Back transformation theory

Willem Vervoort, Michaela Dolk & Floris van Ogtrop

2020-04-15

```
# root dir
knitr::opts_knit$set(root.dir =
                        "D:/Cloudstor/Virtual Experiments/VirtExp")
                        #"C:/Users/rver4657/ownCloud/Virtual Experiments/VirtExp")
knitr::opts_chunk$set(echo = TRUE)
# LOAD REQUIRED PACKAGES # #####
library(pander)
library(tidyr)
library(zoo)
library(mgcv)
library(ggplot2)
```

This rmarkdown document and the resulting pdf are stored on github. All directories (apart from the root working directory) refer to the directories in this repository.

Introduction

This document is related to the manuscript "Disentangling climate change trends in Australian streamflow" (vervoort et al.), submitted.

Back transformation

This document is related to the *3.GAMmodelTest.pdf* document and is to explain the back transformation used in the `gamm()` and `gls()` regressions, which are essentially the following type of regression equation
$log(y) = \beta_1 * trend$
How do we interpret β_1?
Assume a 1 unit change in trend in time, this results in:
$log(y_{t=1}) - log(y_{t=0}) = \beta_1$ or
$log(\frac{y_{t=1}}{y_{t=0}}) = \beta_1$
raising both sides to a power results in:
$\frac{y_{t=1}}{y_{t=0}} = exp(\beta_1)$

According to this link, we need to subtract 1 (for log(y + 1) = 0, presumably) to get the ratio

In other words, $exp(\beta_1) - 1$ is equal to the ratio of the y values, or can be interpreted as a fractional change in y. The reason for substracting one is that we are talking about a 1 unit increase in the trend results in an $exp(\beta_1)$ change in y, so to get to fractional change we need to subtract 1

The addition of a simple scalar to y (1 in this case) makes no difference on the trend interpretation. This simply means we are now looking at a change in y+1, which is essentially the same as the change in y as the addition is a scalar.

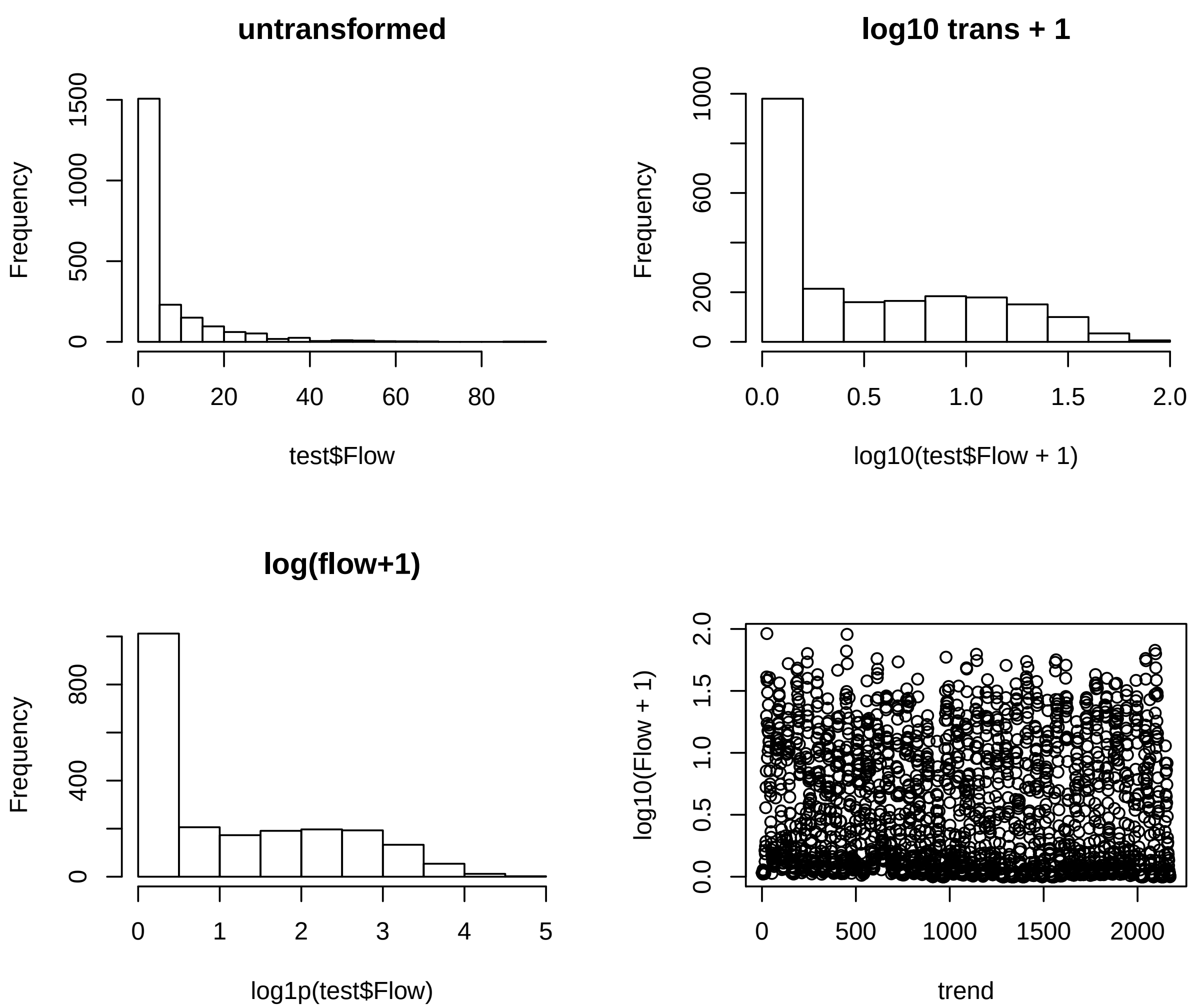
```
load("data/ClimCh_project_MD.Rdata")

test <- flow_rain_maxT_weekly[flow_rain_maxT_weekly$Station=="DOMB",]
```

```r
# add a vector of trend values
test$trend <- 1:nrow(test)


# plot distributions
par(mfrow=c(2,2))
hist(test$Flow, main="untransformed")
hist(log10(test$Flow +1), main ="log10 trans + 1")
hist(log1p(test$Flow), main = "log(flow+1)")

# log10 seems best
# modelling
plot(log10(Flow + 1)~trend, data = test)
```



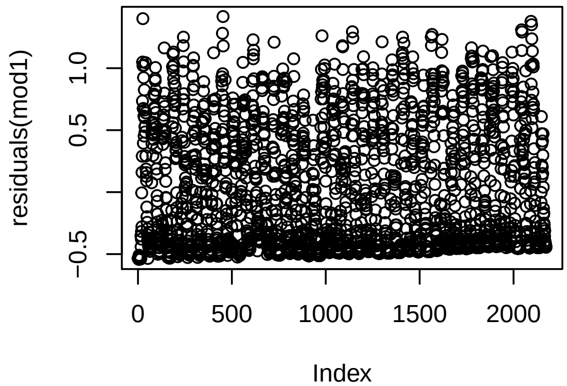

```r
mod1 <- lm(log10(Flow + 1)~trend, data = test)
summary(mod1)
```

```
##
```

```
## Call:
## lm(formula = log10(Flow + 1) ~ trend, data = test)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5424 -0.4402 -0.2247  0.4144  1.4167
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.645e-01  2.163e-02  26.100  < 2e-16 ***
## trend       -5.493e-05  1.723e-05  -3.188  0.00145 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5039 on 2171 degrees of freedom
## Multiple R-squared:  0.004658,   Adjusted R-squared:  0.0042
## F-statistic: 10.16 on 1 and 2171 DF,  p-value: 0.001455
```

```r
plot(residuals(mod1))
# interpret the slope
slope <- exp(coef(mod1)[2])-1
slope
```

```
##         trend
## -5.493294e-05
```



This slope is the fractional change in the flow per week. So if we multiply by 100 we get the “average weekly % change” in the flow. This implies the same % change on an annual scale and can be straight away interpreted as the amplification

This suggest a 0.005% decrease in flow