**More Time Series Analysis!**

**Threshold Models**

A more comprehensive account can be found in Shumway and Stoffer’s textbook, Howell Tong’s book “Threshold Models in Non-linear Time Series Analysis”, and Tong’s book “Non-linear Time Series: A Dynamical System Approach.”

Previously, we have fit models for all t = 1, …, n. This is o.k. to do as long as the dynamics of the “system” from which the data came from does not change. If it does change, “local” models can be fit to the data.

Threshold autoregressive models

Multiple AR(p) models can be fit to segments of the data set. These segments are decided upon by past observations, say xt-1. If xt-1 reaches a “threshold” (i.e., maybe xt-1 ≥ 0.05), a different AR(p) model is fit. There can be multiple thresholds.

Divide the data up into r mutually exclusive and exhaustive regions denoted by R1,…,Rr.

The threshold autoregressive model is r different AR(p) models:

xt = α(j) +  for xt∈Rj for j = 1, .., r

where  ~ ind. (0,)

Notes:

* r = 1 is the “usual” AR(p) model.
* Model estimation, identification, and diagnostics can be done in a similar manner as for r = 1.

Example: Pneumonia and influenza data from Shumway and Stoffer

The data represents the number of United States pneumonia and influenza deaths per 10,000 people from 1968 to 1978. Below is a plot of the first differences with a horizontal line drawn at 0.05.



Shumway and Stoffer investigate the following models where xt denotes the first differences:



with p set to 6.

**Intervention analysis**

A more comprehensive reference is Wei’s textbook.

If you examine this book, please note the different notation that is used:

|  |  |
| --- | --- |
| **Wei** | **My notes** |
| zt and assume μ=0 | xt and assume μ=0 |
| at | wt |
| MA: (1-θ1B) | MA: (1+θ1B) |

We want to be able to account for special events (interventions) when modeling time series data.

Example:

1. OSU started at new campus in Tulsa during the time the enrollment data was collected.
2. A special promotion for a product may have an effect on sales.
3. A new law causes a change in the number of deaths from car accidents or the amount of pollution in the air.

Suppose an intervention occurred at time T. Is there evidence of a change in xt?

To determine change, the mean value of xt for t = 1, …,T-1 and t = T,…,n may be examined. Denote the mean for the first time period as μ1 and for the second time period as μ2. To determine differences of means, a simple two sample t-test could be used. Unfortunately, the “independence” assumption of this type of test is violated and the test can not be used.

Intervention analysis can be used instead.

Types of intervention variables

1. Step function



1. Pulse function



Note that . Therefore, we can always write a pulse function in terms of step functions. Often the notation, It, is used in general to denote an intervention. “I” stands for indicator variable.

The intervention variables are often referred to as "inputs". The responses to intervention variables are often referred to as "outputs".

Possible responses to interventions

1. A fixed unknown impact ω of an intervention is felt b periods after the intervention:

 or 

Example: Let T = 5, , b = 1, and ω = 3. Then the response to the intervention is  Over time, this is what happens:

| t |  |  |
| --- | --- | --- |
| 1 | 0 |  |
| 2 | 0 |  = 0 |
| 3 | 0 |  = 0 |
| 4 | 0 |  = 0 |
| 5 | 1 |  = 0 |
| 6 | 1 |  = 3 |
| 7 | 1 |  = 3 |
|  |  |  |

1. An impact of an intervention is felt b periods after the intervention, but the response is gradual:

 or  where 0 ≤ δ ≤ 1.

Note that

 = 



For δ = 0, the above equations simplify to what is in #1. For δ = 1, the impact increases without bound.

Example: Let T = 5, , b = 1, ω = 3, and δ =0.5. Over time, this is what happens:

| t |  |  |
| --- | --- | --- |
| 1 | 0 |  |
| 2 | 0 | 0 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 1 | 0 |
| 6 | 1 |  = 3 |
| 7 | 1 |  = 4.5 |
|  |  |  |

A figure from Box et al.’s textbook:



Multiple interventions

The response can be written more generally as



where ω(B) = (ω0 - ω1B - - ωsBs) and

 δ(B) = (1 - δ1B - - δrBr)

The ωi's represent the expected initial effects of the intervention. The δj's represent the permanent effects of the intervention, and the roots of δ(B) = 0 are assumed to be on or outside the unit circle.

For multiple interventions, the following model below can be used.

Suppose we have the following ARIMA model . This can be rewritten as



Incorporating multiple interventions produces



where Ijt for j=1, …, K are the intervention variables.

Notes:

1. The seasonal part of a process could also be incorporated into the above.
2. The process without the intervention, , is often called the "noise process" and is denoted by nt.
3. The "noise model" is often found based on the time series before the intervention occurs.
4. Hypothesis tests for the parameters associated with the interventions can be performed to determine if the intervention variable is needed. For example, testing an ω could be done to determine if a new law is effective.

Modeling

Frequently, an ARIMA model is found for the data before the intervention. This model is then fit to the entire data including a regression independent variable to account for the intervention.

In the simplest setting, suppose there are n = 100 observations and an intervention occurs at t = 80. The model with the intervention variable can be fit to all of the observations through using the arima()function by specifying the ARIMA part of the model in the usual way and using the xreg option to include the intervention variable. For example, a simple step invention can be created as

I <- c(rep(0, 79), rep(1, 21))

Outliers

One way to handle outliers is to use an intervention variable.

**Transfer function models**

More comprehensive references are Shumway and Stoffer and Box et al.’s textbooks.

Explanatory variables may influence a time series variable. Transfer function models allow the use of other times series variables to help explain the variability in the time series variable. These models can be thought of as an alternative to regression models with ARMA errors.

Examples:

1. The number of Oklahoma high school graduates could be used to help model OSU enrollment.
2. Advertising expenditures could be used to help model sales.
3. All intervention analysis examples.

General concepts

Let xt denote the value of an input series at time t.

Let yt denote the value of an output series at time t.

Assume that both are stationary and suppose that x influences y.

Typically, influences other than x will affect y. The combined effect on y of these influences is called the “noise” and is denoted by nt at time t (nt is assumed to be stationary).

The transfer function model which relates xt, yt, and nt is:

yt = ν(B)xt + nt

where ν(B) = ν0 + ν1B + ν2B2 + … =  is the transfer function

Notes:

1. xt is assumed independent of nt.
2. xt and nt are assumed to follow an ARMA model.
3. The weights, ν0,ν1,ν2, …, are often called the impulse response weights.
4. Purpose of transfer function modeling is to estimate ν0, ν1, ν2, … and find the model for nt based on the information in xt and yt.

Transfer function

The transfer function can be rewritten as



where

ωs(B) = ω0 - ω1B - ω2B2 - … - ωsBs

δr(B) = 1 - δ1B - δ2B2 - … δrBr (roots outside unit

 circle for δr(B)=0)

b = delay before the input variable (x) produces an effect on the output variable (y).

Compare this representation of the transfer function to the responses to interventions.

Without going into all of the details (and there is a lot), the cross correlation function helps determine the values of r, s, and b.

Example: Gas furnace data

This example comes from Box et al.’s textbook.

xt = input gas rate (transformed for this example)

yt = CO2 output

xt is measured in cubic foot per minute and yt is measured as percentage of output.

From the textbook,



The data is observed in 9 second intervals. Below is the final model used:



The best ARIMA model alone for Yt appears to be ARIMA(3,0,2).

Below are plots of the data along with forecasts.



