Model Specification:

Non-stationarity, and Other Specification Methods

Week X: Video 28

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Video 28 Learning Objectives

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

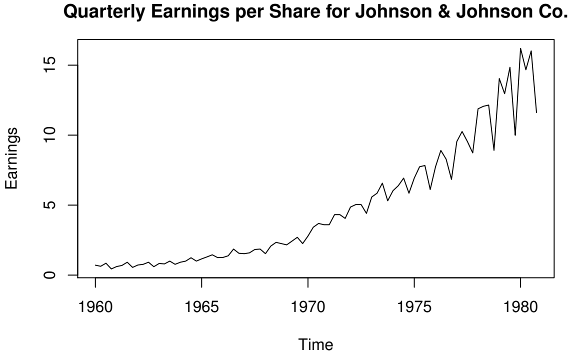
- Explain how to detect and handle non-stationarity during the model specification step
- Describe the potential negative effects of *overdifferencing* a dataset
- Understand how to use AIC and BIC for model specification
- Interpret a best subset ARMA plot to choose an appropriate subset ARMA model

Non-Stationarity: Motivation

In Videos 23-25, we learned how to identify various ARMA models using the sample ACF plot, sample PACF plot, and sample EACF table.

What if our data comes from a *non-stationary* model?

Example:



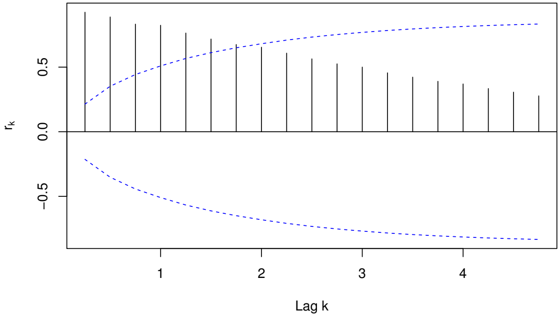
Non-Stationarity: Identification

Non-stationarity will often show up in our model specification plots.

Since non-stationary models often tend to "drift" up or down, the observations are usually highly correlated even for large lags.

So, non-stationary series will usually be identifiable by their very large sample ACF values at high lags. Usually, the sample ACF decreases *linearly* with time.

Example: For the JJ dataset:



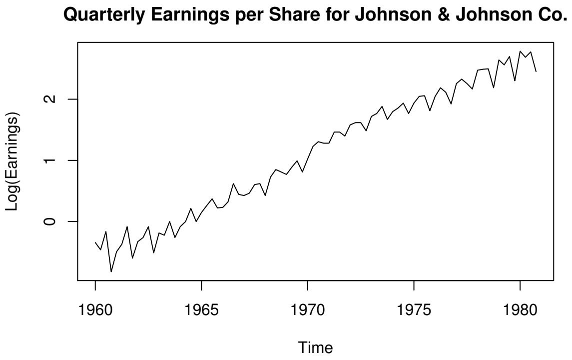
Non-Stationarity: Differencing and/or Transformations

If we identify a non-stationary model (using either its sample ACF plot, or the plot of Y_t vs. t), we can apply our usual transformations to make it stationary:

- Differencing
- Log-transformation
- Power transformation
- A combination of the above (e.g., difference of logs)

Non-Stationarity: Differencing and/or Transformations (cont'd)

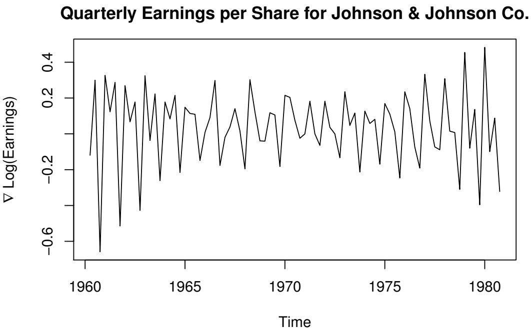
Example: Logs of the JJ data:



This has improved the non-stationarity in the variances.

Non-Stationarity: Differencing and/or Transformations (cont'd)

Example: Differences of the logs of the `JJ` data:

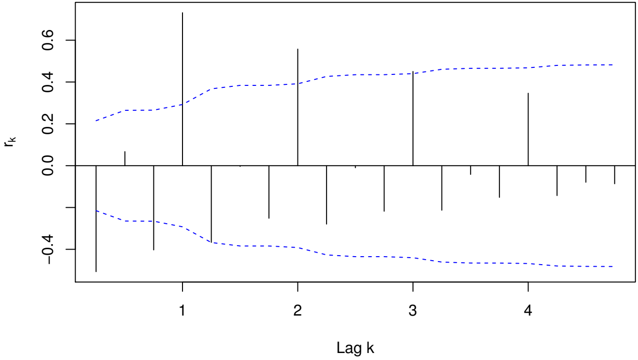


This has removed the linear trend. We now have a process that appears *somewhat* stationary (although there is still some non-constant variance remaining).

Can this process be adequately modelled using an ARMA model?

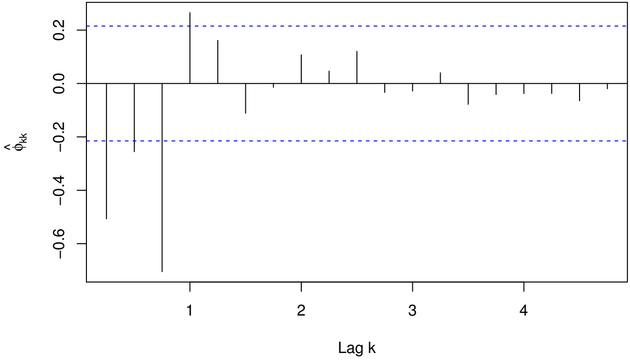
Non-Stationarity: Differencing and/or Transformations (cont'd)

Example: Model specification for the ∇Log(JJ) data:



Non-Stationarity: Differencing and/or Transformations (cont'd)

Example: Model specification for the ∇Log(JJ) data (cont'd):



Non-Stationarity: Overdifferencing

If we choose to difference a time series dataset in an attempt to make it "more stationary", we must be careful not to **overdifference**.

Example: Suppose our dataset comes from a random walk model:

$$Y_t = Y_{t-1} + e_t$$

We know from Ch. 2 that this is a non-stationary model.

Non-Stationarity: Overdifferencing

If we choose to difference a time series dataset in an attempt to make it "more stationary", we must be careful not to **overdifference**.

Example: Suppose our dataset comes from a random walk model:

$$Y_t = Y_{t-1} + e_t$$

We know from Ch. 2 that this is a non-stationary model.

However, its first difference is stationary:

$$W_t = \nabla Y_t = Y_t - Y_{t-1} = e_t$$

The first difference is just a white noise, which is ARMA(0,0).

Non-Stationarity: Overdifferencing (cont'd)

Example: *(cont'd)*

However, suppose we were to take the second difference:

$$W_t' = \nabla^2 Y_t = e_t - e_{t-1}$$

This is now an MA(1) model with the parameter $\theta = 1$.

We now have to estimate an extra parameter! *And*, the model is non-invertible, which makes estimating the parameter difficult.

Non-Stationarity: Overdifferencing (cont'd)

Example: *(cont'd)*

However, suppose we were to take the second difference:

$$W_t' = \nabla^2 Y_t = e_t - e_{t-1}$$

This is now an MA(1) model with the parameter $\theta = 1$.

We now have to estimate an extra parameter! *And*, the model is non-invertible, which makes estimating the parameter difficult.

Therefore, when deciding to difference a dataset, we should:

- Take one difference at a time, carefully looking at the plot of Y_t vs. t, sample ACF plot, etc., each time
- Keep the *principle of parsimony* in mind: "Models should be simple, but not too simple"

Non-Stationarity: Dicky Fuller Unit-Root Test

The **Dicky Fuller Unit-Root Test** is a formal hypothesis test for deciding whether or not a process is difference non-stationary.

- A process is **difference non-stationary** if it is non-stationary, but its first difference *is* stationary.

Hypotheses:

$$\text{H}_0 : Y_t \text{ is difference non-stationary} \quad \text{vs.} \quad \text{H}_a : Y_t \text{ is stationary}$$

More information is on pg. 128-130 of the textbook (§ 6.4).

R code: The function `ADF.test()` in the `uroot` package (examples on pg. 130).

Model Specification: Other Methods

We've seen how the sample ACF plot, sample PACF plot, and sample EACF table, are useful methods for specifying an ARMA model.

What if we are having a difficult time choosing between different plausible models?

Model Specification: Other Methods

We've seen how the sample ACF plot, sample PACF plot, and sample EACF table, are useful methods for specifying an ARMA model.

What if we are having a difficult time choosing between different plausible models?

There are several "model-specification criteria" that are useful for deciding between models in general (not just for time series data).

We will learn two today:

- Akaike's Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

Model Specification with AIC

Akaike's Information Criterion (AIC) says to select the model that minimizes

$$\text{AIC} = -2\log(\text{maximum likelihood of the data}) + 2k$$

where k is the number of estimated parameters in the model.

For an ARMA model: $k = p + q$ (or $p + q + 1$ if there is also a constant trend term).

Model Specification with AIC

Akaike's Information Criterion (AIC) says to select the model that minimizes

$$\text{AIC} = -2\log(\text{maximum likelihood of the data}) + 2k$$

where k is the number of estimated parameters in the model.

For an ARMA model: $k = p + q$ (or $p + q + 1$ if there is also a constant trend term).

The likelihood of the data is the joint pdf of the data, for any given set of parameters.

- So, the "maximum likelihood of the data" is the joint pdf of the data evaluated at the MLEs of the parameters.

This criterion achieves two things:

- Favours models that result in a larger (max.) likelihood of the data
- Penalizes models that have too many parameters

Model Specification with BIC

The Schwarz **Bayesian Information Criterion (BIC)** says to select the model that minimizes

$$\text{BIC} = -2\log(\text{maximum likelihood of the data}) + k\log(n)$$

where k is the number of estimated parameters in the model, and n is the sample size.

This criterion works in a similar way to the AIC, but it has some different properties. For example, it is *consistent* (i.e., the values of p and q it chooses are expected to be equal to the true values, for large enough n).

Best Subset ARMA Selection

What if some of the lower-order ARMA coefficients are equal to zero?

A **subset ARMA(p,q) model** is an ARMA(p,q) model with a subset of its coefficients known to be zero.

Example: A subset ARMA(12,12) model, with all but the order-12 coefficients equal to zero:

$$Y_t = 0.9Y_{t-12} + e_t + 0.4e_{t-12}$$

This type of model may arise from seasonal data, for instance.

However, it may difficult to detect using the sample ACF/PACF/ EACF.

Best Subset ARMA Selection (cont'd)

One way to detect a subset ARMA(p,q) model is by evaluating the BIC (or AIC) for every possible model!

Since this amounts to many, many different models, it is best to plot the several best ones.

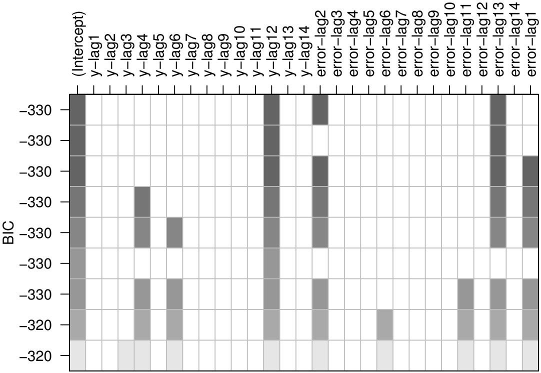
Example: Suppose our data is from $Y_t = 0.9Y_{t-12} + e_t + 0.4e_{t-12}$:

```
# Generate this subset ARMA(12,12) dataset:
set.seed(123123)
phi.vec   <- c(rep(0,times=11), 0.9)
theta.vec <- c(rep(0,times=11), 0.4)
my.data <- arima.sim(model=list(ar=phi.vec, ma=theta.vec),
                     n=200)

# Select and evaluate the best subset ARMA models:
armasubsets.select <- armasubsets(y=my.data, nar=14, nma=14,
                                  y.name='y')

# Plot the subset ARMA selection:
plot(armasubsets.select)
```

Best Subset ARMA Selection (cont'd)



Each row represents one subset ARMA model. The coloured boxes indicate the terms that are included in this model. The models are sorted according to their BIC values, where the lower (i.e., better) values are closer to the top.

Here, we see that the best model is the one with Y_{12}, e_2 and e_{13}. The second-best model is the one with just Y_{12} and e_{13}.

Final Comments

That's all for now!

In this video, we've learned how to detect and handle non-stationarity in model specification.

We also learned about some more useful model specification criteria, including AIC and BIC, as well as best subset ARMA selection.

Next Week in STAT 485/685: The second step in the Model Building Process: *parameter estimation*.

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.