Prosiding SPMIPA. pp. 177-184. 2006 *ISBN : 979.704.427.0*

PENENTUAN FAKTOR UTAMA PENYEBAB GANGGUAN LISTRIK DENGAN METODE BOOTSTRAP (STUDI KASUS DI KOTA SEMARANG)

Tarno
Program Studi Statistika FMIPA UNDIP Semarang
Jl. Prof. Soedarto, Kampus UNDIP Tembalang

Abstrak: Dalam tulisan ini dibahas tentang penentuan faktor utama yang berpengaruh secara signifikan terhadap pemadaman listrik di Semarang dengan menggunakan metode bootstrap. Pada awalnya faktor-faktor yang diduga berpengaruh terhadap pemadaman listrik di kota Semarang adalah: kerusakan jaringan transmisi, kerusakan trafo dan kerusakan fuse (sekering). Dengan melibatkan 3 faktor tersebut dibentuk model regresi yang menyatakan hubungan antara kerusakan jaringan, trafo dan fuse(sekering) terhadap pemadaman listrik di kota Semarang. Prosedur pemilihan model regresi yang terbaik dilakukan dengan menentukan estimasi sesatan prediksi atas semua model yang mungkin yaitu sebanyak 2^p-1 = 7 model, dengan p: banyaknya prediktor. Model yang terpilih adalah model yang memiliki rata-rata sesatan prediksi terkecil dan melibatkan variabel prediktor sesedikit mungkin. Prosedur pemilihan model terbaiknya dilakukan dengan menggunakan metode bootstrap residual (RB) dan bootstrap data berpasangan (PB). Prosedur pemilihan model dengan bootstrap konsisten untuk ukuran sampel n dengan replikasi bootstrap sebanyak B ($n \ll B \ll n^n$). Pemilihan variabel dengan metode bootstrap untuk ukuran sampel bootstrap m dengan $\frac{m}{n} \to 0$ dan $m \to \infty$, berdasarkan bootstrap residual (RB) dan bootstrap data berpasangan (PB) juga konsisten. Berdasarkan simulasi yang dilakukan dengan software R, diperoleh model regresi terbaik dengan melibatkan 2 variabel yaitu kerusakan jaringan dan kerusakan sekering. Dengan demikian faktor utama penyebab pemadaman listrik di kota Semarang adalah kerusakan jaringan dan kerusakan sekering.

Kata Kunci: pemilihan variabel, sesatan prediksi, bootstrap

PENDAHULUAN

PT. PLN (Persero) merupakan Badan Usaha yang memberikan jasa pelayanan listrik kepada masyarakat. Keberhasilan PT. PLN dalam menyediakan jasa pelayanan listrik sangat tergantung pada alat-alat yang digunakan sebagai sarana penyampaian jasa listrik tersebut. Gangguan-gangguan pada peralatan sangat memungkinkan terjadinya pemadaman listrik di suatu wilayah tertentu. Kerusakan peralatan yang dapat menyebabkan gangguan atau pemadaman listrik seringkali terjadi di kota Semarang. Secara geografis kota Semarang terletak di daerah perbukitan, dimana wilayahnya dapat dibedakan menjadi dua bagian yaitu Semarang atas dan Semarang bawah. Terkait dengan kondisi geografis tersebut Semarang atas sering terjadi gangguan cuaca seperti: angin kencang dan petir, sedangkan di Semarang bawah sering terjadi banjir. Faktor-faktor alam tersebut dapat menyebabkan kerusakan pada jaringan transmisi PLN, sedangkan kerusakan trafo, fuse/sekering seringkali disebabkan oleh pemakaian listrik yang berlebihan.

Berdasarkan kondisi di atas, dalam tulisan ini dilakukan pengkajian terhadap data gangguan listrik di kota Semarang yang diduga disebabkan oleh kerusakan peralatan antara lain: kerusakan trafo, sekering dan jaringan transmisi. Diduga jumlah kerusakan peralatan tersebut berpengaruh secara signifikan terhadap jumlah pemadaman listrik, sehingga hubungan fungsional antara jumlah kerusakan peralatan dengan jumlah gangguan listrik dapat dinyatakan dalam suatu model matematika. Adapun model matematika yang diduga sesuai dengan kenyataan tersebut adalah model regresi linier:

$$\mathbf{y}_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \,, i = 1,2,\ldots,n \qquad (1)$$

dengan $\mathbf{y}_i$: respon ke-i (jumlah pemadaman pada kurun waktu ke-i), $\mathbf{x}_i$: 3-vektor variabel prediktor (jumlah kerusakan trafo, sekering dan jaringan transmisi) yang berkaitan dengan $\mathbf{y}_i$, $\boldsymbol{\beta}$: 3-vektor parameter yang tidak diketahui dan $\varepsilon_i \sim N(0,\sigma^2)$.

Untuk mengestimasi parameter dalam model regresi tersebut biasanya digunakan metode kuadrat terkecil. Jika estimasi parameter telah diperoleh berarti telah diperoleh estimasi model untuk respon **y** yang dapat digunakan untuk melakukan prediksi untuk nilai **y** berdasarkan prediktor **x**. Beberapa komponen dari **x** mungkin tidak menghasilkan prediksi yang akurat karena tidak berpengaruh secara signifikan terhadap respon **y**, oleh karena itu perlu dilakukan pemilihan model terbaik. Menurut Shao dan Tu (1995, 1996), pemilihan variabel dalam model regresi linier dapat dilakukan dengan beberapa metode antara lain : Akaike Information Criterion (AIC), Cp (Mallows), Bayesian Information Criterion (BIC) serta metode Validasi-Silang dan bootstrap. Menurut referensi [1], kriteria AIC secara eksak atau pendekatan merupakan estimator tak bias untuk model dengan semua parameternya tak nol, tetapi jika digunakan untuk memilih model dengan komponen parameternya ada yang sama dengan nol, kriteria ini kadang-kadang tidak konsisten (bias). Sedangkan untuk kriteria BIC secara asimptotis tidak konsisten untuk data pengamatan berukuran besar, dan lebih baik apabila diterapkan pada model runtun waktu. Sedangkan menurut referensi [2], [3] dan [4] pemilihan model linier dengan bootstrap memiliki sifat konsisten untuk ukuran sampel yang cukup besar.

Berdasarkan argumen-argumen tersebut, perlu dikaji lebih dalam tentang faktor utama penyebab gangguan listrik sekaligus menentukan model terbaik yang menyatakan hubungan antara jumlah kerusakan peralatan (trafo, fuse dan jaringan) terhadap pemadaman listrik di kota Semarang. Dari model yang mungkin dilakukan pemilihan model regresi terbaik dengan menggunakan metode bootstrap, yaitu metode pembangkitan data pengamatan berbasis komputer untuk mendapatkan sampel berukuran besar, sehingga asumsi-asumsi yang disyaratkan dalam model regresi dapat terpenuhi. Disamping itu sampel yang dikumpulkan di lapangan tidak perlu berukuran besar, sehingga peneliti dapat melakukan efisiensi waktu dan biaya untuk pengumpulan data di lapangan. Untuk mendapatkan sampel berukuran besar, cukup dilakukan pembangkitan data dengan simulasi komputer di laboratorium.

PEMILIHAN VARIABEL DAN SESATAN PREDIKSI

Prediksi nilai respon dengan menggunakan prediktor **x**, mungkin tidak tergantung pada semua komponen **x**, artinya penggunaan semua komponen dari **x** belum tentu menghasilkan prediksi yang akurat. Dibawah model (1): $y_i = x_i'\beta + \varepsilon_i$, $i = 1,2,...,n$,dengan **y** : variabel respon, **x** : p-vektor prediktor , β : p-vektor parameter yang tak diketahui dan ε : sesatan random dengan mean 0 dan variansi σ^2. Karena beberapa komponen dari β mungkin sama dengan 0 maka model yang lebih akurat adalah berbentuk :

$$y_i = x_{i,\alpha}'\beta_\alpha + \varepsilon_i \text{ , } i = 1,2,...,n \text{ dengan } \alpha \subset \{1,2,...,p\}. \tag{2}$$

Jika β_α dan $x_{i,\alpha}$ sebagai subvektor yang memuat komponen-komponen dari β dan x_i berada dalam α, maka terdapat (2^p-1) model berbeda yang mungkin yang berbentuk (2), masing-masing terkait dengan suatu himpunan bagian α dan dinotasikan dengan $\hat{\alpha}$. Dimensi (ukuran) dari $\hat{\alpha}$ adalah banyaknya prediktor dalam $\hat{\alpha}$. Jika ***A*** menyatakan semua himpunan bagian dari {1,2,…,p}, dan diketahui masing-masing komponen dari β adalah 0 atau tidak, maka model-model $\hat{\alpha}$ dapat diklasifikasikan menjadi dua kategori :

- Kategori I (*incorrect model*) : Minimal satu komponen dari β yang tidak nol tidak berada dalam β_α.
- Kategori II (*correct model*) : β_α memuat semua komponen dari β yang tidak nol.

Memilih model dari kategori I berarti menghilangkan minimal satu prediktor yang penting, sedangkan memilih model dari kategori II berarti mengeliminasi semua variabel yang tak terkait dengan variabel respon. Dengan demikian model optimalnya adalah model (2) dengan α_0 sedemikian hingga β_α memuat semua komponen dari β yang semuanya tidak nol, yaitu model dalam kategori II dengan dimensi terkecil.

Model optimal tersebut perlu dipilih dari model (2) berdasarkan data (y_1,x_1), (y_2,x_2), …, (y_n,x_n) yang memenuhi (1). Jika diasumsikan $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ independen dan berdistribusi identik dengan mean 0 dan variansi σ^2, maka dengan Metode Kuadrat Terkecil diperoleh :

$$\hat{\beta}_\alpha = (X_\alpha' X_\alpha)^{-1} X_\alpha' y \text{ dengan } \mathbf{y} = (y_1, y_2, \ldots, y_n)' \text{ dan } X_\alpha = (x_{1\alpha}, x_{2\alpha}, ..., x_{n\alpha}).$$

Jika $\mathbf{y_f}$: nilai respon yang akan datang maka: $\hat{y}_{f\alpha} = x_{f\alpha}'\hat{\beta}_\alpha$ untuk suatu nilai prediktor $\mathbf{x_f}$. Hal ini berakibat bahwa mean dari sesatan prediksi kuadrat $mse(x_f, \alpha)$ adalah :

$$\text{mse}(x_f,\alpha) = E(y_f - \hat{y}_{f\alpha})^2 = \sigma^2 + \sigma^2 x'_{f\alpha}(X'_\alpha X_\alpha)^{-1} x_{f\alpha} + \Delta(x_f,\alpha),$$
$$\text{dengan } \Delta(x_f,\alpha) = [x'_f\beta - x'_{f\alpha}(X'_\alpha X_\alpha)^{-1}X\beta]^2$$

Jika α dalam kategori II maka $X\beta = X_\alpha\beta_\alpha$, $x'_f\beta = x'_{f\alpha}\beta_\alpha$ dan $\Delta(x_f,\alpha) = 0$. Sehingga model optimalnya adalah model α dengan ukuran terkecil. Dengan demikian jika $\text{mse}(x_f,\alpha)$ diketahui, maka model optimal dapat dipilih dengan meminimalkan $\text{mse}(x_f,\alpha)$ atas semua $\alpha \in \boldsymbol{A}$. Model optimal dapat juga ditentukan dengan meminimalkan rata-rata dari sesatan prediksi kuadrat $\text{mse}(x_f,\alpha)$ atas X={$x_1, x_2, \ldots, x_n$}:

$$\overline{mse}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} mse(x_i,\alpha) = \sigma^2 + \frac{\sigma^2 p}{n} + \Delta(\alpha), \text{ dengan } \Delta(\alpha) = \frac{1}{n}\beta' X'(I - H_\alpha)X\beta,$$

$H = X_\alpha (X'_\alpha X_\alpha)^{-1} X'_\alpha$ dan I matriks identitas pxp.

Namun, $\text{mse}(x_f,\alpha)$ dan $\overline{\text{mse}}(\alpha)$ kedua-duanya tidak diketahui. Sehingga mengestimasi $\overline{\text{mse}}(\alpha)$ lebih mudah dari pada mengestimasi $\text{mse}(x_f,\alpha)$ dengan menggunakan $\widehat{\overline{\text{mse}}}(\alpha)$, kemudian memilih model dengan meminimalkan $\widehat{\overline{\text{mse}}}(\alpha)$ atas $\alpha \in \boldsymbol{A}$. Untuk mendapatkan model terbaik, lebih lanjut dilakukan pemilihan model dengan metode bootstrap.

ESTIMASI PARAMETER DENGAN BOOTSTRAP

Jika parameter $\boldsymbol{\beta}$ merupakan parameter regresi yang akan diestimasi dengan $\hat{\boldsymbol{\beta}}$, maka di lingkungan boostrap $\hat{\boldsymbol{\beta}}$ dapat diestimasi dengan $\hat{\boldsymbol{\beta}}^*$. Untuk mengestimasi parameter regresi dapat dilakukan dengan beberapa prosedur bootstrap, antara lain: bootstrap berdasarkan residual, bootstrap pasangan data pengamatan.

Bootstrap Residual

Bootstrap berdasarkan residual (RB) , diusulkan oleh Efron (1979). Jika diketahui model regresi (1) : $y_i = \mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i, i = 1,2,\ldots,n$, dengan $\mathbf{y}_i$ adalah respon ke-i , $\mathbf{x}_i$: p-vektor variabel prediktor yang berkaitan dengan y_i, $\boldsymbol{\beta}$: p-vektor parameter yang tidak diketahui, dengan asumsi bahwa ε_i i.i.d dengan mean 0 dan variansi σ^2, maka untuk memperoleh estimasi parameter dapat dilakukan prosedur bootstrap :

a. Parameter $\boldsymbol{\beta}$ diestimasi dengan $\hat{\boldsymbol{\beta}}$ dan F_ε diestimasi dengan fungsi distribusi empiris $\hat{F}_\varepsilon$ dengan mengambil massa peluang n^{-1} terhadap $r_i - \bar{r}, i = 1,2,\ldots,n$ dengan $r_i = y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}$ merupakan residual ke-i dan $\bar{r} = n^{-1}\sum_{i=1}^{n} r_i$.
b. Bangkitkan data independen dan berdistribusi identik $\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*$ dari $\hat{F}_\varepsilon$ dan didefinisikan $y_i^* = \mathbf{x}'_i\hat{\boldsymbol{\beta}} + \varepsilon_i^*$.
c. Ditentukan $\hat{\boldsymbol{\beta}}^*$ berdasarkan data $(y_1^*, x'_1),(y_2^*, x'_2),\ldots(y_n^*, x'_n)$, yaitu : $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^*$.
d. Ulangi langkah diatas sebanyak B kali dengan $n << B << n^n$, sebagai replikasi bootstrap.

Karena $E_*(\varepsilon_i^*) = 0$, untuk setiap i, maka $E_*(\hat{\boldsymbol{\beta}}^*) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E_*(\mathbf{y}^*) = \hat{\boldsymbol{\beta}}$, (3)

dan juga $\text{var}_*(\hat{\boldsymbol{\beta}}^*) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\text{var}_*(\mathbf{y}^*)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, (4)

dengan $\hat{\sigma}^2 = \text{var}(\varepsilon_i^*) = n^{-1}\sum_{i=1}^{n}(r_i - \bar{r})^2$. Dari persamaan (3) dan (4), prosedur bootstrap ini menghasilkan estimator variansi dan estimator tak bias dari $\hat{\boldsymbol{\beta}}$ yang konsisten.

Bootstrap Data Berpasangan

Bootstrap berpasangan (PB), nampaknya merupakan suatu prosedur yang sangat alami apabila $\mathbf{x}_i$ random dan $(y_i, x'_i), i = 1,2,\ldots,n$, independen dan berdistribusi identik (i.i.d). Dalam kasus ini, untuk mengestimasi parameter regresi dapat dilakukan prosedur sebagai berikut :

a. Dibangkitkan data bootstrap dari fungsi distribusi empiris $(y_i, x_i^{'}), i = 1,2,\ldots,n$ dengan massa peluang n^{-1}, yaitu : $(y_1^*, x_1^*), (y_2^*, x_2^*), \ldots (y_n^*, x_n^*)$.

b. Ditentukan estimasi parameter regresi : $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^*$.

c. Ulangi langkah diatas sebanyak B sebagai replikasi bootstrap.

PEMILIHAN VARIABEL DENGAN BOOTSTRAP

Prosedur pemilihan model dengan bootstrap dapat diturunkan dari estimator untuk $\overline{\text{mse}}\,(\alpha)$. Pandang estimator bootstrap berbentuk :

$$\hat{\overline{mse}}_{BOOT}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha)^2 + \hat{e}(\alpha) \tag{5}$$

dengan $\hat{e}(\alpha)$: estimator bootstrap dari sesatan ekses harapan :

$$e(\alpha) = E\left[\frac{1}{n}\sum_{i=1}^{n}(y_{f,i} - x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha)^2 - \frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha)^2\right] = \frac{2\sigma^2 p_\alpha}{n}$$

dengan $y_{f,i}$ respon yang akan datang pada x yang independen dengan y_i .

Jika $\hat{\boldsymbol{\beta}}_\alpha^*$ estimator dari $\hat{\boldsymbol{\beta}}_\alpha$ di lingkungan bootstrap maka diperoleh :

$\hat{\boldsymbol{\beta}}_\alpha^* = (\mathbf{X}_\alpha^{*'}\mathbf{X}_\alpha^*)^{-1}\mathbf{X}_\alpha^{*'}\mathbf{y}_\alpha^*$ untuk bootstrap berpasangan, dan

$\hat{\boldsymbol{\beta}}_\alpha^* = (\mathbf{X}_\alpha^{'}\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha^{'}\mathbf{y}_\alpha^*$ untuk bootstrap residual,

dengan $\mathbf{y}_\alpha^* = (y_{1\alpha}^*, y_{2\alpha}^*, \ldots, y_{n\alpha}^*)$, $y_{i\alpha}^* = x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha + \varepsilon_i^*$ dan ε_i^* independen dan berdistribusi identik dari $\hat{F}_\varepsilon$ yaitu fungsi distribusi empiris dari $\boldsymbol{\varepsilon}$ dengan massa peluang 1/n. Sehingga diperoleh estimator dari $e(\alpha)$ untuk bootstrap berpasangan adalah :

$$\hat{e}(\alpha) = E_*\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i\alpha}^{*'}\hat{\boldsymbol{\beta}}_\alpha^*)^2 - \frac{1}{n}\sum_{i=1}^{n}(y_i^* - x_{i\alpha}^{*'}\hat{\boldsymbol{\beta}}_\alpha^*)^2\right] \tag{6}$$

dan untuk bootstrap residual :

$$\hat{e}(\alpha) = E_*\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha^*)^2 - \frac{1}{n}\sum_{i=1}^{n}(y_i^* - x_{i\alpha}^{'}\hat{\boldsymbol{\beta}}_\alpha^*)^2\right] \tag{7}$$

Estimator bootstrap untuk $\overline{\text{mse}}\,(\alpha)$ diberikan oleh persamaan (5) dengan $\hat{e}(\alpha)$ yang bersesuaian. Perlu diketahui bahwa $e(\alpha) = e_n(\alpha)$ tergantung pada ukuran sampel n. Untuk bootstrap residual perhitungan secara langsung menghasilkan $\hat{e}(\alpha) = \dfrac{2\hat{\sigma}^2 p_\alpha}{n}$ (8)

yang merupakan estimator tak bias asimptotis dan konsisten untuk $e(\alpha)$. Dan $\hat{e}(\alpha)$ dengan menggunakan bootstrap pasangan akan mendekati ruas kanan pada persamaan (8) tetapi penurunannya cukup rumit.

Lebih lanjut untuk mendapatkan prosedur pemilihan model bootstrap yang konsisten , dapat juga dipilih ukuran sampel m sedemikian hingga $e_m(\alpha)$ dapat diestimasi dengan $\hat{e}_m(\alpha)$ dengan m/n $\rightarrow 0$ dan kemudian meminimalkan :

$$\hat{\overline{\text{mse}}}_{BOOT-m}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{'}\hat{\boldsymbol{\beta}}_\alpha)^2 + \hat{e}_m(\alpha) \text{ atas } \alpha \in \boldsymbol{A}. \tag{9}$$

Untuk mengestimasi $e_m(\alpha)$ pada bootstrap pasangan , terlebih dahulu dibangkitkan sampel berpasangan $(y_1^*, x_1^*), (y_2^*, x_2^*), \ldots, (y_n^*, x_m^*)$ dan menggunakan :

$$\hat{e}_m(\alpha) = E_*\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i\alpha}^{*'}\hat{\boldsymbol{\beta}}_{m,\alpha}^*)^2 - \frac{1}{m}\sum_{i=1}^{m}(y_i^* - x_{i\alpha}^{*'}\hat{\boldsymbol{\beta}}_{m,\alpha}^*)^2\right] \tag{10}$$

dengan $\hat{\beta}_{m,\alpha}^*$ didefinisikan seperti pada $\hat{\beta}_\alpha^*$ yang didasarkan pada m pasangan data bootstrap. Jika m dipilih sedemikian hingga m/n $\rightarrow 0$ dan m $\rightarrow \infty$ maka : $\hat{e}_m(\alpha) = \dfrac{2\hat{\sigma}^2 p_\alpha}{m} + o(m^{-1})$

Pada bootstrap residual, digunakan secara langsung $\hat{e}_m(\alpha) = \frac{2\hat{\sigma}^2 p_\alpha}{m}$ dan mengestimasi $e_m(\alpha)$ dengan $\frac{2\hat{\sigma}^2 p_\alpha}{m}$.

KONSISTENSI BOOTSTRAP

Teorema 1 [2] dan [3]. Jika diasumsikan bahwa ε_i i.i.d dan $\max_{i \le n} h_{i\alpha} \to 0$ untuk semua $\alpha \in \boldsymbol{A}$, dengan $h_{i\alpha} = x'_{i\alpha}(\mathbf{X}'_\alpha \mathbf{X}_\alpha)^{-1} x_{i\alpha}$.

(i) Pandang suatu estimator bootstrap $\overset{\triangle}{mse}_{BOOT}(\alpha)$ dalam persamaan (5) dengan $\hat{e}(\alpha)$ seperti yang diberian pada persamaan (6) untuk PB dan (7) untuk RB. Maka, apabila α dalam kategori I,

$$\overset{\triangle}{mse}_{BOOT}(\alpha) = \overline{mse}(\alpha) + o_p(1); \qquad (11)$$

sedangkan apabila α dalam kategori II, $\overset{\triangle}{mse}_{BOOT}(\alpha) = \frac{\|\varepsilon\|^2}{n} + \frac{2\sigma^2 p_\alpha}{n} - \frac{\varepsilon' H_\alpha \varepsilon}{n} + o_p(n^{-1})$.

(ii) Jika $\overset{\triangle}{mse}_{BOOT-m}(\alpha)$ yang didefinisikan dalam persamaan (9) dengan $\hat{e}_m(\alpha)$ seperti dalam persamaan (10) untuk PB dan $\frac{2\hat{\sigma}^2 p_\alpha}{m}$ untuk RB. Lebih lanjut ukuran sampel bootstrap m dipilih sedemikian hingga

$$\frac{m}{n} \to 0, \quad \frac{n}{m} \max_{i \le n} h_{i\alpha} \to 0 \text{ untuk semua } \alpha \in \boldsymbol{A} \qquad (12)$$

Maka apabila α dalam kategori I, $\overset{\triangle}{mse}_{BOOT-m}(\alpha) = \overline{mse}(\alpha) + o_p(1)$;

sedangkan apabila α dalam kategori II, $\overset{\triangle}{mse}_{BOOT-m}(\alpha) = \frac{\|\varepsilon\|^2}{n} + \frac{\sigma^2 p_\alpha}{m} + o_p(m^{-1})$.

(iii) Lebih lanjut diasumsikan bahwa $\liminf_n \inf_{\alpha \text{ dlm kategori I}} \Delta(\alpha) > 0$.

Maka $\lim_{n\to\infty} P\{\hat{\alpha}_{BOOT} \text{ dalam kategori I}\} = 0$, dan $\lim_{n\to\infty} P\{\hat{\alpha}_{BOOT} = \alpha_0\} < 1$

kecuali bila $\alpha = \{1,2,\ldots,p\}$ dalam kategori II (<u>a correct model</u>); dan $\hat{\alpha}_{BOOT-m}$ konsisten, yaitu $\lim_{n\to\infty} P\{\hat{\alpha} = \alpha_0\} = 1$ berlaku untuk $\hat{\alpha}_{BOOT-m}$, dengan $\hat{\alpha}_{BOOT}$ dan $\hat{\alpha}_{BOOT-m}$ adalah model-model terpilih dengan masing-masing meminimalkan $\overset{\triangle}{mse}_{BOOT}(\alpha)$ dan $\overset{\triangle}{mse}_{BOOT-m}(\alpha)$.

SIMULASI

Untuk memberikan gambaran yang jelas tentang prosedur penentuan factor utama penyebab pemadaman listrik di kota Semarang, pada bagian ini diberikan hasil simulasi terhadap data yang telah dicatat pada kantor PLN Semarang selama 4 tahun belakangan. Adapun variabel-variabel yang terlibat dalam pemodelan ini adalah: Jumlah pemadaman listrik sebagai variabel respon y, dan variabel prediktornya adalah: jumlah kerusakan jaringan (x1), jumlah kerusakan trafo (x2) serta jumlah kerusakan sekering (x3). Namun untuk mengatasi penyimpangan asumsi regresi linier yang mungkin terjadi, variabel-avariabel tersebut diambil transformasi logaritma. Hasil simulasi untuk menentukan estimasi rata-rata sesatan prediksi dengan menggunakan **software 'R'** disajikan dalam Tabel 1 berikut.

Tabel.1: Estimasi mse untuk 7 model yang mungkin dengan
Bootstrap Residual (RB) dan Bootstrap Data Berpasangan (PB)

No	Variabel-variabel dalam Model			Est.MseR B-16 (B-200)	Est.Mse RB-16 (B-500)	Est.Mse PB-16 (B-200)	Est.MseP B-16 (B-500)	Est.Mse PB-10 (B-200)	Est.Mse PB-10 (B-500)
	x1	x2	x3						
1	x1			0.797	0.787	0.310	0.308	0.324	0.320
2		x2		**0.193**	**0.195**	**0.077**	**0.077**	**0.081**	**0.080**
3			x3	0.323	0.318	0.125	0.125	0.130	0.130
4	x1	x2		0.085	0.083	0.029	0.029	0.030	0.030
5	x1		x3	**0.053**	**0.054**	**0.017**	**0.017**	**0.018**	**0.018**
6		x2	x3	0.232	0.232	0.068	0.068	0.069	0.069
7	x1	x2	x3	**0.041**	**0.043**	**0.012**	**0.012**	**0.012**	**0.013**

Berdasarkan hasil perhitungan estimasi mse pada Tabel 1 serta dengan mempertimbangkan prinsip parsimonius, diperoleh estimasi model regresi terbaik sebagai berikut.

ln(y) = 1,765 + 0,394 ln(x1) + 0,555 ln(x3).

KESIMPULAN

Prosedur pemilihan model dengan menggunakan metode Bootstrap konsisten untuk ukuran sampel bootstrap m (1<m<n) dengan $\frac{m}{n} \to 0$ dan $m \to \infty$. Berdasarkan nilai estimasi *mse* yang minimum serta dengan mempertimbangkan prinsip parsimonius, maka ditetapkan model regresi terbaik dengan melibatkan 2 variabel prediktor, adalah:

ln(y) = 1,765 + 0,394 ln(x1) + 0,555 ln(x3)

dengan y: jumlah gangguan listrik, x1: jumlah kerusakan jaringan dan x3: jumlah kerusakan sekering. Dengan demikian faktor utama penyebab gangguan listrik di kota Semarang adalah kerusakan jaringan dan kerusakan sekering.

DAFTAR PUSTAKA

[1]. Hjorth, Urban, *Computer Intensive Statistical Methods*, *Validation Model Selection and Bootstrap,* Chapman and Hall, New York, 1994.

[2]. Shao and Tu, *The Jackknife and Bootstrap*, Springer-Verlag, New York, 1995.

[3]. Shao, Bootstrap Model Selection, *Journal American Statistics Assosiation,* Vol. 91, No. 434, pp. 655-665, 1996.

[4]. Shao, An Asymptotic Theory for Linear Model Selection, *Statistica Sinica*, Vol. 7, pp. 221-264, 1997.

[5]. Tarno dan Subanar, Pemilihan Model Regresi Linier Terbaik dengan Bootstrap, *Jurnal Matematika dan Komputer*, FMIPA UNDIP, 2001.

LAMPIRAN

Data tentang jumlah gangguan listrik (y), jumlah kerusakan jaringan (x1), jumlah kerusakan trafo (x2) serta jumlah kerusakan sekering (x3) pada PT PLN (Persero) kota Semarang selama 4 tahun belakangan.

x1	x2	x3	y		ln(y)	ln(x1)	ln(x2)	ln(x3)
103.00	45.00	127.00	492.00		6.20	4.63	3.81	4.84
100.00	40.00	32.00	532.00		6.28	4.61	3.69	4.88
54.00	28.00	88.00	351.00		5.86	3.99	3.33	4.48
89.00	63.00	215.00	687.00		6.53	4.49	4.14	5.37
61.00	45.00	105.00	524.00		6.26	4.11	3.81	4.65
43.00	12.00	62.00	212.00		5.36	3.76	2.48	4.13
26.00	16.00	76.00	207.00		5.33	3.26	2.77	4.33
73.00	17.00	100.00	406.00		6.01	4.29	2.83	4.61
x1	x2	x3	y		ln(y)	ln(x1)	ln(x2)	ln(x3)
102.00	38.00	40.00	327.00	5.79	4.62	3.64	3.69	
272.00	23.00	88.00	540.00		6.29	5.61	3.14	4.48
377.00	64.00	117.00	724.00		6.58	5.93	4.16	4.76
587.00	27.00	57.00	776.00		6.65	6.38	3.30	4.04
85.00	144.00	289.00	845.00		6.74	4.44	4.97	5.67
247.00	422.00	695.00	1902.00		7.55	5.51	6.05	6.54
311.00	631.00	1076.00	2743.00		7.92	5.74	6.45	6.98
384.00	768.00	1372.00	3373.00		8.12	5.95	6.64	7.22

Lampiran : Listing Program dengan Software 'R'

```
VSde<-function(d, M)
{
    x1 <- c(4.63,4.61,3.99,4.49,4.11,3.76,3.26,4.29,4.62,5.61,5.93,6.38,4.44,5.51,5.74,5.95)
    x2 <- c(3.81,3.69,3.33,4.14,3.81,2.48,2.77,2.83,3.64,3.14,4.16,3.30,4.97,6.05,6.45,6.64)
    x3 <- c(4.84,4.88,4.48,5.37,4.65,4.13,4.33,4.61,3.69,4.48,4.76,4.04,5.67,6.54,6.98,7.22)
    y <- c(6.20,6.28,5.86,6.53,6.26,5.36,5.33,6.01,5.79,6.29,6.58,6.65,6.74,7.55,7.92,8.12)
    b1 <- matrix(0, 2 * 16, nrow = 2)
    b2 <- matrix(0, 2 * 16, nrow = 2)
    b3 <- matrix(0, 2 * 16, nrow = 2)
    b12 <- matrix(0, 3 * 16, nrow = 3)
    b13 <- matrix(0, 3 * 16, nrow = 3)
    b23 <- matrix(0, 3 * 16, nrow = 3)
    b123 <- matrix(0, 4 * 16, nrow = 4)
    s1 <- rep(0, M)
    s2 <- rep(0, M)
    s3 <- rep(0, M)
    s12 <- rep(0, M)
    s13 <- rep(0, M)
    s23 <- rep(0, M)
    s123 <- rep(0, M)
    I <- c(1:16)
    Is <- matrix(0, d * M, nrow = d)
    Ic <- matrix(0, (16 - d) * M, nrow = 16 - d)
    for(j in 1:M) {
        Is[, j] <- sample(I, replace = F, size = d)
        Ic[, j] <- I[ - c(Is[, j])]
        yy <- y[Ic[, j]]
        xx1 <- x1[Ic[, j]]
        xx2 <- x2[Ic[, j]]
        xx3 <- x3[Ic[, j]]
        xxx1 <- x1[Is[, j]]
        xxx2 <- x2[Is[, j]]
        xxx3 <- x3[Is[, j]]
        C <- rep(1, d)
        y1 <- y[Is[, j]]
        b1[, j] <- glm(yy ~ xx1)$coef
        X1 <- cbind(C, xxx1)
        s1[j] <- 1/(d) * sum((y1 - (X1 %*% b1[, j]))^2)
        b2[, j] <- glm(yy ~ xx2)$coef
        X2 <- cbind(C, xxx2)
        s2[j] <- 1/(d) * sum((y1 - (X2 %*% b2[, j]))^2)
        b3[, j] <- glm(yy ~ xx3)$coef
        X3 <- cbind(C, xxx3)
```

```
    s3[j] <- 1/(d) * sum((y1 - (X3 %*% b3[, j]))^2)
    b12[, j] <- glm(yy ~ xx1 + xx2)$coef
    X12 <- cbind(C, xxx1, xxx2)
    s12[j] <- 1/(d) * sum((y1 - (X12 %*% b12[, j]))^2)
    b13[, j] <- glm(yy ~ xx1 + xx3)$coef
    X13 <- cbind(C, xxx1, xxx3)
    s13[j] <- 1/(d) * sum((y1 - (X13 %*% b13[, j]))^2)
    b23[, j] <- glm(yy ~ xx2 + xx3)$coef
    X23 <- cbind(C, xxx2, xxx3)
    s23[j] <- 1/(d) * sum((y1 - (X23 %*% b23[, j]))^2)
    b123[, j] <- glm(yy ~ xx1 + xx2 + xx3)$coef
    X123 <- cbind(C, xxx1, xxx2, xxx3)
    s123[j] <- 1/(d) * sum((y1 - (X123 %*% b123[, j]))^2)
}
cat(  "MSE.1  =", 1/M * sum(s1), "MSE.2  =",1/M * sum(s2), "MSE.3  =",1/M * sum(s3), "\n",
      "MSE.12 =",1/M*sum(s12),"MSE.13 =",1/M*sum(s13),"\n","MSE.23 =",1/M*sum(s23),  "\n",
        "MSE.123  =", 1/M * sum(s123), "\n")]
```