Replication of Lyubchich et al. (2020): A data-driven approach to detecting change points in linear regression models

Emi McGeady

2024-12-11

```
#>
#> Call:
#> lm(formula = EarlySummerAnoxicVol ~ JanAprTNLoad, data = d)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -1.3151 -0.5222  0.0542  0.4533  1.9368
#>
#> Coefficients:
#>                Estimate Std. Error t value Pr(>|t|)
#> (Intercept)  -9.804e-01  4.046e-01  -2.423   0.0214 *
#> JanAprTNLoad  6.903e-06  1.069e-06   6.457 3.37e-07 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.7933 on 31 degrees of freedom
#>   (1 observation deleted due to missingness)
#> Multiple R-squared:  0.5735, Adjusted R-squared:  0.5598
#> F-statistic: 41.69 on 1 and 31 DF,  p-value: 3.367e-07
```

```
#>
#> Call:
#> lm(formula = LateSummerAnoxicVol ~ JanMayTNLoad, data = d)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -1.47911 -0.45241 -0.09363  0.51997  1.21833
```

```
#>
#> Coefficients:
#>                Estimate Std. Error t value Pr(>|t|)
#> (Intercept)  -2.174e-01  4.265e-01  -0.510 0.613848
#> JanMayTNLoad  5.596e-06  1.360e-06   4.115 0.000265 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.6894 on 31 degrees of freedom
#>   (1 observation deleted due to missingness)
#> Multiple R-squared:  0.3533, Adjusted R-squared:  0.3324
#> F-statistic: 16.94 on 1 and 31 DF,  p-value: 0.0002649
```

Regression Models

- (25%) Estimates in equations (8) and (9); retype the equations in this format in your report

Try to soft code estimates into the equations Have to get the correct estimations.

```
coeff8 <- summary(m8)$coefficients[,1]
coeff9 <- summary(m9)$coefficients[,1]
se8 <- summary(m8)$coefficients[, 2]
se9 <- summary(m9)$coefficients[, 2]
```

$$\hat{y}_{1t} = -0.98\ (0.405) + 6.903101e-06\ (1.069109e-06)\ JanAprTNLoad_t, \tag{8}$$

$$\hat{y}_{2t} = -0.217\ (0.426) + 5.596 \cdot 10^{-6}\ (1.360 \cdot 10^{-6})\ JanMayTNLoad_t, \tag{9}$$

where $\hat{y}_{1t}$ is Chesapeake Bay early summer anoxic volumes, and $\hat{y}_{2t}$ is late summer anoxic volumes, $JanAprTNLoad_t$ is total nitrogen load from Susquehanna and Potomac Rivers during January–April, $JanMayTNLoad_t$ is total nitrogen load from Susquehanna River during January–May, and standard errors of the coefficients are shown in parentheses.

- (25%) Figure 3 (can provide a different look but preserve the superscripts in axes labels and labels for the lines in A and D; format as subfigures and not separate figures)
- (25%) Figure 4 (this figure can be presented as separate figures)
- (25%) Bootstrapped p-values (closely) corresponding to the two lines for CART in Table 2.