To compare the variance of regression coefficients by using different resampling techniques

Anujit, Ipsita, Pooja

November 3, 2010

1 Introduction

Consider the following regression models

$$y_i = \alpha + \beta x_i + \epsilon_i, i = 1, \cdots, n$$

$$y_i = \alpha + \beta x_i + x_i \epsilon_i, i = 1, \cdots, n$$

where α=1.5, β=7.5, n=15, x_i's are generated from $Uniform(0, 20)$ distribution,and ϵ_i's are generated from one of the following distribution
let $\hat{\beta}$ be the OLS estimator of β. For each of the above cases,compare

- The true variance of $\hat{\beta}$ (estimated)
- The estimated variance using delete-1 jackknife
- The estimated variance using residual bootstrap
- The estimated variance using paired bootstrap

2 Results

Table 1: Results of model 1

Distribution	standard normal	Double exponential	t-distribution with d.f 2	Logistic
Estimated Variance of beta - hat by simulation	.054100064	.08149226	.05461763	.07262628
Estimated Variance of beta - hat by jackknife sampling	.05205357	.08352675	.0497562	.07139115
Estimated Variance of beta - hat by residual bootstrap	.05365725	.09632952	.05114177	.06413493
Estimated Variance of beta - hat by paired bootstrap	.05222555	.08326425	.04760964	.05591118

Table 2: Result of model 2

Distribution	standard normal	Double exponential	t-distribution with d.f 2	Logistic
Estimated Variance of beta - hat by simulation	.04227235	.86834482	.6635769	1.22626803
Estimated Variance of beta - hat by jackknife sampling	.02731309	.07073187	.4044742	.08270864
Estimated Variance of beta - hat by residual bootstrap	.03903361	.52648465	.03214410	.03903361
Estimated Variance of beta - hat by paired bootstrap	.03479961	.57495839	.3733652	.03479961

Figure 1 : Boxplot for double exponential (**model 1**)

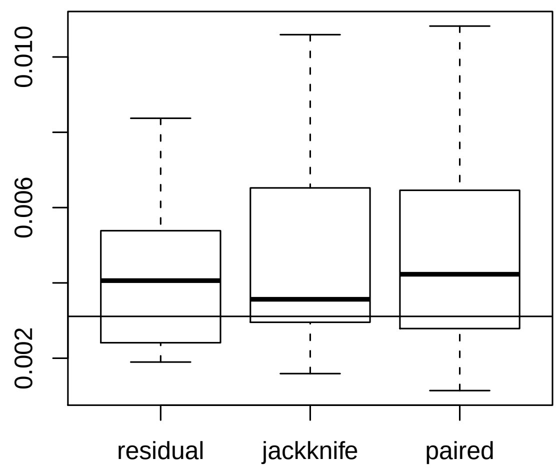


Figure 2 : Boxplot for double exponential (**model 2**)

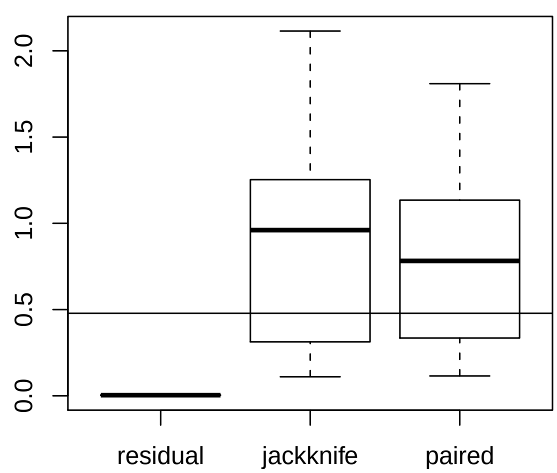


Figure 3 : Boxplot for logistic with parameter 0,1 (**model 1**)

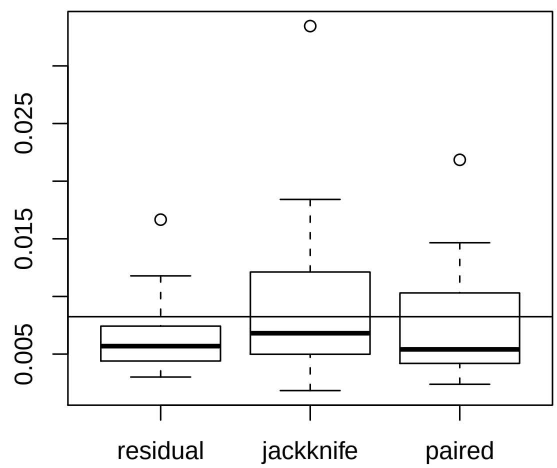


Figure 4 :Boxplot for logistic with parameter 0,1 (**model 2**)

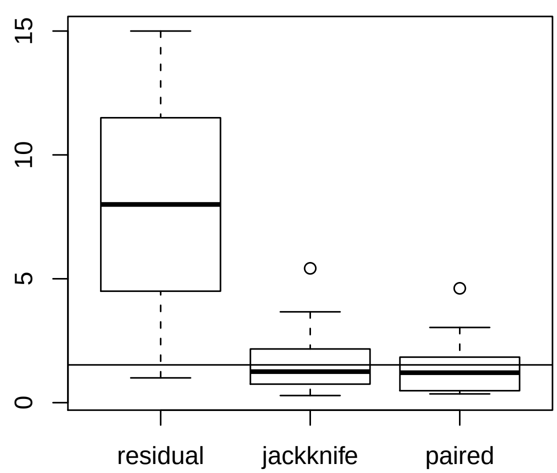


Figure 5 :Boxplot for normal with parameter 0,1 (**model 1**)

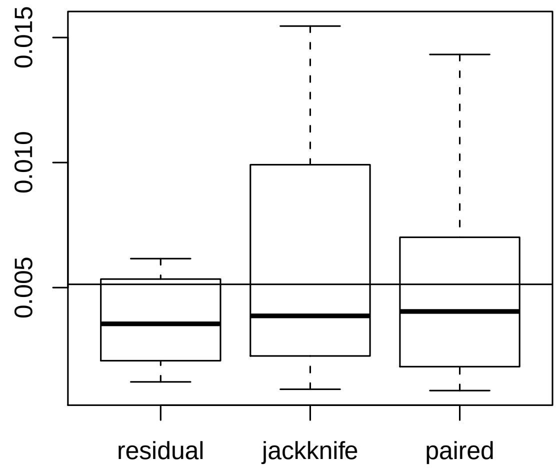


Figure 6 :Boxplot for normal with parameter 0,1 (**model 2**)

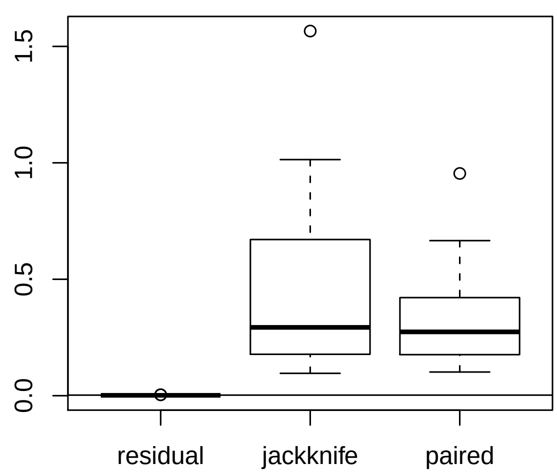


Figure 7 :Boxplot for t-distribution with d.f 2 (**model 1**)

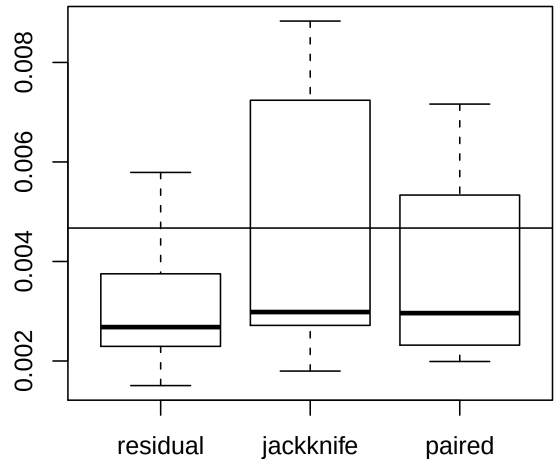
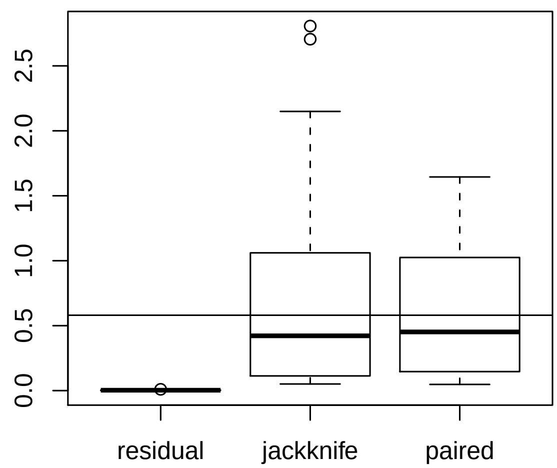


Figure 8 :Boxplot for t-distribution with d.f 2 (**model 2**)



3 Methodology

We considered two linear models

$$y_i = \alpha + \beta x_i + \epsilon_i, i = 1, \cdots, n \tag{1}$$

$$y_i = \alpha + \beta x_i + x_i\epsilon_i, i = 1, \cdots, n \tag{2}$$

where α=1.5, β=7.5, n=15,x_i-s are generated from U(0,20) distribution,and ϵ_i-s are generated from one of the following distribution.

- standard normal
- t_5
- double exponential with rate 1
- logistic(with location parameter 0 and scale parameter 1)

Now we want to estimate β for these models ,and then to find variance of estimated β by these resampling techniques.We adopted 4 resampling plans to obtain $\hat{\beta}$ and its variance.

3.1 Simulation

In first techniques we fitted linear model and then calculated value of β and then simply simulated the process to find estimate variance of $\hat{\beta}$.

3.2 Residual Bootstrap

We fitted a least square regression equation and define the usual residuals as

$$\epsilon_{in} = y_i - x_i'\beta_n, i = 1, \cdots, n$$

We used

$$\epsilon_{in}^{\sim} = \epsilon_{in} - n^{-1}\Sigma_{i=1}^{n}\epsilon_{in}$$

Let F_n be the EDF that puts mass n^{-1} at each $\{\epsilon_{in}^{\sim}\}, i = 1, \cdots, n$
Now let $\{\epsilon_1^*, \cdots, \epsilon_n^*\}$ be i.i.d. F_n and
let

$$y_i^* = x_i'\beta_n + \epsilon_i^*, \qquad i = 1, \cdots, n.$$

where $\{\epsilon_1^*, \cdots, \epsilon_n^*\}$ is a random sample drawn with replacement from $\{\epsilon_{1n}^{\sim}, \cdots, \epsilon_{nn}^{\sim}\}$. then using y^* instead of y in the linear model and we estimated beta .

3.3 Paired Bootstrap

Consider the ECDF F_n which puts mass n^{-1} at each pair $\{y_i, x_i\}, i = 1, \cdots, n$.The bootstrap random samples are drawn with replacement from y_i, x_i and hence the pair y_i^*, x_i^* are i.i.d. F_n. then using these pairs we estimated β and its variance

3.4 delete-1 Jackknife

In this method we deleted 1 pair among $\{y_i, x_i\}$ and using the remaining pairs we estimated β .Repeating this n times we calculated its variance.which is given by

$$V(\hat{\beta}) = n^{-1}\Sigma_{i=1}^{n}(\beta_i - \bar{\beta})^2$$

where $\bar{\beta} = \frac{1}{n}\Sigma_{i=1}^{n}\beta_i$.

3.5 Variance consistent

Any resampling plan ,say R is said to be variance consistent if its variance estimates say V_R,satisfies $V(\beta_n)/V_R \to 1$

4 Conclusion

4.1 Model 1

We see from our result that overall residual bootstrap is approximates our simulated result i.e residual bootstrap is *variance consistent*

4.2 Model 2

In model 2 where we have heteroscedasticity no resampling plans go near to simulated result of variance. But from theory we know that paired bootstrap resampling technique is robust ,but our results deviate.