ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 3, Tahun 2015, Halaman 603 - 612
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PEMODELAN DATA INDEKS HARGA SAHAM GABUNGAN MENGGUNAKAN REGRESI *PENALIZED SPLINE*

Novia Agustina[1], Suparti[2], Moch. Abdul Mukid[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro
E-mail: noviaagustina11@gmail.com[1], supartisudargo@yahoo.co.id[2], mamukid@yahoo.com[3]

ABSTRACT

Indonesia Composite Index (IHSG) is an indicator of stock price changes in Indonesia Stock Exchange. IHSG is time series data that can be modeled with parametric models. But there are some assumptions for parametric model, while the fluctuated IHSG data usually doesn't occupy these assumptions. Another alternative for this study is nonparametric regression. Penalized spline regression is one of nonparametric regression method that can be used. The optimal penalized spline models depends on the determination of the optimal smoothing parameter λ and the optimal number of knots, that has a minimum value of Generalized Cross Validation (GCV). The best model in this study is penalized spline degree 1 (linear) with 1 knot, that is 5120,625, smoothing parameter λ value is 41590, and GCV value is 1567,203. R^2 value for in sample data is 83,2694% and R^2 value for out sample data is 96,4976% show that the model have a very good performance. MAPE values for in sample data is 0,5983% and MAPE values for out sample data is 0,4974%. Because the value of MAPE in sample and out sample is less than 10%, it means that the performance of the model and forecasting are very accurate.

Keywords: Indonesia Composite Index, Nonparametric Regression, Penalized Spline Regression, GCV, MAPE

1. PENDAHULUAN

Saat ini investasi di Indonesia sedang mengalami perkembangan. Salah satu investasi yang cukup menjanjikan adalah investasi saham. Saham adalah sertifikat yang menunjukkan bukti kepemilikan suatu perusahaan, dan pemegang saham memiliki hak klaim atas penghasilan dan aktiva perusahaan (Bapepam, 2003). Sebelum mengambil keputusan untuk mulai berinvestasi saham, perlu dikumpulkan informasi sebanyak mungkin karena investasi selalu memiliki kemungkinan keuntungan dan kerugian. Salah satu informasi yang diperlukan tersebut adalah Indeks Harga Saham Gabungan (IHSG).

Data IHSG merupakan data runtun waktu yang dapat dimodelkan dengan model parametrik. Namun dalam model parametrik terdapat asumsi yang harus dipenuhi, yaitu asumsi stasioneritas dan *white noise*. Sedangkan data IHSG biasanya tidak memenuhi asumsi tersebut. Oleh karena itu dilakukan pemodelan data IHSG menggunakan analisis yang tidak memerlukan asumsi-asumsi yang harus dipenuhi, salah satunya adalah regresi nonparametrik. Regresi nonparametrik digunakan jika bentuk kurva data tidak diketahui sebelumnya. Regresi nonparametrik memiliki fleksibilitas yang tinggi karena bentuk estimasi kurva regresinya dapat menyesuaikan datanya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Eubank, 1999).

Regresi *spline* merupakan salah satu metode regresi nonparametrik yang dapat digunakan untuk memodelkan data. Regresi *spline* adalah suatu pendekatan ke arah pencocokan data dengan tetap memperhitungkan kemulusan kurva. *Spline* merupakan potongan polinomial tersegmen yang digabungkan oleh titik-titik knot yang dapat menjelaskan karakteristik dari data (Eubank, 1999).

Dalam regresi *spline*, pemilihan banyak dan letak knot merupakan isu yang penting. Untuk menentukan knot yang optimal, perlu dilakukan perhitungan sebanyak kombinasi banyaknya knot dari banyaknya data. Kemudian dipilih model optimal berdasarkan kriteria tertentu, misalnya nilai GCV minimum. Hal ini membutuhkan waktu yang lama dan jika dilakukan menggunakan software memerlukan memori yang besar. Karena itu diperlukan alternatif untuk mengatasi masalah ini, yaitu regresi *penalized spline* dimana knot terletak di titik-titik kuantil dari nilai *unique* (tunggal) variabel prediktor (Ruppert *et al*., 2003).

Estimator regresi *penalized spline* diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS) yang terdiri dari jumlah kuadrat residual dan penalti kekasaran yang membuat model menjadi lebih mulus berdasarkan nilai parameter penghalus λ (Ruppert *et al*., 2003). Parameter penghalus λ yang optimal diperoleh dengan kriteria *Generalized Cross-Validation* (GCV) minimum.

Dengan menggunakan regresi *penalized spline*, dapat dilakukan pemodelan nilai IHSG yang akan berguna bagi para investor dan pelaku usaha sebagai bahan pertimbangan dalam mengambil keputusan investasi saham.

2. TINJAUAN PUSTAKA

2.1. Indeks Harga Saham Gabungan

Indeks Harga Saham Gabungan (IHSG) merupakan salah satu jenis indeks yang dikeluarkan oleh Bursa Efek Indonesia. IHSG menggunakan seluruh saham yang tercatat di Bursa Efek Indonesia sebagai komponen penghitungan indeks. IHSG pertama kali diperkenalkan pada tanggal 1 April 1983 sebagai indikator pergerakan harga-harga saham yang tercatat di bursa (BEI, 2010).

Pergerakan IHSG menjadi indikator penting bagi para investor untuk memperkirakan apakah mereka akan menjual, menahan, atau membeli suatu atau beberapa saham. Pergerakan nilai IHSG biasanya menunjukkan perubahan situasi pasar yang terjadi. Seiring dengan perkembangan dan dinamika pasar, IHSG mengalami periode naik dan turun. Nilai IHSG yang mengalami kenaikan menggambarkan kondisi pasar yang sedang aktif**.** Kondisi inilah yang biasanya menunjukkan keadaan yang diinginkan. Nilai IHSG yang tetap menunjukkan keadaan pasar yang stabil, sedangkan nilai IHSG yang menurun menggambarkan kondisi pasar yang sedang lesu (Anoraga dan Pakarti, 2006).

2.2. Dekomposisi Cholesky

Menurut Ruppert *et al.* (2003), jika **A** adalah matriks simetris yang definit positif, maka dekomposisi dari matriks **A** dinyatakan sebagai

$$\mathbf{A} = \mathbf{R}^T\mathbf{R}$$

dimana **R** adalah matriks segitiga atas dengan elemen-elemen diagonal positif.

2.3. Dekomposisi Nilai Singular

Menurut Leon (1998) dekomposisi nilai singular adalah suatu pemfaktoran matriks dengan mengurai suatu matriks ke dalam tiga matriks **U**, **Σ**, dan **V**. Jika diketahui suatu matriks adalah matriks **A** berukuran m×n, maka dekomposisi dari matriks **A** dinyatakan sebagai

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T$$

dimana **U** merupakan matriks orthogonal berukuran m×m, **V** merupakan matriks orthogonal berukuran n×n, sedangkan **Σ** adalah matriks berukuran m×n yang semua elemen di luar diagonalnya adalah 0, dan elemen-elemen diagonalnya memenuhi $\sigma_1 \geq \sigma_2, \geq \ldots \geq \sigma_r \geq 0$ dimana r=min(m,n) (Leon, 1998).

2.4. Regresi Parametrik

Regresi parametrik merupakan metode statistika yang menyatakan hubungan antara variabel respon dengan beberapa variabel prediktor. Salah satu model regresi parametrik adalah regresi linier yang digunakan jika hubungan antara variabel respon dengan beberapa variabel prediktor berbentuk linier (Montgomery dan Peck, 1992). Misalkan diketahui data sebanyak n dengan k variabel prediktor maka model regresi linier dapat dituliskan sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$
$$y_i = \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \text{ ; i = 1,2,…,n} \quad (1)$$

dimana

y_i = variabel respon untuk pengamatan ke-i

x_{ij} = variabel prediktor ke-j untuk pengamatan ke-i dengan $x_{0i} = 1$

β_j = koefisien regresi pada x_j

ε_i = residual ke-i yang diasumsikan independen, berdistribusi normal dengan rata-rata 0 dan varian σ^2.

2.5. Regresi Nonparametrik

Regresi nonparametrik digunakan jika bentuk kurva data tidak diketahui sebelumnya. Model regresi nonparametrik dinyatakan sebagai

$$y_i = f(x_i) + \varepsilon_i$$

dimana fungsi $f(x_i)$ adalah fungsi regresi yang tidak diketahui bentuknya. Regresi nonparametrik memiliki fleksibilitas yang tinggi karena bentuk estimasi kurva regresinya dapat menyesuaikan datanya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Eubank, 1999).

2.6. Regresi *Spline*

Menurut Eubank (1999), regresi *spline* adalah suatu pendekatan ke arah pencocokan data dengan tetap memperhitungkan kemulusan kurva. *Spline* merupakan potongan polinomial tersegmen yang digabungkan oleh titik-titik knot yang dapat menjelaskan karakteristik dari data. Knot merupakan titik perpaduan bersama yang memperlihatkan terjadinya perubahan perilaku dari fungsi *spline* pada interval-interval yang berbeda.

Secara umum fungsi *spline* dengan derajat p dan m knot dinyatakan sebagai

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{m} \beta_{p+j} (x - k_j)_+^p$$

dengan fungsi *truncated* adalah sebagai berikut:

$$(x - k_j)_+^p = \begin{cases} (x - k_j)^p & , x \geq k_j \\ 0 & , x < k_j \end{cases}$$

dan $k_1, k_2, \ldots, k_m$ merupakan titik-titik knot.

Sehingga model regresi *spline* adalah

$$y_i = f(x_i) + \varepsilon_i$$
$$y_i = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{m} \beta_{p+j} (x - k_j)_+^p + \varepsilon_i$$

2.7. Regresi *Penalized Spline*

Menurut Ruppert *et al*. (2003), estimator regresi *penalized spline* diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS) sebagai berikut:

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda^{2p} \sum_{j=1}^{m} \beta_{p+j}^2 \quad , \lambda \geq 0 \quad (2)$$

Suku pertama pada persamaan (3) adalah jumlah kuadrat residual dan suku keduanya adalah penalti kekasaran yang membuat model menjadi lebih halus berdasarkan nilai parameter penghalus λ (Ruppert *et al.*, 2003). Dalam bentuk matriks persamaan (2) dinyatakan sebagai

$$\mathbf{L} = (\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta}) + \lambda^{2p}\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{D\beta} \qquad (3)$$

dimana λ = parameter penghalus

$$\mathbf{X}=\begin{bmatrix} 1 & x_1 & \dots & x_1^p & (x_1-k_1)_+^p & \dots & (x_1-k_m)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^p & (x_n-k_1)_+^p & \dots & (x_n-k_m)_+^p \end{bmatrix}$$

$$\boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \beta_{p+1} \\ \vdots \\ \beta_{p+m} \end{bmatrix}$$

$$\mathbf{D}=\begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(p+1)x(p+1)} & \mathbf{0}_{(p+1)xm} \\ \mathbf{0}_{mx(p+1)} & \mathbf{I}_{mxm} \end{bmatrix} = \mathrm{diag}(\mathbf{0}_{p+1}, \mathbf{1}_m).$$

Pendugaan parameter dilakukan dengan meminimumkan persamaan (3) dengan prinsip $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta}} = 0$ sehingga diperoleh estimasi parameter β sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda^{2p}\boldsymbol{D})^{-\mathbf{1}}\boldsymbol{X}^T\boldsymbol{Y}$$

Sedangkan estimasi parameter untuk y adalah:

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$
$$\hat{\boldsymbol{Y}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda^{2p}\boldsymbol{D})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$
$$\hat{\boldsymbol{Y}} = \boldsymbol{S}_\lambda\boldsymbol{Y}$$

dengan $\boldsymbol{S}_{\boldsymbol{\lambda}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda^{2p}\boldsymbol{D})^{-1}\boldsymbol{X}^T$ dinamakan matriks hat yang bersifat simetris dan definit positif.

2.8. Pemilihan Parameter Penghalus Optimal

Parameter penghalus λ yang optimal diperoleh dengan menghitung nilai *Generalized Cross-Validation* (GCV) minimum. Menurut Green dan Silverman (1994), fungsi GCV dapat dinyatakan sebagai berikut:

$$\mathrm{GCV} = n^{-1}\frac{\mathrm{RSS}(\lambda)}{[1\text{-}n^{-1}tr(\boldsymbol{S}_\lambda)]^2}$$

dengan $\mathrm{RSS}(\lambda) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

$\boldsymbol{S}_{\boldsymbol{\lambda}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda^{2p}\boldsymbol{D})^{-1}\boldsymbol{X}^T$

Demmler-Reinsch ortogonalization merupakan suatu metode yang dapat mempermudah dalam memilih GCV optimal untuk setiap parameter penghalus λ. Algoritma *Demmler-Reinsch ortogonalization* adalah sebagai berikut:

1. Menghitung dekomposisi cholesky dari $\mathbf{X}^T\mathbf{X}$ untuk mendapatkan matriks **R** sedemikian sehingga $\mathbf{X}^T\mathbf{X} = \mathbf{R}^T\mathbf{R}$.
2. Menghitung dekomposisi nilai singular matriks $(\mathbf{R}^{-1})^T\,\mathbf{D}\mathbf{R}^{-1}$ untuk mendapatkan **U** diag(s) $\mathbf{U}^T$ sedemikian sehingga $(\mathbf{R}^{-1})^T\,\mathbf{D}\mathbf{R}^{-1} = \mathbf{U}\,\mathrm{diag}(\mathbf{s})\,\mathbf{U}^T$.

3. Menghitung matriks $\mathbf{A} = \mathbf{X}\mathbf{R}^{-1}\mathbf{U}$ dan vektor $\mathbf{b}=\mathbf{A}^T\mathbf{Y}$.
4. Menghitung $\widehat{\boldsymbol{Y}} = \boldsymbol{A}\left(\frac{\boldsymbol{b}}{\mathbf{1}+\lambda^{2p}\boldsymbol{s}}\right)$

   dimana $\mathbf{1} = \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}_{(p+m+1)x1}$ dan $\mathbf{s} = \begin{bmatrix}s_1\\s_2\\\vdots\\s_{p+m+1}\end{bmatrix}_{(p+m+1)x1}$

   dengan $\mathrm{tr}(\boldsymbol{S_\lambda}) = \sum_{i=1}^{p+m+1}(1+\lambda^{2p}s_i)^{-1}$

2.9. Pemilihan Knot Optimal

Banyaknya knot (m) merupakan banyaknya titik dimana terjadi perubahan perilaku fungsi pada interval yang berlainan. Dalam regresi penalized *spline*, knot terletak pada titik kuantil dari nilai *unique* (tunggal) variabel prediktor $\{x_i\}_{i=1}^n$. Dengan knot ke-j adalah kuantil ke-$\frac{j}{m+1}$ dari nilai *unique* variabel prediktor $\{x_i\}_{i=1}^n$.

Algoritma yang digunakan untuk menentukan banyak knot optimal dalam regresi *penalized spline* adalah algoritma *full-search*. Dalam algoritma *full-search*, dihitung nilai GCV untuk banyak knot m=1,2, … sampai banyak knot tertentu yang dicobakan, dengan ketentuan $m < (n_{unique} - p - 1)$, dimana n_{unique} adalah banyaknya nilai *unique* dari variabel prediktor $\{x_i\}_{i=1}^n$ dan p adalah derajat polinomial. Kemudian dipilih banyak knot yang memiliki nilai GCV terkecil (Ruppert *et al*., 2003).

2.10. Regresi Nonparametrik untuk Data Runtun Waktu

Menurut Hardle (1990), sifat statistik dari regresi penghalus secara umum dianalisis dalam kerangka struktur observasi yang berdistribusi identik dan independen (iid). Asumsi $\{(X_i,Y_i),\ i=1,2,...,n\}$ adalah sampel yang independen digunakan untuk penyederhanaan permasalahan. Namun data yang diperoleh dalam kehidupan sehari-hari seringkali tidak memenuhi asumsi bahwa observasi $(X_1,Y_1), (X_2,Y_2), \ldots (X_n,Y_n)$ independen. Oleh karena itu perlu disusun suatu pemodelan data yang asumsi independensi datanya tidak dipenuhi. Terdapat tiga konsep dasar matematika yang mendasari pemodelan ini, yaitu:

1. Model (S): Suatu barisan stasioner $\{(X_i,Y_i),\ i=1,2,...,n\}$ adalah hasil observasi dan akan diestimasi $f(x) = E(Y|X=x)$
2. Model (T): Suatu runtun waktu $\{Z_i,\ i=2,3,...n\}$ adalah hasil observasi dan digunakan untuk memprediksi Z_{n+1} dengan $f(x) = E(Z_{n+1}|\ Z_n = x)$ □
3. Model (C): Error observasi $\{e_{in}\}$ dalam model regresi dengan rancangan tetap Y_{in} □□ $f(i/n) + e$, membentuk barisan variabel random yang berkorelasi.

Permasalahan model runtun waktu (T) dapat digambarkan ke dalam model (S) dengan mendefinisikan dalam runtun waktu $\{Z_i,\ i=2,3,...n\}$, nilai lag Z_{i-1} sebagai X_i dan nilai Z_i sebagai Y_i. Selanjutnya masalah prediksi Z_{n+1} dari $\{Z_i\}$, $i=1,2,...,n$ dapat dianggap sebagai masalah penghalusan regresi untuk $\{(X_i,Y_i),\ i=2,3,...n\} = \{(Z_{i-1},Z_i),\ i=2,3,...n\}$.

2.11. Ketepatan Metode Peramalan

Kinerja model yang digunakan dalam peramalan dapat dilihat berdasarkan nilai koefisien determinasi (R^2) dan *Mean Absolute Percentage Error* (MAPE). Koefisien determinasi adalah persentase nilai Y yang dapat dijelaskan oleh garis regresi (Algifari, 2000). Nilai koefisien determinasi (R^2) dapat dinyatakan sebagai berikut:

$$R^2 = \frac{\sum_{i=1}^n(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n(Y_i - \bar{Y})^2}$$

dimana n = banyaknya data

Y_i = data aktual periode ke-i
$\hat{Y}_i$ = data hasil prediksi periode ke-i
$\bar{Y}$ = rata-rata data aktual

MAPE merupakan rata-rata dari keseluruhan persentase kesalahan (selisih) antara data aktual dengan data hasil peramalan. MAPE dapat dijabarkan sebagai berikut:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i} x\ 100\%$$

dimana n = banyaknya data yang diprediksi
Y_i = data aktual periode ke-i
$\hat{Y}_i$ = data hasil prediksi periode ke-i

Interpretasi mengenai nilai MAPE yang disarankan oleh Lewis (1982) dalam Chen *et al.* (2008) adalah sebagai berikut:

1. Persentase nilai MAPE yang kurang dari 10% menghasilkan hasil peramalan yang sangat akurat.
2. Persentase nilai MAPE yang berada di antara 10% sampai 20% menghasilkan peramalan yang baik
3. Persentase nilai MAPE yang berada di antara 20% sampai 50% menghasilkan peramalan yang proporsional atau wajar.
4. Persentase nilai MAPE yang lebih dari 50% menghasilkan peramalan yang tidak akurat.

3. METODOLOGI PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data Indeks Harga Saham Gabungan harian periode 1 Oktober 2014 sampai dengan 27 Februari 2015. Data tersebut dibagi menjadi dua yaitu data pada tanggal 1 Oktober 2014 sampai 31 Januari 2015 sebagai data *in sample* untuk menyusun model, dan data pada tanggal 1 Februari 2015 sampai 27 Februari 2015 sebagai data *out sample* untuk mengetahui ketepatan model. Data tersebut diperoleh dari website yahoo finance (www.finance.yahoo.com).

3.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah nilai Indeks Harga Saham Gabungan harian periode 1 Oktober 2014 sampai dengan 27 Februari 2015. Data tersebut dimodifikasi menjadi dua variabel yaitu variabel prediktor (X_i) dan variabel respon (Y_i). Nilai lag yang keluar pada plot PACF pada data tersebut adalah lag 1. Sehingga variabel prediktor pada penelitian ini adalah Y_{t-1}.

3.3. Software yang Digunakan

Software statistika yang digunakan dalam penelitian ini adalah Microsoft Excel 2007, MINITAB 14, dan R i386 3.0.3.

3.4. Teknik Pengolahan Data

Langkah-langkah dalam analisis data adalah sebagai berikut:

1. Mengumpulkan data Indeks Harga Saham Gabungan per hari periode 1 Oktober 2014 sampai dengan 27 Februari 2015.
2. Memodifikasi bentuk data Indeks Harga Saham Gabungan menjadi variabel respon dan variabel prediktor berdasarkan plot PACF.
3. Menentukan derajat spline yang digunakan.
4. Menentukan banyaknya knot yang digunakan.

5. Menghitung nilai GCV untuk setiap parameter penghalus λ yang mungkin untuk setiap nilai knot dengan menggunakan algoritma *Demmler – Reinsch Ortogonalization*.
6. Menentukan parameter penghalus λ optimal dengan melihat nilai GCV paling minimum.
7. Menentukan banyak titik knot optimal dengan algoritma *full search*.
8. Mendapatkan model regresi *penalized spline* optimal berdasarkan banyak titik knot dan parameter penghalus λ optimal.
9. Menghitung nilai prediksi Indeks Harga Saham Gabungan.
10. Menguji ketepatan prediksi.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data IHSG yang digunakan memiliki nilai minimum sebesar 4913,05 terjadi pada 13 Oktober 2014. Sedangkan nilai maksimum sebesar 5323,88 terjadi pada 23 Januari 2015. Sehingga range dari data tersebut adalah sebesar 410,83. Nilai rata-rata dari data tersebut adalah 5113,1176 dan standar deviasi sebesar 94,38462.

Data IHSG dapat dimodifikasi menjadi dua variabel yaitu variabel prediktor (X_i) dan variabel respon (Y_i). Nilai lag yang keluar pada plot PACF pada data tersebut adalah lag 1. Sehingga variabel prediktor pada penelitian ini adalah Y_{i-1}. Variabel prediktor (X_i) terdiri dari data IHSG pada waktu ke 1,2,…,i-1. Sedangkan variabel respon (Y_i) terdiri dari data IHSG ke 2,3,…,i.

4.2. Regresi *Penalized Spline*

Pemilihan model regresi *penalized spline* optimal dilakukan dengan menentukan parameter penghalus λ dan banyak knot optimal dengan kriteria GCV minimum.

4.2.1. Pemodelan Regresi *Penalized Spline* dengan Derajat 1

Secara umum fungsi *spline* dengan derajat p=1 dan m knot dinyatakan sebagai

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{m} \beta_{1+j}(x - k_j)_+$$

dengan fungsi *truncated* adalah sebagai berikut:

$$(x - k_j)_+ = \begin{cases} (x - k_j), & x \geq k_j \\ 0, & x < k_j \end{cases}$$

Pemilihan model optimal dilakukan dengan algoritma *full search*. Banyak knot yang digunakan adalah 1,2,…,20. Diperoleh parameter penghalus λ optimal sebesar 41590 dan banyak knot optimal adalah satu knot yang terletak pada nilai 5120,625. GCV optimal yang diperoleh adalah 1567,203.

4.2.2. Pemodelan Regresi *Penalized Spline* dengan Derajat 2

Secara umum fungsi *spline* dengan derajat p=2 dan m knot adalah:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{j=1}^{m} \beta_{2+j}(x - k_j)_+^2$$

dengan fungsi *truncated* adalah sebagai berikut:

$$(x - k_j)_+^2 = \begin{cases} (x - k_j)^2, & x \geq k_j \\ 0, & x < k_j \end{cases}$$

Pemilihan model optimal dilakukan dengan algoritma *full search*. Banyak knot yang digunakan adalah 1,2,…,72. Kemudian dipilih banyak knot optimal yang memiliki GCV minimum. Diperoleh parameter penghalus λ optimal sebesar 226 dan banyak knot optimal adalah 6 knot yang terletak pada nilai 5001,09; 5048,933; 5099,927; 5142,403; 5165,95; dan 5214,75. GCV optimal yang diperoleh adalah 1594,313.

4.2.3 Pemodelan Regresi *Penalized Spline* dengan Derajat 3

Secara umum fungsi *spline* dengan derajat p=3 dan m knot dinyatakan sebagai

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^{m} \beta_{3+j}(x - k_j)_+^3$$

dengan fungsi *truncated* adalah sebagai berikut:

$$(x - k_j)_+^3 = \begin{cases} (x - k_j)^3 & , x \geq k_j \\ 0 & , x < k_j \end{cases}$$

Pemilihan model optimal dilakukan dengan algoritma *full search*. Banyak knot yang digunakan adalah 1,2,…,65. Kemudian dipilih banyak knot optimal yang memiliki GCV minimum. Diperoleh parameter penghalus λ optimal sebesar 214 dan banyak knot optimal adalah 11 knot yang terletak pada nilai 4964,783; 5025,16; 5036,87; 5066,83; 5092,565; 5120,625; 5144,163; 5159,67; 5174,108; 5209,475; dan 5230,905. GCV optimal yang diperoleh adalah 1603.231.

4.3 Pemilihan Model Regresi *Penalized Spline* Terbaik

Pemilihan model regresi *penalized spline* terbaik dilakukan dengan melihat nilai GCV terkecil. Nilai GCV optimal dari setiap model dengan derajat 1, 2, dan 3 adalah sebagai berikut:

Tabel 1. Perbandingan Nilai GCV Optimal

Derajat	Banyak Knot	Titik Knot	Parameter Penghalus λ Optimal	GCV Optimal
1	1	5120,625	41590	1567,203
2	6	5001,09; 5048,933; 5099,927; 5142,403; 5165,95; 5214,75	226	1594,313
3	11	4964,783; 5025,16; 5036,87; 5066,83; 5092,565; 5120,625; 5144,163; 5159,67; 5174,108; 5209,475; 5230,905	214	1603,231

Berdasarkan Tabel 1 dapat dilihat bahwa nilai GCV optimal untuk model dengan derajat 1 lebih kecil daripada GCV optimal untuk model dengan derajat 2 dan derajat 3. Sehingga model regresi *penalized spline* terbaik adalah model dengan derajat 1 dengan GCV sebesar 1567,203. Model tersebut memiliki persamaan sebagai berikut:

$$\hat{f}(x) = 494{,}5769 + 0{,}9031503x + 0{,}06646313(x - 5120{,}625)_+$$

4.4. Ketepatan Metode Peramalan

Model regresi *penalized spline* terbaik yang telah diperoleh digunakan untuk memprediksikan data *out sample*, yaitu data IHSG pada periode 1 Februari 2015 sampai dengan 27 Februari 2015. Kinerja model yang digunakan dalam peramalan dapat dilihat berdasarkan nilai koefisien determinasi (R^2) dan *Mean Absolute Percentage Error* (MAPE) untuk data *in sample* dan data *out sample*.

Diperoleh nilai R^2 *in sample* yang cukup tinggi yaitu sebesar 83,2694%. Sementara nilai R^2 *out sample* sebesar 96,4976%. Karena nilai R^2 *in sample* maupun R^2 *out sample* tersebut mendekati 100%, artinya kinerja model sangat baik untuk data *in sample* dan data *out sample*. Sedangkan nilai MAPE *in sample* yang diperoleh sebesar 0,5983% dan MAPE *out sample* sebesar 0,4974%. Karena nilai MAPE *in sample* dan MAPE *out sample* sangat kecil dan kurang dari 10%, artinya kinerja model sangat akurat.

Dilakukan prediksi nilai Indeks Harga Saham Gabungan (IHSG) harian untuk periode 2 Maret 2015 sampai 31 Maret 2015 menggunakan model terbaik yang telah diperoleh. Karena nilai IHSG yang sebenarnya sudah diketahui, hal ini dapat digunakan untuk membandingkan nilai IHSG aktual dengan nilai IHSG prediksinya seperti berikut:

Tabel 2. Nilai IHSG Aktual dan Prediksi IHSG Periode 2 Maret 2015 sampai 31 Maret 2015

Periode	Nilai Aktual	Nilai Prediksi
2 Maret 2015	5477,83	5438,919
3 Maret 2015	5474,62	5465,622
4 Maret 2015	5448,06	5462,509
5 Maret 2015	5450,95	5436,756
6 Maret 2015	5514,79	5439,558
9 Maret 2015	5444,63	5501,459
10 Maret 2015	5462,93	5433,431
11 Maret 2015	5419,57	5451,174
12 Maret 2015	5439,83	5409,132
13 Maret 2015	5426,47	5428,776
16 Maret 2015	5435,27	5415,822
17 Maret 2015	5439,15	5424,355
18 Maret 2015	5413,15	5428,117
19 Maret 2015	5453,85	5402,907
20 Maret 2015	5443,06	5442,37
23 Maret 2015	5437,1	5431,908
24 Maret 2015	5447,65	5426,129
25 Maret 2015	5405,49	5436,359
26 Maret 2015	5368,8	5395,48
27 Maret 2015	5396,85	5359,905
30 Maret 2015	5438,66	5387,102
31 Maret 2015	5518,67	5427,642

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Model terbaik yang diperoleh adalah model regresi *penalized spline* derajat 1 dengan satu knot pada 5120,625 dengan parameter penghalus λ optimal sebesar 41590 dan GCV optimal sebesar 1567,203. Sehingga diperoleh estimasi model regresi *penalized spline* terbaik sebagai berikut:
$$\hat{f}(x) = 494{,}5769 + 0{,}9031503x + 0{,}06646313(x - 5120{,}625)_+$$
2. Nilai ketepatan model terbaik dapat dilihat dari besarnya nilai koefisien determinasi (R^2) dan MAPE. Diperoleh nilai R^2 *in sample* maupun R^2 *out sample* mendekati 100% sehingga kinerja model sangat baik. Sementara nilai MAPE *in sample* dan MAPE *out*

sample yang diperoleh kurang dari 10%, artinya kinerja model dan peramalan yang diperoleh sangat akurat.

DAFTAR PUSTAKA

Algifari, 2000, *Analisis Regresi Teori, Kasus, dan Solusi Edisi Kedua*. Yogyakarta: BPFE-Yogyakarta.

Chen, R.J.C., Blomfield, P., Cubbage, F.W., 2008, *Comparing Forecasting Models in Touris*. Journal of Hospitality and Tourism Research, Vol. 32, No. 1.

Eubank, R.L., 1999, *Spline Smoothing and Nonparametric Regression Second Edition*. New York: Marcel Dekker.

Green, P. J., and *Silverman*, B. W., *1994*, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.

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

Leon, S, 1998, *Aljabar Linear dan Aplikasinya Edisi Kelima*. Alih Bahasa Alit Bondan. Jakarta: Erlangga.

Montgomery, D.C., and Peck, E.A., 1992, *Introduction to Linear Regression Analysis Second Edition*. New York: John Wiley & Sons.

Ruppert, D., Wand, M.P., and Carroll, R.J., 2003, *Cambridge Series in Statistical and Probabilistic Mathematics: Semiparametric Regression*. New York: Cambridge University.

www.finance.yahoo.com (diakses pada tanggal 3 Maret 2015).

[Bapepam] Badan Pengawas Pasar Modal, 2003, *Panduan Investasi di Pasar Modal Indonesia*. Jakarta: Bapepam.

[BEI] Bursa Efek Indonesia, 2010, *Buku Panduan Indeks Harga Saham*. Jakarta: BEI.