

Tutorials for the R package *chnngpt*

Youyi Fong

October 28, 2017

This tutorial illustrates the use of *chnngpt* (Fong et al., 2017a) with code examples. Estimation and hypothesis testing methods are described in Fong (2017); Fong et al. (2017b) and Fong et al. (2015), respectively, and references therein.

1 Continuous threshold regression models

1.1 Linear regression example

To estimate a thresholded linear regression model with a segmented-type change point in *Girth* for the *trees* dataset, we call

```
fit=chnngptm(formula.1=Volume~1, formula.2=~Girth, family="gaussian", data=trees,
type="segmented", var.type="bootstrap", weights=NULL)
```

- `formula.2` and `formula.1`: threshold variable and the rest of the model
- `type`: type of threshold model to fit
- `var.type`: *bootstrap* method is recommended for confidence interval
- `weights` can be supplied
- `est.method` defaults to *fastgrid* and is recommended

Calling `summary(fit)`, we get

```
Coefficients:
              Est      p.value   (lower   upper)
(Intercept) -24.614440 NA -37.580354 -11.648527
Girth        3.993966 NA  2.785558   5.202373
(Girth-chnngpt)+ 4.266618 NA  1.765770   6.767467

Threshold:
  74.2% (lower upper)
  16.0  12.9  19.1
```

Calling `plot(fit)`, we get Figure 1.

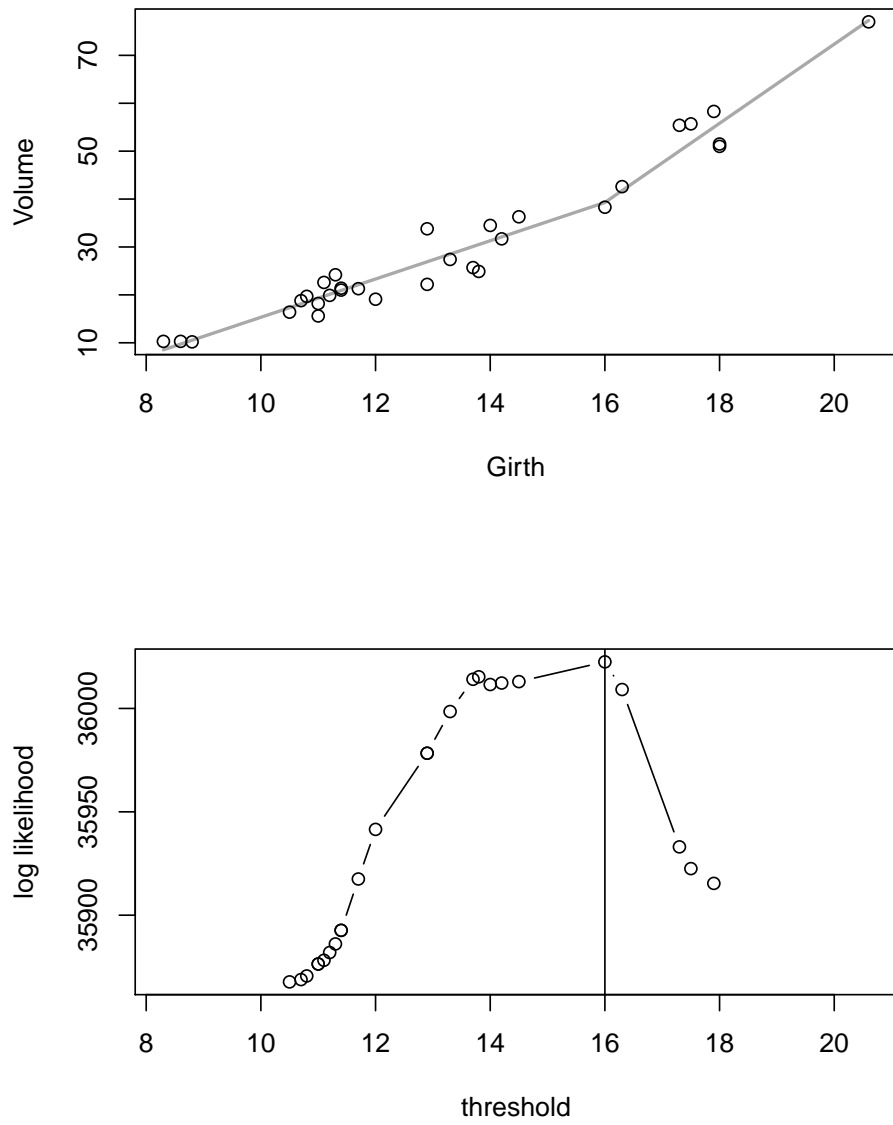


Figure 1: (a) Scatterplot of timber volume vs girth. The gray line shows the fitted segmented model. (b) Log likelihood of the submodel versus threshold parameter.

To test whether there is a change point, we call

```
test=chngpt.test(formula.null=Volume~1, formula.chngpt=~Girth, trees,  
family="gaussian")
```

When printed, we get

```
Maximum of Likelihood Ratio Statistics  
data:  trees  
Maximal statistic = 17.694, change point = 15.388, p-value = 0.00014  
alternative hypothesis:  two-sided
```

The first line gives the type of test carried out, and it is maximal likelihood ratio test here, which is the default. In addition, a plot function can be called on the test object to show the score or likelihood ratio statistic as a function of candidate change points.

1.2 Logistic regression example

To estimate a logistic regression model with a hinge-type change point in *NAb_SF162L* for the MTCT dataset, we call

```
library(splines)
fit=chnngptm(formula.1=y~birth, formula.2=~NAb_SF162LS, family="binomial", dat.mtct,
type="hinge", est.method="smoothapprox", var.type="robust", aux.fit=glm(y~birth
+ ns(NAb_SF162LS,3), dat.mtct, family="binomial"), weights=NULL)
```

- `formula.2` and `formula.1`: threshold variable and the rest of the model
- `type`: type of threshold model to fit
- `est.method`: *smoothapprox* is recommended
- `var.type`: *robust* is recommended for confidence interval
- `aux.fit`: required for *robust* variance estimation
- `weights` can be supplied

Calling `summary(fit)`, we get

```
Coefficients:
                                OR      p.value    (lower    upper)
(Intercept)          0.7026523  0.341429662  0.3388366  1.4571044
birthVaginal          1.2397649  0.523159883  0.6393632  2.4039809
(NAb_SF162LS-chnngpt)+ 0.6712371  0.001332547  0.5270730  0.8548327

Threshold:
      26.3%    (lower    upper)
7.373374  5.472271  8.186464
```

To test whether there is a change point, we call

```
test=chnngpt.test(formula.null=y~birth, formula.chngpt=~NAb_SF162LS, dat.mtct,
type="hinge", family="binomial", main.method="score")
```

When printed, we get

```
Maximal Score Test
data:  dat.mtct
Maximal statistic = 3.3209, change point = 7.0347, p-value = 0.00284
alternative hypothesis: two-sided
```

The first line gives the type of test carried out, and it may be maximal likelihood ratio test. In addition, a plot function can be called on the test object to show the score or likelihood ratio statistic as a function of candidate change points.

2 Discontinuous threshold regression models

The package also provides some support for estimation and hypothesis testing under discontinuous threshold regression models. What is missing, though, is confidence intervals for parameter estimates.

References

- Fong, Y. (2017), “Fast Bootstrap Confidence Intervals for Continuous Threshold Linear Models,” *Department of Biostatistics, University of Washington Technical Report*.
- Fong, Y., Di, C. and Permar, S. (2015), “Change point testing in logistic regression models with interaction term,” *Statistics in medicine*, 34, 1483–1494.
- Fong, Y., Huang, Y., Gilbert, P. and Permar, S. (2017a), “chngt: threshold regression model estimation and inference,” *BMC Bioinformatics*, 18, 454.
- Fong, Y., Chong, D., Huang, Y. and Gilbert, P. (2017b), “Model-robust Inference for Continuous Threshold Regression Models,” *Biometrics*, 73, 452–462.