JURNAL GAUSSIAN, Volume 1, Nomor 1, Tahun 2012, Halaman 93-102
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

ANALISIS INDEKS HARGA SAHAM GABUNGAN (IHSG) DENGAN MENGGUNAKAN MODEL REGRESI KERNEL

Icha Puspitasari[1], Suparti[2], Yuciana Wilandari[3]
[1]Mahasiswa Jurusan Statistika FSM UNDIP
[2,3]Staff Pengajar Jurusan Statistika FSM UNDIP

ABSTRAK

Saham merupakaninvestasi yang banyak dipilih para investor, salah satu indikator yang menunjukkan pergerakan harga saham adalah Indeks Harga Saham Gabungan (IHSG). IHSG merupakan data runtun waktu sehingga untuk menganalisisnya dapat menggunakan metode runtun waktu klasik. Namun dengan metode tersebut banyak asumsi yang harus dipenuhi, sehingga diperlukan metode alternatif salah satunya metode regresi nonparametrik karena dalam model regresi nonparametrik tidak ada asumsi khusus sehingga model ini merupakan metode alternatif yang dapat digunakan dalam analisis IHSG. Dalam makalah ini dibandingkan nilai MSE yang dihasilkan dari analisis runtun waktu klasik, regresi parametrik linier sederhana dan regresi nonparametrik kernel. Data IHSG yang digunakan adalah periode minggu pertama Januari 2011 sampai dengan minggu ke empat Februari 2012. Data tersebut merupakan data *closing price* saham mingguan pada periode perdagangan terakhir. Hasil perbandingan nilai MSE dari dataIHSG yang sering fluktuatif pada tiga analisis didapatkan nilai MSE terkecil adalah pada analisis menggunakan regresi nonparametrik kernel dengan fungsi triangle dan badwidth h sebesar 58.2 dengan nilai MSE = 6987.787. Model terbaik tersebut dapat digunakan untuk memprediksikan nilai IHSG selanjutnya.

Kata Kunci : IHSG, *time series,* regresi parametrik, regresi nonparametrik, kernel.

1. PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan ekonomi yang terjadi di Indonesia saat ini cukup pesat dan telah mengubah pola pikir masyarakat di bidang ekonomi umumnya dan bidang investasi pada khususnya. Investasi dalam bentuk saham merupakan investasi yang banyak dipilih para investor. Salah satu indikator yang menunjukkan pergerakan harga saham adalah Indeks Harga Saham Gabungan (IHSG).MenurutSunariyah (2003)**,**Indeks Harga Saham Gabungan (IHSG) adalah suatu rangkaian informasi historis mengenai pergerakan harga saham gabungan sampai tanggal tertentu dan mencerminkan suatu nilai yang berfungsi sebagai pengukuran kinerja suatu saham gabungan di bursa efek.IHSG sering digunakan sebagai acuan para investor guna melihat representasi harga saham keseluruhan sehingga untuk menganalisis kemungkinan kenaikan atau penurunan harga saham diperlukan suatu metode analisis.

Data *time series* IHSG yang sering fluktuatif apabila dianalisis dengan menggunakan runtun waktu ditemukan banyak permasalahan karena adanya asumsi-asumsi yang harus dipenuhi. Selain dapat dianalisis dengan menggunakan analisis runtun waktu, dapat juga dianalisis dengan menggunakan analisis regresi.Pada regresi parametrik ditemukan permasalahan karena adanya asumsi-asumsi yang harus dipenuhi atau apabila semua asumsi terpenuhi akan menghasilkan error yang besar. Sehingga perlu digunakan analisis yang tidak ada asumsi-asumsi yang harus dipenuhi, salah satunya adalah regresi nonparametrik.Berdasarkan beberapa metode model regresi tersebut maka dalam skripsi ini akan dibahas tentang regresi kernel dengan Estimator Nadaraya-Watson dan menggunakan beberapa macam fungsi kerneldengan meminimalkan nilai CV untuk mendapatkan *bandwidth* optimal dan sebagai pembanding dilakukan analisis dengan menggunakan analisis *time series* dan regresi linier klasik.

1.2 Tujuan Penulisan

1. Menggunakan model regresi kernel untuk menganalisis data IHSG yang fluktuatif.
2. Menentukan parameter penghalus yang optimal (*bandwidth optimum*) dengan meminimalkan *cross validation*.
3. Membandingkan nilai MSE dari analisis *time series,* regresi parametrik dan regresi nonparametrik kernel.
4. Melakukan prediksi IHSG dengan menggunakan regresi kernel.

2. TINJAUAN PUSTAKA

2.1 Indeks Harga Saham

Indeks saham adalah harga saham yang dinyatakan dalam angka indeks.Indeks saham digunakan untuk tujuan analisis dan menghindari dampak negatif dari penggunaan harga saham dalam rupiah.Setiap bursa efek akan menetapkan angka basis indeks yang berbeda, yaitu ada yang dimulai dengan basis 100, 500, atau 1.000**.**Penentuan indeks harga saham dapat dibedakan menjadi dua macam, yaitu yang disebut dengan Indeks Harga Saham Individu dan Indeks Harga Saham Gabungan.Situasi pasar secara umum baru dapat diketahui jika Indeks Harga Saham Gabungan diketahui. Indeks Harga Saham Gabungan (IHSG) berubah setiap hari karena perubahan harga pasar setiap hari dan adanya saham tambahan.

2.2 Runtun Waktu

Dasar pemikiran *time series* adalah pengamatan sekarang (Z_i) tergantung pada satu atau beberapa pengamatan sebelumnya (Z_{i-k}).

Model-Model Runtun Waktu

1. Model Autoregresif (AR)

Bentuk umum suatu proses autoregresif tingkat p $\{AR(p)\}$ adalah

$$Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + ... + \phi_p Z_{i-p} + \varepsilon_i$$

2. Model Moving Average (MA)

Bentuk umum model Moving Average tingkat q atau MA(q) adalah

$$Z_i = \varepsilon_i + \theta_1 \varepsilon_{i-1} + \theta_2 \varepsilon_{i-2} + \cdots + \theta_q \varepsilon_{i-q}$$

3. Model ARMA(p,q) (Autoregresif orde p dan Moving Average orde q)

Suatu perluasan yang dapat diperoleh dari model AR dan MA adalah model campuran yang berbentuk

$$Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + ... + \phi_p Z_{i-p} + \varepsilon_i + \theta_1 \varepsilon_{i-1} + \theta_2 \varepsilon_{i-2} + ... + \theta_q \varepsilon_{i-q}$$

4. Model ARIMA(p,d,q) (Autoregresif orde p, Integrate orde d, dan Moving Average orde q)

Persamaan umum

$$\phi_p(B)(1-B)^d Z_i = \theta_q(B)\varepsilon_i$$

2.3 Regresi Parametrik

Salah satu model regtesi parametrik adalah regresi linier sederhana. Model regresi linier sederhana dapat ditulis :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{`} \qquad , i = 1,2,3,\ldots,n$$

dengan asumsi sebagai berikut:

$$\varepsilon_i \sim NID(0, \sigma^2)$$

Estimasi dari model regresi linier tersebut adalah

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

Koefisien regresi β_0 dan β_1 dapat ditentukan dengan metode kuadrat terkecil didapat

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

2.4 Regresi Nonparametrik

Model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linier maupun nonlinier dikarenakan tidak adanya asumsi yang harus dipenuhi. Ada beberapa teknikpendugaan nilai peubah respons dalam regresi nonparametrik, yakni Kernel, Spline, Polinomial Lokal, Deret Fourier, dan Wavelet.

Model regresi nonparametrik secara matematis dapat ditulis:

$$Y_i = m(x_i) + \varepsilon_i$$

Dengan ε_i adalah galat yang diasumsikan terdistribusi di sekitar 0, $m(x)$ adalah sebuah fungsi yang mewakili perilaku intrinsik dari data. Regresi nonparametric yang digunakan dalam makalah ini adalah regresi kerne.

Fungsi Kernel

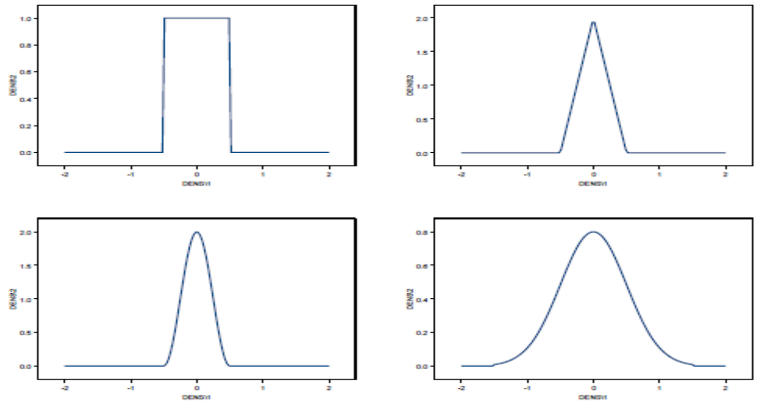
Fungsi kernel dinotasikan *K(u)* merupakan suatu fungsi yang kontinu, simetris, terbatas, dan $\int_{-\infty}^{\infty} K(u)du = 1$.

Berikut empat macam fungsi kernel pada S-Plus :

Tabel 2.1 Tabel Fungsi Kernel

Nama Kernel	Bentuk Fungsi	Nama Kernel	Bentuk Fungsi
Box	$K_{box}(u) = \begin{cases} 1, & \lvert u \rvert \le 0{,}5 \\ 0, & \lvert u \rvert > 0{,}5 \end{cases}$	**Parzen**	$K_{par}(u) = \begin{cases} (9/8) - (3/2)\lvert u \rvert + u^2/2, & 1/2 \le \lvert u \rvert \le 3/2 \\ (3/4) - u^2, & \lvert u \rvert \le 1/2 \\ 0, & \lvert u \rvert \ge 3/2 \end{cases}$
Triangle	$K_{tri}(u) = \begin{cases} 1 - \lvert u \rvert, & \lvert u \rvert \le 1 \\ 0, & \lvert u \rvert > 1 \end{cases}$	**Normal**	$K_{nor}(u) = (1/\sqrt{2\pi})\exp(-u^2/2)$

Plot dari ke empat fungsi kernel tersebut adalah sebagai berikut:



Gambar 2.1 Gambar fungsi kernel pada S-Plus yakni: box, triangle, parzen dan gaussian

(Carmona, 2004)

Regresi Kernel

Penggunaan fugsi kernel untuk mengestimasi fungsi regresi terbukti cukup berguna. Fungsi yang "kasar" dengan kata lain fungsi yang acak harus menjadi fungsi yang "*smoothed out*".Analisis data menggunakan kernel memungkinkan menggunakan beberapa *bandwidth* dan memilih estimator terakhir didasarkan pada penilaian kualitatif dari hasil estimasi. Membandingkan perhitungan estimasi juga dimungkinkan dengan estimasi $\hat{m}$ dipilih sesuai kriteria minimal kuadrat error.

$$MSE(\hat{m}) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}(x_i)\right)^2$$

Namun cara tersebut bukan merupakan cara yang tepat, karena semakin kecil nilai MSE yang dihasilkan maka *bandwidth* yang dihasilkan juga akan semakin kecil dan akan memberikan penghalusan yang *undersmooth*. Cara yang tepat adalah dengan memilih *bandwidth* yang optimal dengan meminimalkan estimasi pendekatan regresi kernel salah satunya adalah *Cross Validation* (CV).

(Ogden,1997)

Estimator Nadaraya-Watson

Untuk mengkonstruksi penduga Nadaraya Watson (N-W) diasumsikan bahwa baik variabel bebas maupun variabel target, keduanya adalah variabel random. Misalkan $f(x)$ adalah densitas untuk variabel random X, $f(y)$ adalah densitas untuk variabel random Y dan $f(x,y)$ adalah densitas gabungan untuk variabel random (X,Y), maka:

$$m(x) = \frac{1}{f_X(x)}\int_{-\infty}^{\infty} y f(x,y)dy$$

Dengan mengadopsi estimator densitas Kernel, yaitu salah satu metode yang paling sederhana dari pendugaan $f(x,y)$ dan $f_X(x)$.

$$\hat{f}(x,y) = \frac{1}{nh_xh_y}\sum_{i=1}^{n} K_x\left(\frac{x - X_i}{h_x}\right)K_y\left(\frac{y - Y_i}{h_y}\right)$$

$$\hat{f}_X(x) = \frac{1}{nh_x}\sum_{i=1}^{n} K_x\left(\frac{x - X_i}{h_x}\right)$$

Dimana $K_x(.)$ dan $K_y(.)$ adalah sebuah fungsi Kernel, sedangkan h_x dan h_y adalah sebuah bilangan positif yang disebut dengan *bandwidth*. Selanjutnya karena

$$\hat{m}(x) = \frac{1}{\hat{f}_X(x)} \int_{-\infty}^{\infty} y\hat{f}(x,y)dy$$

Maka didapatkan estimator Nadaraya-Watson sebagai berikut:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^{n} K_i\left(\frac{x-X_i}{h}\right)}$$

$$\hat{m}(x) = \sum_{i=1}^{n} W_i(x) Y_i$$

Dimana $W_i(x) = \frac{K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}$

$$\sum W_i(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} = 1$$

Jelas bahwa penduga yang disediakan oleh N-W merupakan sebuah rataan terbobot X_i dari $\{Y_i\}$.

(Takezawa, 2006)

Menurut Hardle (1990) pemilihan Kernel itu sebenarnya tidak terlalu berpengaruh terhadap nilai prediksinya, dan yang paling berpengaruh adalah pemilihan *bandwidth*nya. Jika h mendekati nol maka fungsi dugaan akan *undersmoothed* dan memiliki varian yang besar. Jika h adalah sebuah bilangan besar / dekat dengan rentang nilai-nilai X maka fungsi dugaan akan *oversmoothed* dan biasnya tinggi.

Regresi Nonparametrik untuk Data *Time Series*

Model (T): *time series*, $(Z_i\,,\ i \geq 1)$ adalah hasil observasi dan dalam memprediksi Z_{n+1} dengan $m(x) = E(Z_{n+1}|Z_n = x)$. Satu langkah unutuk memprediksi masalah *time series* (T) satu dimensi dapat dipetakan atau digambarkan ke model pertama. Dengan menetapkan untuk *time series* stasioner $\{Z_i\,, i \geq 1\}$. Nilai lag Z_{i-1} sebagai X_i dan nilai Z_i sebagai Y_i, kemudian masalah pendugaan Z_{n+1} dari $(Z_i)_{i=2}^{n}$ dapat dianggap sebagai masalah regresi penghalusan untuk $\{(X_i, Y_i)\}_{i=2}^{n} = \{(Z_{i-1}, Z_i)\}_{i=2}^{n}$. Permasalahan prediksi untuk *time series*$\{Z_i\}$ adalah sama seperti estimasi $m(x) = E(Y|X = x)$ untuk dua dimensi *time series*$\{(X_i, Y_i)\}_{i=1}^{n}$.

Pemilihan *Bandwidth Optimum*

Salah satu metode pendekatan yang mungkin dilakukan untuk menemukan sebuah estimasi yang tidak bias adalah *Cross Validation* (CV). Menurut Hardle (1990) metode *Cross Validation* atau sering disebut CV adalah metode penggunaan data untuk menunjukkan apa yang harus dilakukan jika pengulangan observasi tersedia. Langkah pertama, satu observasi ke-j dikeluarkan (n-1) data yang tersisa digunakan untuk memperoleh penghalusan pada:

$$\hat{m}_{h,j}(X_j) = n^{-1} \sum_{i \neq j} W_{hi}(X_j) Y_i$$

Langkah kedua yaitu mengerjakan seperti langkah pertama untuk j=1,2,...,n dan diperoleh fungsi:

$$CV(h) = n^{-1} \sum_{j=1}^{n} \left[Y_j - \hat{m}_{h,j}(X_j)\right]^2$$

3. METODOLOGI PENELITIAN

3.1 Sumber Data

Dalam studi kasus ini, data bersumber dari *yahoo finance*. Data historis diambil dari data *Composite Indeks* (IHSG) periode minggu pertama Januari 2011 sampai dengan minggu ke empat Februari 2012. Data tersebut merupakan data *closing price* saham mingguan pada periode perdagangan terakhir.

3.2 Variabel Penelitian

Variabel yang mendukung penelitian ini adalah variabel yang menyusun data *historical prices* IHSG yaitu harga penutup (*close*).

3.3 Software yang Digunakan

Software yang digunakan dalam pengerjaan tugas akhir ini adalah S-plus 2000, SPSS 16, Minitab 14, dan Eviews 4.

3.4 Langkah Analisis

1. Analisis Time series klasik
2. Analisis dengan Regresi Parametrik Linier
3. Analisis dengan Regresi Nonparametrik Kernel
4. Membandingkan nilai MSE dari ketiga model
5. Prediksi dengan menggunakan model terbaik

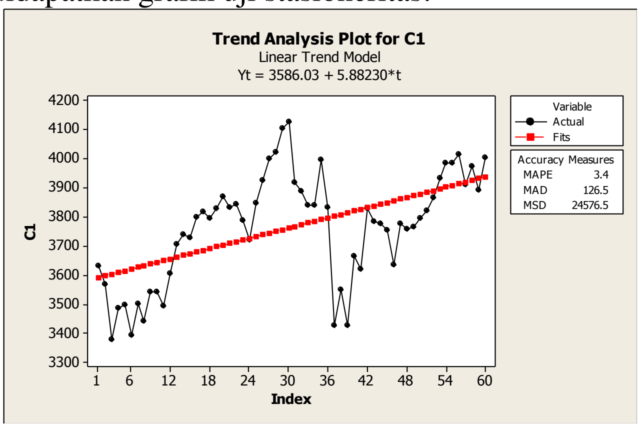
4. ANALISIS DAN PEMBAHASAN

4.1 Deskripsi Data

Data yang digunakan adalah data historis mingguan IHSG pada minggu pertama bulan Januari 2011 sampai dengan minggu ke empat bulan Februari 2012. Data tersebut merupakan data *time series* yang berupa harga *clossing price* IHSG.

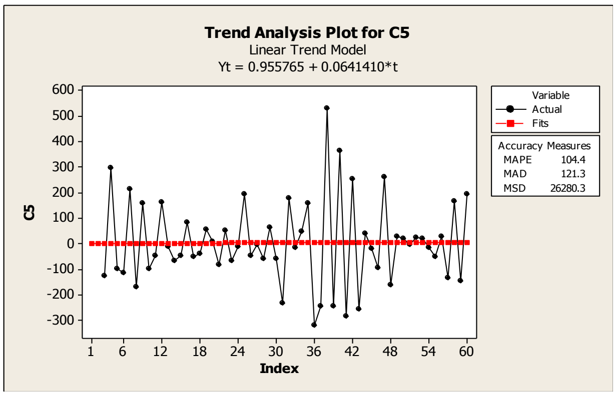
4.2 Analisis Runtun Waktu

Dari data IHSG didapatkan grafik uji stasioneritas:



Gambar 4.1 Grafik Stasioneritas Data IHSG

Dari Gambar 4.1 dapat dilihat data IHSG belum stasioner sehingga perlu dilakukan differensi.Setelah differensi dua kali didapatkan grafik uji stasioneritas:



Gambar 4.2 Grafik Stasioneritas Data IHSG dengan Differensi Dua Kali

Dapat dilihat pada Gambar 4.2 setelah differensi dua kali data sudah stasiner dan didapatkan model awal adalah ARIMA (0,2,1), ARIMA (1,2,1), ARIMA (2,2,1), ARIMA (1,2,0), ARIMA (2,2,0).Dari uji sinifikansi parameter model yang signifikan adalah ARIMA (0,2,1), ARIMA (1,2,0), ARIMA (2,2,0).Dari uji normalitas residual ketiga model yang sinifikan memiliki residual berdistribusi normal.Dari uji independensi residual, model yang memenuhi asumsi tidak ada korelasi residual antar lag adalah model ARIMA (0,2,1) dengan MSE = 11917.

4.3 Regresi Linier Klasik

4.4 Tabel 4.1 Tabel Coeffidients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	566.262	271.020		2.089	.041
	X	.851	.072	.843	11.827	.000

a. Dependent Variable: y

Dari analisis regresi linier klasik dan didapatkan model regresi Y=566.262+0.581X.

Tabel 4.2 Tabel Anova

ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	1477079.954	1	1477079.954	139.878	.000[a]
	Residual	601905.714	57	10559.749		
	Total	2078985.667	58			

a. Predictors: (Constant), x

b. Dependent Variable: y

Dari semua uji asumsi yang dilakukan semua asumsi terpenuhi dan menghasilkan nilai MSE = 10559.749.

4.5 Regresi Nonparametrik Kernel

Pada analisis regresi nonparametrik kernel dilakukan dengan cara meminimalkan *cross validation* untuk mendapatkan *bandwidth* optimal, dari bandwidth optimal yang dihasilkan dicari nilai MSE. Dari regresi nonparametrik kernel normal, box, parzen dan triangle didapatkan hasil:

Pembanding	Regresi Nonparametrik			
	Normal	**Box**	**Parzen**	**Triangle**
CV min	0.00128741108801994	1.41765082352386	0.000128136624583882	7.6774675323911e-006
H	23.6	75.5	46.3	58.2
MSE	7081.074	7483.123	7141.503	**6987.787**

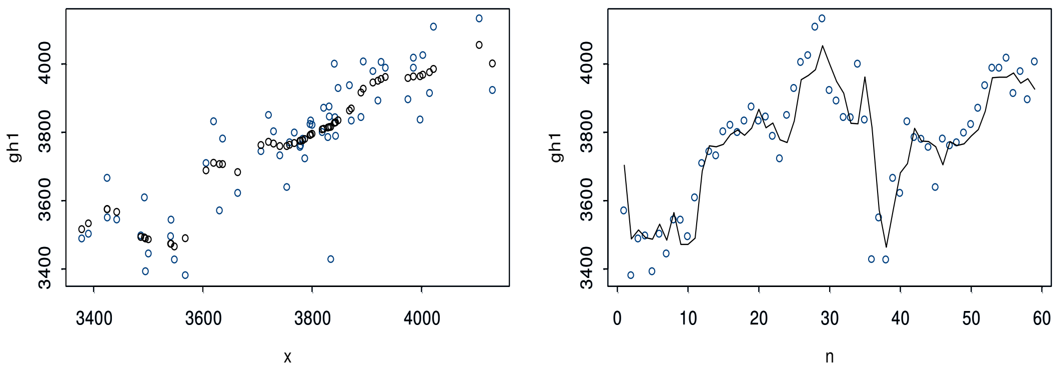
model terbaik adalah analisis dengan menggunakan kernel triangle dengan nilai MSE = 6987.787

4.6 Model Terbaik Keseluruhan

Model terbaik didapat dari perbandingan nilai MSE dari analisis *time series*, regresi linier, dan regresi nonparametrik kernel. Analisis yang menghasilkan nilai MSE terkecil akan menghasilkan model terbaik. Perbandingan nilai MSE masing-masing analisis adalah :

Analisis	***Time series***	**Regresi Parametrik**	**Regresi Nonparametrik**
Model	ARIMA(0,2,1)	Regresi Linier	Triangle
MSE	11917	10559.749	**6987.787**

Nilai MSE terkecil didapat dari fungsi kernel Triangle dengan nilai MSE = 6987.787 sehingga model dari fungsi triangle yang digunakan dalam prediksi regresi kernel dan dihasilkan grafik pemulusan kernel triangle sebagai berikut:



a b

ooo Data aktual ooo Hasil pemulusan — Hasil pemulusan terhadap n

Gambar 4.3a.Grafik pemulusan kernel triangle terhadap X dengan h = 58.2.
b.Grafik pemulusan kernel triangle terhadap n dengan h = 58.2.

4.6 Prediksi

Model regresi yang terbaik dapat digunakan untuk memprediksikan nilai Y yang akan datang. Pada kasus ini nilai Y yang selanjutnya sudah diketahui, hal ini dapat digunakan untuk membandingkan nilai Y aktual dengan Y prediksinya seperti berikut :

Data ke	Y	Y prediksi
60	4004.87	3925.25200109682
61	3991.54	3966.76572504831
62	4082.54	3961.29971590358
63	4041.56	3985.71635974196
64	4121.55	3994.58912445223
65	4166.37	4020.01897857607
66	4159.28	3921.64
67	4181.37	3955.49725888325
68	4163.98	3921.64
69	4216.68	3929.98713737529

5. KESIMPULAN

1. Metode regresi nonparametrik kernel dapat digunakan sebagai alternatif untuk menyelesaikan permasalahan data yang fluktuatif dikarenakan pada regresi nonparametrik tidak diperlukan asumsi-asumsi khusus yang harus dipenuhi.
2. Pada metode regresi kernel, pemilihan *bandwidth* optimal lebih penting dibandingkan pemilihan fungsi yang digunakan.
3. Hasil perbandingan nilai MSE dari dataIHSG pada model*time series*klasik,regresi linier dan regresi kernel didapatkan nilai MSE terkecil pada model regresi kernel dengan fungsi triangle dengan badwidth h optimal sebesar 58.2 dan nilai MSE =6987.787

6. DAFTAR PUSTAKA

Eubank, R.L.1988. *Spline Smoothing and Nonparametric Regression.* Texas: Department of Stasistics Southern Methodist University Dallas.

Gujarati, D. 1997. *Ekonometrika Dasar*. Terjemahan Sumarno Zain. Jakarta:Erlangga.

Hardle, W. 1990. *Applied Nonparametric Regression* .New York: Cambridge University Press.

Ogden, R.T. 1997. *Essential Wavelets for Statistical Applications and Data Analysis.* Boston.

Takezawa, K. 2006. *Introduction to Nonparametric Regression.* New Jersey: John Wiley & Sons,Inc.