INDR 422/522

Fikri Karaesmen

Spring 2023

Simple time series forecasts - 3

March 14, 2023

Reminders

- Blackboard page is becoming active
  - Last Week’s slides
  - Last year’s lecture slides
  - Class Exercise Solutions
  - Will be uploading the current slides as we proceed
- First lab available, please take a look and work on the exercises
- Second lab will be available this Friday
- Participation taken. Please participate in polls.
- Please follow announcements

Class Exercise from last lecture

CLASS EXERCISE, March 9, 2023

1. Let $Y_t = \mu + \epsilon_t$ where ϵ_t are i.i.d random variables with mean zero and variance σ^2. Consider the forecast: $\hat{y}_{t+1} = (y_t + y_{t-1})/2$.

   (a) $\hat{Y}_{t+1}$ and Y_{t+1} are independent
   Solution: $\hat{Y}_{t+1}$ depends Y_t and Y_{t-1} but by definition of the process (ϵ_t are independent) does not depend on Y_{t+1}. Therefore, the are independent.

   (b) $\hat{Y}_{t+1}$ and Y_t are independent
   Solution: $\hat{Y}_{t+1}$ depends on Y_t. They are dependent and positively correlated.

   (c) $\hat{Y}_{t+1}$ and $\hat{Y}_{t+2}$ are independent
   Solution: $\hat{Y}_{t+1}$ and $\hat{Y}_{t+2}$ both depend on Y_t. Therefore, they are dependent and positively correlated.

   (d) Denote the error by $E_t = Y_t - \hat{Y}_t$. What is the variance of the error.
   Solution:

$$Var(E_t) = Var(Y_t - \hat{Y}_t) = \sigma^2 + (-1)^2\sigma^2/2 = 3\sigma^2/2.$$

Class Exercise from last lecture

(c) $\hat{Y}_{t+1}$ and $\hat{Y}_{t+2}$ are independent
Solution: $\hat{Y}_{t+1}$ and $\hat{Y}_{t+2}$ both depend on Y_t. Therefore, they are dependent and positively correlated.

(d) Denote the error by $E_t = Y_t - \hat{Y}_t$. What is the variance of the error.
Solution:

$$Var(E_t) = Var(Y_t - \hat{Y}_t) = \sigma^2 + (-1)^2\sigma^2/2 = 3\sigma^2/2.$$

2. Consider the following demand process: $Y_t = \mu_1 + \epsilon_t$ if t is odd and $Y_t = \mu_2 + \epsilon_t$ if t is even, where ϵ_t are i.i.d random variables with mean zero and variance σ^2 and $\mu_1 > \mu_2$.

(a) Consider $\hat{y}_{t+1} = y_t$. Is this unbiased?
Solution: If t is even then $E[\hat{Y}_{t+1}] = \mu_2$ but $E[Y_{t+1}] = \mu_1$ since $t+1$ is odd. Therefore, this forecast is not unbiased.

(b) Consider $\hat{y}_{t+1} = (y_t + y_{t-1} + y_{t-2})/3$. Is this unbiased?
Solution: If t is even then $E[\hat{Y}_{t+1}] = (2\mu_2 + \mu_1)/3$ but $E[Y_{t+1}] = \mu_1$ since $t+1$ is odd. Therefore, this forecast is not unbiased.

(c) Consider $\hat{y}_{t+1} = y_{t-1}$. Is this unbiased?
Solution: If t is even then $E[\hat{Y}_{t+1}] = \mu_1$ and $E[Y_{t+1}] = \mu_1$ since $t+1$ is odd. A symmetric argument holds when t is odd. Therefore, this forecast is unbiased.

Summary so far

- We can have estimators (forecasts) based on observations
    - Focusing on most recent observations (i.e. naïve forecasts, moving averages)
- We can control the variance by estimating required parameters based on multiple observations
    - Simple averages or exponential smoothing
- But there is a trade-off between adapting to changes (more focus on most recent observations – higher variance) and stability (focust on long-term averages - lower variance)
- We have systematic ways of estimating basic patterns from recent data.

$$f(x+h) = f(x) + hf'(x) + h^2\frac{f''(x)}{2} + ...$$

We can now infer an unbiased forecast:

$$\hat{y}_{T+h} = y_T + +h(y_T - y_{T-1}) + h^2((y_T - y_{T-1}) - (y_{T-1} - y_{T-2}))/2$$

Time series: simple forecasts: error performance

- We use the term error to denote the difference between an observed value and its forecast:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

  (where the training set is $\{y_1, y_2, ..y_T\}$ and the test set is $\{y_{T+1}, y_{T+2}, ..\}$
- Note that the term residual refers to the training set and the term error refers to the test set.

Time series: simple forecasts: error performance

- Absolute error measures:

$$1.\ \text{MAE} = \sum_{t=1}^{T} \frac{|e_t|}{T}$$

$$2.\ \text{MSE} = \sum_{t=1}^{T} \frac{|e_t^2|}{T}$$

$$3.\ \text{RMSE} = \sqrt{\text{MSE}}$$

- Percentage error measure:

$$4.\ \text{MAPE} = \sum_{t=1}^{T} \frac{|p_t|}{T}$$

where $p_t = 100e_t/y_t$

Time series: simple forecasts: error performance

- Scaled error measure: we express the error in comparison to simple but reasonable benchmark, for instance the naive forecast $\hat{y}_T = y_{T-1}$. Let $e'(t)$ denote the error of benchmark forecast in period T and MAE(benchmark) its mean absolute error. We define

$$q_t = \frac{|e_t|}{\text{MAE(benchmark)}}$$

- and the Mean Absolute Scaled Error is then:

$$5.\ \text{MASE} = \sum_{t=1}^{T} \frac{q_t}{T}$$

Scaled Errors

- We’ll look at more sophisticated methods with more parameters. But it’s not always the case that having more parameters lead to a better forecast.
- We should always compare performance with respect to simple benchmarks (i.e. Moving averages, double exponential smoothing etc).
- If the benchmarks perform better than the sophisticated forecast, we should use the benchmark.

Time series: simple forecasts: error distribution

- For planning purposes (inventory control, capacity setting, staffing etc), we have to use the probability distribution of the forecast and not only its point estimate (for the mean).
- Let $Y(t) = f(t) + \epsilon_t$. If the distribution of ϵ_t is known then we know the distribution of the forecast $\hat{Y}_t$.
- If all is well (i.e. our forecast is unbiased) and we are lucky, ϵ_t is independent and normally distributed with variance σ_2. If so our forecast is also normally distributed with mean $\mu = f(t)$ and variance σ^2.
- We are then able to write a nice prediction interval around our point forecast $\hat{y}_t$:

$$\hat{y}_t \pm 1.96\sigma$$

This is a 95% prediction interval: $1.96 = \Phi^{-1}(0.975)$.

Time series: simple forecasts: error distribution

- In practice, we don't know the true model that generates the demand but we have observed the residuals. We can use them to estimate the forecast variance $\hat{\sigma}$.
- Note that

$$e_t = y_t - \hat{y}_t$$

  Therefore,

$$Var(E_t) = Var(Y_t - \hat{Y}_t) = Var(Y_t) + Var(\hat{Y}_t) > Var(Y_t)$$

- But we make efforts to make the $Var(\hat{Y}_t)$ relatively small. For instance, if we take as a test model $Y_t = c + \epsilon_t$ and as an unbiased forecast a 5-period moving average, we saw that $Var(\hat{Y}_t) = \sigma_2/5$. Then $Var(E_t) = \sigma^2 + \sigma_2/5$. So, $\sigma_E \approx 1.1\sigma$.

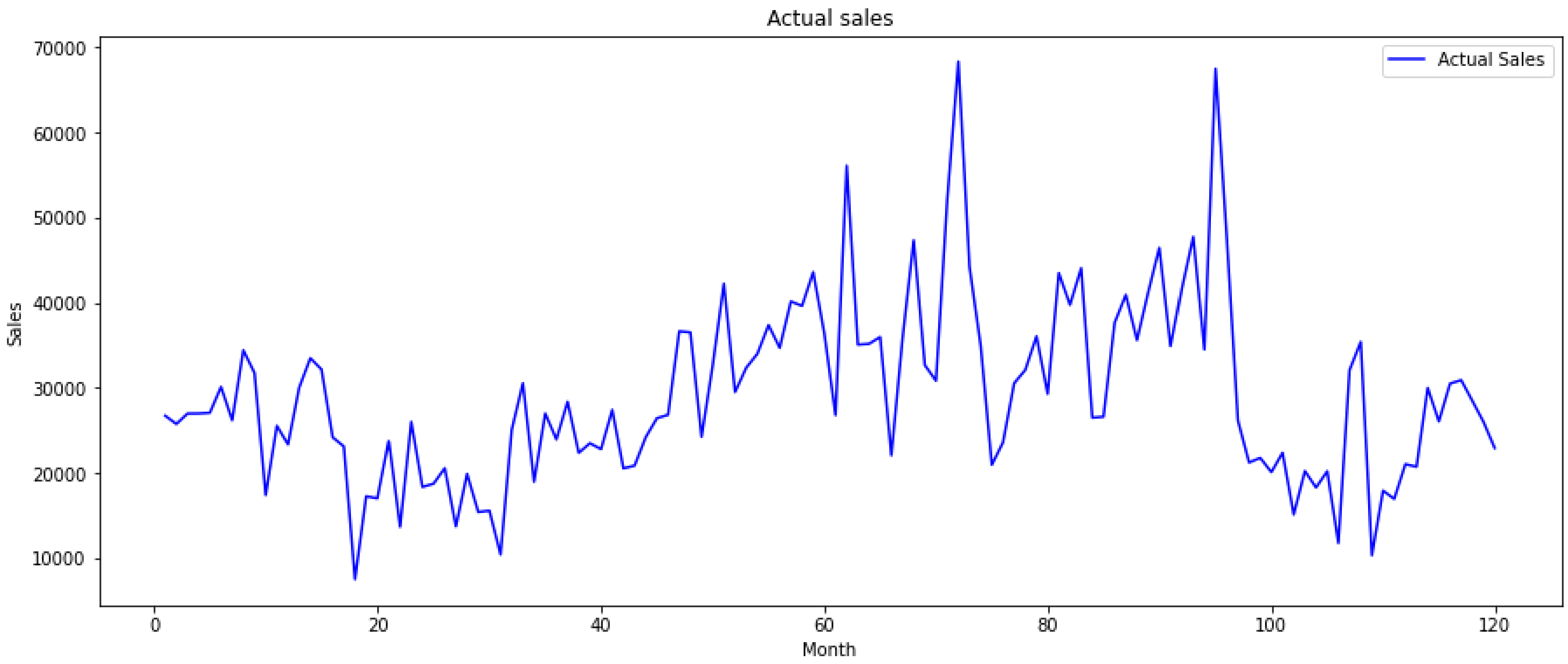
Time series: simple forecasts: error distribution

- We therefore proceed with the distribution of the residuals to approximate distribution of the forecast itself (i.e. the distribution of the forecast is a point estimator plus a random variable represented by the residuals).
- If the residuals are normally distributed, we have approximate prediction intervals given by:

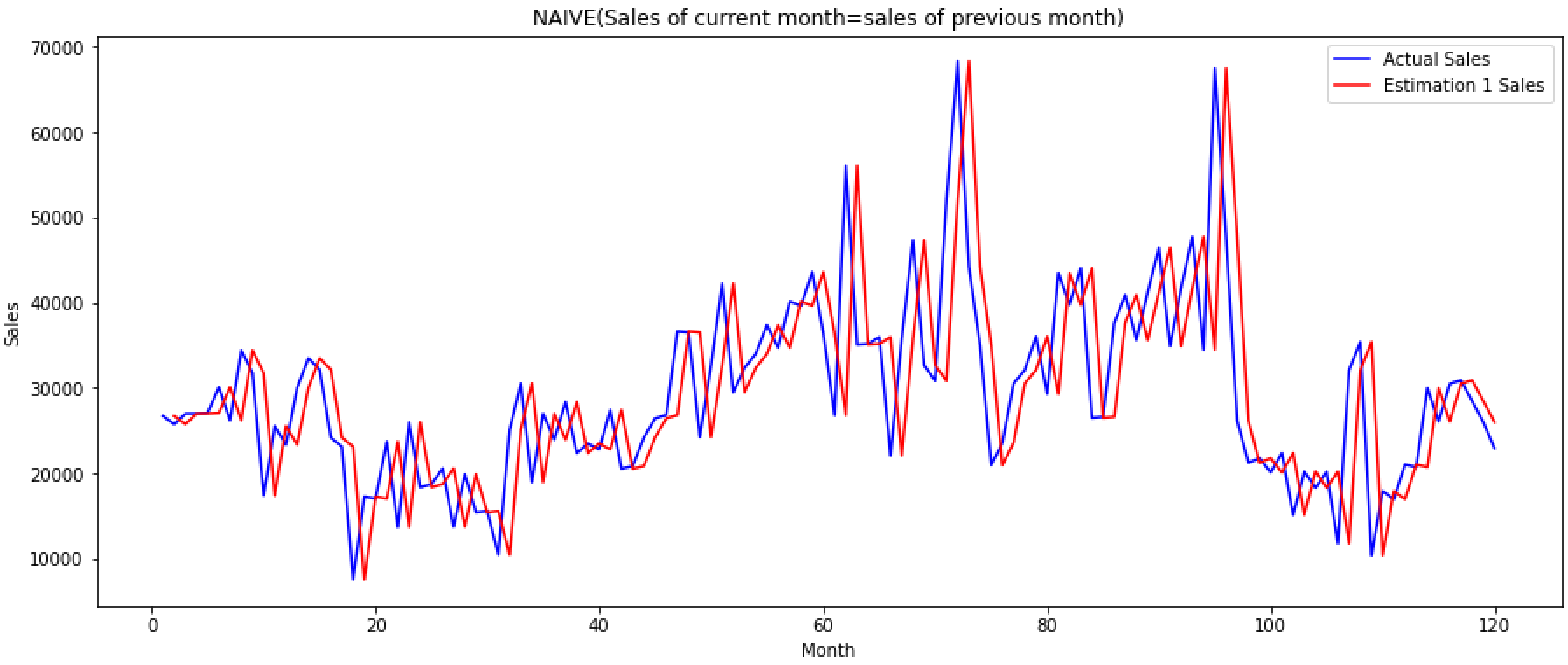
$$\hat{y}_t \pm 1.96\sigma_e$$

- If the residuals are not normally distributed, we need to more work (use MLE to fit a distribution or use the sample as is).
- Residuals are critical for planning purposes. Please note that if we need prediction intervals or a probability distribution for a multi-step ahead forecast, we need to consider multi-step ahead residuals. Typically, the variance of the residuals increase as the horizon increases.

Example (LAB 1): monthly dishwasher sales in Turkey (Jan 93 to Dec. 2002)

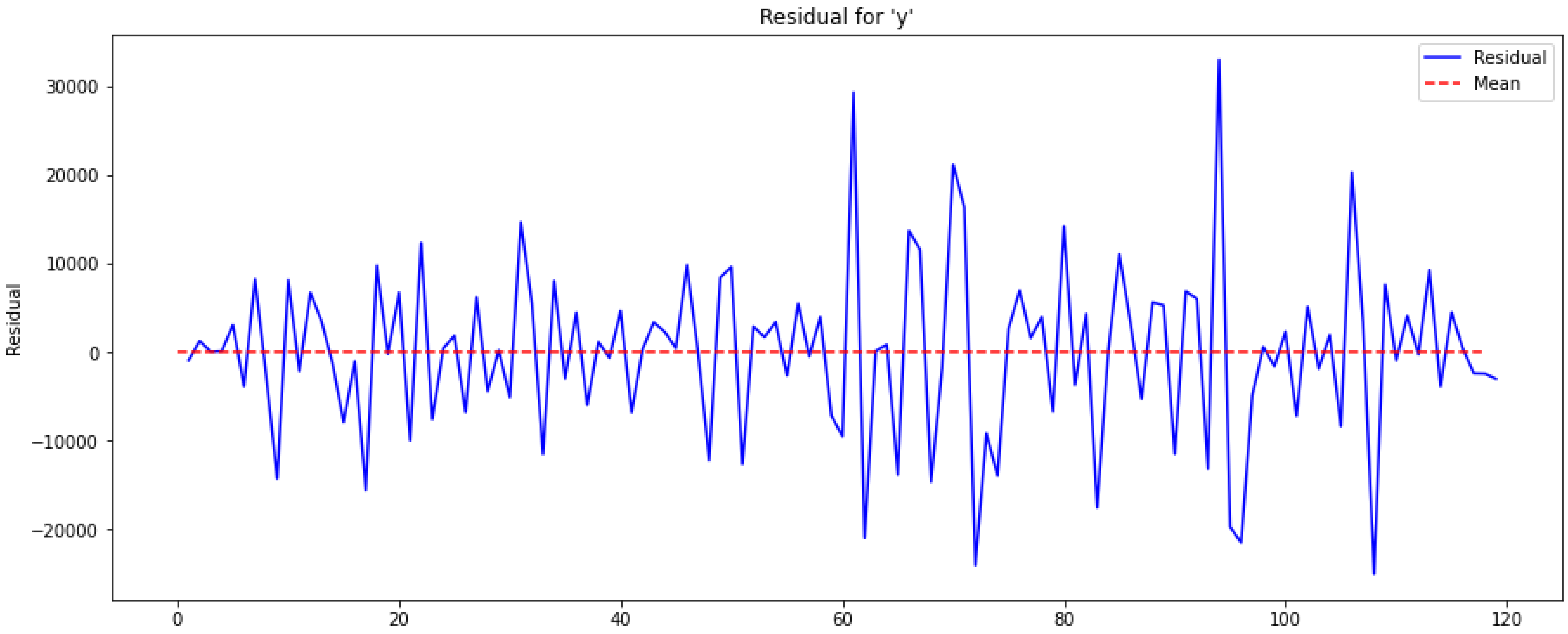


Example (LAB 1): monthly dishwasher sales: naïve forecast



MAE1 = 7002.49
MAPE1 = 26.95%
MSE1 = 92464126.68
RMSE1 = 9615.83

Example (LAB 1): monthly dishwasher sales: naïve forecast - residuals



Mean of Residual: -31.88235294117641
S.D. of Residual: 9615.77403013423

Example (LAB 1): monthly dishwasher sales: naïve forecast - residuals

- Unbiasedness: average residual = -31.882, but is this statistically significant?
- H_o: $\mu_R=0$, H_1:$\mu_R \neq 0$.
- We can run a two-sided test to check the hypothesis
- Recall that $\bar{r} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$ is a (1-α)% confidence interval for μ_R.
- For our case, the half width of the interval is:

$$\frac{9615.77}{\sqrt{119}} t_{1-\alpha/2}$$

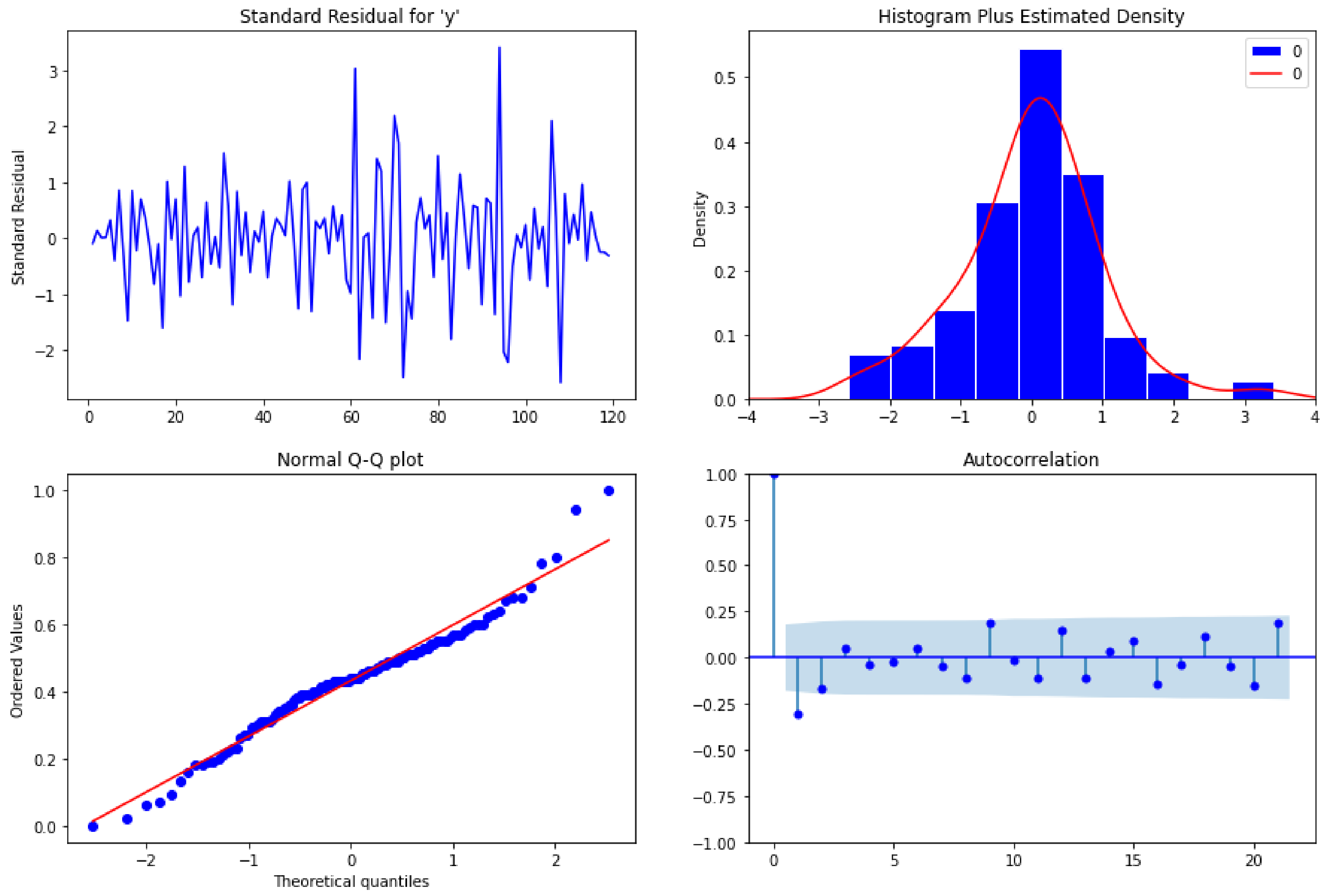
Mean of Residual: -31.88235294117641
S.D. of Residual: 9615.77403013423

We conclude that all reasonable confidence intervals include zero. Therefore, we cannot distinguish that mean of the residuals from zero (i.e. on this data set the forecast is statistically unbiased)

Example (LAB 1): monthly dishwasher sales: naïve forecast - residuals

- The hypothesis tests are weak unless the sample size is very large.
- We should check unbiasedness but most forecasts will pass the statistical test (i.e. cannot reject the null hypothesis that $\mu_R=0$).
- We should quickly check unbiasedness but focus on the error measures!
  - This is the reason for focusing on the variance of the estimator or errors

Example (LAB 1): monthly dishwasher sales: naïve forecast – residual diagnostics (thanks to Bijan Bibak)



$$z_t = \frac{r_t - \bar{r}}{s} = \frac{r_t - (-31.882)}{9615.774}$$

Mean of Residual: -31.88235294

S.D. of Residual: 9615.77403013

Example (LAB 1): monthly dishwasher sales: comparisons

Out[22]:

	MAE	MAPE (%)	MSE	RMSE
Naive	7,002.49	26.95	92,464,126.68	9,615.83
Moving Average	6,715.89	26.20	84,111,407.60	9,171.23
Short Term Trend	10,653.37	40.41	219,873,900.77	14,828.15
Seasonal	10,848.66	43.82	180,296,748.29	13,427.46
Exponential	6,575.95	25.29	79,934,590.74	8,940.61

There’s some room for improvement. We’ll see if we can do any better with more sophisticated methods.

Data Transformations

- It is useful to transform data to have more efficient (and maybe better) forecasts.
- We like to work with forms where there is some pattern plus some random additional error term. We prefer that error term is iid. This means we would like to see stable fluctuations (i.e. constant variance in time).
- We also like stationary processes where we have simpler tools (averaging, exponential smoothing principles, and auto-regressive tools).
- To stabilize the variance, we use power or log transformations:

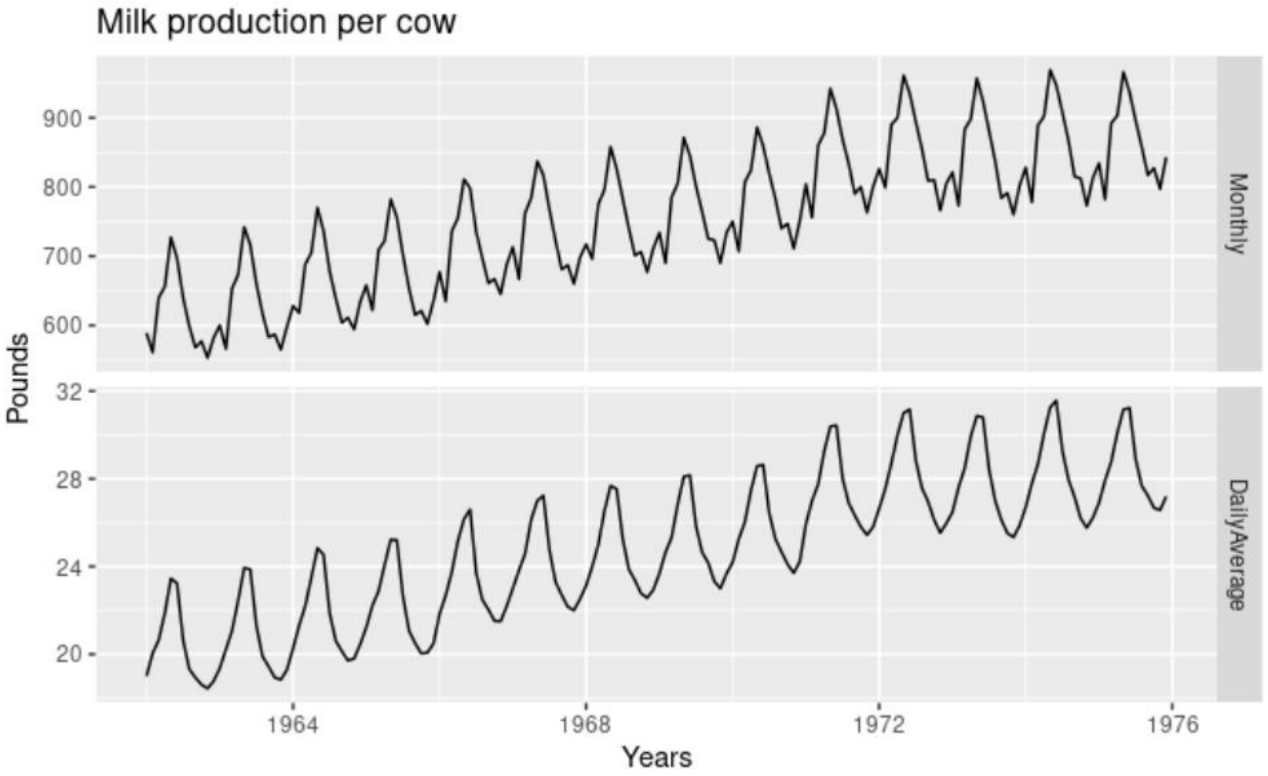
$$
\begin{aligned}
z_t &= \log(y_t) \text{ for all } t \\
z_t &= \sqrt{y_t} \text{ for all } t \\
z_t &= (y_t)^{1/3} \text{ for all } t
\end{aligned}
$$

Data Transformations

- We can the forecast for the transformed series and undo the transformation to have a prediction for the original model.
- Note that these transformations are non-linear and we need to be careful in computing prediction intervals (i.e. first compute them for the transformed model then convert them to the original scale).

Data Transformations: Example

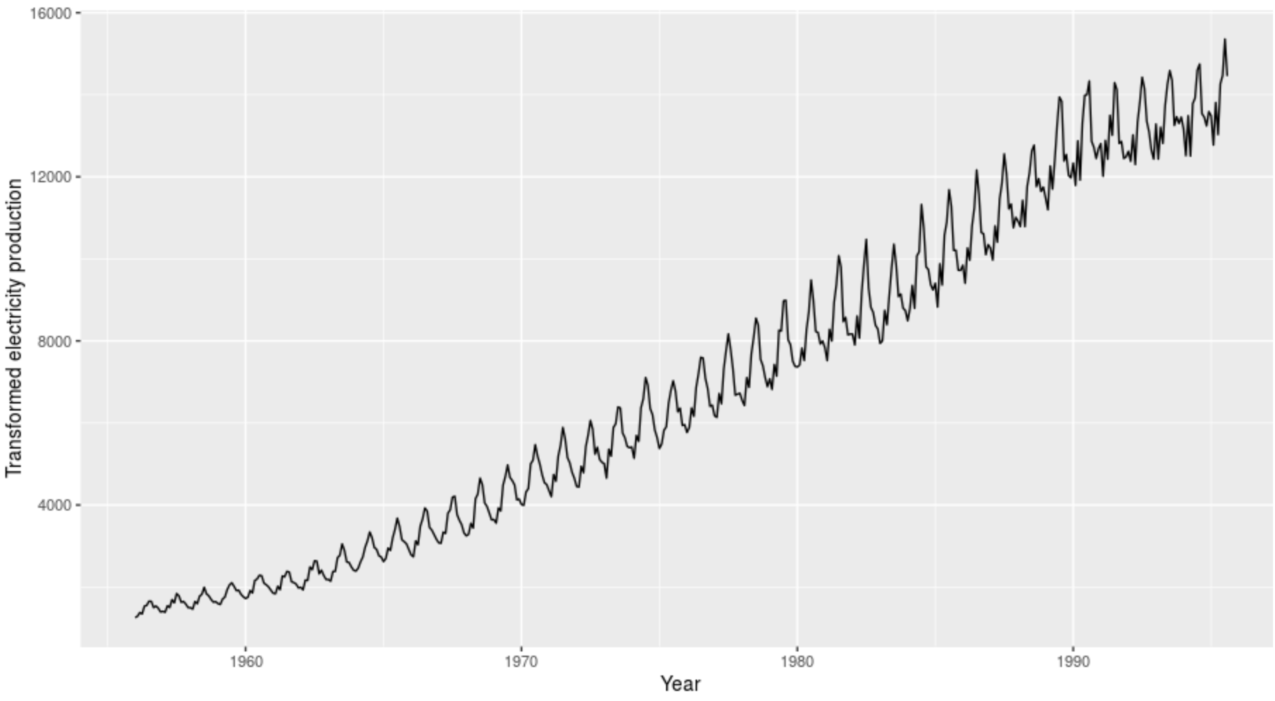
Calendar adjustments: monthly data is sometimes deceitful because each month does not have the same number of days. Convert months to a Daily average.



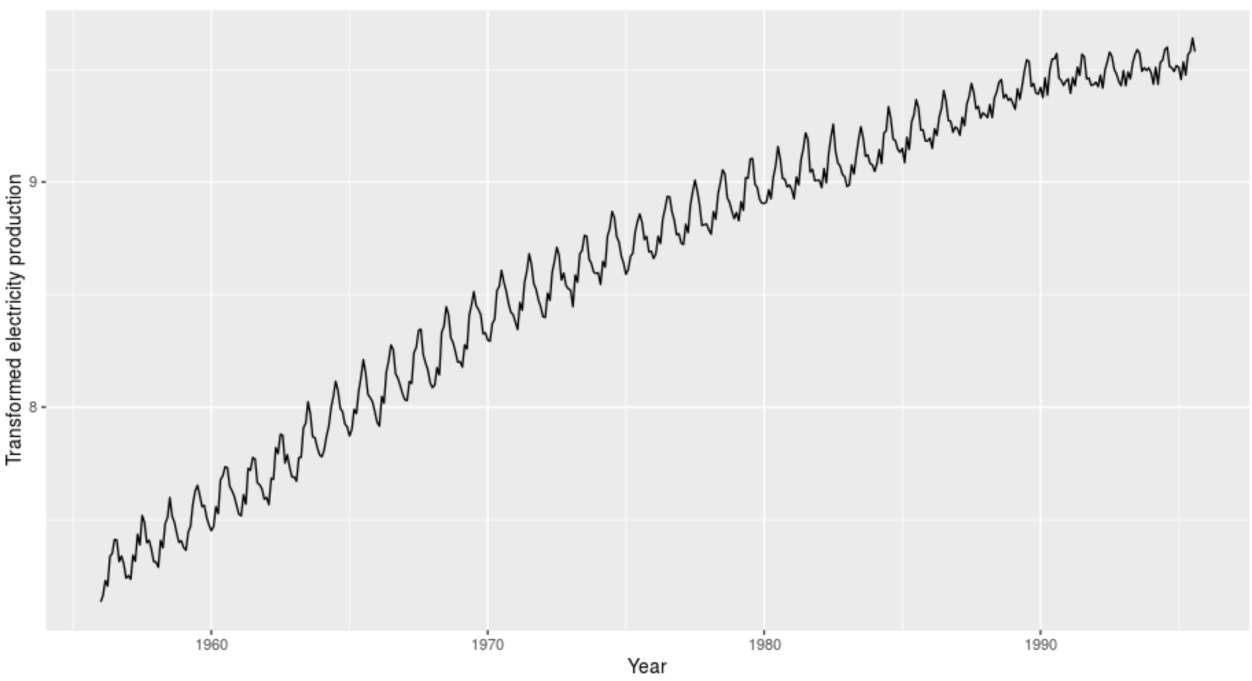
Source: Hydman and Athanasopoulos

Data Transformations: Example

Australian Monthly Electricity Production



Log Transformation



Source: Hydman and Athanasopoulos

Data Transformations: detrending (differencing)

- We can also transform the data to convert the original series to a stationary series.
- Take the example of:

$$Y_t = c + bt + \epsilon_t$$

- Now consider the transformation:

$$W_t = Y_t - Y_{t-1} = b + \epsilon_t - \epsilon_{t-1} = b + \epsilon'_t$$

- This gives an alternative approach for forecasting. Find a forecast $\hat{w}_t$ for W_t and convert it to a forecast for Y_t by reverting the transformation.

Data Transformations

- This gives an alternative approach for forecasting. Find a forecast $\hat{w}_t$ for W_t and convert it to a forecast for Y_t by reverting the transformation:

$$\hat{y}_{t+1|t} = \hat{w}_{t+1|t} + y_t$$

Data Transformations: deseasonalizing

- The transformation $W_t = Y_t - Y_{t-1}$ is known as simple detrending, it removes a linear trend.
- A similar transformation that applies to (simple) seasonal data is called deseasonalizing:

$$W_t = Y_t - Y_{t-m}$$

for a season length of m. A reasonable forecast is then

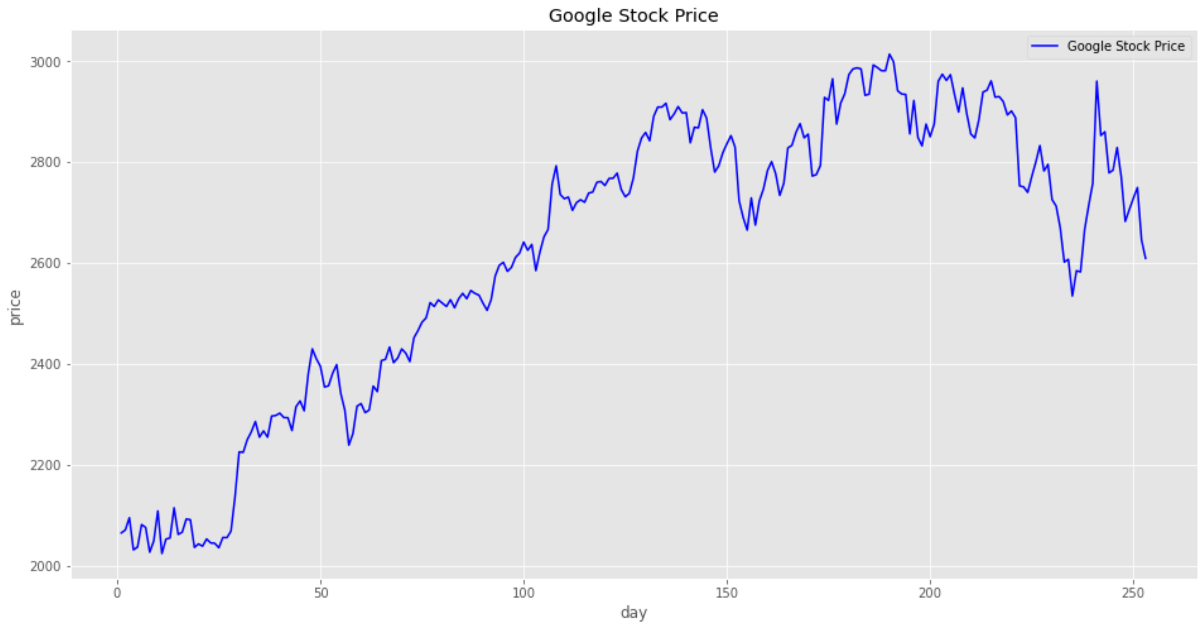
$$\hat{y}_{t+1|t} = \hat{w}_{t+1|t} + y_{t+1-m}$$

Data Transformations

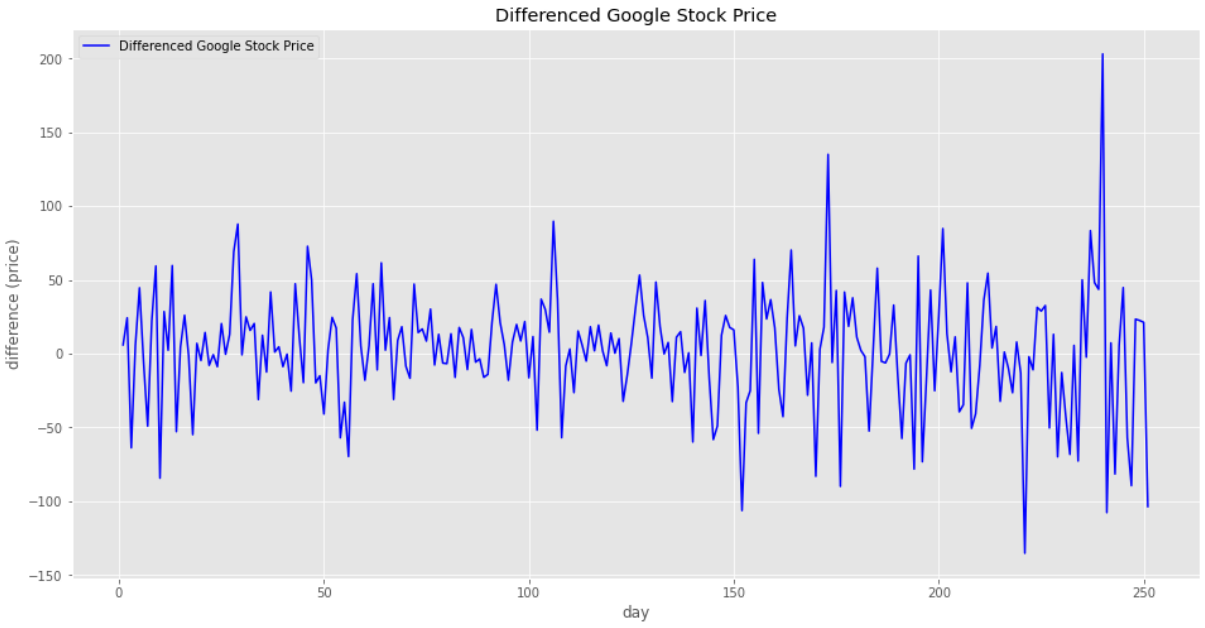
- One can combine these transformations: first deseasonalize and then detrend $W_t = Y_t - Y_{t-m}$ and $V_t = W_t - W_{t-1}$.
- We can difference twice to remove quadratic trend: $W_t = Y_t - Y_{t-1}$ and $V_t = W_t - W_{t-1}$.
- and difference k times to remove a pattern with a leading term of t^k.

Data Transformations: Google Stock Price Example

Google (Alphabet) Stock Price over the last year



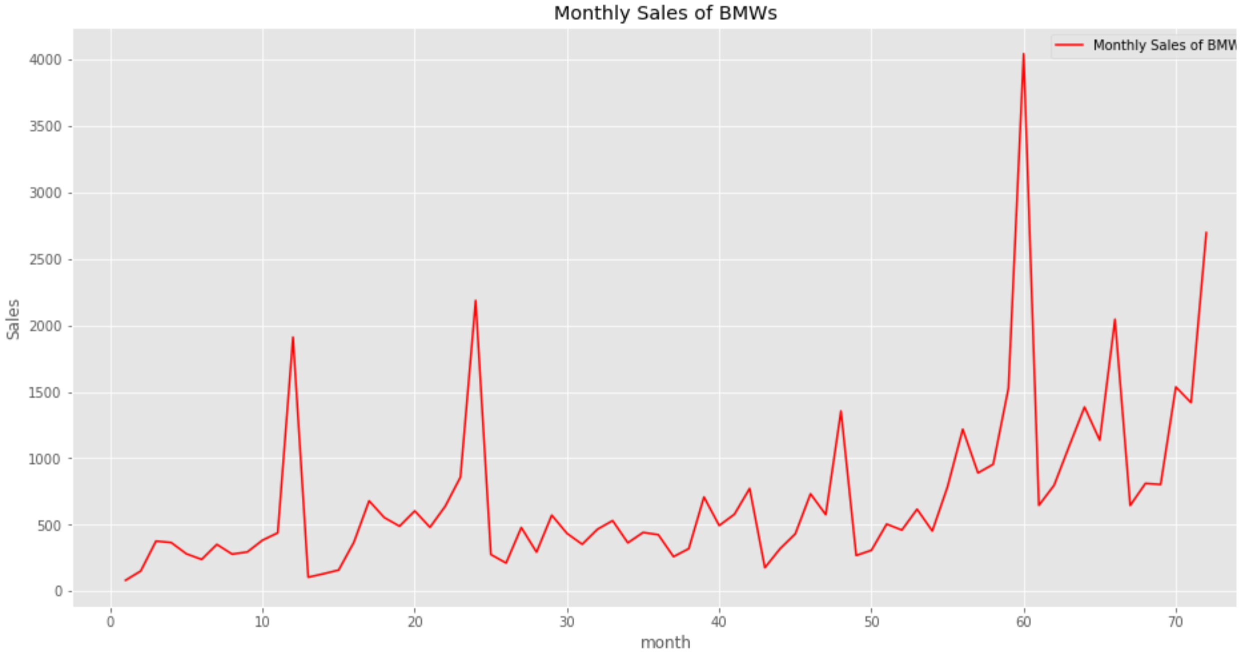
Differenced Google (Alphabet) Stock Price over the last year



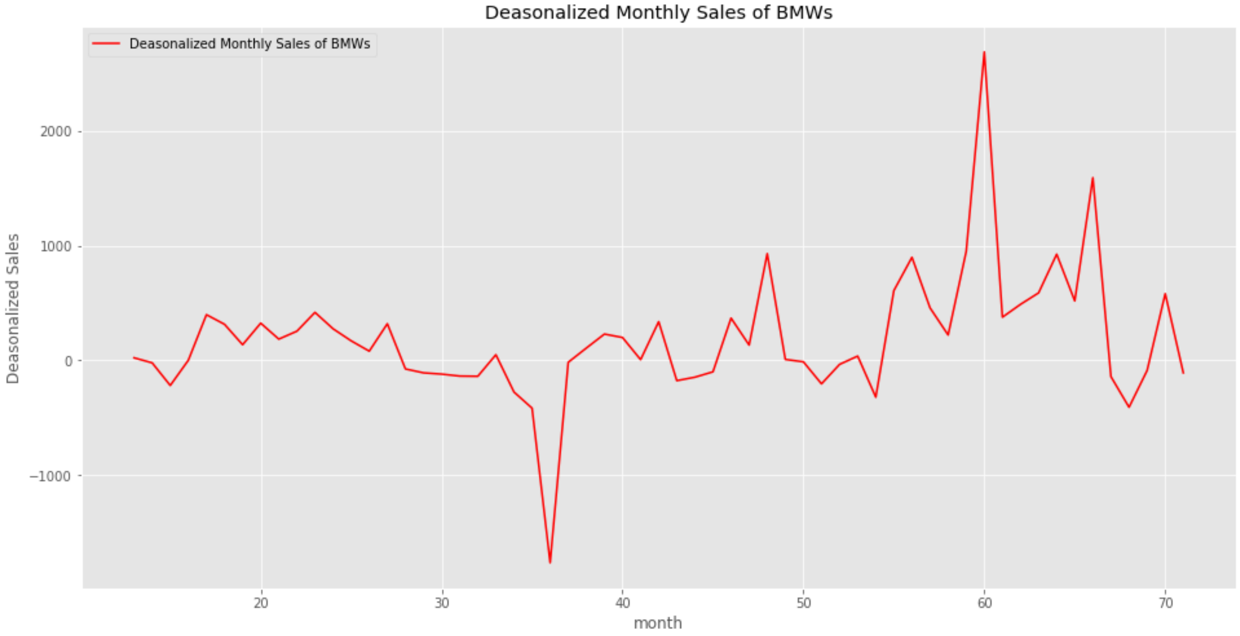
Data compiled on Feb. 22, 2022

Data Transformations: BMW Sales Example

Monthly BMW Sales since 2016



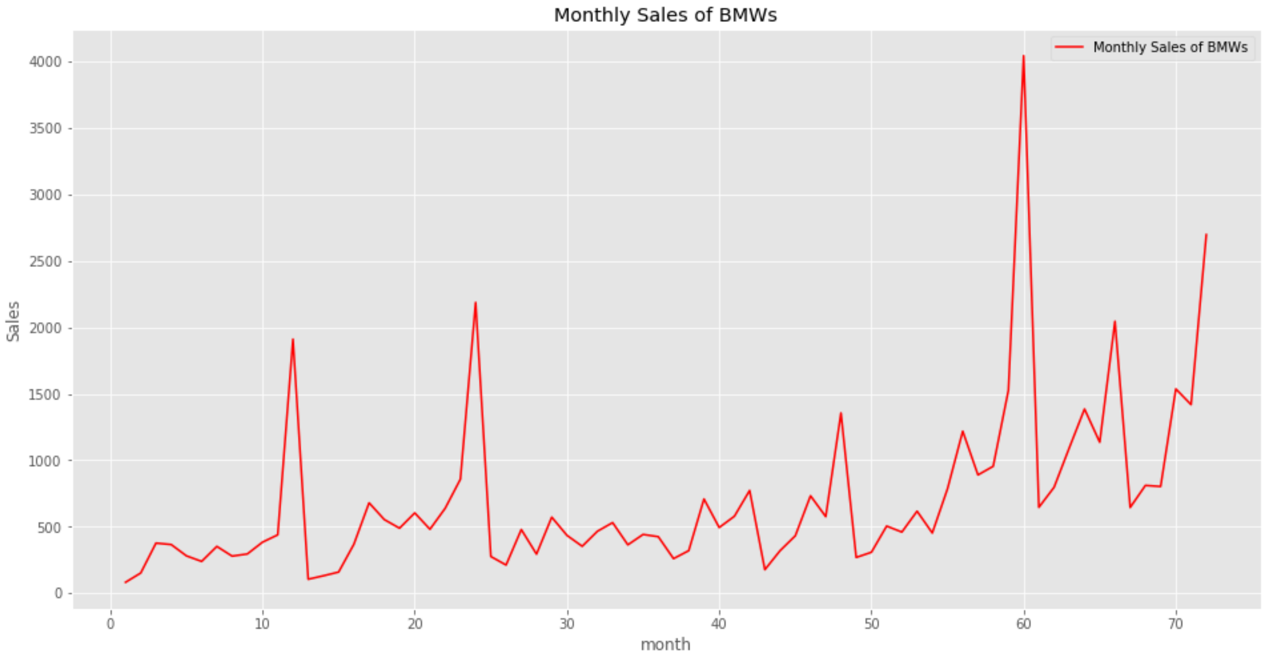
Deseasonalized Monthly BMW Sales



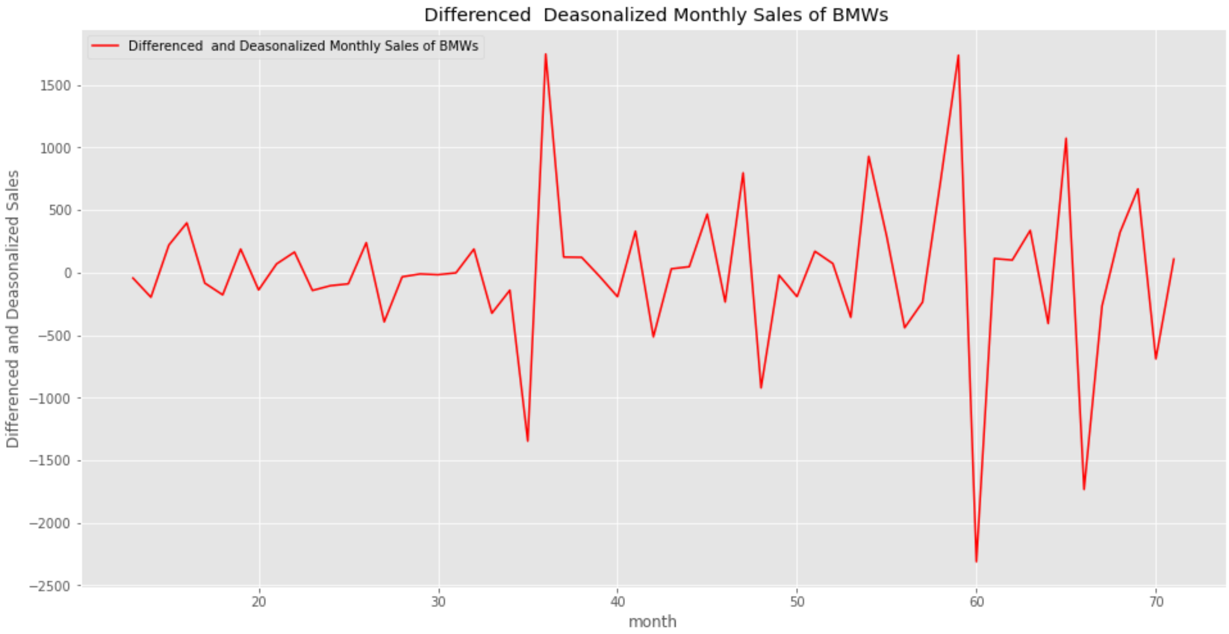
Data until Dec. 2021 (included)

Data Transformations: BMW Sales Example

Monthly BMW Sales since 2016

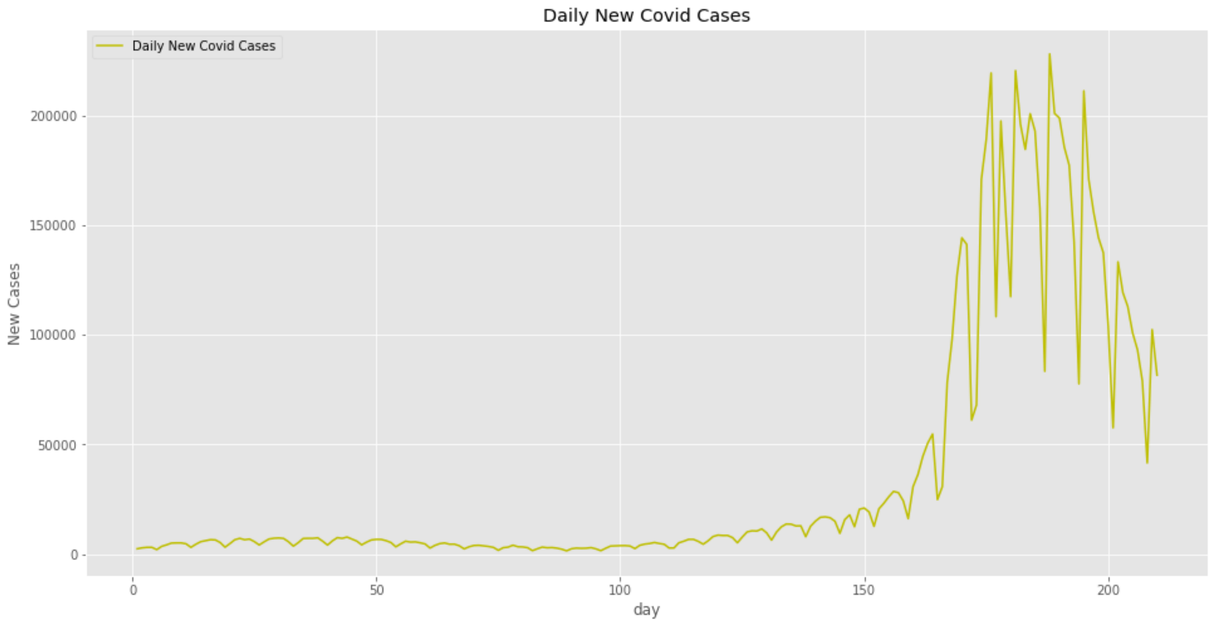


Deseasonalized and Differenced Monthly BMW Sales

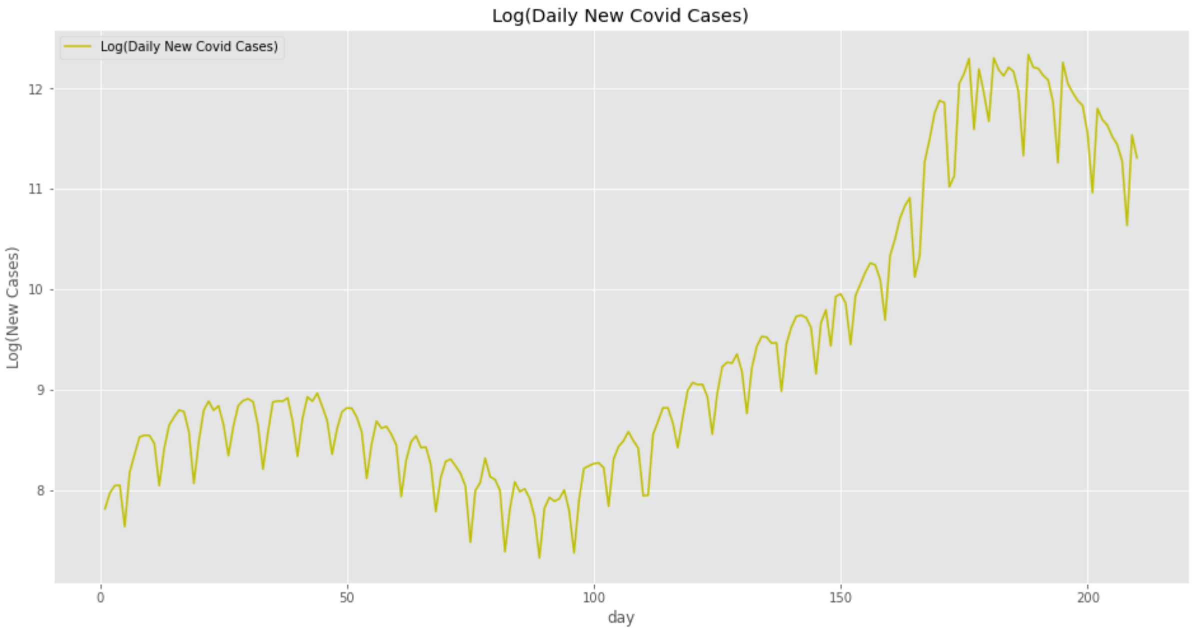


Data Transformations: Daily New Covid Cases in Italy (2021)

Daily New Covid Cases



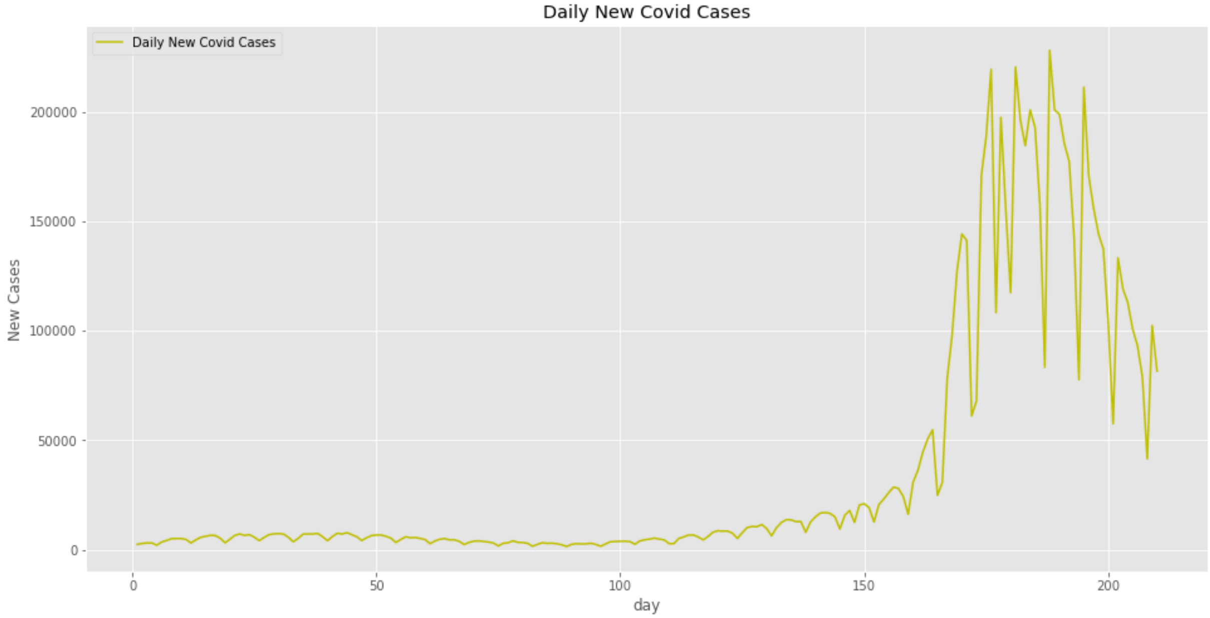
Log(Daily New Covid Cases)



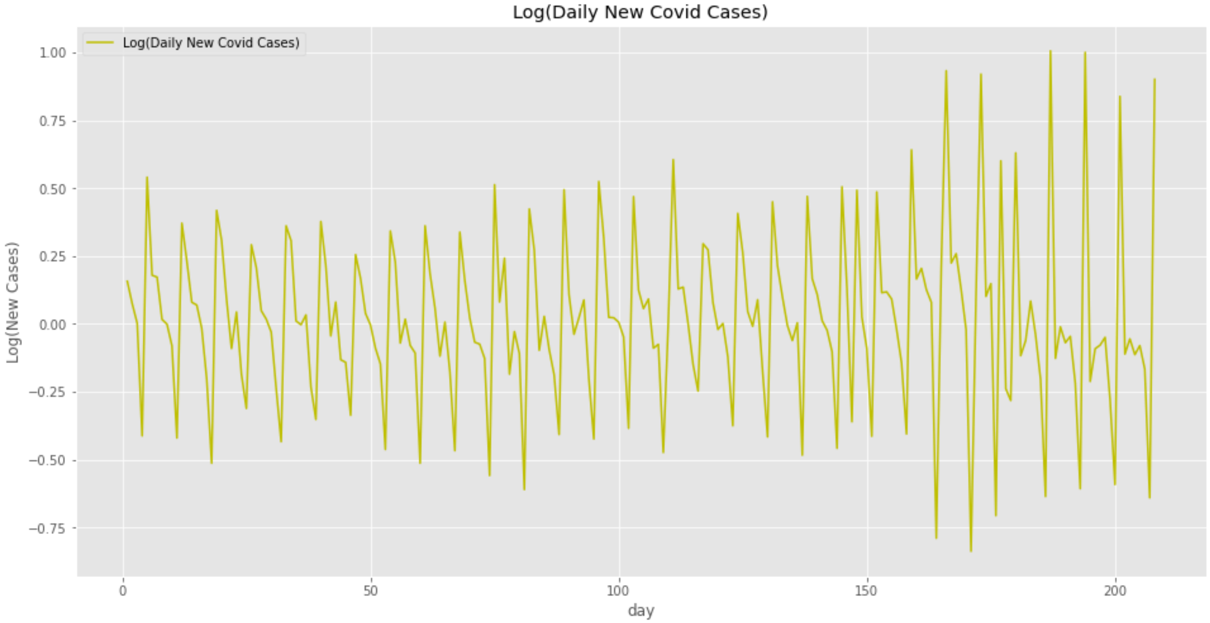
Data from mid 2021

Data Transformations: Daily New Covid Cases in Italy (2021)

Daily New Covid Cases



Diff(Log(Daily New Covid Cases))

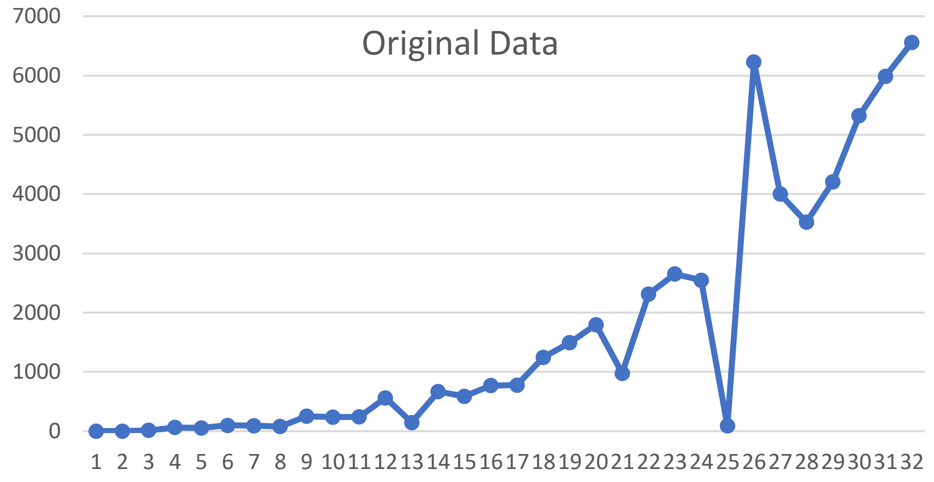


Short Case Study: Predicting daily covid cases

- The goal here is not to reach sophisticated predictions that can guide public policy but to perform a rapid analysis using actual data.
  - using what we learned in the course so far
- I’ll look at data from Italy (appears reliable)
  - Regulary updated
- This analysis looks at the beginning of the pandemic. Official data until March 21, 2020 (included).
- I start at the time of first positive case in Italy.

Predicting daily covid cases: Italy

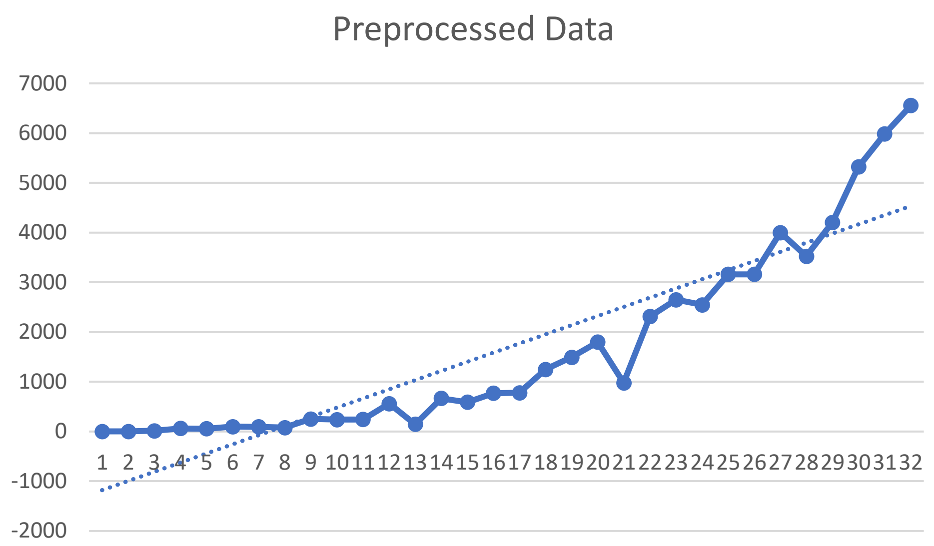
- 32 days of data.
- First some preprocessing:



- Apparently, some observation/reporting problem took place on day 25 where only 90 cases were reported. On day 24 there were 2547 cases and day 26 there were 6130 cases!
- I'll replace the data on day 25 by (2547+6130)/2, the average of the previous day and the next day.
  - There could be other ways of correcting. My choice of correction increases the total number of cases.

Predicting daily covid cases: Italy

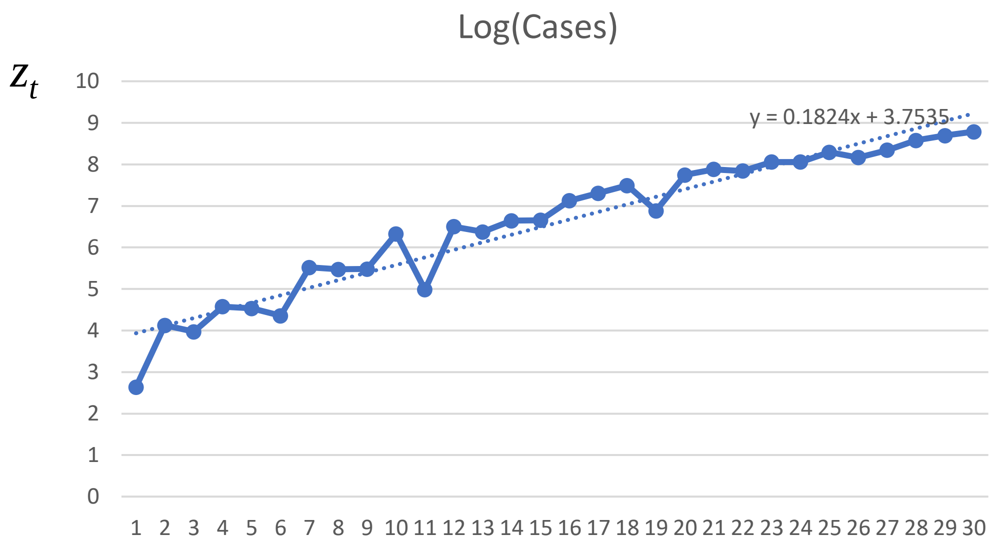
- This is the cleaned data.



- As we keep hearing, there is an exponential increase.
- Let’s do a log transformation.

Predicting daily covid cases: Italy

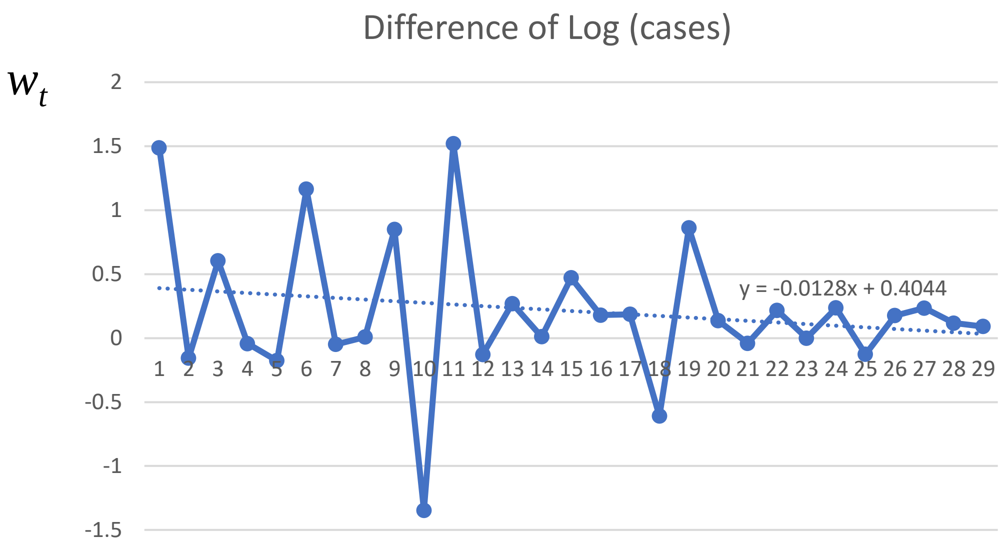
- $z_t = \ln(y_t)$



- There is a significant linear trend. This confirms the exponential growth of the original data. We'll take differences to remove it.

Predicting daily covid cases: Italy

- $w_t = z_t - z_{t-1}$



- There is a small decreasing linear trend after differencing. We can run a linear regression to decide whether that's statistically significant. If so, we might want to take a second difference.

Predicting daily covid cases: Italy

- We test the following linear model to see whether w_t has trend:

$$w_t = b_0 + b_1 t + \varepsilon_t$$

- The results are as follows:
- The slope is statistically insignificant (p-value->0.324).
- So first order differencing was sufficient to remove the trend.

Linear regression model:
 y ~ 1 + x1

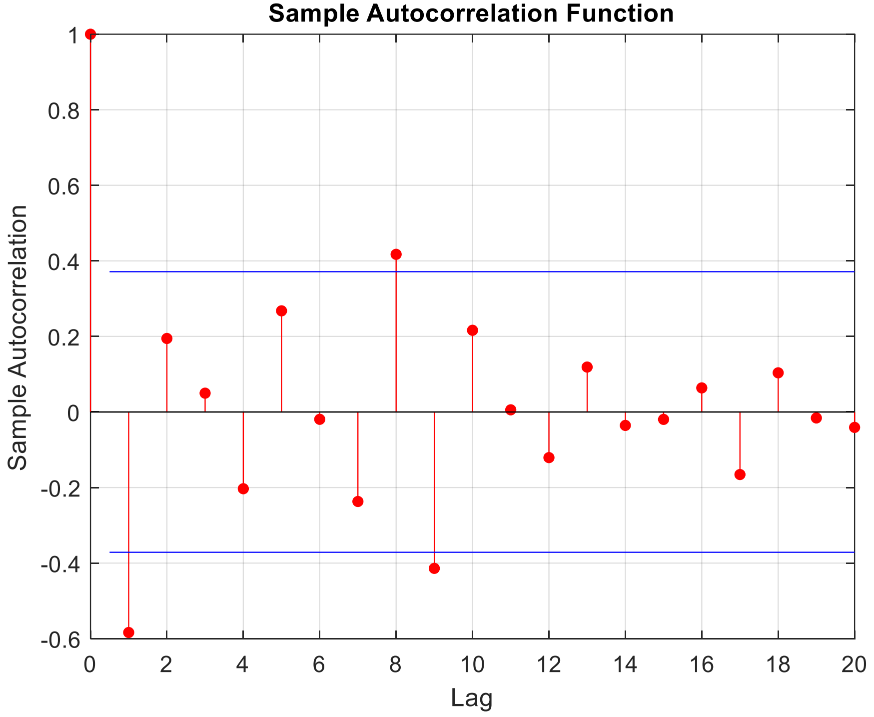
Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	0.4044	0.21934	1.8437	0.076231
x1	-0.012824	0.012771	-1.0042	0.32422

Number of observations: 29, Error degrees of freedom: 27
Root Mean Squared Error: 0.575
R-squared: 0.036, Adjusted R-Squared 0.000298
F-statistic vs. constant model: 1.01, p-value = 0.324

Predicting daily covid cases: Italy

- Finally, we can check the autocorrelation plot of w_t.



- There is some AC at lags 1, 8 and 9. But our data set is too small to fit an AR model so we'll ignore that and fit a simple model.

Predicting daily covid cases: Italy

- Proposed model:

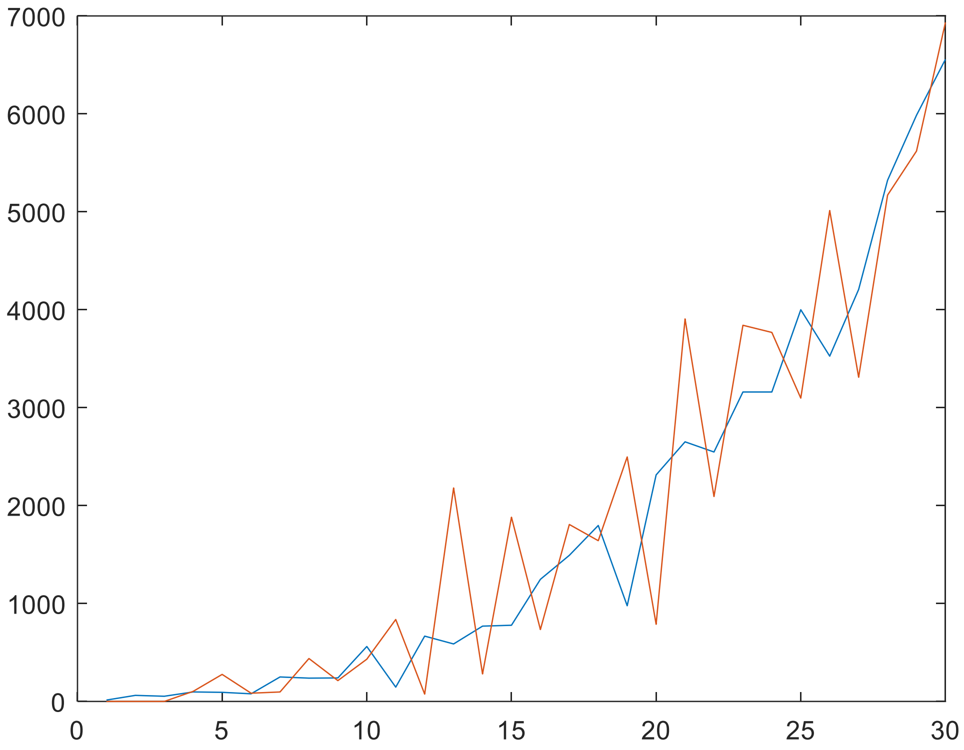
$$G_t = z_{t\text{-}1} + ((z_{t\text{-}1} - z_{t\text{-}2}) + (z_{t\text{-}3} - z_{t\text{-}4}))/2$$

$$F_t = e^{Gt}$$

- G_t estimates the current slope by averaging the last two independent differences and uses the last observation as an estimator of the current level.
- The final forecast F_t reverses the log transformation.
- Note that F_t is an unbiased estimator for the mean of: $y_t = e^{a+bt+\varepsilon t}$

Predicting daily covid cases: Italy

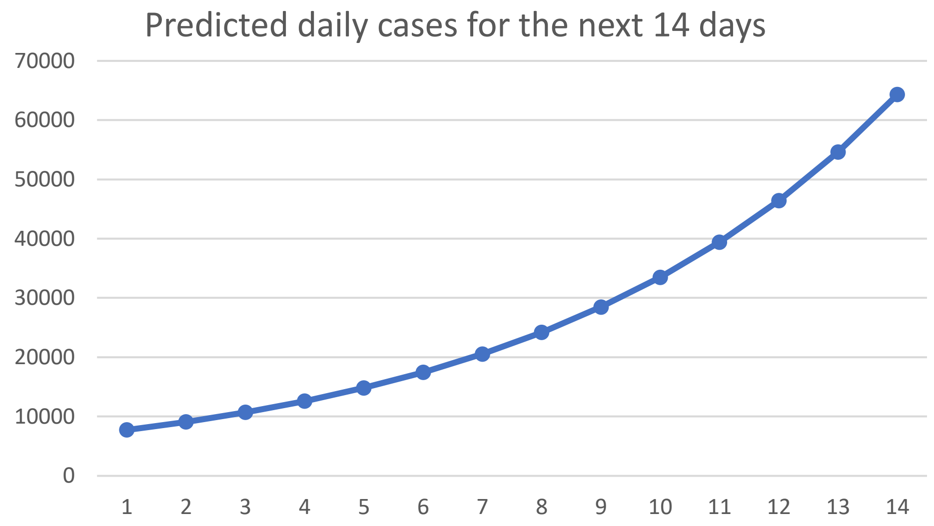
- Blue: observed new daily covid cases
- Orange: prediction from the model



Root Mean Squared Error = 627.983

Predicting daily covid cases: Italy

- We are hoping that the rate of infection growth slows soon. If it does not, our model predicts that in a week the new daily number of infections would be around 20000 and in two weeks they would be around 64000.



- “The lockdown should hopefully prove these predictions wrong.” was my comment from March 2020. (This turned out to be true for the first peak)