Models for Non-Stationary Time Series:

Transformations

Week VII: Video 21

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Introduction

In Videos 19 & 20, we've seen how *differencing* can be used to obtain a stationary process from a non-stationary one.

Differencing can be thought of as a type of *transformation* on the original process.

In this video, we'll learn about a few other useful transformations.

Video 21 Learning Objectives

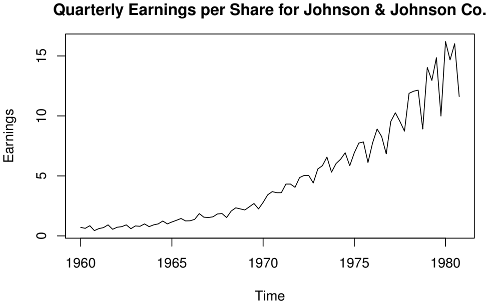
By the end of this video, we should be able to:

- Identify the motivation behind the *log-transformation*, and how it can help solve the problem of unequal variances
- Identify the cases in which the *difference of logs* may be a useful transformation, and how it can help us obtain stationarity
- Describe the *power transformation*, and choose a value of its parameter λ
- Implement each of the above transformations in R, and describe our findings

Log-Transformations: Motivation

In many real-life datasets, the variability of the time series increases as the mean of the process increases.

Example:

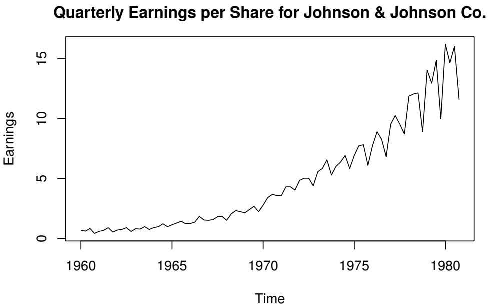



```
> data(JJ)
> plot(JJ, ylab='Earnings', xlab='Time', type='l', main='Quarterly
      Earnings per Share for Johnson & Johnson Co.')
```

Log-Transformations: Motivation

In many real-life datasets, the variability of the time series increases as the mean of the process increases.

Example:




```
> data(JJ)
> plot(JJ, ylab='Earnings', xlab='Time', type='l', main='Quarterly
      Earnings per Share for Johnson & Johnson Co.')
```

Taking the log-transformation of the data may be useful for stabilizing the variance.

Log-Transformations: What They Accomplish

Suppose that $Y_t > 0$ for all t, and that

$$E(Y_t) = \mu_t$$
$$\sqrt{Var(Y_t)} = \mu_t \sigma$$

In other words, the standard deviation of the process is proportional to the mean value of the process at any time t.

Log-Transformations: What They Accomplish

Suppose that $Y_t > 0$ for all t, and that

$$E(Y_t) = \mu_t$$
$$\sqrt{Var(Y_t)} = \mu_t \sigma$$

In other words, the standard deviation of the process is proportional to the mean value of the process at any time t.

Suppose that we take the (natural) logarithm of the data: $\log(Y_t)$. Note that the Taylor approximation of the function $\log(x)$ about a value c is:

$$\begin{aligned} \log(x) &= \log(c) + \frac{x-c}{1!}\left(\frac{d}{dx}\log(x)\right)\bigg|_{x=c} + \frac{(x-c)^2}{2!}\left(\frac{d^2}{dx^2}\log(x)\right)\bigg|_{x=c} + \dots \\ &\approx \log(c) + \frac{x-c}{c} \end{aligned}$$

So:

$$\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}$$

Log-Transformations: What They Accomplish (cont'd)

Then:

$$E[\log(Y_t)] \approx E\left[\log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}\right] = \log(\mu_t) + \frac{1}{\mu_t}[E(Y_t) - \mu_t] = \log(\mu_t)$$

Log-Transformations: What They Accomplish (cont'd)

Then:

$$E[\log(Y_t)] \approx E\left[\log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}\right] = \log(\mu_t) + \frac{1}{\mu_t}[E(Y_t) - \mu_t] = \log(\mu_t)$$

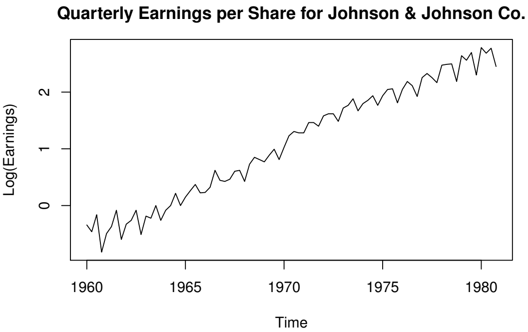
and

$$Var[\log(Y_t)] \approx Var\left[\log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}\right] = \frac{1}{\mu_t^2} Var(Y_t) = \frac{\mu_t^2 \sigma^2}{\mu_t^2} = \sigma^2$$

So, transforming the series has resulted in an approximately constant variance over time!

Log-Transformations: Example

Example: Let's take the logs of the Johnson & Johnson data:




```
> plot(log(JJ), ylab='Log(Earnings)', xlab='Time', type='l',
      main='Quarterly Earnings per Share for Johnson & Johnson Co.')
```

We can see that transforming the series has resulted in an approximately constant variance over time!

Differences of Logs: Motivation

Suppose that $\{Y_t\}$ is non-stationary, but its *percentage change* is some well-behaved, stationary process:

$$Y_t = (1 + X_t)Y_{t-1}$$

where $100X_t$ is the percentage change from Y_{t-1} to Y_t.

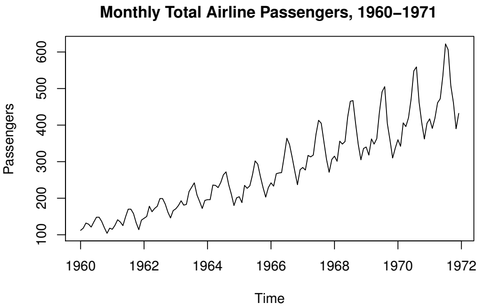
Differences of Logs: Motivation

Suppose that $\{Y_t\}$ is non-stationary, but its *percentage change* is some well-behaved, stationary process:

$$Y_t = (1 + X_t)Y_{t-1}$$

where $100X_t$ is the percentage change from Y_{t-1} to Y_t.

Example:




```
> data(airpass)
> plot(airpass, ylab='Passengers', xlab='Time', type='l', main='Monthly
      Total Air Passengers, 1960-1971')
```

Differences of Logs: What They Accomplish

Now:

$$
\begin{aligned}
Y_t &= (1 + X_t) Y_{t-1} \\
\frac{Y_t}{Y_{t-1}} &= 1 + X_t \\
\log\left(\frac{Y_t}{Y_{t-1}}\right) &= \log(1 + X_t) \\
\log(Y_t) - \log(Y_{t-1}) &= \log(1 + X_t) \\
\nabla[\log(Y_t)] &= \log(1 + X_t)
\end{aligned}
$$

Differences of Logs: What They Accomplish

Now:

$$\begin{aligned} Y_t &= (1 + X_t) Y_{t-1} \\ \frac{Y_t}{Y_{t-1}} &= 1 + X_t \\ \log\left(\frac{Y_t}{Y_{t-1}}\right) &= \log(1 + X_t) \\ \log(Y_t) - \log(Y_{t-1}) &= \log(1 + X_t) \\ \nabla[\log(Y_t)] &= \log(1 + X_t) \end{aligned}$$

Suppose that the percentage changes are not too large, e.g. at most $\pm 20\%$. Then, $|X_t| < 0.2$, and so: $\log(1 + X_t) \approx X_t$.

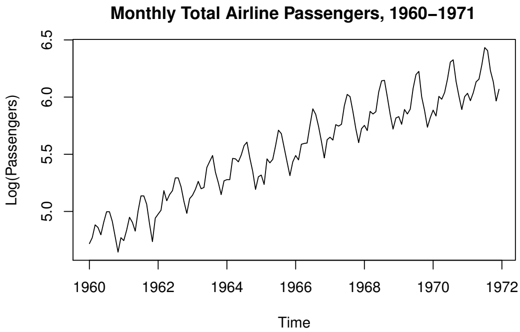
Therefore:

$$\nabla[\log(Y_t)] \approx X_t$$

So, the differences of the logs of the data might be well-modelled by a stationary process. This is a very commonly used approach!

Differences of Logs: Example

Example: Let's take the logs of the airline passenger data:

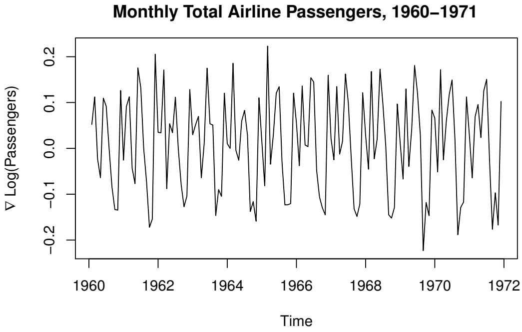



```
> plot(log(airpass), ylab='Log(Passengers)', xlab='Time', type='l',
      main='Monthly Total Air Passengers, 1960-1971')
```

We can see that logging the series has resulted in an approximately constant variance over time.

Differences of Logs: Example (cont'd)

Example: Let's take the differences of the logs of the airline passenger data:




```
> plot(diff(log(airpass)),
      ylab=expression(paste(nabla,'Log(Passengers)')), xlab='Time',
      type='l', main='Monthly Total Air Passengers, 1960-1971')
```

We can see that the differenced log of the series looks a lot more stationary.
(Note: There may still be a seasonal effect here that needs to be accounted for.)

Comment on Log-Transformations

We have made the assumption that $Y_t > 0$ for all t. Otherwise, the logs will be undefined.

This assumption is usually OK, since we most commonly use log-transformations on Poisson-like data (count data), which is > 0.

However, if we wish to use log-transformations on data where some values are negative or zero, we can:

1. Add a positive constant c to each value of Y_t, such that they all become positive
2. Subtract c from the resulting fitted values, to obtain our estimate of $\hat{Y}_t$

Power Transformations

A more general family of transformations, which *includes* the log-transformation, is the **power transformation** family.

It is defined as follows, for some fixed value of the parameter λ:

$$\text{Power Transformation of } Y_t : \begin{cases} \frac{(Y_t)^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(Y_t) & \text{for } \lambda = 0 \end{cases}$$

Power Transformations

A more general family of transformations, which *includes* the log-transformation, is the **power transformation** family.

It is defined as follows, for some fixed value of the parameter λ:

$$\text{Power Transformation of } Y_t : \begin{cases} \frac{(Y_t)^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(Y_t) & \text{for } \lambda = 0 \end{cases}$$

Note that: $\lim_{\lambda \to 0} \left(\frac{x^\lambda - 1}{\lambda} \right) = \log(x)$, so the function changes smoothly as λ approaches 0.

Some other special cases:

- For $\lambda = \frac{1}{2}$: Square-root transformation (often useful with Poisson-like data, i.e. count data)
- For $\lambda = -1$: Reciprocal transformation

Note: Just as for log-transformations, we are assuming that $Y_t > 0$ for all t.

Power Transformations: Choosing λ

$$\text{Power Transformation of } Y_t : \begin{cases} \frac{(Y_t)^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(Y_t) & \text{for } \lambda = 0 \end{cases}$$

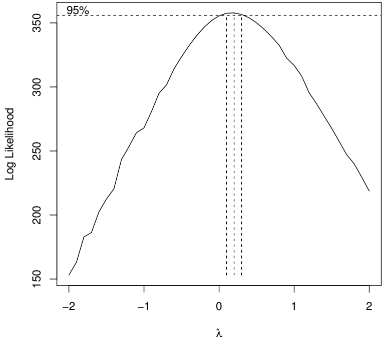
In most cases, we don't need precise estimation of λ.

R allows us to try out a range of values, and evaluates the log-likelihood for the data at each value of λ.

Larger values of the log-likelihood imply that the corresponding value of λ "matches" the data well.

Power Transformations: Choosing λ (cont'd)

Example: For the Johnson & Johnson earnings dataset:




```
> lambda.estimation <- BoxCox.ar(JJ)
> lambda.estimation$lambda[which.max(lambda.estimation$loglike)]
0.2
```

The maximum likelihood occurs at $\lambda = 0.2$, suggesting that the best power transformation might be $\frac{(Y_t)^{0.2}-1}{0.2}$.

Final Comments

That's all for now!

In this video, we've seen several different transformations that can help us obtain a stationary process from a non-stationary one: logs, differences of logs, and power transformations.

We've learned the mathematical motivations behind each of these transformations, and seen some examples of how to implement the transformations in R.

Coming Up Next: Constant terms in ARIMA models.

Thank you!

References:

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

[2] Chan, K. S., & Ripley, B. (2020). TSA: Time Series Analysis. R package version 1.2.1.