PENGGUNAAN *BOOTSTRAP* DATA DEPENDEN UNTUK MEMBANGUN SELANG KEPERCAYAAN PADA PARAMETER MODEL PERAMALAN DATA STASIONER

Siana Halim, Herman Mallian
Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Kristen Petra Surabaya
Email: halim@petra.ac.id

ABSTRAK

Bootstrap merupakan area penelitian yang terus berkembang. Ada banyak ide dan proposal-proposal yang berbeda telah diberikan oleh para peneliti. Namun demikian, dalam makalah ini hanya akan diulas secara singkat beberapa metode *Bootstrap* untuk data independen maupun data dependen. Akhirnya akan diberikan sebuah contoh kasus penggunaan Bootstrap untuk membangun selang kepercayaan pada peramalan data stasioner.

Kata kunci: *Bootstrap*, *resampling*, peramalan

ABSTRACT

The Bootstrap is a lively research area. A lot Of ideas are around and have let to quiet different proposals. In this paper we sketch briefly some Bootstrap methods for independent and dependent data. Finally we give an Bootstrap example for constructing confidence interval in the forecasting for stationer data.

Keywords: *Bootstrap, resampling, forecasting.*

1. PENDAHULUAN

Pada saat ini *Bootstrap* sudah menjadi metode standard dalam ilmu statistika modern. Penelitian ini jauh bermula pada tahun tujuh puluh-an dari ide *resampling*. Karya *seminal* dari Efron (Efron, 1979) memberikan sintesa beberapa ide awal *resampling* dan tak dapat dipungkiri memberikan acuan baru dalam simulasi berdasarkan analisa statistik. Ide dasar dari bootstrap adalah membangun data bayangan (*pseudo* data) dengan menggunakan informasi dari data asli. Namun demikian, kita tetap harus memperhatikan sifat-sifat dari data asli tersebut, sehingga data bayangan akan memiliki karakteristik semirip mungkin dengan data asli.

Dalam makalah ini, akan diulas sekilas tentang *Bootstrap* untuk data independen dan data dependen beserta aplikasi dari Bootstrap pada data dependen untuk membangun selang kepercayaan pada estimasi parameter dari sebuah *time series* stasioner.

2. *BOOTSTRAP* UNTUK DATA INDEPENDEN

Bootstrap merupakan metode simulasi yang berbasis pada data dan seringkali digunakan sebagai alat dalam statistika inferensia. Penggunaan kata *Bootstrap* ini diambil dari frase *„to pull oneself up by one's bootstrap"* (Efron, Tibshirani, 1993). *Resampling* untuk data indenden (iid – *independent identical distributed*) merupakan metode *Bootstrap* yang paling sederhana. Misalkan $X_1, X_2, \ldots, X_n$ yang berdistribusi *P*. *Resample* untuk data *iid* dilakukan dengan cara melakukan mengambilan sampel dari data asli secara acak dengan pengembalian (*replacing sample*). *Resample* ini $X_1^*, X_2^*, \ldots, X_n^*$ dan proses resample ini dapat dilakukan secara berulang-ulang, misalnya B kali dimana $B > n$. Secara formal, proses resample ini dapat dibangun dengan membangkitkan $X_1^*, X_2^*, \ldots, X_n^*$ secara independen bersyarat (diberikan himpunan data asli) dan memiliki distribusi bersyarat $\hat{P}_n$ yang merupakan distribusi empiris. $\hat{P}_n(A) = n^{-1}\sum_{i=1}^{n} I(X_i \in A)$, dimana I merupakan fungsi indikator. Bootstrap estimate dari statistik $T(P)$ didefinisikan sebagai *plug-in* estimate dari $T(\hat{P}_n)$. Contoh menggunakan *Bootstrap* untuk mendapatkan standard error diberikan pada (Efron, Tibshirani, 1993) dan dapat diringkas dalam langkah-langkah sebagai berikut:

1. Menentukan jumlah B sampel independen *Bootstrap* $X^{*1}, X^{*2}, ..., X^{*B}$ di mana masing-masing sampel berisi n data yang diperoleh dari **x** (data awal).
2. Mengevaluasi replikasi yang ada pada masing-masing sampel *Bootstrap*

   $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$ $\qquad$ $b = 1, 2, ..., B$
3. Mengestimasi standar error $se_{F B}(\hat{\theta})$ dengan menggunakan standar deviasi untuk *Bootstrap* yang direplikasi B kali.

$$\overline{se}_B = \left\{ \frac{\sum_{b=1}^{B}\left[\hat{\theta}^*(b) - \hat{\theta}^*(.)\right]^2}{B-1} \right\}^{1/2} \qquad (1)$$

dimana: $\hat{\theta}^{*}(.) = \sum_{b=1}^{B} \hat{\theta}^*(b)/B$

Beberapa alasan mengapa Bootstrap dapat diaplikasikan dalam masalah-masalah statistik diberikan pada (Mammen, 1992).

3. *BOOTSTRAP* UNTUK DATA DEPENDEN

Bootstrap untuk data dependen merupakan area riset yang sangat berkembang. Ada banyak ide dan proposal dalam membangun *Bootstrap* untuk data dependen. Hal ini karena r*esampling* pada data *dependen* harus dibangun sedemikian rupa sehingga struktur ketergantungan antara data tidak hilang. Beberapa proposal tentang boostrap untuk data dependen antara lain, *block bootstrap* (Kuensch, 1989), *bootstrap* untuk ARMA model, (Franke, 1992), bootstrap untuk model-model *nonparametric smoothing* (Franke, 2002a and 2002b).

Dari beberapa kemungkinan di atas, hal ini paling mudah dilakukan pada kasus model-model klasik ARMA (*Auto Regressive Moving Average*) yang berdimensi hingga dengan residual *i.i.d.* (Franke, 1992). Sebuah contoh untuk model linear *autoregressive*

$$X_t - \mu_X = \sum_{j=1}^{p} \rho_j (X_{t-j} - \mu_X) + \varepsilon_t, \quad t \in Z \qquad (2)$$

di mana $\mu_X = E(X_t)$ adalah mean pengamatan (*observation mean*) dan $\{\varepsilon_t\}$ adalah deret inovasi yang bersifat *i.i.d* dengan sifat $E(\varepsilon_t) = 0$ dan ε_t independen terhadap $\{X_s, s < t\}$. *Parameter-parameter* $\rho_1, .., \rho_P$ *dapat diestimasi dengan menggunakan least square* ataupun dengan menggunakan persamaan-persamaan Yule Walker. Nilai residual dapat dicari melalui persamaan berikut

$$\tilde{\varepsilon}_t = X_t - \hat{\mu}_X - \sum_{j=1}^{p} \hat{\rho}_j (X_{t-j} - \hat{\mu}_X) \qquad (3)$$

dimana $\hat{\mu}_X = n^{-1}\sum_{t=1}^{n} X_t$ dan $\hat{\rho}_1,...,\hat{\rho}_p$ adalah nilai estimasi dari parameter-parameter tersebut. *Bootstrap resample* dilakukan dengan membangkitkan

$$X_t^* - \hat{\mu}_X = \sum_{j=1}^{p} \hat{\rho}_j (X_{t-j}^* - \hat{\mu}_X) + \varepsilon_t^* \qquad (4)$$

dimana ε_t^* dibangkitkan dengan pengembalian (*replacement*) dari residual terpusat (*centered residuals*) $\hat{\varepsilon}_t = \tilde{\varepsilon}_t - n^{-1}\sum_{i=1}^{n} \tilde{\varepsilon}_i$.

Salah satu aplikasi dari *bootstrap* dari data dependen ini adalah untuk mencari selang kepercayaan (*confidence interval*) dari parameter-parameter model peramalan yang bersesuaian. Adapun langkah-langkah yang dilakukan untuk melakukan metode *bootstap* adalah sebagai berikut:

Algoritma 1

Langkah 1: Memberikan nilai indeks 1 sampai n pada *error* hasil peramalan. Melakukan *resampling* dengan pengembalian pada index *error*. Kemudian index *error* diganti dengan nilai *error* yang sebenarnya.

Langkah 2: Menggunakan hasil perhitungan *error* pada Langkah 1 untuk membangun sejumlah 1000 sampel *Bootstrap error* $\varepsilon^{*1}, \varepsilon^{*2},..., \varepsilon^{*1000}$. Masing-masing sampel berisi n buah *random sampling error*.

Langkah 3: Membangun 1000 *time series* baru dengan menggunakan formulasi

Untuk AR: $Y_t^{*B} = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + ... + \rho_p Y_{t-p} + \varepsilon_t^{*B}$ t = 1,2,...,n

Untuk MA: $Y_t^{*B} = \varepsilon_t^{*B} - \theta_1 \varepsilon_{t-1}^{*B} - \theta_2 \varepsilon_{t-2}^{*B} - ... - \theta_q \varepsilon_{t-q}^{*B}$ t = 1,2,...,n

Untuk ARMA: $Y_t^{*B} = \varepsilon_t^{*B} + \rho_1 Y_{t-1} - \theta_1 \varepsilon_{t+l-1}^{*B}$

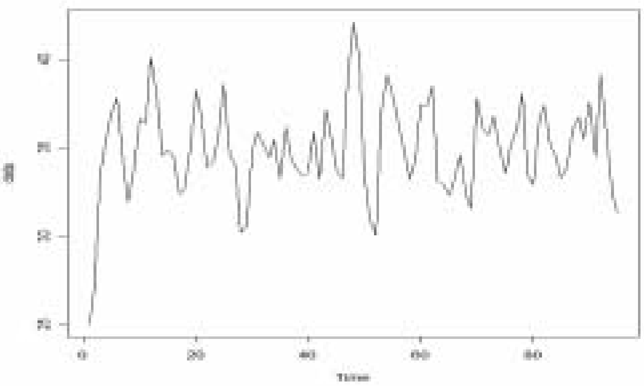
Langkah 4: Mengestimasi nilai-nilai parameter *time series* baru yang dibangun pada Langkah 3. Parameter yang dihasilkan adalah parameter yang baru dan berjumlah 1000 buah.

Langkah 5: Melakukan pengurutan nilai-nilai parameter dari yang terkecil hingga yang terbesar.

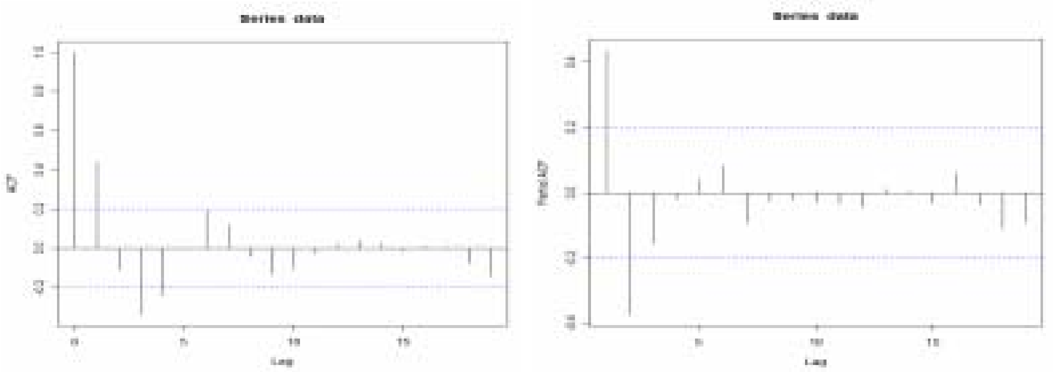
Langkah 6: Memperoleh 95% *confidence interval* dengan cara membuang sejumlah 2,5% pada urutan parameter bagian atas dan sejumlah 2,5% pada urutan parameter bagian bawah. Parameter yang baru memiliki tingkat kepercayaan 95%.

4. STUDI KASUS

Data untuk studi kasus yang digunakan pada makalah ini adalah data viskositas (Bowerman and O'Connel, 1993, p.471). Plot dari data beserta sample *Autocorrelation Function* (ACF) dan sample *Partial Autocorrelation Function* (PACF) diberikan pada gambar berikut.



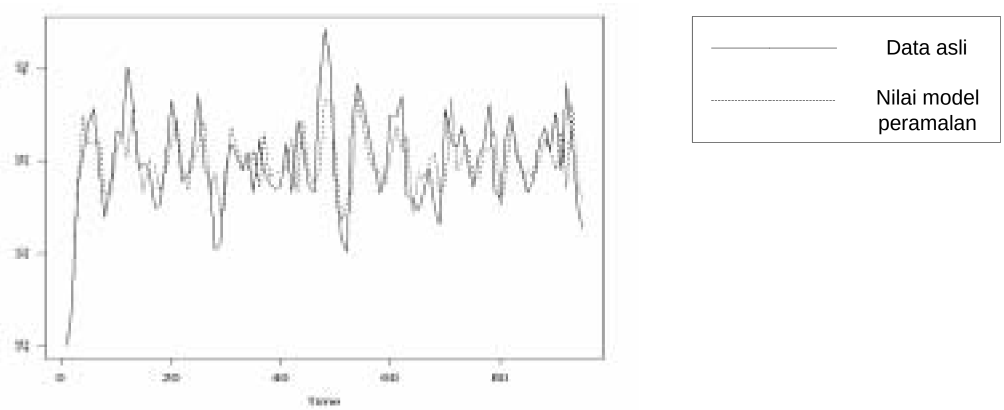
Gambar 1. Plot data viskositas (Bowerman and O'Connel, 1993, p.471)



Gambar 2. (kiri) Plot Sample ACF, (kanan) Plot sample PACF

Digunakan program R untuk mendapatkan model terbaik berdasarkan nilai minimum AIC *Akaike Information Criterion*. (Mallian,2006). Model peramalan yang terbaik untuk data di atas adalah AR(2) dengan nilai parameter-parameternya adalah sebagai berikut

- Parameter 1 (ρ_1) : 0,6821, dengan *confidence interval* [0.459, 0.9053]
- Parameter 2 (ρ_2) : -0,4333, dengan *confidence interval* [-0.6695, -0.197]



Gambar 3. Plot Data Awal dengan Nilai Dari Persamaan Model Peramalan

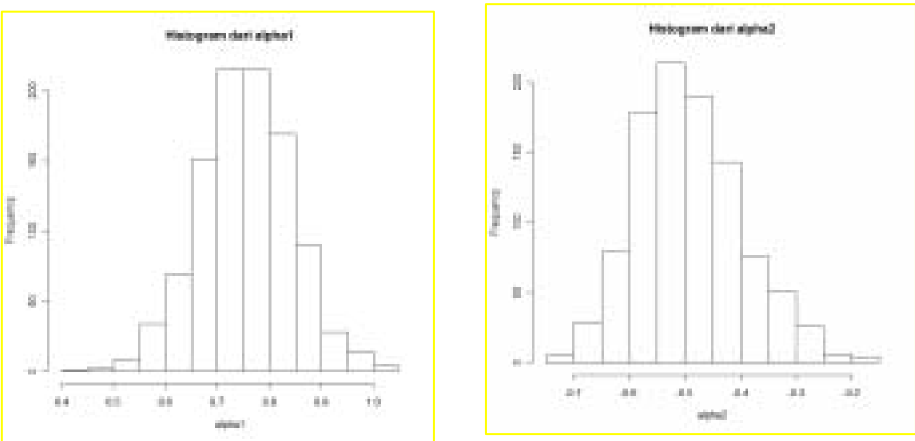
Tabel 1. Selang Kepercayaan Parameter Model Peramalan dan Nilai Peramalan Masa Mendatang Hasil Analitik

	Nilai	Selang Kepercayaan	Range
Parameter 1	0.6821	[0.459,0.9053]	0.4463
Parameter 2	-0.433	[-0.6695,-0.197]	0.4725
1 Periode Mendatang	33.56	[29.2944,37.8207]	8.5263
2 Periode Mendatang	35.59	[30.4253,40.7462]	10.3209
3 Periode Mendatang	35.98	[30.8221,41.1465]	10.3244

Selanjutnya dari model terbaik untuk contoh kasus ini, pembangunan selang kepercayaan dengan Bootstrap untuk model AR(2). Hal ini dapat dilakukan dengan mengikuti Algoritma 1, dengan mengganti Langkah 3 menjadi

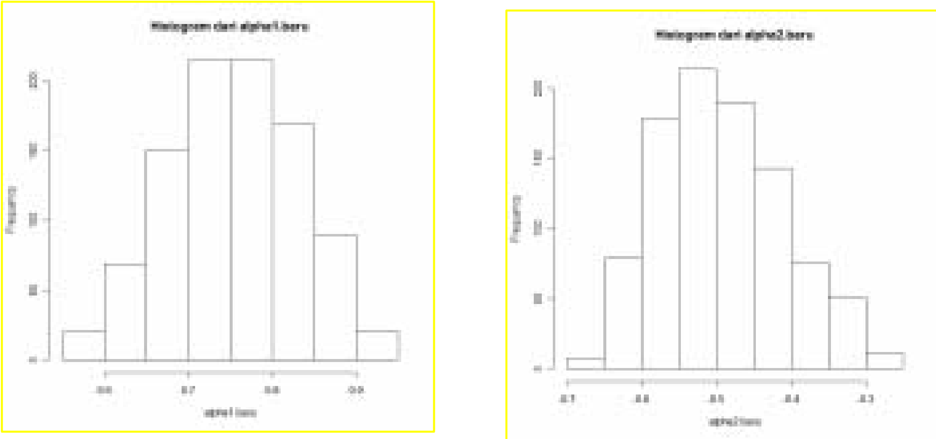
$$Y_t^{*B} = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \varepsilon_t^{*B} \quad t = 1,2,...,n$$

Parameter-parameter model peramalan hasil metode *Bootstrap* dapat dianalisa dengan melihat histogramnya. Hasil dari histogram di atas menunjukkan bahwa parameter ρ_1, ρ_2 terpusat di tengah dan menyerupai kurva normal. Namun histogram masih tampak sedikit miring. Untuk itu *Bootstrap* parameter setelah diurutkan, nilainya dibuang 2.5% di kiri dan 2.5% di kanan. Histogram untuk nilai-nilai parameter setelah pemotongan ini dapat dilihat pada Gambar 5.



Gambar 4. Hasil Simulasi Nilai Parameter dengan Menggunakan *Bootstrap*

Selanjutnya selang kepercayaan dari parameter-parameter ini dilihat dari nilai terkecil dari parameter-parameter *Bootstrap* setelah dipotong di 2.5% di kiri dan 2.5% di kanan. Selang kepercayaan untuk parameter-parameter beserta dengan nilai peramalan untuk 3 masa mendatang dapat di lihat pada Tabel. 2. Dari sini terlihat bahwa *range* dari selang kepercayaan *Bootstrap* lebih sempit bila dibandingan dengan selang kepercayaan yang dihitung secara analitik (lihat Box, 1976). Berdasarkan analisa statistik hal ini berarti, bila selang kepercayaan semakin sempit maka daerah penolakan dari uji hipotesa normal akan semakin lebar, maka dapat disimpulkan bahwa selang kepercayaan yang dibangun dengan *Bootstrap* lebih akurat.



Gambar 5. Hasil Simulasi Nilai Parameter dan Nilai Peramalan Menggunakan *Bootstrap* yang telah dipotong 2.5% di kiri dan 2.5% di kanan.

Tabel 2. Selang Kepercayaan Parameter Model Peramalan dan Nilai Peramalan Masa Mendatang Hasil *Bootstrap*

	Nilai	Selang Kepercayaan	Range
Parameter 1	0.6821	[0.65248,0.95759]	0.3051
Parameter 2	-0.433	[-0.69518,-0.34386]	0.3513
1 Periode Mendatang	33.56	[32.8918, 34.1026]	1.211
2 Periode Mendatang	35.59	[34.5967,36.5785]	1.9818
3 Periode Mendatang	35.98	[35.336, 36.8765]	1.5405

5. KESIMPULAN

Pada makalah ini telah diulas secara singkat *Bootstrap* untuk data *i.i.d* maupun data dependen, beserta aplikasinya untuk membangun selang kepercayaan pada ARMA(p,q). Pada kasus di atas 95% selang kepercayaan yang diberikan melalui *Bootstrap* lebih akurat bila dibandingkan dengan perhitungan secara analitik. Sifat-sifat secara analitik dari *Bootstraping* model-model ARMA dapat dilihat pada (Franke, 1992).

DAFTAR PUSTAKA

Bowerman, Bruce L. and O'Connell, Richard T., 1993, *Forecasting And Time Series: An Applied Approach*, 3P^{rdP} edition, Duxbury Press.

Box, G.E.P, Jenkins, G.M, 1976, *Time Series Analysis and Control*, Revised Edition, Holden-Day, California.

Efron, B., 1979, "Bootstrap methods : Another look at jackknife", *Annals Statistics*, 7:1-26.

Efron, B. dan Tibshirani, R.J, 1993, *An Introduction to the Bootstrap*, Chapman and Hall, London.

Franke, J. and Kreiss, J.P., 1992, ”Bootstraping ARMA models”, *Journal of Time Series Analysis*, 13:297-317.

Franke, J,. Kreiss, J.P. and Mammen, E., 2002a. ”Bootstrap of kernel smoothing in nonlinear time series”, *Bernoulli*, 8:1-37.

Franke,J,. Kreiss, J.P., Mammen, E., and Neumann, M.H, 2002b, ”Properties of the Nonparametric Autoregressive Bootstrap”, *Journal of Time Series Analysis*, 23:555-585.

Kuensch, H.R., 1989, “The Jackknife and the bootstrap for general stationary observations”, *Annals of Statistics*, 17:1217-1241.

Mallian, H., 2006, *Studi Literatur Tentang Model Peramalan ARMA (p,q) dan Selang Kepercayaan Parameter Model dengan Menggunakan Bootstrap*, Tugas Akhir Jurusan Teknik Industri, Universitas Kristen Petra.

Mammen, E., 1992, *When does bootstrap work ? Asymptotic results and simulations,* Springer Lecture Notes in Statistics 77, Springer, Heidelberg, Berlin.