Prosiding SPMIPA. pp. 157-162. 2006 *ISBN : 979.704.427.0*

PREDIKSI PRODUKSI JAGUNG DI JAWA TENGAH DENGAN *ARIMA* DAN *BOOTSTRAP*

Sri Rahayu, Tarno
Jurusan Matematika FMIPA UNDIP Semarang
Jl. Prof. Soedarto, Kampus UNDIP Tembalang, Semarang

Abstrak: Dalam tulisan ini dibahas tentang penerapan metode *ARIMA* dan *Bootstrap* untuk prediksi produksi jagung di Jawa Tengah sampai dengan tahun 2009. Jagung merupakan salah satu bahan pangan alternative pengganti beras. Data jumlah produksi jagung dari masa ke masa merupakan data runtun waktu sehingga untuk memprediksikan jumlah produksi yang akan datang digunakan teknik-teknik analisis runtun waktu. Salah satu metode yang paling sering digunakan dalam pemodelan time series untuk peramalan adalah *Autoregressive Integrated Moving Average (ARIMA)* Box-Jenkins. Namun seringkali data yang diperoleh berukuran relatif kecil, sehingga sulit untuk menjamin dipenuhinya asumsi-asumsi dalam analisis statistik klasik. Sebagai akibatnya inferensi statistik tidak dapat dilakukan dengan baik terhadap parameter model. Oleh sebab itu, diperlukan suatu pendekatan non parametrik yang bebas asumsi, salah satunya adalah *metode Bootstrap*. Metode *Bootstrap* merupakan metode berbasis komputer yang berguna untuk menaksir berbagai kuantitas statistik seperti mean, standar error, dan bias suatu estimator atau untuk estimasi interval serta untuk mengestimasi distribusi suatu statistik. Penerapan kedua metode: ARIMA dan Bootstrap untuk pemodelan dan prediksi produksi jagung di Jawa Tengah sampai dengan tahun 2009 memberikan hasil yang hampir sama yaitu untuk estimasi parameter model maupun standar errornya, sehingga hasil prediksi dengan kedua metode tersebut juga hampir sama. Metode bootstrap mempunyai kelebihan dibandingkan dengan ARIMA karena distribusi parameter modelnya dapat ditentukan.

Kata Kunci: Prediksi, *ARIMA*, *Bootstrap*

PENDAHULUAN

Pangan merupakan kebutuhan pokok yang harus selalu terpenuhi untuk menjaga kelangsungan hidup manusia. Jagung merupakan salah satu bahan makanan pokok setelah beras. Jawa Tengah merupakan salah satu propinsi penghasil jagung yang cukup besar di Indonesia. Beberapa tahun ke depan produksi jagung dipandang perlu untuk diprediksikan agar dapat dibuat suatu perencanaan yang matang terkait dengan ketersediaan dan kebutuhan akan jagung sebagai bahan pangan alternative pengganti beras. Data jumlah produksi jagung dari masa ke masa merupakan data runtun waktu sehingga untuk memprediksikan jumlah produksi yang akan datang digunakan teknik-teknik analisis runtun waktu.

Salah satu metode yang paling sering digunakan dalam pemodelan runtun waktu untuk peramalan adalah *Autoregressive Integrated Moving Average (ARIMA)* Box-Jenkins. Agar model ARIMA Box-Jenkins menghasilkan ramalan yang optimal, maka model tersebut harus memenuhi asumsi residual white noise dan berdistribusi normal. Namun kadangkala data yang diperoleh berukuran relatif kecil, sehingga sulit untuk menjamin dipenuhinya asumsi-asumsi dalam analisis statistik klasik. Sebagai akibatnya inferensi statistik tidak dapat dilakukan terhadap parameter model (http://digilib.its.ac.id).

Untuk mengatasi masalah tersebut para statistikawan menempuh dengan cara memperbesar ukuran sampel. Padahal penambahan sampel ini kadang-kadang sulit dan bahkan tidak memungkinkan untuk dilakukan sehingga dinilai kurang efisien. Oleh karena itu permasalahan yang akan diuraikan disini adalah bagaimana menentukan suatu model terbaik dalam pemodelan time series dengan jumlah data yang berukuran relatif kecil.

Masalah inferensi statistik ini sering kali mengakibatkan estimasi beberapa aspek dari suatu sebaran peluang F didasarkan pada suatu sampel acak yang ditarik dari $\hat{F}_n$. Fungsi sebaran empiris $\hat{F}_n$, adalah estimasi sederhana dari sebaran F. Suatu cara yang tepat untuk mengestimasi beberapa aspek yang diinginkan dari F, seperti mean, median, atau korelasi adalah dengan menggunakan aspek dari $\hat{F}_n$. Hal inilah yang disebut prinsip penggantian (*plug-in*). Metode bootstrap adalah suatu aplikasi langsung dari prinsip penggantian tersebut. Metode Bootstrap merupakan suatu metode pendekatan non parametrik yang bebas asumsi, berbasis komputer dan berguna untuk menaksir berbagai kuantitas statistik seperti mean, standar error, dan bias suatu estimator atau untuk membentuk interval konfidensi serta untuk mengestimasi distribusi suatu statistik. Oleh karena itu, dalan tulisan ini dibahas tentang penerapan metode Bootstrap untuk prediksi produksi jagung di Jawa Tengah selama beberapa tahun ke depan serta membandingkan hasilnya dengan menggunakan metode ARIMA Box-Jenkins.

METODE ARIMA BOX-JENKINS

Model-model *Autoregressive Integrated Moving Average (ARIMA)* telah dipelajari secara mendalam oleh George Box dan Gwilym Jenkins (1976), dan nama mereka sering disinonimkan dengan proses ARIMA yang telah diterapkan untuk analisis time series. Menurut Box-Jenkins terdapat beberapa tahapan dalam memodelkan suatu data time series yaitu : identifikasi *(identification)*, estimasi *(estimation)* , verifikasi *(diagnostic checking)*, dan peramalan *(forecasting).*

Identifikasi Model

Langkah ini bertujuan untuk mengetahui apakah model runtun waktu stasioner atau nonstasioner. Jika prosesnya nonstasioner maka untuk menjadikannya stasioner harus dilakukan differensi/transformasi baru kemudian diidentifikasi. Berdasarkan pengamatan terhadap grafik f.a.k dan f.a.k.p yang diperoleh dari data runtun waktu maka diharapkan dapat dikenali pola runtun waktu yang kemudian dituangkan kedalam model umum.

Adapun gambaran perincian secara umum model AR(p), MA(q), ARMA(p,q) disajikan pada tabel 1 berikut.

Tabel 1. Berbagai ciri bentuk ACF dan PACF untuk model AR, MA, ARMA

Model	ACF	PACF
AR(p)	Menurun dengan cepat secara exponensial atau seperti gelombang sinus yang melemah	Terpotong setelah lag p
MA(q)	Terpotong setelah lag q	Menurun dengan cepat secara exponensial atau seperti gelombang sinus yang melemah
ARMA(p,q)	Menurun dengan cepat setelah lag (q-p)	Menurun dengan cepat setelah lag (p-q)

Estimasi Parameter

Setelah diperoleh satu atau beberapa model sementara maka langkah selanjutnya adalah mencari estimasi untuk parameter-parameter dalam model itu. Hasil estimasi parameter yang diperoleh kemudian diuji untuk mengetahui apakah parameter tersebut signifikan atau tidak.

Adapun hipotesisnya yaitu :

H_0 : Parameter = 0

H_1 : Parameter ≠ 0

dengan statistik uji :

$$T_{hitung} = \frac{parameter \quad estimasi}{SE \quad parameter \quad estimasi}$$

Sedangkan kriteria ujinya adalah :

$\left|T_{hitung}\right| < T_{\alpha/2, n-1}$, H_0 diterima

$\left|T_{hitung}\right| > T_{\alpha/2, n-1}$, H_0 ditolak

Bila H_0 diterima, berarti parameternya tidak signifikan, dan jika H_0 ditolak, berarti parameternya cukup signifikan.

Verifikasi Model

Langkah ini adalah langkah pemeriksaan apakah model yang diestimasi cukup cocok dengan data yang ada. Tahap verifikasi didasarkan pada analisis residual $\hat{\varepsilon}_t$. Asumsi dasar dalam model ARIMA adalah bahwa ε_t merupakan variabel random independen berdistribusi normal dengan mean nol dan varian konstan. Sehingga dalam data runtun waktu diharapkan tingkah laku residual $\hat{\varepsilon}_t$ mirip dengan ε_t.

Untuk memeriksa apakah $\hat{\varepsilon}_t$ berdistribusi normal dapat digunakan uji Box-Ljung. Uji ini menggunakan autokorelasi sample dari residual untuk memeriksa hipotesis null. Adapun hipotesisnya adalah :

$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$ (autokorelasi residual tidak signifikan)

$H_1 : \rho_1 \neq \rho_2 \neq \ldots \neq \rho_k \neq 0$ (autokorelasi residual signifikan)

dengan statistik uji :

$$Q = n(n+2)\sum_{k=1}^{K} \frac{r_k^{\;2}}{n-k},$$ dengan:

n	:	banyaknya data
K	:	lag maksimum
$r_k^{\;2}$	:	nilai kuadrat dari koefisien autokorelasi pada periode k

Jika $Q < \chi^2_{(\alpha, K-p-q)}$ atau $p - value > \alpha$ maka H_0 diterima dan ini berarti nilai autokorelasi residual sama dengan nol sehingga residual independen.

Dalam pemilihan model terbaik didasarkan pula pada nilai MSE terkecil dan prinsip parsimoni yaitu model dengan parameter sesedikit mungkin lebih disenangi daripada model dengan parameter yang banyak.

Peramalan

Setelah diperoleh model terbaik, selanjutnya model tersebut digunakan untuk meramalkan keadaan pada masa yang akan datang.

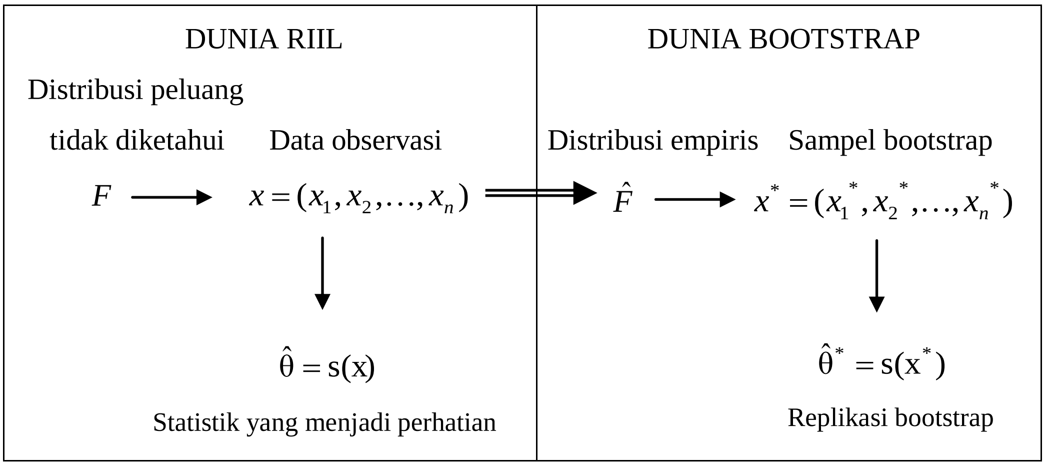
METODE BOOTSTRAP

Teori Bootstrap

Metode bootstrap pertama kali diperkenalkan oleh Bradley Efron pada tahun 1979. Metode Bootstrap pada dasarnya adalah melakukan pengambilan sampel (*resampling*) dengan pengembalian dari sampel hasil observasi dengan replikasi B kali ($n << B << n^n$) dengan n adalah ukuran sampel. Metode Bootstrap merupakan suatu metode pendekatan nonparametrik untuk menaksir berbagai kuantitas statistik seperti mean, standar error, dan bias suatu estimator atau untuk membentuk interval konfidensi dengan memanfaatkan kecanggihan teknologi komputer. Metode bootstrap dapat juga digunakan untuk mengestimasi distribusi suatu statistik. Distribusi ini diperoleh dengan menggantikan distribusi populasi yang tidak diketahui dengan ditribusi empiris berdasarkan data sampel, kemudian melakukan pengambilan sampel (*resampling*) dengan pengembalian dari distribusi empiris yang selanjutnya dipergunakan untuk mencari penaksir bootstrap. Dengan metode bootstrap tidak perlu melakukan asumsi distribusi dan asumsi-asumsi awal untuk menduga bentuk distribusi dan pengujian-pengujian statistiknya.

Masalah inferensi statistik sering kali mengakibatkan pengestimasian beberapa aspek dari suatu sebaran peluang F berdasarkan pada suatu sampel acak yang ditarik dari F. Fungsi sebaran empiris $\hat{F}_n$, adalah estimasi sederhana dari sebaran F. Suatu cara yang tepat untuk mengestimasi beberapa aspek yang diinginkan dari F, seperti mean, median, atau korelasi adalah dengan menggunakan aspek dari $\hat{F}_n$. Inilah yang disebut prinsip penggantian (plug-in). Metode bootstrap adalah suatu aplikasi langsung dari prinsip penggantian ini.

Untuk menjelaskan metode bootstrap dapat dibayangkan sebagai suatu masalah real (nyata) dan suatu masalah buatan yang sangat mirip atau bisa dikatakan identik. Masalah buatan inilah yang disebut dengan masalah bootstrap. Skema berikut dapat menjelaskan gambaran dari metode bootstrap tersebut.



Skema dari metode bootstrap untuk kasus satu sampel. Dalam dunia real distribusi peluang yang tidak diketahui F memberikan data x = (x_1, x_2, ..., x_n) melalui resampling random, dari x dihitung statistik yang menjadi perhatian $\hat{\theta} = s(x)$. Dalam dunia bootstrap, F membangkitkan x^* melalui resampling random, memberikan $\hat{\theta}^* = s(x^*)$

Dalam masalah real terdapat data observasi $x_1, x_2, \ldots, x_n$ sebagai sampel acak berukuran n dari populasi dengan fungsi distribusi kontinu $F(x)$ yang tidak diketahui. Didefinisikan suatu parameter θ dan diestimasi dengan $\hat{\theta} = s(x_1, x_2, \ldots, x_n)$, jelas bahwa kuantitas ini diketahui karena data $x_1, x_2, \ldots, x_n$ telah diobservasi. Sedangkan masalah bootstrap meniru masalah real tetapi distribusi $F(x)$ diganti dengan fungsi distribusi empiris $\hat{F}_n(x)$ dan parameternya didefinisikan melalui fungsi distribusi empiris

$$\hat{F}_n(x) = \frac{1}{n}\{\#(x_i \le x), \quad 1 \le i \le n\} \qquad untuk\ -\infty < x < \infty$$

Lebih lanjut disimulasikan dalam jumlah yang sama observasi independent dari $\hat{F}_n(x)$ dan dinyatakan sebagai sampel bootstrap, $x_1^*, x_2^*, \ldots, x_n^*$ dengan $x_i^* \in \hat{F}_n(x)$

Sampel bootstrap didefinisikan sebagai sampel random berukuran n yang diambil dari $\hat{F}_n(x)$ dengan pengembalian. Jadi bila dipunyai sampel random berukuran n akan didapat kemungkinan sampel bootstrap sebanyak n^n, dari tiap sampel bootstrap ini dapat ditentukan $\hat{\theta}^*$. Seluruh kemungkinan sampel ini dinamakan jumlah sampel bootstrap ideal.

Data ini dibangkitkan dari $\hat{F}_n(x)$ dengan pengembalian, kemudian dihitung estimasi parameter

$$\hat{\theta}^* = s\left(x_1^*, x_2^*, \ldots, x_n^*\right)$$

dengan estimator yang sama seperti dalam masalah real. Penekanan pada hasil pembootstrapan adalah sebaran frekuensi relatif dari $\hat{\theta}^*$ yang dihitung dari resampel merupakan dugaan dari sebaran resampling $\hat{\theta}$.

Perhitungan $\hat{\theta}^*$ berdasarkan semua kemungkinan sampel bootstrap memerlukan waktu yang cukup lama. Sehingga untuk mencapai efisiensi dalam perhitungan digunakan metode pendekatan yaitu simulasi montecarlo, dengan simulasi tersebut prosedur resampling pada metode bootstrap dapat dikurangi menjadi $n \le B \le n^n$, sejumlah B yang cukup besar tetapi jauh lebih kecil jika dibandingkan dengan jumlah sampel bootstrap ideal.

Secara umum langkah-langkah dasar metode Bootstrap menurut Efron yaitu:

1. Menentukan distribusi empiris $\hat{F}_n(x)$ bagi sampel dengan peluang 1/n untuk masing-masing x_i
2. Menentukan sampel bootstrap $x_1^*, x_2^*, \ldots, x_n^*$ yang diambil dari x_i dengan pengembalian
3. Menentukan replikasi bootstrap $\hat{\theta}^*$ berdasarkan sampel bootstrap
4. Ulangi langkah 2 dan 3 sebanyak B kali, untuk B yang cukup besar

5. Berikan probabilitas untuk $B\hat{\theta}^*$ dengan menempatkan peluang 1/B bagi masing-masing $\theta_1^*, \theta_2^*, \ldots, \theta_B^*$. Distribusi ini adalah estimasi bootstrap untuk distribusi sampling $\hat{\theta}$.

Metode Bootstrap untuk Time Series

Pada metode bootstrap untuk time series digunakan dua pendekatan yaitu residual resampling dan moving blocks bootstrap [1]. Pendekatan residual resampling pada dasarnya adalah melakukan pengambilan sampel (*resampling*) dengan pengembalian dari sampel residual dengan replikasi B kali ($n << B << n^n$) dengan n adalah ukuran sampel. Jika diketahui model runtun waktu :

$$Z_t = \phi Z_{t-1} + \varepsilon_t \qquad \ldots \qquad (3.1.1)$$

dengan Z_t adalah nilai pada periode t, Z_{t-1} : nilai pada periode $t-1$, ϕ : parameter yang tidak diketahui $(-1 < \phi < 1)$, ε_t : error random yang diasumsikan berasal dari distribusi F yang tidak diketahui dengan nilai harapan 0 maka untuk memperoleh estimasi parameter dalam model runtun waktu dapat dilakukan prosedur bootstrap sebagai berikut :

a. Model runtun waktu (3.1.1) mempunyai 2 komponen yang tidak diketahui yaitu (ϕ, F_ε) dengan F_ε merupakan distribusi dari ε_t yang tidak diketahui
b. Parameter ϕ diestimasi dengan $\hat{\phi}$ dengan menggunakan metode kuadrat terkecil (least square). Jika $\hat{\phi}$ telah diketahui maka kita dapat menghitung $\hat{\varepsilon}_t$ yaitu $\hat{\varepsilon}_t = Z_t - \hat{\phi} Z_{t-1}$. Sehingga F_ε dapat diestimasi dengan fungsi distribusi empiris dari $\hat{\varepsilon}_t$ yaitu dengan mengambil masa peluang 1/n terhadap $\hat{\varepsilon}_t - \bar{\varepsilon}_t$ dimana $\bar{\varepsilon}_t = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t$. $\hat{F}_\varepsilon$ dipusatkan pada 0 karena F_ε mempunyai mean 0.
c. Data Bootstrap dibangkitkan dari model tersebut dengan (ϕ, F_ε) diganti dengan ($\hat{\phi}, \hat{F}_\varepsilon$). Dengan kata lain dibangkitkan data independen dan berdistribusi identik $\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*, \ldots, \varepsilon_n^*$ dari $\hat{F}_\varepsilon$ dan didefinisikan $Z_t^* = Z_t + \varepsilon_t^*$
d. Dengan metode kuadrat terkecil dihitung $\hat{\phi}^*$ berdasarkan data $(Z_1^*, Z_1), (Z_2^*, Z_2), \ldots, (Z_n^*, Z_n)$ yaitu $\hat{\phi}^* = (Z^T Z)^{-1} Z^T z$ dengan $z = (Z_1^*, Z_2^*, \ldots, Z_n^*)$
e. Ulangi langkah diatas sebanyak B kali sebagai replikasi Bootstrap.

SIMULASI

Sebagai suatu implementasi secara praktis dilakukan simulasi terhadap “data produksi jagung” yang diambil dari BPS Jawa Tengah dengan menggunakan metode ARIMA Box-Jenkins dan metode bootstrap residual. Dari hasil pengolahan data dengan minitab diperoleh model terbaik yang sesuai dengan data jumlah produksi jagung di Jawa Tengah yaitu model ARIMA (2,1,0) yaitu :

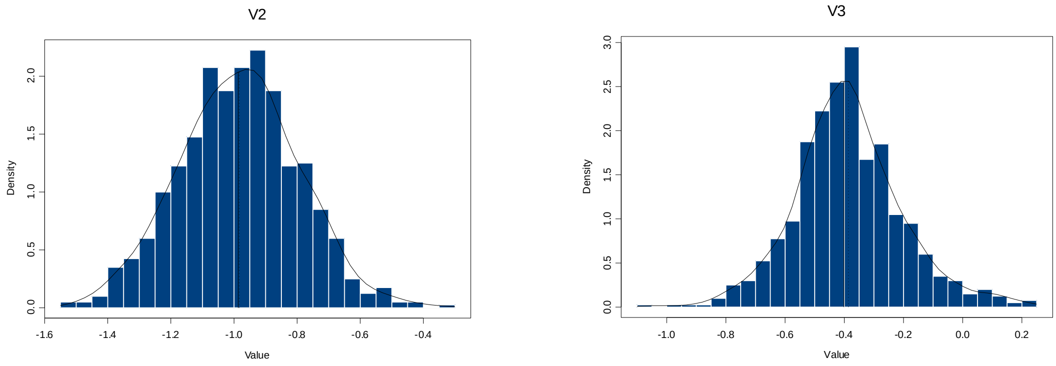
$$W_t = -0.9881 W_{t-1} - 0.4031 W_{t-2}$$
$$Z_t = 0.0119 Z_{t-1} + 0.5850 Z_{t-2} + 0.4031 Z_{t-3}$$

Selanjutnya sesuai dengan metode Bootstrap dilakukan pembangkitan data independen berdistribusi identik ε_t^* sehingga sampel Bootstrap Z^* dapat diperoleh dengan mensubstitusikan residual ε_t^* kedalam persamaan baru hasil estimasi. Dalam hal ini jumlah replikasi yang diambil adalah sebesar 1000 karena nilai ini dianggap jauh lebih besar dari jumlah data asli (n=30) namun jauh lebih kecil dari jumlah sampel Bootstrap ideal yaitu n^n.

Dari hasil pengolahan data dengan menggunakan S-Plus diperoleh model terbaik yang representatif yaitu model :

$$W_t = -0.9864 W_{t-1} - 0.4014 W_{t-2}$$
$$Z_t = 0.0136 Z_{t-1} + 0.585 Z_{t-2} + 0.4014 Z_{t-3}$$

Selain itu dengan metode Bootstrap dapat diketahui juga jenis distribusi dari masing-masing parameternya seperti dalam gambar berikut :



Gambar 1. Distribusi dari estimasi parameter model dengan bootstrap.

Hasil lengkap estimasi parameter model dengan metode ARIMA dan Bootstrap disajikan dalam tabel 2 berikut

Tabel 2. Estimasi parameter model dan Standart Errornya.

	Parameter	Metode ARIMA	Metode Bootstrap
Koefisien	ϕ_1	-0.9881	-0.9864
	ϕ_2	-0.4031	-0.4014
SE koefisien	ϕ_1	0.1788	0.1945
	ϕ_2	0.1823	0.1814
MSE		70773738751	70800546563

Sedangkan hasil peramalan untuk 5 tahun kedepan dengan kedua metode diberikan pada table 3 berikut

Tabel 3. Hasil prediksi produksi jagung sampai dengan tahun 2009

Tahun	Metode ARIMA	Metode Bootstrap
2005	1755653	1760104
2006	1871557	1871230
2007	1789514	1790740
2008	1823860	1826994
2009	1822994	1822650

KESIMPULAN

Berdasarkan hasil pengolahan data dengan metode Bootstrap ternyata didapatkan nilai standar error dan taksiran parameter serta MSE yang secara statistik tidak berbeda dengan hasil ARIMA Box-Jenkins. Namun dengan metode Bootstrap dapat diketahui distribusi dari parameternya.

DAFTAR PUSTAKA

[1]. Efron, Bradley, and Tibshirani RJ, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.

[2]. Makridakis. S, Wheelwright, and McGee, *Metode dan Aplikasi Peramalan*, Erlangga, Jakarta, 1993.

[3]. Soejoeti, Zanzawi, *Materi Pokok Runtun Waktu*, Karunika, Jakarta, 1987.

[4]. Tarno dan Subanar, Pemilihan Model Regresi Linier dengan Bootstrap, *Jurnal Matematika dan Komputer*, UNDIP Semarang, Vol. 4, No. 1, 46-58, 2001.