Differencing

Jared Fisher

Lecture 4a

Announcements

Announcements

- Homework 2: will be due on Wednesday, February 17 (tomorrow)
- Midterm 1: Thursday, February 25

Midterm

- Covers everything from weeks 0-4 (everything up to and including Friday's lab): lecture, lab, homework
- Open book, open note, open Google/internet
- Cannot ask questions or discuss with anyone other than course staff
- Exam is online on Gradescope: multiple choice, T/F, number input, short response, etc.
- Timed at 60 minutes
- Two time slots: during class (2:15-3:15ish) or in the evening 8:00-9:00pm. Sign up for your slot with the link provided on the announcement.

Brief Recap

CODE

Didn't get to look over R code last time

Definition: Linear Time Invariant Filter

A linear time-invariant filter with coefficients $\{a_j\}$ for $j = \ldots, -2, -1, 0, 1, 2, 3, \ldots$ transforms an input time series $\{U_t\}$ into an output time series $\{V_t\}$ via

$$V_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}.$$

In the above definition, the coefficients $\{a_j\}$ are often assumed to satisfy $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.

An equal-weighted, two-sided filter

Our smoother from last time is a special case of a linear smoother, where $a_j = \frac{1}{2q+1}$ for $j = -q, ..., 0, ...q$:

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t-j}$$

- (Sometimes called a “rectangular kernel” as weights are all a constant positive value or zero)
- (For a more general definition, see TSA4e 4.7)

Other Filter Choices

- Exponential Smoothing

$$\hat{m}_t = \frac{1-\alpha}{\alpha}\left[\alpha Y_t + \alpha^2 Y_{t-1} + \alpha^3 Y_{t-2} + \ldots\right]$$

- What constraint should there be on α?

Other Filter Choices

- Binomial Weights

$$a_j = 2^{-q} \binom{q}{q/2 + j} \quad \text{for } j \in -\frac{q}{2}, -\frac{q}{2} + 1, ..., -1, 0, 1, ..., \frac{q}{2}$$

- q=2

```
## [1] 0.25 0.50 0.25
```

- q=4

```
## [1] 0.0625 0.2500 0.3750 0.2500 0.0625
```

Differencing

Differencing

- Remember that our main goal is to mathematically manipulate the raw time series so that what remains can be assumed to be stationary.
- Until now we have been estimating the trend and seasonality through models or smoothing
- Instead, we can simply remove the trend and/or seasonality without estimating it by differencing!

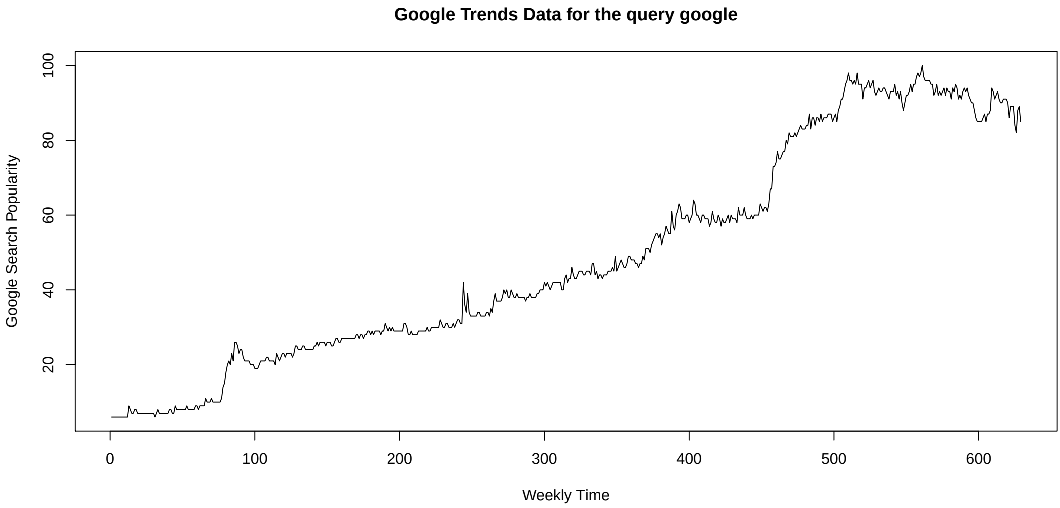
Differencing

- Assume a simple trend model $Y_t = m_t + W_t$, where m_t is the deterministic trend and W_t is white noise.
- When m_t is approximately linear, we can look at the differenced series instead:

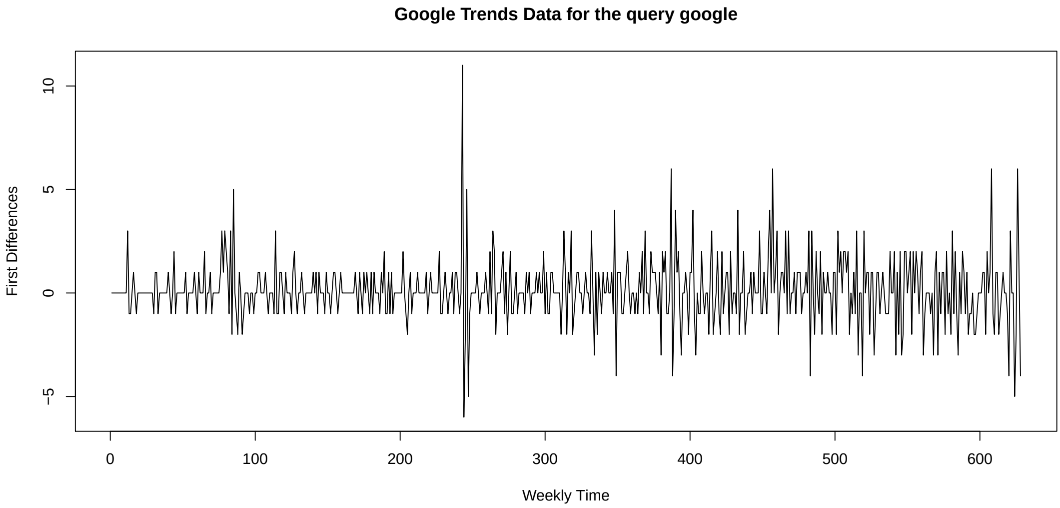
$$V_t = \nabla Y_t = Y_t - Y_{t-1} = \underbrace{m_t - m_{t-1}}_{\approx \text{const.}} + \underbrace{W_t - W_{t-1}}_{\text{no trend}}$$

- When m_t is approximately polynomial, consider higher ordered differences: $\nabla^d Y_t$.
- The connection: this is a filter with $a_0 = 1$ and $a_1 = -1$, 0 otherwise.

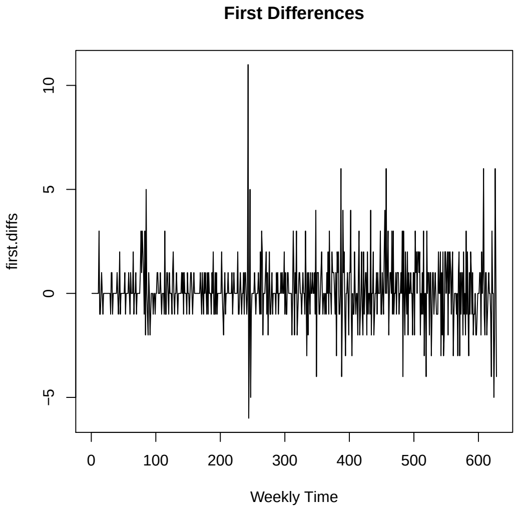
Recall Googling “Google”

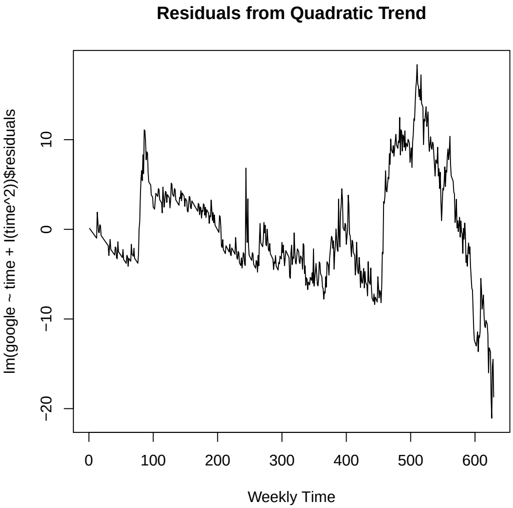


Google's First Differences

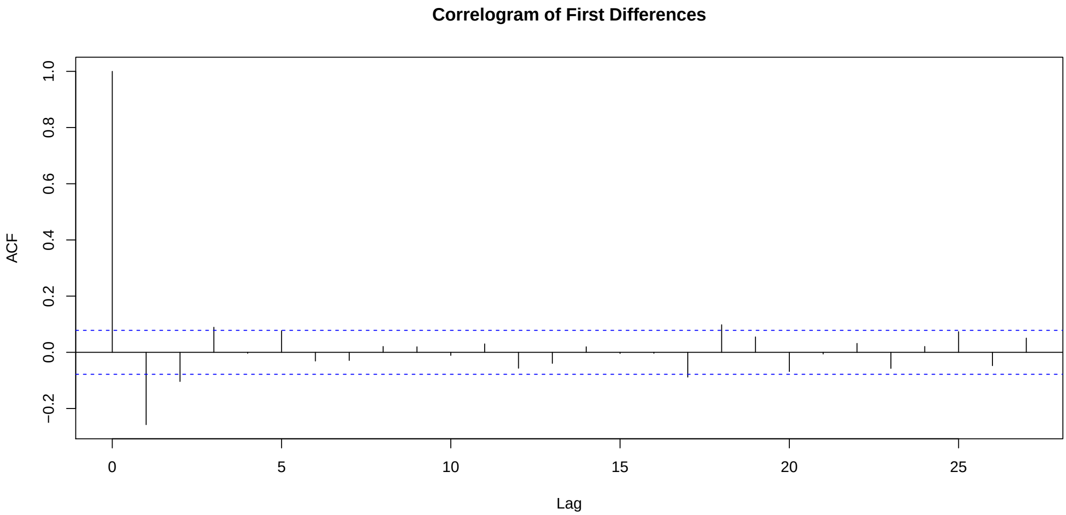


More stationary than quadratic model?





Google First Differences



What Happened?

If $Y_t = m_t + W_t$, and $m_t = \alpha + \beta t$

$$
\begin{aligned}
\nabla Y_t &= Y_t - Y_{t-1} \\
&= \alpha + \beta t + W_t - (\alpha + \beta(t-1) + W_{t-1}) \\
&= \beta + W_t - W_{t-1}
\end{aligned}
$$

Using Differences to Forecast

- As per the last slide, if $V_t = \nabla Y_t$ looks like white noise, but with nonzero mean, then that mean is our estimate for β!
- Then what is our forecast for Y_{t+1}, if we know $Y_1, ..., Y_t$?
- Recall $\nabla Y_t = Y_t - Y_{t-1}$, so $\nabla Y_{t+1} = Y_{t+1} - Y_t$
- $\Rightarrow Y_{t+1} = \nabla Y_{t+1} + Y_t$, so we need an estimate of ∇Y_{t+1}
- Using the equation on the last slide:

$$E(\nabla Y_{t+1}) = E(\beta + W_{t+1} - W_t) = \beta + 0 + 0 = \beta$$

- So, $E(Y_{t+1}) = Y_t + \beta$
- $\hat{Y}_{t+1} = Y_t + \bar{V}$, where $\bar{V}$ is the sample mean of the first differences

Higher Order Differences

What if there is a trend in ∇Y_t? We can try taking the differences again!

$$\begin{aligned}
\nabla^2 Y_t &= \nabla(\nabla Y_t) \\
&= \nabla(Y_t - Y_{t-1}) \\
&= \nabla(Y_t) - \nabla(Y_{t-1}) \\
&= Y_t - 2Y_{t-1} + Y_{t-2}
\end{aligned}$$

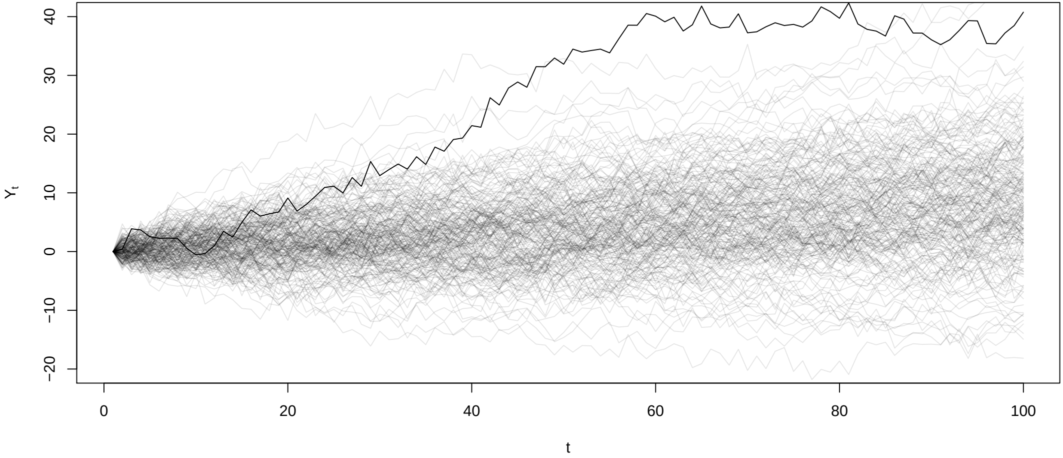
Higher Order Differences

$$
\begin{aligned}
\nabla^2 Y_t &= Y_t - 2Y_{t-1} + Y_{t-2} \\
\nabla^3 Y_t &= Y_t - 3Y_{t-1} + 3Y_{t-2} - Y_{t-3} \\
\nabla^4 Y_t &= Y_t - 4Y_{t-1} + 6Y_{t-2} - 4Y_{t-3} + Y_{t-4} \\
&\vdots
\end{aligned}
$$

- These remove quadratic trends, cubic trends, quartic trends, and so forth

An Example of "When would I use this?"

- For $Y_t = m_t + X_t$, say $m_t = m_{t-1} + \delta + W_t$ for $\delta \in R$ and W_t is white noise.
- Below are 200 examples, where a single possibility is highlighted

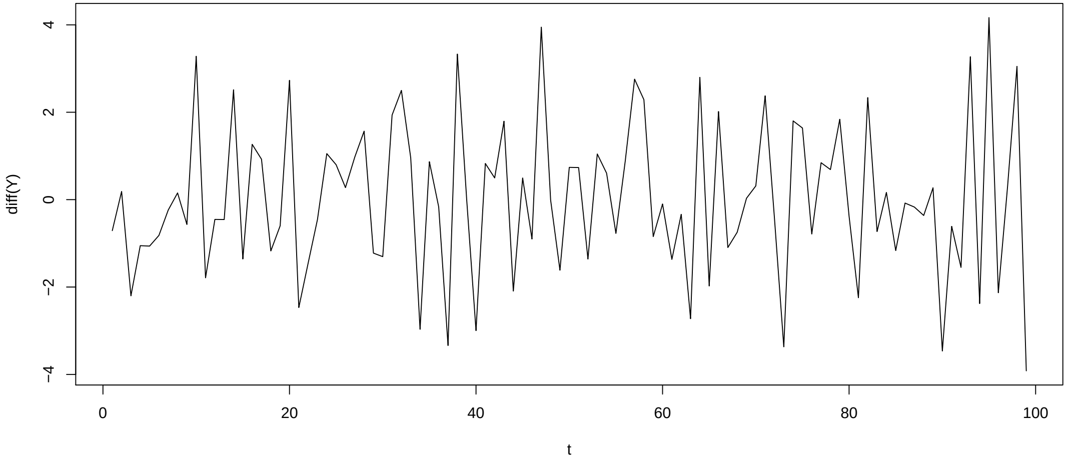


An Example of "When would I use this?"

- This is called a state space model, where the trend has a "random walk with drift",
- This can be a complicated model!
- But note that there's no trend (m_t) in the first difference, just constants and noise:

$$
\begin{aligned}
\nabla Y_t &= Y_t - Y_{t-1} \\
&= m_t + X_t - m_{t-1} - X_{t-1} \\
&= (m_{t-1} + \delta + W_t) + X_t - m_{t-1} - X_{t-1} \\
&= \delta + W_t + X_t - X_{t-1}
\end{aligned}
$$

First Differences



Differencing

- Useful for removing trends that are approximately linear (or polynomial for higher order differences)
- We do get forecasts!
- However, we get no estimate of the trend (because we removed it without estimating it!)

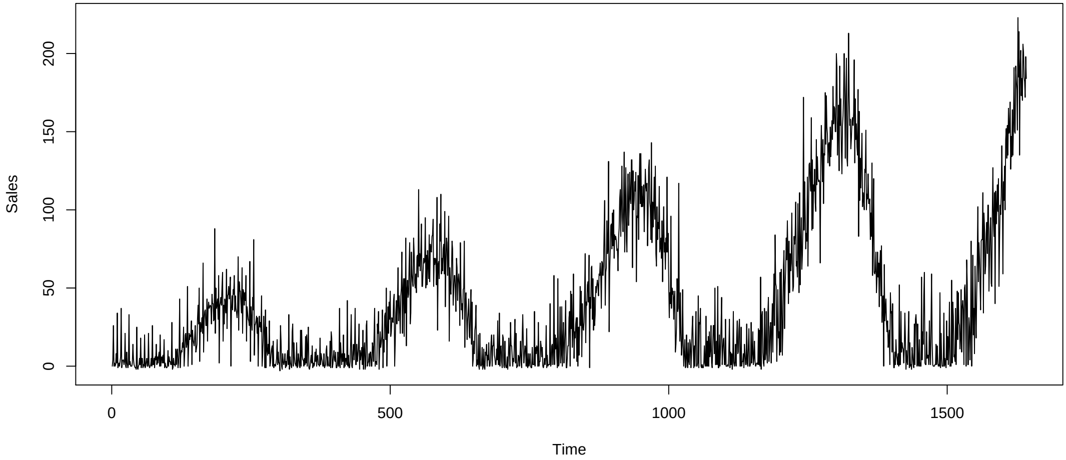
Seasonal Differencing

We can obtain residuals adjusted for seasonality without fitting a seasonality function by differencing:

$$\begin{aligned}\nabla_d Y_t &= Y_t - Y_{t-d} \\ &= s_t - s_{t-d} + X_t - X_{t-d} \\ &= X_t - X_{t-d}\end{aligned}$$

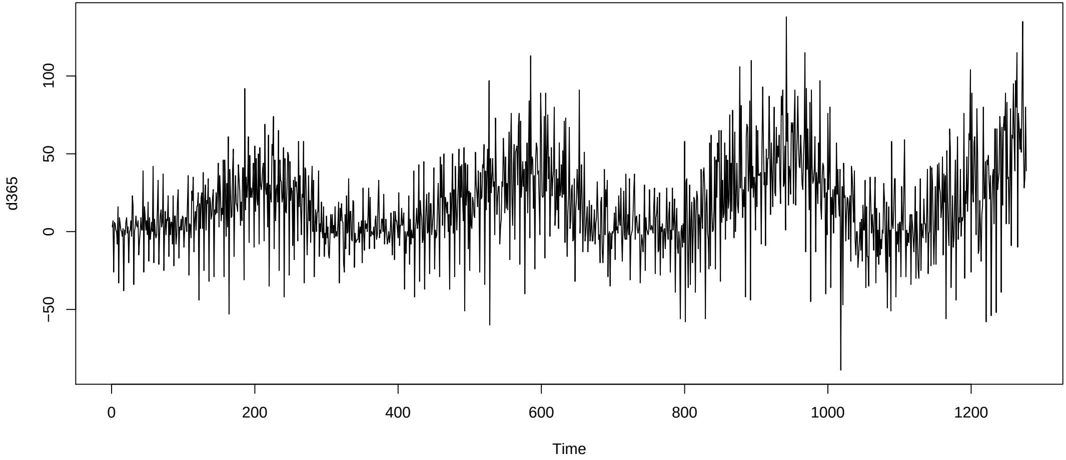
... because $s_t = s_{t-d}$!

Air Conditioning Sales



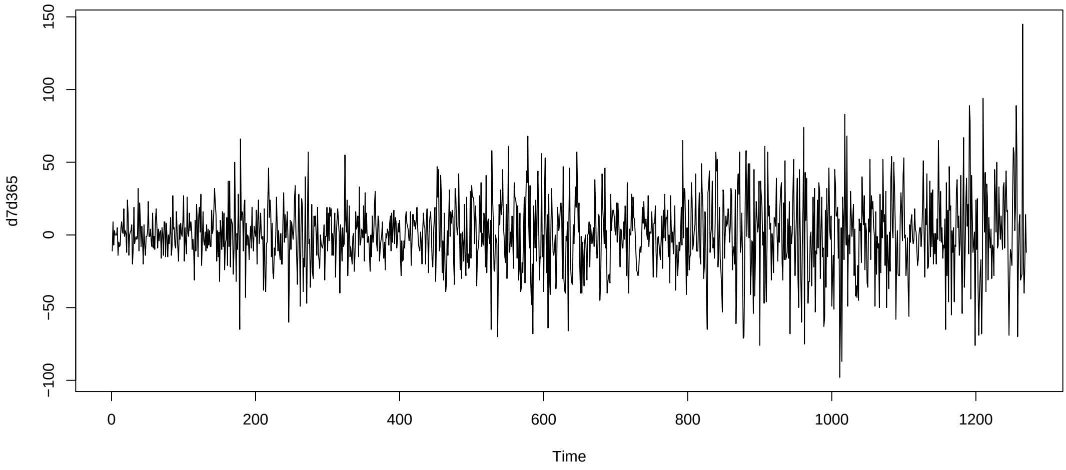
Air Conditioning Sales - $\nabla_{365} Y_t$

```r
d365 = diff(Sales,lag = 365)
plot.ts(d365)
```



Air Conditioning Sales - $\nabla_7 \nabla_{365} Y_t$

```r
d7d365 = diff(d365,lag = 7)
plot.ts(d7d365)
```



Improvement?

- While it appears that the variance fluctuates in year cycles, this is obviously much closer to stationary than what we started with!
- Note however that we lose data when we difference (365+7 datapoints to be precise)

```r
length(Sales)
```

```
## [1] 1642
```

```r
length(d7d365)
```

```
## [1] 1270
```

- This is okay for this problem though as we still have many years of data, and our primary interest is forecasting.

Forecasting with Seasonal Differences

- Let's assume $V_t = \nabla_7 \nabla_{365} Y_t$ shown previously is sufficiently stationary. How would we forecast Y_{t+1} given all data $Y_1, ..., Y_t$?
- Note

$$\begin{aligned}
\nabla_7 \nabla_{365} Y_t &= \nabla_7 (Y_t - Y_{t-365}) \\
&= (Y_t - Y_{t-365}) - (Y_{t-7} - Y_{t-365-7}) \\
&= Y_t - Y_{t-365} - Y_{t-7} + Y_{t-365-7} \\
&= Y_t - Y_{t-7} - Y_{t-365} + Y_{t-372}
\end{aligned}$$

Forecasting with Seasonal Differences

- Thus $\nabla_7\nabla_{365} Y_{t+1} = Y_{t+1} - Y_{t-6} - Y_{t-364} + Y_{t-371}$
- And $Y_{t+1} = Y_{t-6} + Y_{t-364} - Y_{t-371} + \nabla_7\nabla_{365} Y_{t+1}$
- Hence $E(Y_{t+1}|Y_1, ..., Y_t) = Y_{t-6} + Y_{t-364} - Y_{t-371} + E(\nabla_7\nabla_{365} Y_{t+1}|Y_1, ..., Y_t)$
- And we estimate
$$\hat{Y}_{t+1} = Y_{t-6} + Y_{t-364} - Y_{t-371} + \bar{V}$$
where $\bar{V}$ is the sample mean of $\nabla_7\nabla_{365} Y_{t+1}$