PROSEDUR ESTIMASI PARAMETER MODEL REGRESI MENGGUNAKAN RESAMPLING BOOTSTRAP DAN JACKKNIFE

ESTIMATION OF PARAMETER REGRESION MODEL USING BOOTSTRAP AND JACKKNIFE

Hedi

(Staf Pengajar UP MKU Politeknik Negeri Bandung)

ABSTRAK

Penelitian ini menerapkan metode *resampling* pada pasangan data pengamatan untuk mendapatkan estimasi parameter model regresi dengan metode *resampling* bootstrap dan *jackknife*. Metode *bootstrap* didasarkan teknik pengambilan sampel dengan pengembalian, sedangkan Jackknife didasarkan penghapusan satu sampel pada setiap pengambilan sampel dengan pengembalian. Berdasarkan data simulasi, ditentukan *bias* estimasi parameter dan panjang interval dengan tingkat kepercayaan 95% serta *standard error*. Selanjutnya, dibandingkan reliabilitas yang diterapkan pada kedua metode tersebut. Berdasarkan *bias*, panjang interval, dan *standard error* estimasi parameter model regresi, metode *resampling bootstrap* lebih kecil daripada metode *resampling jackknife*.

Kata Kunci: resampling bootstrap dan jackknife, bias, standard error, dan panjang interval.

ABSTRACT

This research estimates parameters regresion model in pair observation data using bootsrtap and jackknife resampling methods. Bootstrap approach was based on the resampling observation and Jackknife was based on the delete—one resampling observation. Based on the simulation data, parameters estimation bias, the length of interval with 95% trust level, and standard error were determined. Then, the reliability applied to the two methods was compared. Based on the bias, interval length, and standard error of the parameters regresion model estimation, it was obtained that resampling method using bootstrap was less than using jackknife.

Keywords: bootstrap and jackknife resampling methods, bias, standard errors, interval length.

PENDAHULUAN

Model persamaan regresi linear

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

dengan ε diasumsikan berdistribusi normal dengan rata-rata 0 dan varians, σ^2 $\hat{\beta}$ sebagai penaksir tak bias dari β diperoleh dengan menggunakan ordinary least square. Dalam implementasinya, dihadapkan kita sering pada dengan permasalahan hanya mendapatkan jumlah sampel yang kecil dalam suatu pemodelan. Hal ini dapat berakibat pada parameter yang diperoleh bias, underestimate, atau overestimate. Masalah lain yang sering muncul yaitu distribusi sampling tidak selalu memenuhi distribusi normal. Biasanya diperlukan penurunan formulasi analitis yang sulit dilakukan sehingga dapat menyebabkan akurasi estimatornya tidak valid.

Metode resampling dapat digunakan untuk mengatasi permasalahan tersebut. Dalam metode resampling, sampel dianggap sebagai populasi. Metode resampling Bootstrap dan Jacknife adalah teknik resampling parametrik dan nonparametrik yang bertujuan untuk mengestimasi standard error dan confidence interval dari parameter populasi seperti rata-rata, median, proporsi, odds rasio, koefisien korelasi, atau perhitungan koefisien regresi tanpa asumsi-asumsi distribusi (Efron, 1982).

Berdasarkan uraian tersebut, identifikasi masalah pada penelitian adalah bagaimana menentukan estimasi parameter model regresi dengan tidak dipenuhinya asumsi distribusi normal pada residual dan bagaimana *resampling*

yang terbaik pada pasangan data pengamatan.

Penelitian ini bertujuan menentukan estimasi parameter model regresi dengan menerapkan *resampling bootstrap* dan *jackknife* pada pasangan data pengamatan dan membandingkan metode *resampling bootstrap* dan jackknife dengan menggunakan data simulasi.

TINJAUAN PUSTAKA

1. Analisis Regresi Berganda dengan Prediktor Random

Diberikan variabel random dalam bentuk vector $Z=(Z_1,\ldots,Z_r)^t,$ dan *output* variabel acak Y dengan (r+1) variabel tersebut berdistribusi gabungan P(Z,Y) dengan *mean* $E(Z)=\mu_Z$ dan $E(Y)=\mu_Y$ Jika prediksi Y dinyatakan sebagai fungsi dari Z yakni f(Z), penyimpangan diberikan oleh L(Y,f(Z)) dan ekspektasinya disebut sebagai fungsi risiko R(f)

$$R(f) = E(L(Y, f(Z))) = E((Y - f(Z))^{2})$$

$$R(f) = E_{X}(E_{Y|X}(Y - f(Z))^{2}|Z)$$
(2)

Selanjutnya dengan menerapkan identitas

$$Y - f(Z) = Y - \mu(Z) + \mu(Z) - f(Z) \qquad (3)$$
 Diperoleh

$$E_{Y|Z}\{(Y - f(Z))^2 | Z = z\} =$$

 $E_{Y|Z}\{(Y - \mu(z))^2 | Z = z\} + (\mu(z) - f(z))^2$ (4)
dan persamaan 4 minimum jika

$$f(\mathbf{z}) = \mu(\mathbf{z}) = E_{Y|z} \{ Y \mid \mathbf{Z} = \mathbf{z} \},\$$

dengan f^* adalah fungsi dari z sedemikian sehinga R(f) minimum, fungsi risiko untuk $f = f^*$, $R(f^*)$ disebut risiko bayes.

$$R(f^*) = \min R(f) = E\{(Y - \mu(z))^2\}$$
 (5)

Dengan demikian, prediksi terbaik Y di Z = z, menggunakan mean kuadrat error yang diberikan oleh $\mu(z)$.

Jika persamaan 1, diasumsikan ϵ tidak berkorelasi dengan Z_1, \dots, Z_r maka fungsi regresi dengan Z linier diberikan oleh

$$\mu(\mathbf{x}) = \mathbf{X}^{\mathsf{t}} \, \mathbf{\beta} \tag{6}$$

dengan $\beta=(\beta_0,\ldots,\beta_r)^t$ adalah vector koefisien regresi ukuran (r+1),dan $X=(1,\ Z^t)^t$ adalah vector ukuran (r+1). Selanjutnya, dipilih $\beta_0,\ \beta_1,\ \ldots,\ \beta_r$ yang meminimumkan risiko Bayes. Dengan memperhatikan persamaan berikut

$$S(\beta) = E\{(Y-X^{t} \beta)^{2}\}$$
 (7)

nilai minimum $s(\beta)$ dapat terjadi, jika nilai $\beta = \beta^*$, dengan

$$\beta^* = [E(XX^t)]^{-1} E(XY)$$
 (8)

sehingga dengan metode *Least square*, estimasi β yakni

$$\hat{\beta}_{=}$$
 (9)

2. Prosedur Estimasi Parameter Model Regresi dengan *Resampling* Bootstrap

Jika regresor random, metode bootstrap yang dipergunakan yakni melakukan resampling pada pasangan data pengamatan (observasi) dengan probabilitas yang terambil sebesar 1/n untuk setiap observasi. Resampling dilakukan sebanyak B kali dengan jumlah B disyaratkan cukup besar=hingga diperoleh estimasi parameter yang konvergen atau bahkan sampai sejumlah n^n sampel. Dengan jumlah B

yang cukup besar ini, diharapkan estimasi parameter regresi yang dihasilkan akan lebih kuat (*robust*). Dengan ketentuan bahwa Vektor

$$w_i = (Y_i, X_{ji})'$$
 dengan
 $Y_i = (y_1, y_2, ..., y_n)'$
dan $X_{jj} = (x_{j1}, x_{j2}, ..., x_{jn})'$.

Dalam metode ini, vektor $(w_1, w_2, ..., w_n)$ sebagai observasi.

Berdasarkan *resampling* observasi , langkah-langkah *bootstrap* dinyatakan sebagai berikut.

1. Diambil sampel bootstrap berukuran $n\left(w_1^{(b)}, w_2^{(b)}, ..., w_n^{(b)}\right)$ dengan $w_i^{(b)} = \left(Y_i^{(b)}, X_{ji}^{(b)}\right)',$ dengan $Y_i^{(b)} = \left(y_1^{(b)}, y_2^{(b)}, ..., y_n^{(b)}\right)',$ dan

$$X_{ji}^{(b)} = (x_{j1}^{(b)}, x_{j2}^{(b)}, ..., x_{jn}^{(b)})'$$
2. Estimasi parameter β dengan OLS dari

$$\hat{\beta}^{(b_1)} \!=\! \left(X^{(b)} \! \cdot \! X^{(b)} \right)^{\!-1} X^{(b)} \! \cdot \! Y^{(b)}$$

sampel bootstrap

- Diulangi langkah 1 dan 2 sebanyak r=1,2,...,B dengan B adalah banyaknya pengulangan
- 4. Dari langkah 3 diperoleh estimasi $\hat{\beta}^{(b_1)}, \hat{\beta}^{(b_2)}, ..., \hat{\beta}^{(b_B)}$.

Menurut Fox (1997), estimasi *bootstrap* untuk koefisien regresi adalah rata-rata dari nilai $\hat{\beta}^{(b_1)}, \hat{\beta}^{(b_2)}, ..., \hat{\beta}^{(b_B)}$ dengan

$$\hat{eta}^{(b)} = \sum_{b=1}^{B} \hat{eta}^{(b_r)} / B = \overline{\hat{eta}}^{(b_r)}$$

Persamaan regresi *bootstrap* $\hat{Y} = X \hat{\beta}^{(b)} + \varepsilon$ dengan $\hat{\beta}^{(b)}$ adalah penaksir tak bias dari β (Shao,1995)

3. Bias, Varians, Interval confidence, dan Persentile bootstrap

Penaksir bias untuk regresi bootstrap

$$bi\hat{a}s_b = \hat{\beta}^{(b)} - \hat{\beta}$$

Penaksir varians regresi bootstrap

$$\operatorname{var}(\hat{\beta}^{(b)}) = \sum_{i=1}^{B} \left[(\hat{\beta}^{(br)} - \hat{\beta}^{(b)}) (\hat{\beta}^{(br)} - \hat{\beta}^{(b)}) \right] / (B - 1)$$

Selang kepercayaan regresi bootstrap:

$$\begin{split} \hat{\beta}^{(b)} - t_{n-p,\alpha/2} * S_e \Big(\hat{\beta}^{(b)} \Big) < \beta < \hat{\beta}^{(b)} + t_{n-p,\alpha/2} * S_e \Big(\hat{\beta}^{(b)} \Big) \\ \text{Jika n} &\geq 30, \text{ dan tingkat kepercayaan } \alpha \\ \text{\%, digunakan nilai distribusi } Z_{\frac{1-\alpha}{2}} \text{ dalam} \end{split}$$

mengestimasi confidence interval (Diciccio and Tibshirani, 1987). Interval persentil bootstrap dapat dibangun dari quantile distribusi sampling bootstrap $\hat{\beta}^{(b)}$. Interval persentil $(\alpha/2)\%$ dan $(1-\alpha/2)\%$ adalah

$$\hat{eta}^{(b_r)}_{(lower)} < \beta < \hat{eta}^{(b_r)}_{(upper)}$$

dengan $\hat{\beta}^{(b_r)}$ adalah penaksir koefisien regresi *bootstrap* dari persamaan sebelumnya dengan batas bawah $(\alpha/2)B$ dan batas atas $(1-\alpha/2)B$.

4. Prosedur Estimasi Parameter Model Regresi dengan *Resampling* Jackknife

Resampling Jacknife pada pasangan data pengamatan (observasi), didasarkan penghapusan satu sampel atau sekelompok sampel dari sampel awal yang dianggap sebagai populasi. Pada satu tahap dan pada tahap selanjutnya, sampel yang telah dihapus tersebut dikembalikan dan dilakukan penghapusan satu atau sekelompok sampel dan seterusnya sampai semua sampel dari populasi mendapat kesempatan untuk dihapus.

Langkah-langkah metode resampling Jacknife hapus-satu sebagai berikut.

- 1. Diambil sampel berukuran n dari populasi secara acak dan beri nama dengan vektor $w_i = (Y_i, X_{ji})$ 'dengan $Y_i = (y_1, y_2, ..., y_n)$ ' dan $X_{ji} = (x_{j1}, x_{j2}, ..., x_{jn})$ 'untuk j=1,2,...,k dan i=1,2,...,n
- 2. Dihapus baris pertama dari vector $w_i = (Y_i, X_{ii})'$ dan ini merupakan vector dengan ukuran sampel (n-1) vector y sehingga menjadi $Y_i^{(j)} = (y_2^{(j)}, y_3^{(j)}, ..., y_n^{(j)})'$ dan $X_{ii}^{(j)} = (x_{i2}^{(j)}, x_{i3}^{(j)}, ..., x_{in}^{(j)})'$ sebagai sampel dengan jackknife, hapus satu $(w_1^{(j)})$ dan taksir koefisien regresi $\hat{\beta}^{(j_1)}$ dari $(w_l^{(j)})$. Selanjautnya, hapus baris kedua dari vector $w_i = (Y_i, X_{ii})$ 'dan kemudian beri label ukuran sampel n-1 dengan himpunan observasi $Y_i^{(j)} = (y_1^{(j)}, y_3^{(j)}, ..., y_n^{(j)})$ 'dan $X_{ji}^{(j)} = (x_{j1}^{(j)}, x_{j3}^{(j)}, ..., x_{jn}^{(j)})'$ sebagai $(w_2^{(j)})$ dan taksir koefisien regresi $\hat{\beta}^{(j_2)}$. Hapus satu per satu sampel dari n observasi dan taksir

koefisien regresi $\hat{\beta}^{(j_i)}$ dengan nilai $\hat{\beta}^{(j_i)}$ adalah koefisien regresi *Jackknife*. Setelah penghapusan pengamatan ke-i dari w_i , diperoleh nilai taksir koefisien regresi Jackknife $\hat{\beta}^{(j_1)}, \hat{\beta}^{(j_2)}, ..., \hat{\beta}^{(j_n)}$

3. Dihitung koefisien regresi Jacknife yang merupakan rata-rata dari

$$\hat{\beta}^{(j_1)}, \hat{\beta}^{(j_2)}, ..., \hat{\beta}^{(j_n)}$$

$$\hat{\beta}^{(j)} = \sum_{i=1}^n \hat{\beta}^{(j_i)} / n = \overline{\hat{\beta}}^{(j_i)}$$

sehingga model persamaan regresi untuk regresi *Jacknife* hapus satu sebagai berikut

$$\hat{Y} = X \hat{\beta}^{(j)} + \varepsilon$$

5. Bias, Varians, Interval confidence, dan persentile jackknife

Penaksir bias untuk regresi jackknife

$$bi\hat{a}s_{j}(\hat{\beta}) = (n-1)(\hat{\beta}^{(j)} - \hat{\beta})$$
$$var(\hat{\beta}^{(j)}) = \frac{(n-1)}{n} \sum_{i=1}^{n} (\hat{\beta}^{(ji)} - \hat{\beta}^{(j)})(\hat{\beta}^{(ji)} - \hat{\beta}^{(j)})$$

Penaksir varians regresi *jackknife*Confidence interval regresi jackknife

$$\hat{\beta}^{(j)} - t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(j)}) < \beta < \hat{\beta}^{(j)} + t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(j)})$$

Interval persentil *jackknife* dapat disusun dari *quantile* distribusi *sampling jackknife* $\hat{\beta}^{(j)}$.

Interval persentil $(\alpha/2)\%$ dan $(1-\alpha/2)\%$ adalah

$$\hat{\beta}^{(j)}_{(lower)} < \beta < \hat{\beta}^{(j)}_{(upper)}$$

dengan $\hat{\beta}^{(j)}$ adalah penaksir koefisien regresi *jackknife* dengan batas bawah $(\alpha/2)n$ dan batas atas $(1-\alpha/2)n$.

HASIL DAN PEMBAHASAN

Hasil simulasi yang dibangkitkan oleh persamaan $Y_i = -2 + 0.2X_i + 0.5X_i + \varepsilon_i$ dengan $i = 1, 2, 3, ..., 100, dan <math>\varepsilon_i$ berdistribusi normal dengan rata-rata 0 dan varians 1 disajikan pada Tabel.

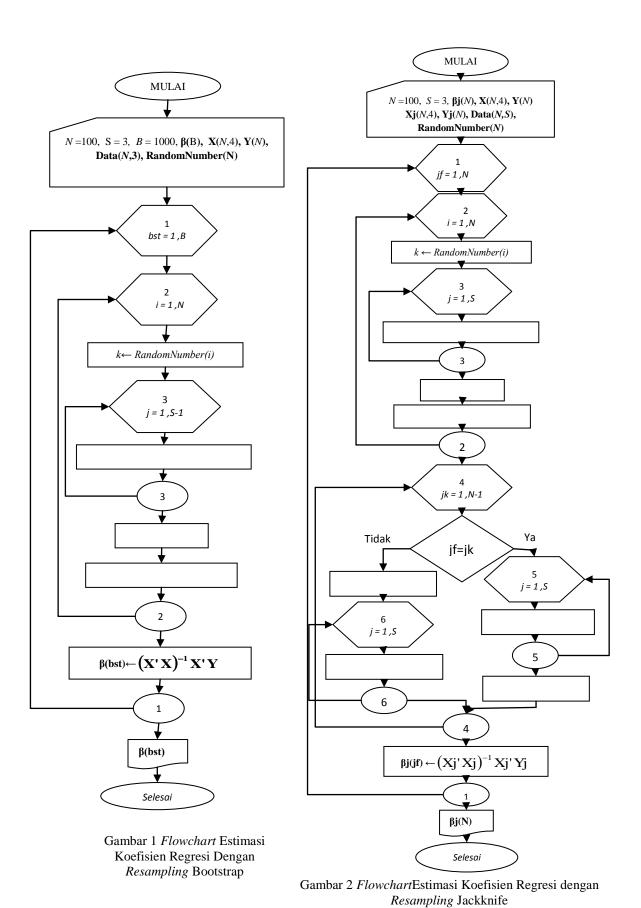
Tabel 1 Data Hasil Simulasi

No	X_1	Χ,	Y	No	X_1	Χ,	Y	No	X_1	Χ,	Y
1	10,4	3,1	0,73	37	23	12,9	8,83	73	21,9	2,9	4,20
2	13,3	3,22	2,98	38	22	10,6	8,07	74	21	4,67	4,09
3	17,9	4,3	2,45	39	12	3,6	1,73	75	28,9	3,89	7,15
4	15,9	4,15	3,53	40	33	4,25	8,35	76	25	3,95	4,75
5	19,7	5,1	0,99	41	23	5,26	5,98	77	27	4,67	4,28
6	22,1	5	4,92	42	22	5,27	4,77	78	35	5,6	8,60
7	23,5	3	4,72	43	12	5,17	2,29	79	23	5,78	5,23
8	26	2	3,12	44	13	6,18	3,81	80	22	6,89	6,63
9	13	4	2,80	45	15	8,17	4,25	81	15	5,98	4,26
10	33	6	8,95	46	12	9,16	5,80	82	28	7,25	6,48

11	22,6	7	6,27	47	36	8,19	9,15	83	38,8	8,25	8,73
12	13	8	4,67	48	27	11,7	9,96	84	36	4,9	6,97
13	24	9	7,28	49	9,8	15,7	6,27	85	30	5,28	8,12
14	14	4,51	1,38	50	9	6,8	2,84	86	20	6,89	4,19
15	25	4,34	4,54	51	11	6,9	4,17	87	20,6	2,95	2,99
16	24	4,26	4,83	52	11,9	6,96	4,44	88	20,9	4,87	3,62
17	35	8,12	7,54	53	18	8,34	5,22	89	28,9	6,48	7,08
18	23	9,9	8,14	54	34	6,99	8,30	90	33,5	3,78	7,04
19	12	7,4	5,13	55	22	8,5	6,83	91	34,7	8,98	10,62
20	23	7,23	6,70	56	23	8,23	6,97	92	28,9	8,23	5,34
21	22	6,33	6,14	57	34	6,13	7,42	93	10,5	8,1	4,33
22	33	5,12	6,05	58	34	6,8	7,31	94	9,5	7,15	2,51
23	12	4,77	4,96	59	23	7,12	6,98	95	8,5	5,24	1,20
24	22	7,9	6,41	60	17	7,5	5,20	96	11,5	5,76	4,06
25	33	10,2	10,03	61	18	8,1	6,25	97	27,5	4,67	6,92
26	22	11,7	6,99	62	35	8,12	8,71	98	28,6	6,78	6,56
27	12	6,7	5,22	63	34	7,11	5,90	99	32,8	7,88	9,84
28	22,3	11	8,71	64	22	9,05	7,92	100	36,5	6,99	7,59
29	12	7,9	4,08	65	12	9,23	3,89				
30	23,4	6,7	6,74	66	11	8,9	3,16				
31	22,4	11,2	6,89	67	33	8,13	8,95				
32	18	9	5,20	68	15,8	8,25	5,31				
33	10,9	5,4	4,08	69	15,9	9,07	5,28				
34	19,8	9,5	6,09	70	18,8	10,5	7,36				
35	18,9	9,44	4,65	71	17	10,34	6,72				
36	34,5	11,7	13,58	72	19,8	11,8	5,88				

Prosedur resampling bootstrap, pada pasangan data pengamatan (Tabel 1). Sebanyak B=1000 kali dengan pengembalian untuk mendapatkan sampel bootstrap yang masing-masing berukuran N=100. Selanjutnya langkah mengestimasi parameter model regresi berdasarkan sampel bootstrap dinyatakan dalam bentuk flowchart pada Gambar 1.

Berdasarkan prosedur resampling Jackknife, pada pasangan data pengamatan sebanyak N = 100 kali dengan pengembalian, didapat sampel Jackknife yang masing-masing berukuran *N*=99. Selanjutnya, langkah mengestimasi parameter model regresi berdasarkan jackknife sampel ditunjukkan dalam flowchart pada Gambar 2.



Berdasarkan simulasi yang telah dilakukan, diperoleh nilai-nilai estimasi parameter, *bias*, selang kepercayaan, dan *standard error* metode *resampling bootstrap* dan *jackknife* yang ditunjukkan pada Tabel 2 dan Tabel 3. Berdasarkan tabel tersebut , diperoleh

data bahwa bias estimasi parameter regresi metode *resampling bootstrap* lebih kecil dibandingkan dengan menggunakan metode *jackknife*. Begitu juga untuk selang kepercayaan, dengan menggunakan metode *bootstrap* selang kepercayaan menjadi lebih pendek.

Tabel 2. Estimasi Parameter Model Regresi dengan Resampling Bootstrap

Variabel		Rata-rata	Bias	95% Selang	Kepercayaan	Persentil		
		Kata-rata		Bawah	Atas	5%	95%	SE
β_0	-1,66023	-1,66231	-0,000208	-1,66926	-1,65536	-1,85252	-1,48985	0,112012
β_1	0,19855	0,19864	0,0001	0,19843	0,19886	0,19339	0,20476	0,003407
β_2	0,50289	0,50284	-0,00005	0,50218	0,5035	0,48517	0,52121	0,010648

Tabel 3. Estimasi Parameter Model Regresi dengan Resampling Jackknife

Variabel		Rata-rata	Bias	95% Selang	Kepercayaan	Persentil		
		Kata-rata		Bawah	Atas	5%	95%	SE
β_0	-1,66023	-1,66329	-0,00306	-1,68621	-1,64038	-1,84191	-1,48314	0,115492
β_1	0,19855	0,19876	0,00021	0,19802	0,19949	0,19231	0,20539	0,003709
β_2	0,50289	0,50295	-0,00006	0,50092	0,50498	0,48559	0,52038	0,010228

SIMPULAN DAN SARAN SIMPULAN

Resampling Bootstrap, pada regresi berganda model yang regresornya random, diterapkan melalui resampling pada pasangan data pengamatan. Dari N pasangan data pengamatan, diambil sampel berukuran n secara acak dengan pengembalian. Selanjutnya, dilakukan pengulangan sebanyak B kali yang dilanjutkan dengan estimasi parameter regresi sebanyak B kali sehingga didapat estimasi parameter regresi masing-masing banyaknya B. Dengan demikian, dapat dihitung bias,

standar *error*, dan panjang interval dari estimasi parameter regresi.

Metode Resampling Jackknife didasarkan penghapusan satu sampel setiap kali melakukan pengambilan sampel dan sampel yang dihapus tersebut dikembalikan. Dari n sampel, dipilih *n*-1 sampel dan dilakukan pengulangan sebanyak n kali sehingga diperoleh estimasi parameter regresi masing-masing banyaknya n. Penghapusan sampel dari suatu himpunan data, yang akan dianalisis, dilakukan dengan fix secara bergantian; tidak dianjurkan untuk ukuran sampel yang kecil.

Berdasarkan nilai bias estimasi parameter regresi, standard error, dan panjang interval, kasus ini menunjukkan bahwa metode resampling bootstrap lebih kecil dibandingkan dengan resampling jackknife. Ini berarti metode bootstrap lebih baik daripada metode jackknife.

SARAN

Tidak ada standar yang jelas tentang berapa kali sebaiknya dilakukan pengulangan (berapa B yang ideal). Jumlah maksimum pengulangan B yang mungkin dilakukan adalah hingga mencapai n^n kali dari sampel ukuran n. Oleh karena itu, perlu dilakukan penelitian lebih lanjut.

DAFTAR PUSTAKA

- Alan J. Izenman 2008. "Modern Multivariate Statistical Techniques". Springer, pp 109-111,New York
- DiCiccio, T., Tibshirani, R.1977.

 "Bootstrap Confidence Intervals and Bootstrap Approximations", J.

 Amer. Statist. Assoc., 82, pp. 161-169, 1987
- DamodarN.Gujarati. 2009. *Basic Econometrics*. The McGraw-Hill Companies.
- Efron, B dan Robert, T. 1993. An Introduction to The Bootstrap.

 New York: Chapman & Hall, Inc.

- Hogg.Mckean Craig. 2005.

 Introduction to Mathematical
 Statistics, Ed ke 6. New Jersey.
- Suat Sahinler and Dervis Topuz. 2007.
 "Bootstrap and Jackknife
 Resampling Algorithms for
 Estimation of Regression
 Parameters", Journal of Applied
 Quantitative Method. Vol 2 ,No
 2
- Shao, J., Tu, D.1995. *The Jackknife and Bootstrap*. New York: Springer.