each variable, then the estimated regression function that best fits the data must minimize the sum of squared residuals. That is, we choose $a, b_1, b_2, \dots, b_k$ to minimize,

$$\sum_{i=1}^{n} (y_i - a - b_1 x_{1i} - b_2 x_{2i} - \cdots - b_k x_{ki})^2$$

LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- In the case of a linear regression with two regressors, the estimated regression function is a surface,



LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- Given a multiple regression model,

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

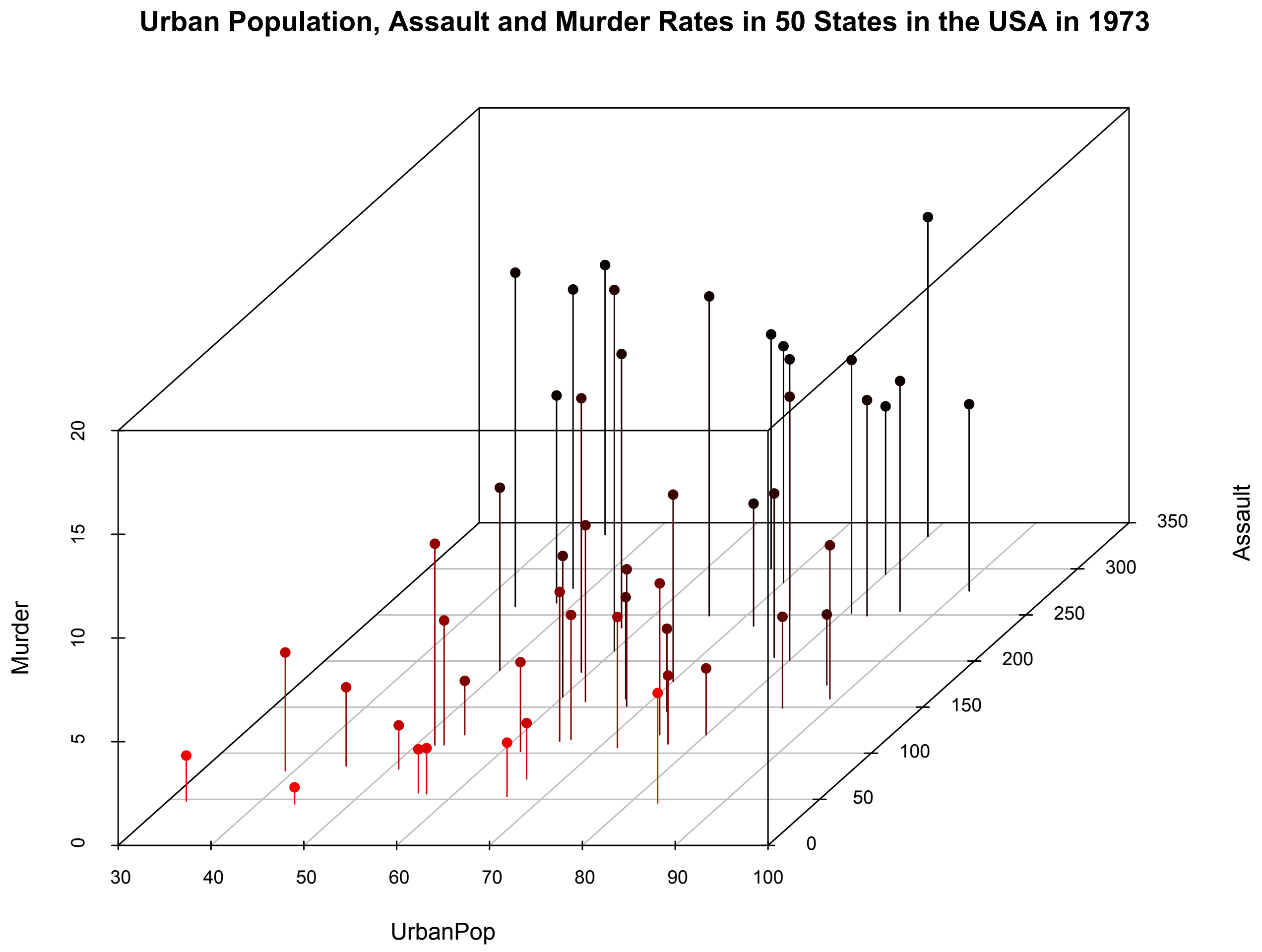
- If the following conditions hold:

  1. $E\left[\varepsilon_i \middle| X_{1i}, X_{2i}, \dots, X_{ki}\right] = 0$
  2. $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i), i = 1, \dots, n$, *are independently and identically distributed draws from their joint distribution.*
  3. $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i)$ *have nonzero finite fourth moments.*
  4. $There\ is\ no\ perfect\ multicollinearity$

- Then in large samples, the OLS estimators, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$,…, $\hat{\beta}_k$ are jointly normally distributed and each $\hat{\beta}_j$ is distributed $N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$ for, $j = 0, \dots, k$

LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- To illustrate, let's consider a dataset that comprises of statistics in arrests per 100,000 residents for assault and murder in each of the 50 US states in 1973 as well as the percentage of the population living in urban areas.



LINEAR REGRESSION WITH MULTIPLE REGRESSORS

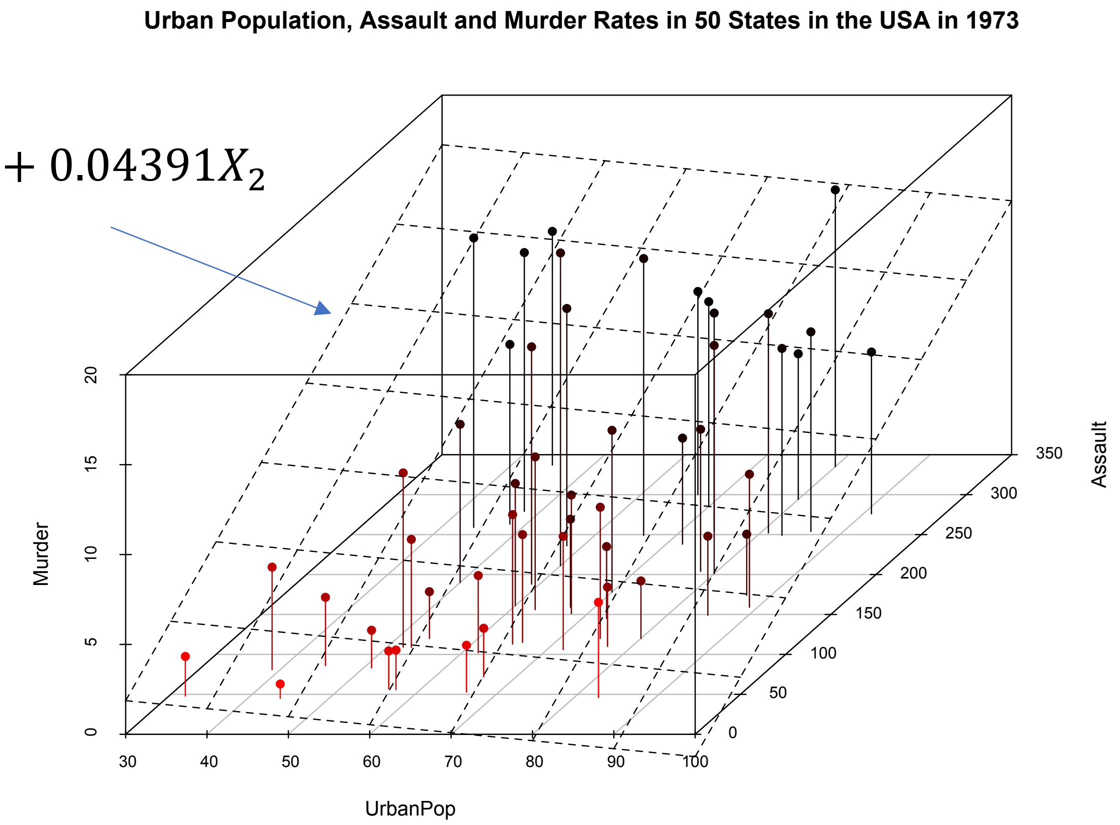
- Let's now compute a linear regression in which we set the murder rate as our dependent variable and assault rate and the proportion of the population living in urban areas as our explanatory variables.

```
> linreg3 <- lm(formula = Murder ~ Assault + UrbanPop)
> summary(linreg3)

Call:
lm(formula = Murder ~ Assault + UrbanPop)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5530 -1.7093 -0.3677  1.2284  7.5985 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.207153   1.740790   1.842   0.0717 .  
Assault      0.043910   0.004579   9.590 1.22e-12 ***
UrbanPop    -0.044510   0.026363  -1.688   0.0980 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.58 on 47 degrees of freedom
Multiple R-squared:  0.6634,	Adjusted R-squared:  0.6491 
F-statistic: 46.32 on 2 and 47 DF,  p-value: 7.704e-12
```

LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- The fitted surface can be visualized as,

$$Y = 3.207153 - 0.04451X_1 + 0.04391X_2$$



LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- Reading and interpreting regression output from a multiple regression is essentially the same as the simple linear regression case. There are just a couple of extra statistics to pay attention to.
- In a multiple regression, the R^2 increases whenever a regressor is added, unless the estimated coefficient on the added regressor is exactly zero.
- The adjusted R^2, or $\bar{R}^2$ is a modified version of R^2 that does not necessarily increase when a new regressor is added.

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSE}{SST}$$

LINEAR REGRESSION WITH MULTIPLE REGRESSORS

- The F statistic is used to test joint hypotheses about regression coefficients.
- The overall regression F statistic that is reported in the output tests the joint hypothesis that all slope coefficients are zero against the alternative that at least one of the coefficients is not zero.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1: \ \beta_j \neq 0, \text{ at least one } j, \ j = 1, \ldots, k$$

- It is computed as,

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

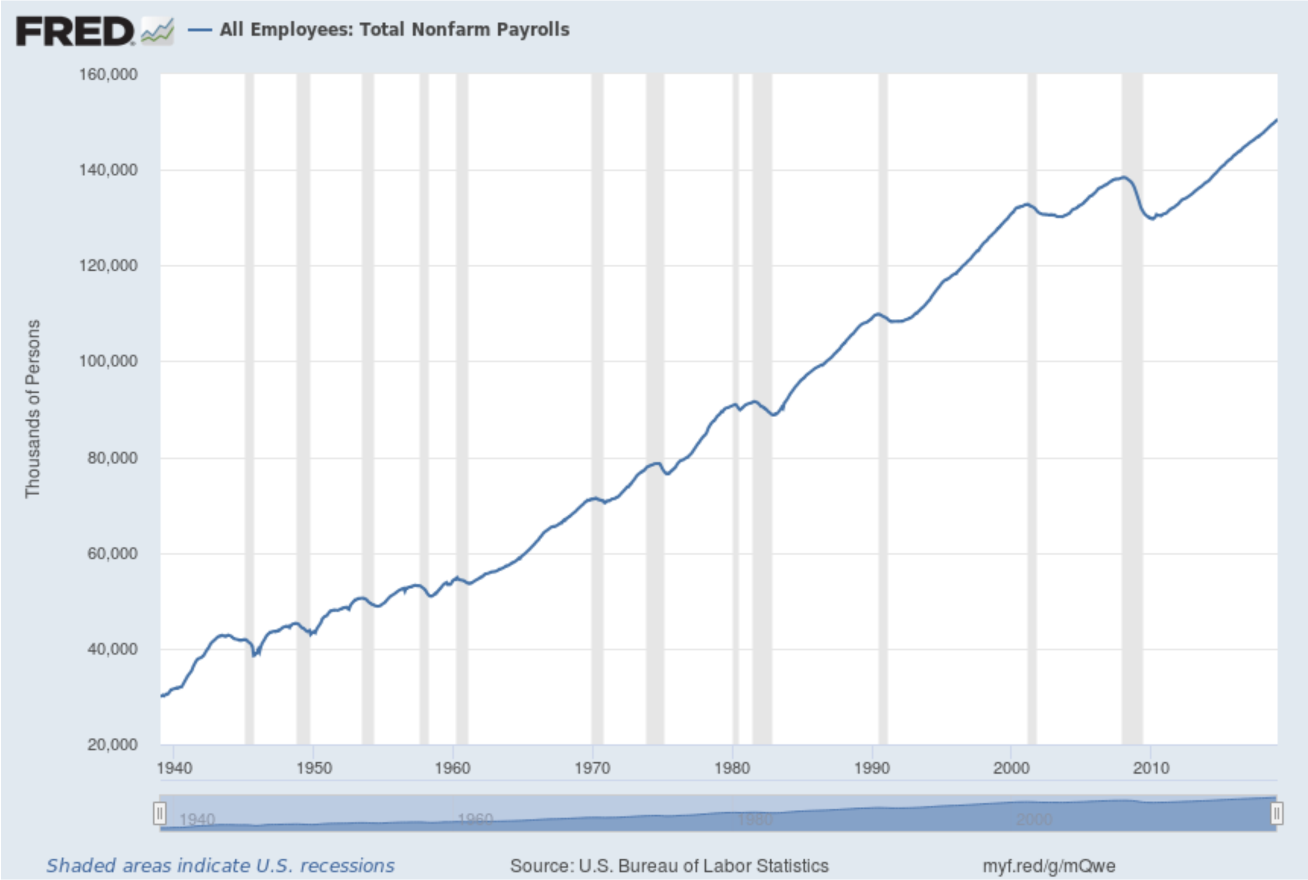
FURTHER READING AND REFERENCES

- Two good references for multiple regression are:
  - Stock, J. and Watson, M. ***Introduction to Econometrics***, Pearson.
  - Wooldridge, Jeffrey M. ***Introductory Econometrics: A Modern Approach***, Cengage Learning.

TIME SERIES AS UNOBSERVED COMPONENTS

- This time series is a measure of the number of U.S. workers in the economy that excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed. This measure accounts for approximately 80 percent of the workers who contribute to Gross Domestic Product (GDP).

- How would you describe this series? What are its most obvious features?



TIME SERIES AS UNOBSERVED COMPONENTS

- Often the time series that we are interested in forecasting are generated by very complex underlying processes.

- In other courses, you will spend time developing theoretical models that will specify the deep structural relationships between variables that govern their evolution through time.

- In this course we will focus instead on specifying statistical models of variables based on their ***time series characteristics***.

- We will conceive of time series as comprising of a set of unobserved underlying components.

TIME SERIES AS UNOBSERVED COMPONENTS

- In general, we will write all time series as comprising of fluctuations that occur over distinct ***time scales.*** For instance, we could write:

$$Y_t = T_t + S_t + C_t + \varepsilon_t$$

- Where

$$T_t = trend\ component$$

Longest time scale

$$C_t = cyclical\ component$$

$$S_t = seasonal\ component$$

Shortest time scale

- This is an additive model, we will consider the multiplicative model as an alternative in the next lecture!

DETERMINISTIC TRENDS

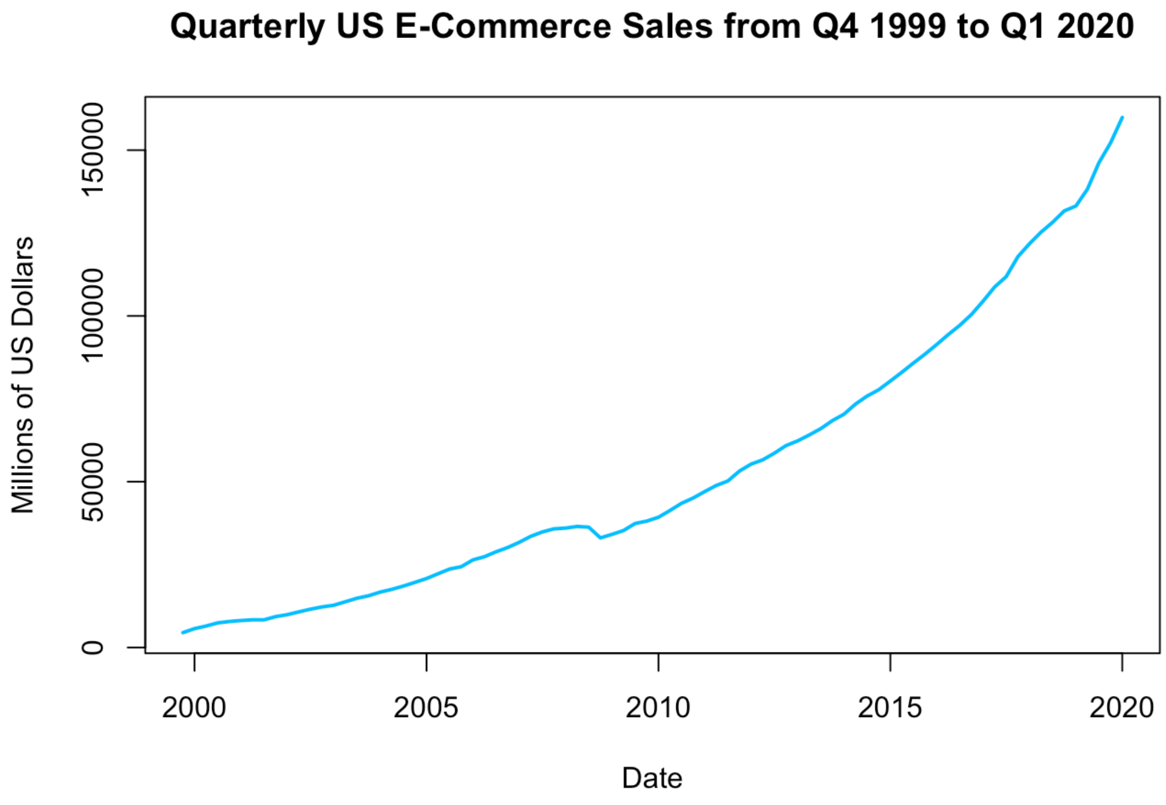
- The trend of a time series can be characterized as the slow, long-run evolution of the series.
- In business, finance and economics, trends are produced by the evolution of preferences, technologies, institutions and demographics. These changes typically take place over several years and decades.
- In this lecture we will be focusing our attention on trends that evolve in a perfectly predictable way, otherwise known as ***deterministic trends***.
- We will discuss stochastic trends in a future lecture.

A LINEAR DETERMINISTIC TREND

- Let's consider quarterly e-commerce sales in United States from the last quarter of 1999 to the first quarter of 2020.

- From visual inspection, we can see that the data clearly appears to be trending upwards over time.

- As a starting point, let's consider a deterministic linear trend model.



A LINEAR DETERMINISTIC TREND

- We would specify the linear deterministic trend model in the following way,

$$y_t = T_t + \varepsilon_t$$

$$T_t = \beta_0 + \beta_1 TIME_t$$

- The variable $TIME$ is known as a time dummy (or time index) and takes value 1 in the first period of the sample, 2 in the second period and so on.
- For a sample of size T, $TIME = (1,2,3, \dots, T-1, T)$.
- The parameters β_0 and β_1 are the intercept and slope of the time trend respectively.

A LINEAR DETERMINISTIC TREND

- We can estimate the time trend using OLS. That is, given the model,

$$y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t$$

- We solve,

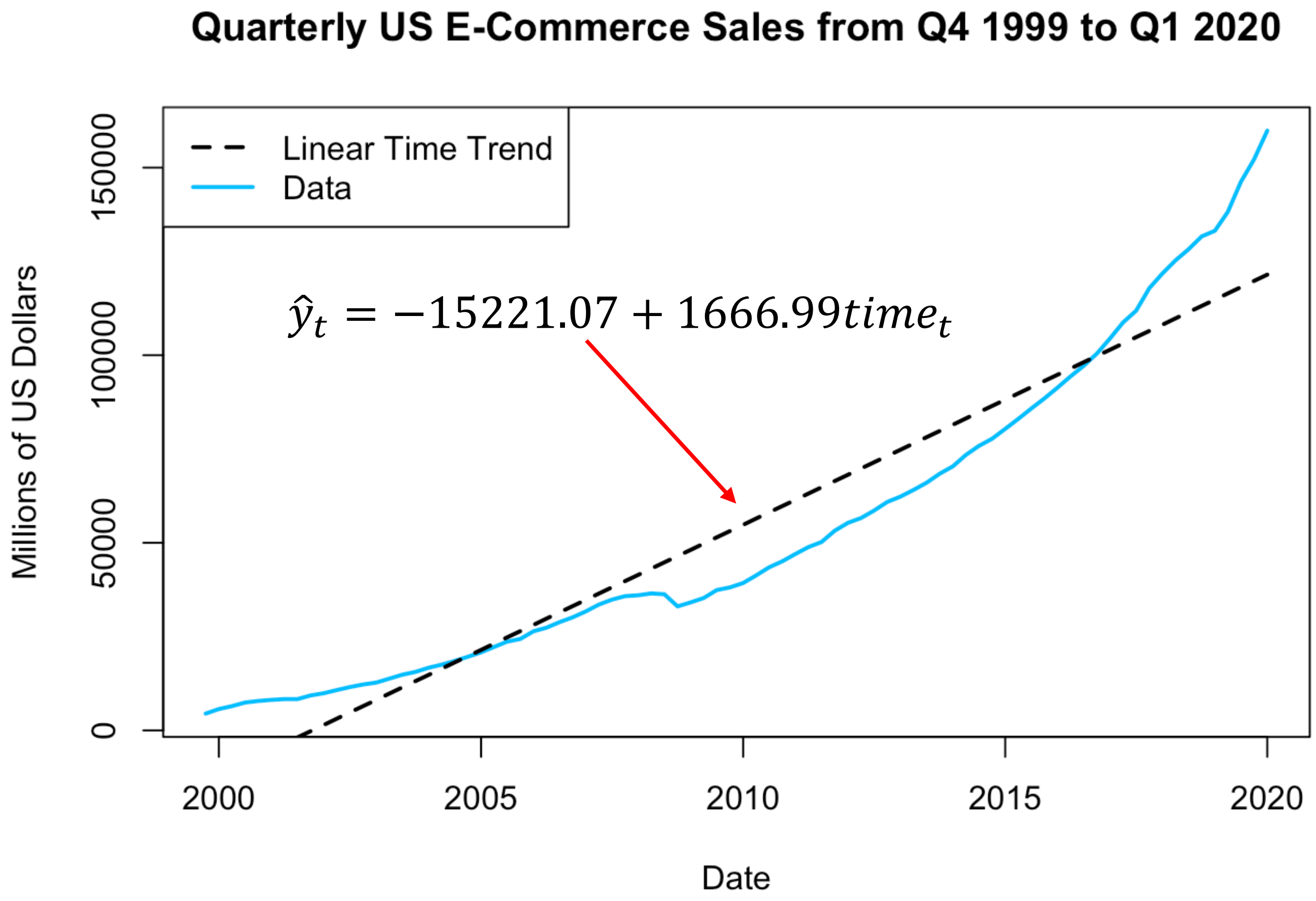
$$min_{\beta_0,\beta_1} \sum_{i=1}^{T} (y_t - \ \beta_0 - \beta_1 TIME_t)^2$$

- In practice, all we need to do is create a time trend variable in ***R*** and estimate a simple linear regression!

A LINEAR DETERMINISTIC TREND

- The estimated results are:

```
## Call:
## lm(formula = ecomdata$ecomsa ~ time)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -15509 -11315  -2088   7483  38381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15221.07    2762.05  -5.511 4.23e-07 ***
## time          1666.99      57.81  28.834  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12390 on 80 degrees of freedom
## Multiple R-squared:  0.9122, Adjusted R-squared:  0.9111
## F-statistic: 831.4 on 1 and 80 DF,  p-value: < 2.2e-16
```

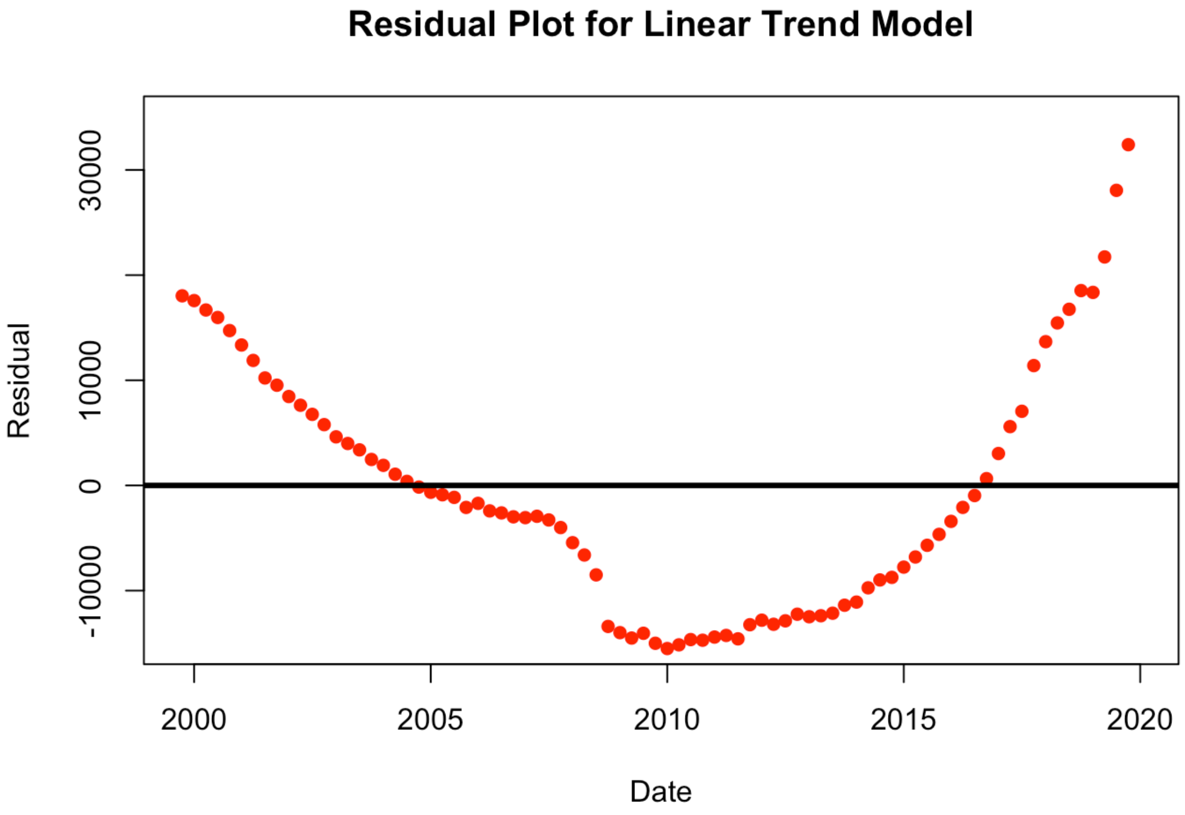


A LINEAR DETERMINISTIC TREND

- We can inspect the residual plot of the regression to get a sense of whether we have left anything out of the regression. Recall that the residuals are computed as:

$$e_t = y_t - \widehat{y_t}$$

- We can see a clear pattern in the residuals. What do you think this is due to?



A QUADRATIC DETERMINISTIC TREND

- In many instances (such as the case of our retail sales data!), the trend does not change at a constant rate.

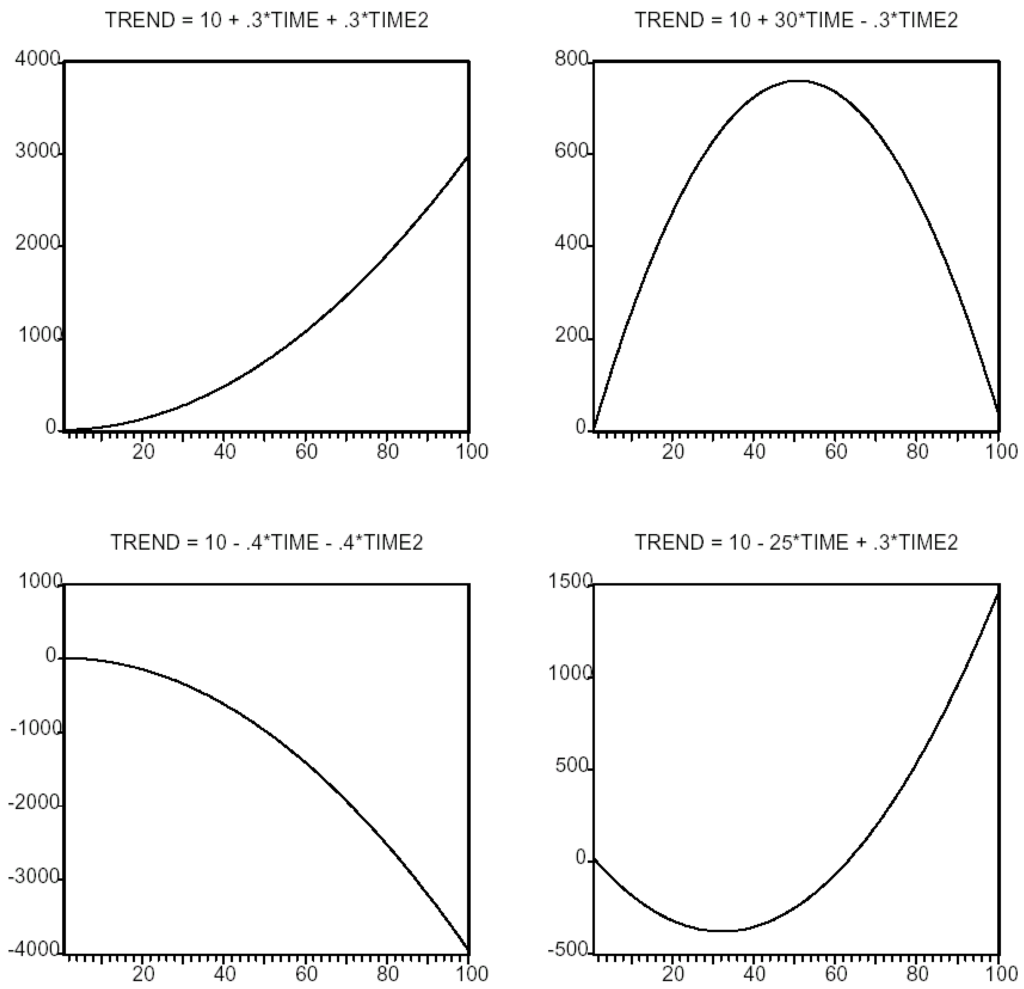
- We can incorporate a quadratic term into our time series specification in order to accommodate some curvature.

$$y_t = \beta_0 + \beta_1 TIME + \beta_2 TIME^2 + \varepsilon_t$$

- A variety of nonlinear trend shapes are possible, depending on the signs and sizes of the coefficients.

A QUADRATIC DETERMINISTIC TREND

- If $\boldsymbol{\beta_1 > 0}$ and $\boldsymbol{\beta_2 > 0}$, the trend is ***increasing***, but at an ***increasing rate***.
- If $\boldsymbol{\beta_1 < 0}$ and $\boldsymbol{\beta_2 < 0}$, the trend is ***decreasing***, but at an ***increasing rate***.
- If $\boldsymbol{\beta_1 < 0}$ and $\boldsymbol{\beta_2 > 0}$, the trend is ***U-shaped***.
- If $\boldsymbol{\beta_1 > 0}$ and $\boldsymbol{\beta_2 < 0}$, the trend has an ***inverted U-shape***.



A QUADRATIC DETERMINISTIC TREND

- To estimate a quadratic trend, we simply create a new variable $TIME^2$ and incorporate it into our linear regression. That is, given the model,

$$y_t = \beta\_0 + \beta_1 TIME_t + \beta_2 TIME^2 + \varepsilon_t$$

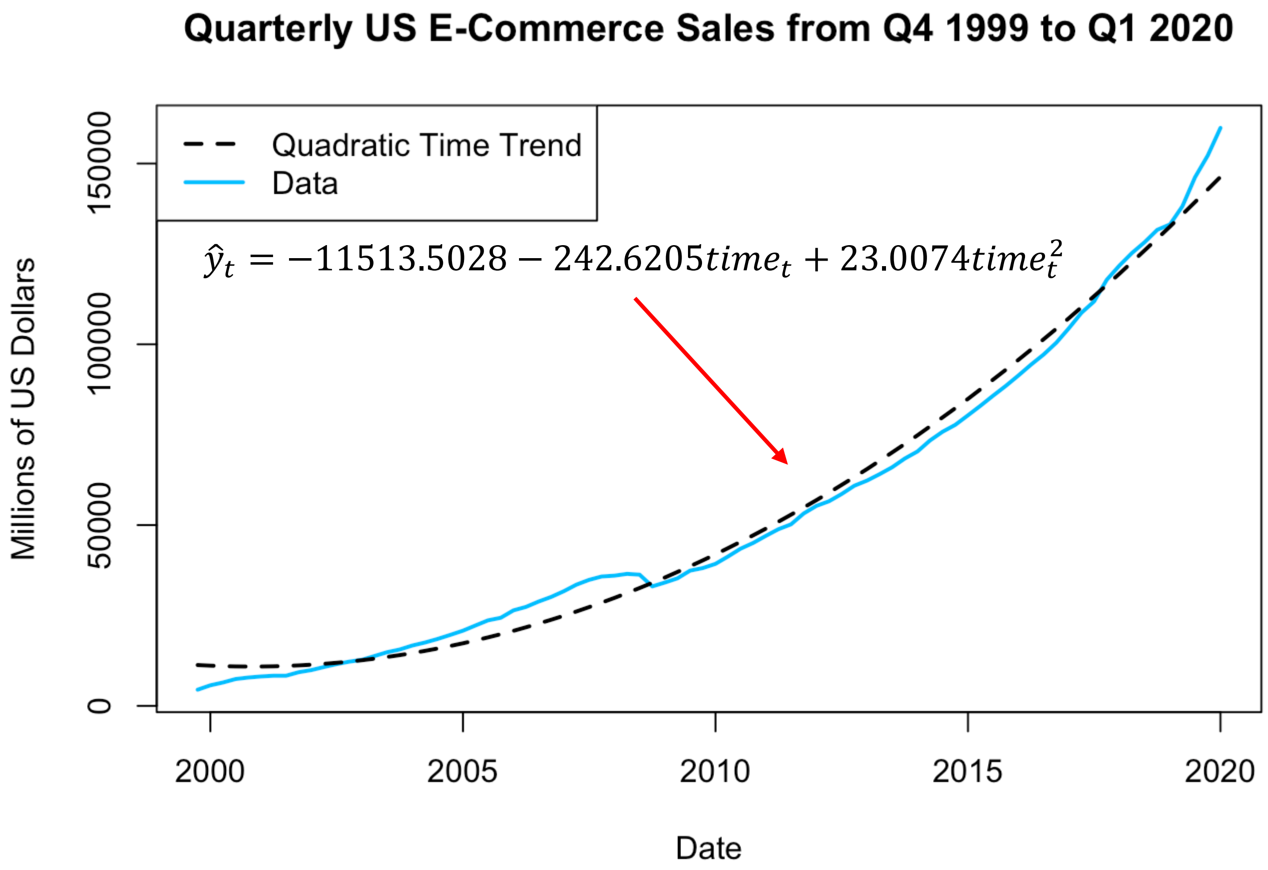
- We solve,

$$min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{T} (y_t - \ \beta_0 - \beta_1 TIME_t - \beta_2 TIME^2)^2$$

A QUADRATIC DETERMINISTIC TREND

- The estimated coefficients are:

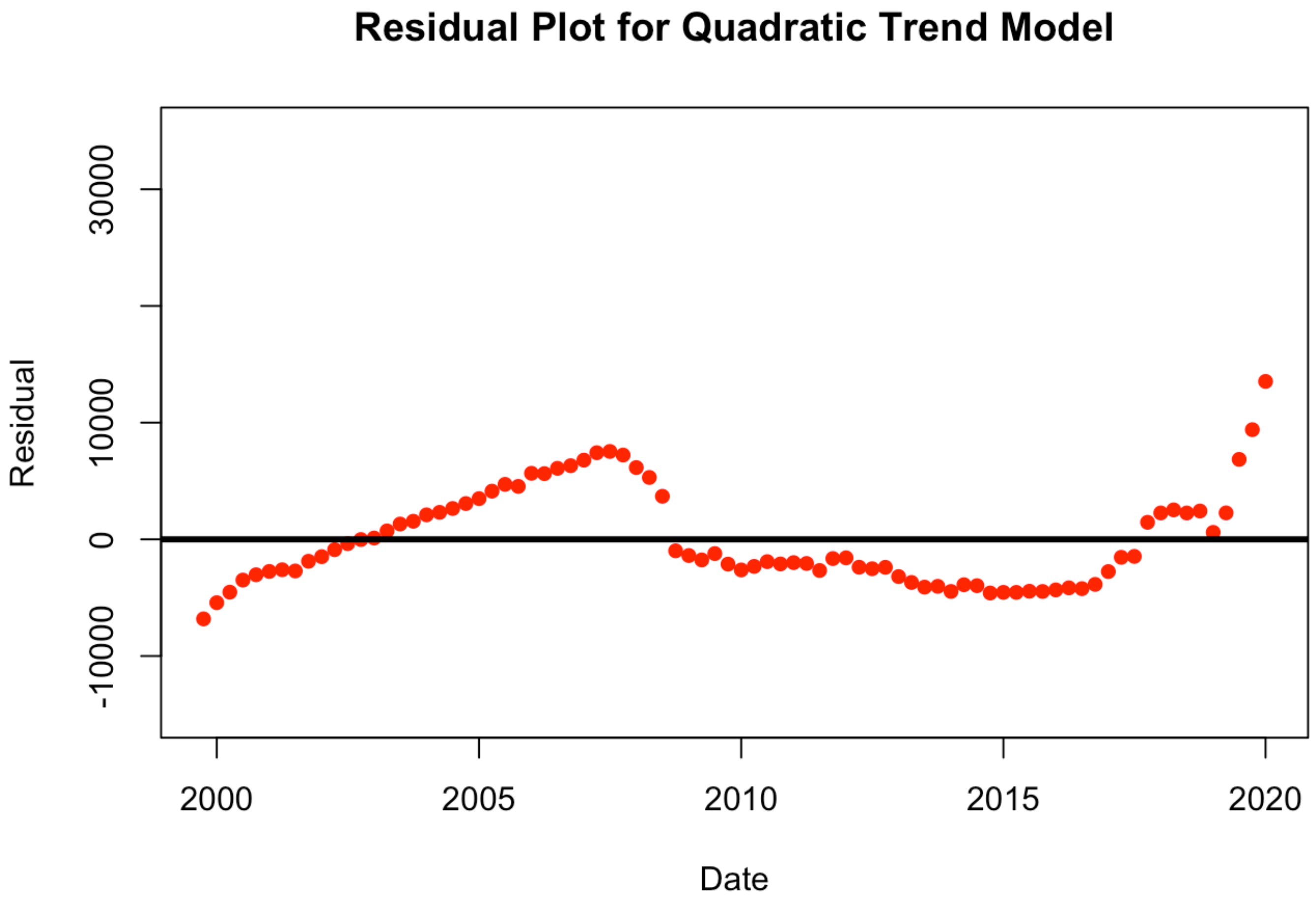
```
## Call:
## lm(formula = ecomdata$ecomsa ~ time + time2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -6818  -2968  -1512   2499  13533
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11513.5028  1423.9597   8.086 5.94e-12 ***
## time         -242.6205    79.1819  -3.064  0.00299 **
## time2          23.0074     0.9244  24.889  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4194 on 79 degrees of freedom
## Multiple R-squared:  0.9901, Adjusted R-squared:  0.9898
## F-statistic:  3939 on 2 and 79 DF,  p-value: < 2.2e-16
```



A QUADRATIC DETERMINISTIC TREND

- Looking at the residuals:

How do these residuals compare to the ones from the linear trend case?



AN EXPONENTIAL DETERMINISTIC TREND

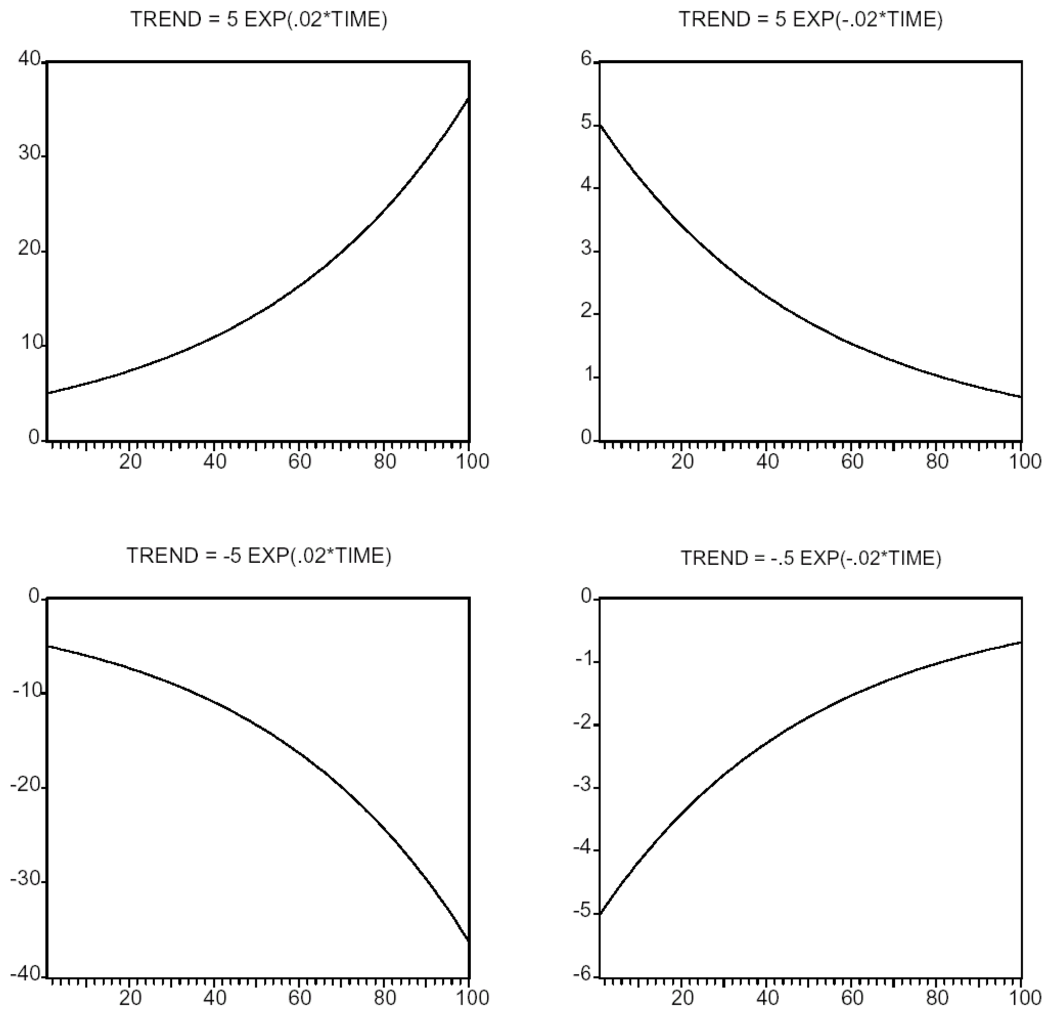
- Another popular nonlinear trend specification is the exponential trend which is written as,

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

- Exponential trends arise when a variable displays a constant percentage growth rate (e.g. three percent per year).
- As with a quadratic trend, depending on the signs and sizes of the parameter values, an exponential trend can achieve a variety of patterns.

AN EXPONENTIAL DETERMINISTIC TREND

- If $\boldsymbol{\beta_0 > 0}$ and $\boldsymbol{\beta_1 > 0}$, the trend is ***increasing***, but at an ***increasing rate***.
- If $\boldsymbol{\beta_0 < 0}$ and $\boldsymbol{\beta_1 < 0}$, the trend is ***increasing***, but at an ***decreasing rate***
- If $\boldsymbol{\beta_0 < 0}$ and $\boldsymbol{\beta_1 > 0}$, the trend is ***decreasing***, but at an ***increasing rate***.
- If $\boldsymbol{\beta_0 > 0}$ and $\boldsymbol{\beta_1 < 0}$, the trend is ***decreasing***, but at a ***decreasing rate***.



AN EXPONENTIAL DETERMINISTIC TREND

- When we specify an exponential trend model, our regression function is no longer linear in its parameters.

$$y_t = \beta_0 e^{\beta_1 TIME_t} + \varepsilon_t$$

- And we would have to solve a nonlinear least squares problem.

$$min_{\beta_0,\beta_1} \sum_{i=1}^{T} \left(y_t - \beta_0 e^{\beta_1 TIME_t}\right)^2$$

- Computers are very good at doing this, but can we get around this?

AN EXPONENTIAL DETERMINISTIC TREND

- We can exploit the property of logarithms and transform our nonlinear model into a linear one! That is,

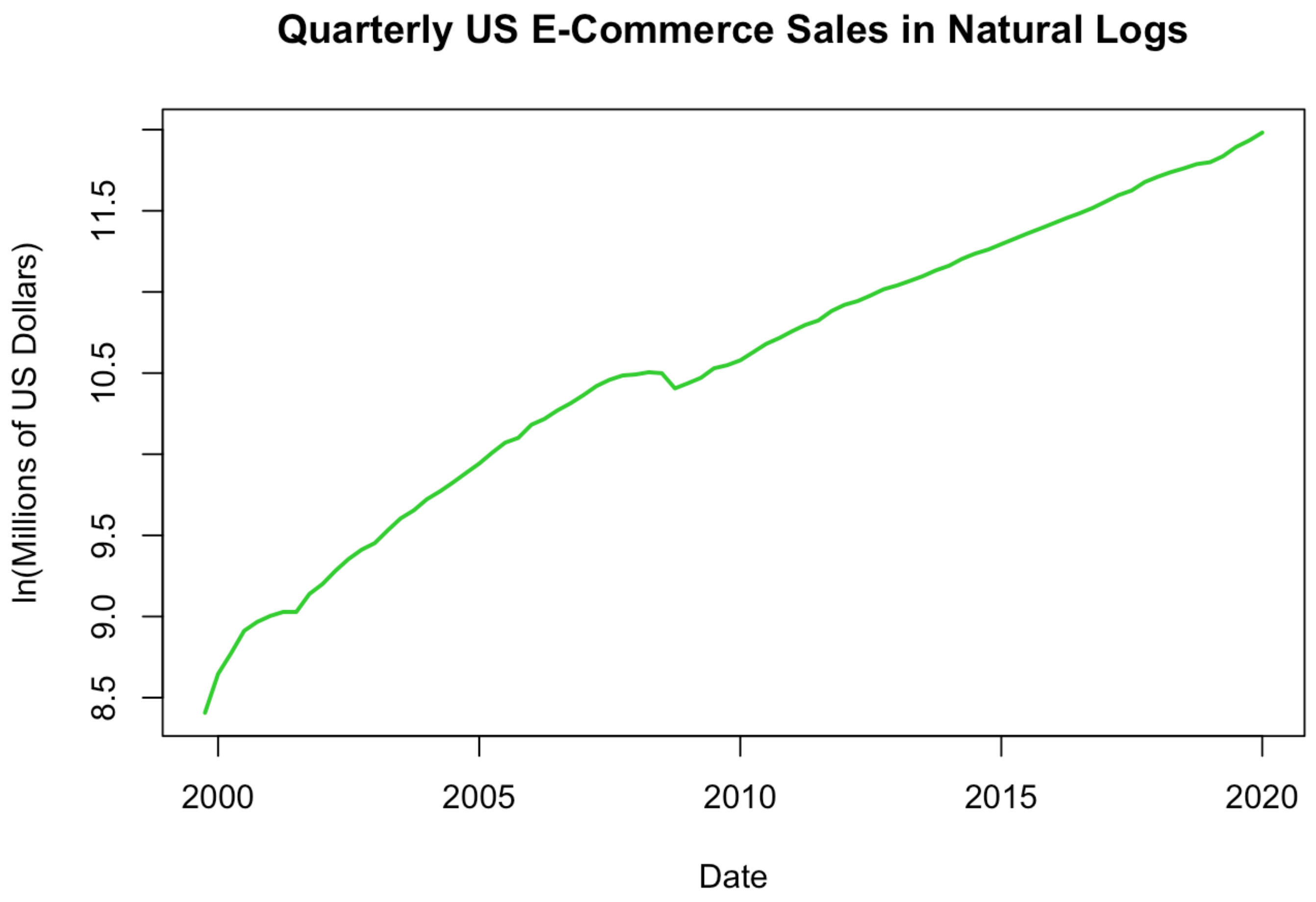
$$\ln(y_t) = \ln(\beta_0) + \beta_1 TIME_t + \epsilon_t$$

- And then we solve:

$$min_{\beta_0,\beta_1} \sum_{i=1}^{T} (\ln(y_t) - \ln(\beta_0) - \beta_1 TIME_t)^2$$

AN EXPONENTIAL DETERMINISTIC TREND

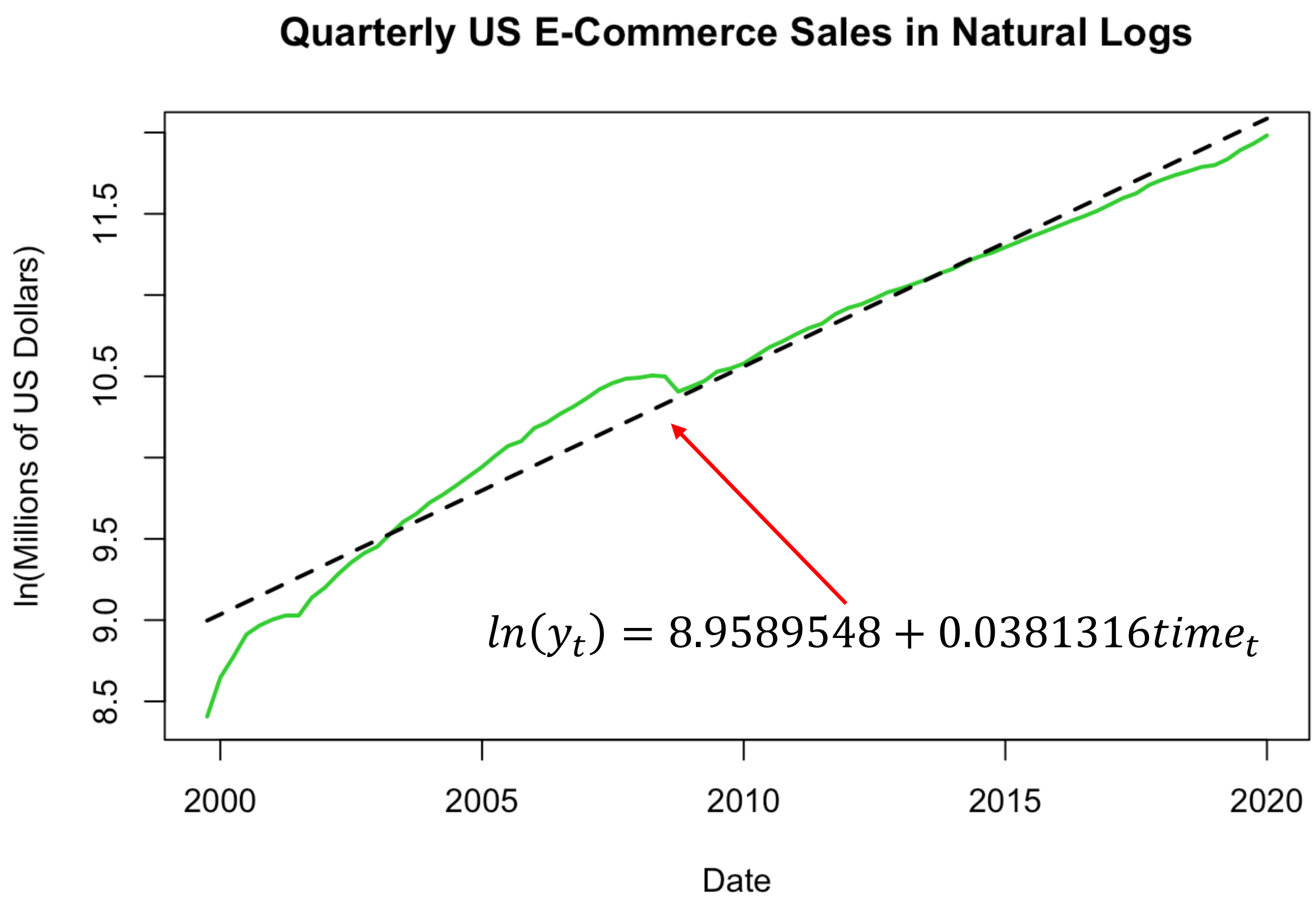
- Our logged data looks like this:



AN EXPONENTIAL DETERMINISTIC TREND

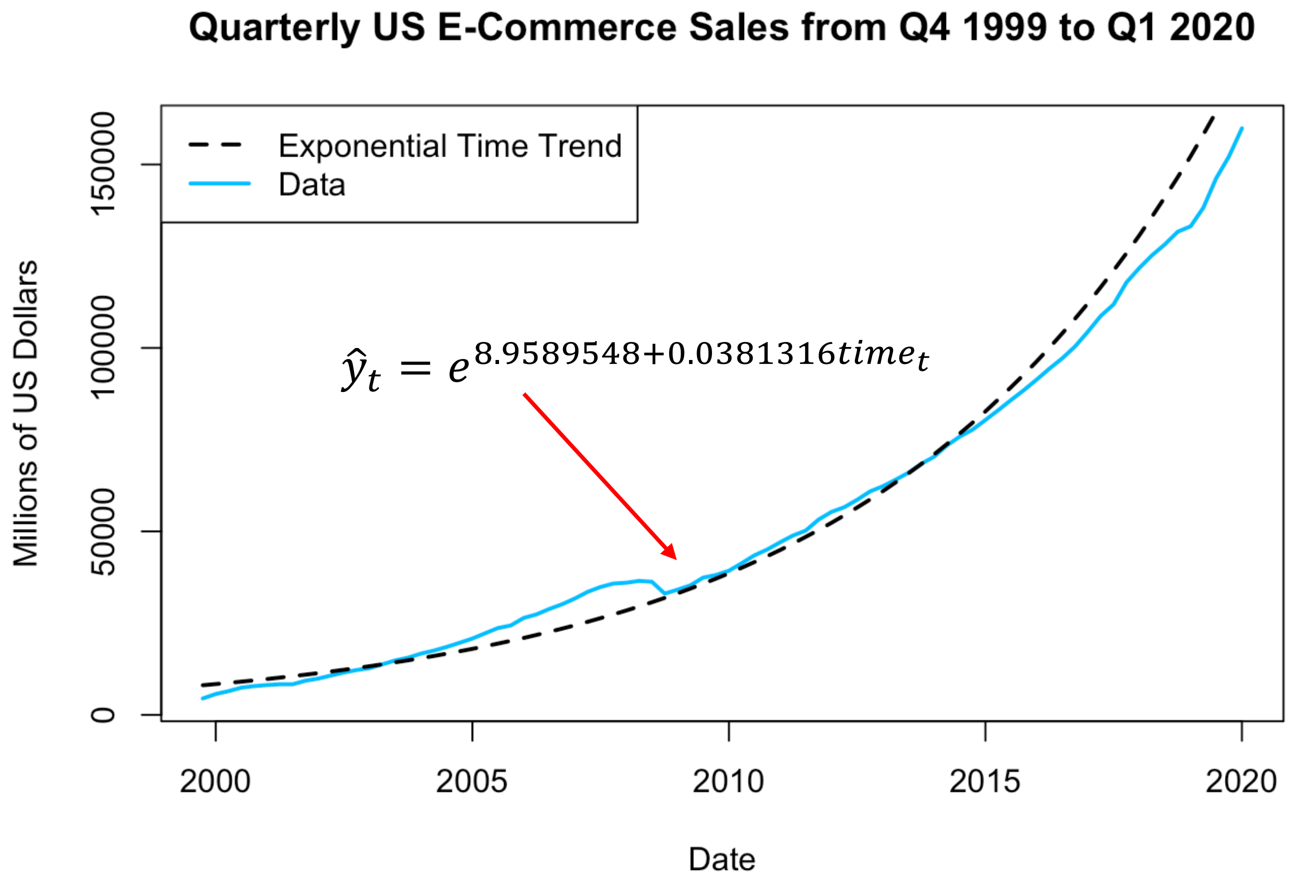
- The estimated coefficients and fitted values are:

```
## lm(formula = ln.ecomsa ~ time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59060 -0.07227  0.00180  0.05281  0.27938
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.9589548  0.0338976  264.29   <2e-16 ***
## time        0.0381316  0.0007095   53.74   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
## Residual standard error: 0.1521 on 80 degrees of freedom
## Multiple R-squared:  0.973,  Adjusted R-squared:  0.9727
## F-statistic:  2888 on 1 and 80 DF,  p-value: < 2.2e-16
```



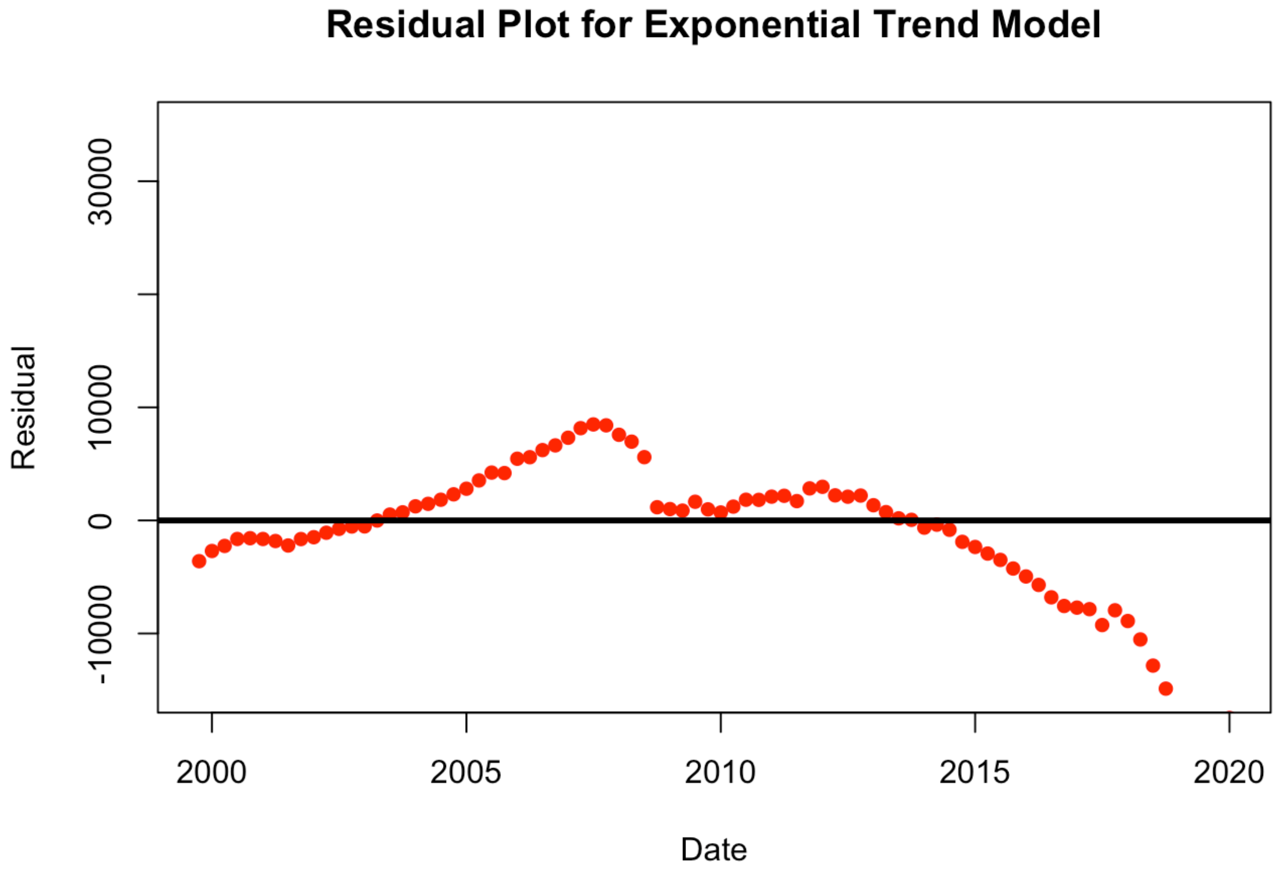
AN EXPONENTIAL DETERMINISTIC TREND

- Transforming back to the levels, we obtain:



AN EXPONENTIAL DETERMINISTIC TREND

- Looking at the residuals:



FORECASTING AND REGRESSION ANALYSIS

- Now that we have figured out how to specify and estimate a deterministic trend model we can now proceed to perform forecasts!
- To understand the connection between regression analysis and forecasts, we have to recognize that a regression model is a ***model of the conditional mean of the dependent variable***. That is, by specifying

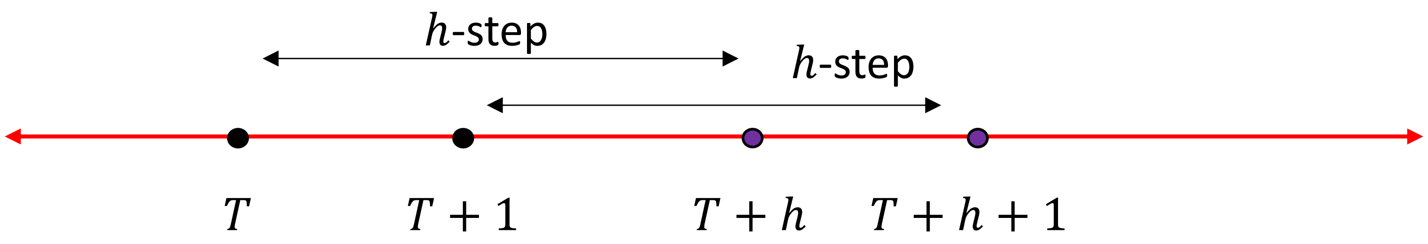
$$Y_t = \alpha + \beta X_t + \varepsilon_t \qquad E\left[\varepsilon_t \middle| X_t\right] = 0$$

- It must be true that,

$$E\left[Y_t \middle| X_t\right] = \alpha + \beta X_t$$

FORECASTING AND REGRESSION ANALYSIS

- Therefore we can think of a forecast of Y as the ***expectation*** of future Y conditional on available information!
- The ***forecast horizon*** is defined as the number of periods between the point in time in which the forecast is made and the date of the forecast. We will characterize it as the ***h-step ahead forecast*.**
- Therefore the h-step ahead forecast of a variable made at time period T produces a forecast for the variable at time $T + h$



FORECASTING AND REGRESSION ANALYSIS

- In mathematical notation, the computation of a h-step ahead forecast at time period t is written as:

$$E\left[Y_{t+h}\middle|\Omega_t\right]$$

- Where Ω_t represents all the information that is known or observed up to and including period t. To illustrate, let's consider the linear trend model:

$$y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t$$

- The h-step ahead forecast is therefore given by

$$E\left[y_{t+h}\middle|\Omega_t\right] = \beta_0 + \beta_1 TIME_{t+h} + E\left[\varepsilon_{t+h}\middle|\Omega_t\right] = \beta_0 + \beta_1 TIME_{t+h}$$

FORECASTING AND REGRESSION ANALYSIS

- Since the trend coefficients β_0 and β_1 are unknown, we use their OLS estimates,

$$E\left[y_{t+h} \middle| \Omega_t\right] = \hat{\beta}_0 + \hat{\beta}_1 TIME_{t+h}$$

- This produces what is known as a ***point forecast*** as the calculation results in only a single value. To illustrate, consider our estimated linear trend regression:

$$\hat{y}_t = -15221.07 + 1666.99TIME_t$$

- The last period in the sample is Q1 2020 and there are 82 observations in our sample. This means that the 1-step ahead forecast for Q2 2020 is given by:

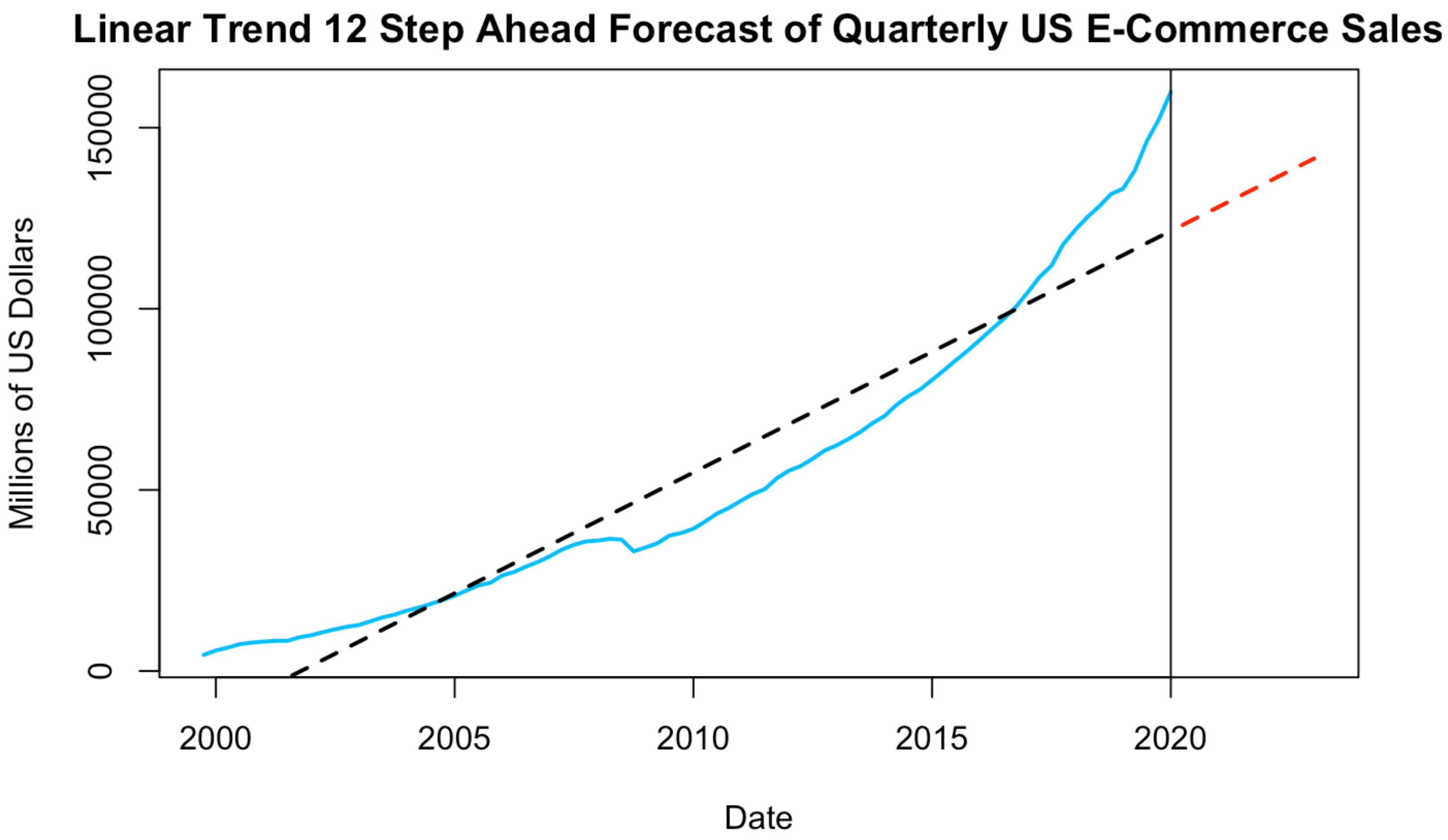
$$E\left[y_{83} \middle| \Omega_{82}\right] = \widehat{y_{83}} = -15221.07 + 1666.99TIME_{83} = -15221.07 + 1666.99(83) = 123139.1$$

- Point forecasts are concise and easy to understand, however they do not convey any sense of how precise the forecast is. We will discuss how to construct interval forecasts next week!

FORECASTING AND REGRESSION ANALYSIS

- Suppose we were interested in generating the 12-step ahead point forecast for our linear trend model.

- Does the linear trend model do a good job?



MODEL SELECTION

- From our example we can easily discern how the three different trend models perform relative to each other in terms of their forecasting accuracy.
- However, simply eyeballing a chart is not a particularly rigorous approach to model selection! We will use two statistics to help guide us:

$$AIC = -2loglikelihood + 2K$$

$$BIC = SIC = -2loglikelihood + log(T)K$$

- Where

$$loglikelihood\ is\ larger\ when\ SSE = \sum_{t=1}^{T} e_t^2\ is\ smaller$$

$$K = number\ of\ parameters$$

MODEL SELECTION

- Intuitively, regression models that fit the data well will have a relatively low SSE or equivalently a high R^2. However, we know that we can always achieve this by adding an arbitrarily large number of regressors.

- Both the BIC and AIC of regression models are penalized functions of the mean squared residual SSE where the penalties are functions of the degrees of freedom used in fitting the model, $T - K$. We will choose the models that produce the lowest realizations of these criteria.

- While they are quite similar, the BIC and AIC have different asymptotic properties which are beyond the scope of this course.

MODEL SELECTION

- The adjusted R^2, AIC and BIC from our linear, quadratic and exponential trend models are:

MODEL	$Adj. R^2$	AIC	BIC
Linear	0.9111	1782.343	1789.563
Quadratic	0.9898	1605.632	1615.632
Exponential	NA	1527.634	1534.855

- We can't easily compare the log-linear trend model to the other trend models because the dependent variable is measured on a different scale. We will have to estimate the exponential trend model in the levels via nls in order to generate the appropriate AIC and BIC.