ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 3, Tahun 2015, Halaman 527 - 532
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PEMODELAN REGRESI NONPARAMETRIK DATA LONGITUDINAL MENGGUNAKAN POLINOMIAL LOKAL (Studi Kasus: Harga Penutupan Saham pada Kelompok Harga Saham Periode Januari 2012 – April 2015)

Izzuddin Khalid[1], Suparti[2], Alan Prahutama[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staff Pengajar Jurusan Statstika FSM Universitas Diponegoro
izkha93@gmail.com, supartisudargo@yahoo.co.id, alan.prahutama@gmail.com

ABSTRACT

Stocks are securities that can be bought or sold by individuals or institutions as a sign of participating or possessing a company in the amount of its proportions. From the lens of market capitalization values, stocks are divided into 3 groups: large capitalization (Big-Cap), medium capitalization (Mid-Cap) and small capitalization (Small-Cap). Longitudinal data is observation which is conducted as n subjects that are independent to each subject observed repeatedly in different periods dependently. Smoothing technique used to estimate the nonparametric regression model in longitudinal data is local polynomial estimator. Local polynomial estimator can be obtained by WLS (Weighted Least Square) methods. Local polynomial estimator is very dependent on optimal bandwidth. Determination of the optimal bandwidth can be obtained by using GCV (Generalized Cross Validation) method. Among the Gaussian kernel, Triangle kernel, Epanechnikov kernel and Biweight kernel, it is obtained the best model using Gaussian kernel. Based on the application of the model simultaneously, it is obtained coefficient of determination of 97,80174% and MSE values of 0,03053464. Using Gaussian kernel, MAPE out sample of data is obtained as 11,74493%.

Keywords: Longitudinal Data, Local Polynomial, Stocks

1. PENDAHULUAN

Analisis regresi merupakan salah satu teknik yang paling umum digunakan dalam statistik. Tujuan dari analisis ini adalah untuk mengeksplorasi hubungan antara variabel prediktor dan respon, yaitu untuk menilai kontribusi variabel prediktor dan untuk mengidentifikasi dampaknya terhadap variabel respon [3]. Apabila tidak ada informasi apapun tentang bentuk fungsi, maka pendekatan yang digunakan adalah pendekatan nonparametrik. Hal ini dikarenakan pada pendekatan tersebut tidak bergantung pada asumsi bentuk kurva tertentu , sehingga memiliki fleksibilitas yang lebih besar [2].

Data longitudinal merupakan pangamatan yang dilakukan sebanyak n subjek yang saling independen dengan setiap subjek diamati secara berulang dalam kurun waktu berbeda yang saling dependen [8]. Terdapat beberapa pendekatan untuk mengestimasi kurva regresi, salah satunya adalah dengan menggunakan estimator polinomial lokal. Kelebihan dari polinomial lokal adalah kemampuannya dalam beradaptasi terhadap data yang artinya membagi data tersebut kedalam suatu wilayah kemudian melakukan estimasi terhadap wilayah yang sudah ditetapkan nilainya tersebut [3]. Estimator polinomial lokal dapat diperoleh dengan metode WLS (*Weighted Least Square*). Sedangkan untuk mengestimasi parameter penghalus (*bandwidth*) dengan menggunakan metode GCV (*Generalized Least Square*).

Kemajuan perekonomian suatu negara dapat ditandai dengan adanya pasar modal yang tumbuh dan berkembang dengan baik [4]. Pasar modal memfasilitasi berbagai sarana dan prasarana kegiatan jual-beli surat-surat berharga dan kegiatan terkait lainnya. Surat berharga bersifat pemilikan dikenal dengan nama saham. Saham merupakan surat berharga

sebagai bukti penyertaan atau kepemilikan seseorang atau badan usaha dalam suatu perusahaan. Jika dilihat dari nilai kapitalisasi pasar, saham dapat dibagi menjadi 3 kelompok yaitu kapitalisasi besar (*Big-Cap*), kapitalisasi sedang (*Mid-Cap*) dan kapitalisasi kecil (*Small-Cap*).

Dalam penelitian ini, akan digunakan estimator polinomial lokal untuk mengestimasi kurva regresi nonparametrik pada data longitudinal. Data yang digunakan adalah data harga penutupan saham (*closing price*) bulanan pada masing-masing kelompok harga saham berdasarkan nilai kapitalisasi pasar pada bulan Januari 2012 – April 2015.

2. TINJAUAN PUSTAKA

2.1 Saham

Saham merupakan salah satu komoditas keuangan yang diperdagangkan di pasar modal yang paling populer. Saham merupakan surat berharga yang dapat dibeli atau dijual oleh perorangan atau lembaga di pasar tempat surat tersebut diperjualbelikan [4]. Sejalan dengan pertumbuhan industri keuangan, saham mengalami pembagian menjadi beberapa jenis salah satunya dilihat berdasarkan nilai kapitalisasi pasar. Adapun pembagian dari nilai kapitalisasi pasar menurut terdapat 3 kelompok yaitu:

a. Kapitalisasi Besar (*Big Cap*)
   Kapitalisasi besar sering disebut juga dengan *blue chip* yang merupakan sekelompok saham unggulan/ saham lapis pertama yang ditransaksikan di bursa efek. Saham yang tergolong dalam kelompok ini adalah saham yang nilai kapitalisasi pasarnya > Rp 5 triliun.
b. Kapitalisasi Sedang (*Mid Cap*)
   Kelompok saham ini sering disebut saham lapis kedua. Kapitalisasi pasar saham kelompok ini berkisar antara Rp 1 triliun – Rp 5 triliun. Umumnya, perusahaan yang sahamnya berada dalam lapis ini adalah perusahaan yang cukup besar dan berada di industrinya untuk waktu cukup lama.
c. Kapitalisasi Kecil (*Small Cap*)
   Kelompok saham ini sering disebut saham lapis ketiga. Umumnya, saham lapis ini memiliki *return on investment* tinggi karena harganya yang relatif murah namun risikonya yang besar. Nilai kapitalisasi pasar saham pada kelompok ini, yaitu kurang dari Rp 1 triliun.

2.2 Data Longitudinal

Data longitudinal merupakan data yang diperoleh dari pengukuran atau pangamatan yang dilakukan sebanyak n subjek yang saling independen dengan setiap subjek diamati secara berulang dalam kurun waktu berbeda yang saling dependen [8]. Jika y_{ij} menyatakan pengamatan untuk subjek ke-i dan waktu ke-j, t_{ij} menyatakan waktu pengamatan untuk subjek ke-i dan waktu ke-j, n menyatakan banyaknya subjek, dan n_i menyatakan banyaknya ulangan pada subjek ke-i dalam kurun waktu berbeda maka diberikan data longitudinal:

$$(t_{ij}, y_{ij}), \quad j = 1,2,\dots,n_i \; ; \; i = 1,2,\dots,n$$

Model regresi nonparametrik untuk data longitudinal dapat dimodelkan sebagai berikut [8]:

$$y_{ij} = \eta(t_{ij}) + e_{ij} \; , \; j = 1,2,...,n_i \; ; \; i = 1,2,...,n$$

dengan $\eta(t_{ij})$ adalah fungsi yang tidak diketahui bentuknya dari data longitudinal dan e_{ij} adalah error pengukuran pada subjek ke-i pada waktu ke-j.

2.3 Estimator Kernel

Fungsi kernel dinotasikan sebagai K(x) merupakan suatu fungsi yang kontinu, simetris dan terbatas. Secara umum fungsi kernel K dengan parameter penghalus (*bandwidth*) h didefinisikan sebagai berikut:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

Berikut terdapat beberapa jenis fungsi kernel yaitu [6]:

1. Kernel Gaussian
   $$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, -\infty < x < \infty$$
2. Kernel Segitiga
   $$K(x) = \begin{cases} 1 - |x|, & |x| \le 1 \\ 0, & x\ yang\ lain \end{cases}$$
3. Kernel Epanechnikov
   $$K(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| \le 1 \\ 0, & x\ yang\ lain \end{cases}$$
4. Kernel *Biweight*
   $$K(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2, & |x| \le 1 \\ 0, & x\ yang\ lain \end{cases}$$

2.4 Polinomial Lokal pada Data Longitudinal

Melalui deret Taylor, $\eta(t_{ij})$ dapat didekati secara lokal oleh polinomial berderajat *p* sebagai berikut [8] :

$$\eta(t_{ij}) \approx \eta(t) + (t_{ij} - t)\eta^{(1)}(t) + \cdots + (t_{ij} - t)^p \eta^{(p)}(t)/p!$$

Misalkan $\beta_r(t) = \eta^{(r)}(t)/r!$, $r = 0,1,2,\ldots, p$, maka persamaan (4) dapat ditulis menjadi:

$$\eta(t_{ij}) \approx \beta_0(t) + (t_{ij} - t)\beta_1(t) + \cdots + (t_{ij} - t)^p \beta_p(t)$$

Parameter $\boldsymbol{\beta}$ dapat diestimasi dengan menggunakan metode *Weighted Least Square* (WLS), yaitu dengan meminimumkan:

$$L = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta})^2 K_{ho}(t_{ij} - t)$$

dengan $\boldsymbol{x}_{ij} = [1, (t_{ij} - t), \ldots, (t_{ij} - t)^p]^T$ dan $K_{ho}(t_{ij} - t) = \frac{1}{h} K\left(\frac{t_{ij} - t}{h}\right)$

Persamaan $\boldsymbol{L}$ dalam bentuk matriks dapat dituliskan sebagai berikut:

$$\boldsymbol{L} = (\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{K}_h (\boldsymbol{y} - \boldsymbol{X\beta})$$

dengan $\boldsymbol{y} = \left[y_{11}, \dots, y_{1n_i}, \dots, y_{n1}, \dots, y_{nn_i}\right]^T$, $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X}_n \end{bmatrix}$,

$\boldsymbol{K}_{ih} = diag\left(K_{ho}(t_{i1} - t), \dots, K_{ho}\left(t_{in_i} - t\right)\right)$, $\boldsymbol{K}_h = diag(\boldsymbol{K}_{1h}, \dots, \boldsymbol{K}_{nh})$.

Sehingga diperoleh estimasi untuk $\widehat{\boldsymbol{\beta}}$ sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T K_h X})^{-1} \boldsymbol{X^T K_h y}$$

2.5 *Generalized Cross Validation* (GCV)

Pemilihan *bandwidth* yang optimal merupakan hal yang sangat penting dalam analisi regresi. Pemilihan *bandwidth* yang terlalu besar mengakibatkan plot hasil estimasi model akan menjauhi plot data awal sehingga menjadi sangat halus (*oversmoothing*). Pemilihan *bandwidth* yang terlalu kecil mengakibatkan plot hasil estimasi model yang berliuk-liuk (*undersmoothing*).Nilai dari *generalized cross validation* (GCV) dapat dihitung dengan rumus sebagai berikut[8]:

$$GCV_i(h) = \frac{n_i^{-1} \sum_{j=1}^{n_i} [y_{ij} - \hat{y}_{ij}]^2}{\{1 - tr(A_\rho)/n\}^2} \qquad , i = 1,2, \dots, n$$

Nilai *bandwidth* (h) yang optimal adalah nilai yang bersesuaian dengan GCV(h) yang minimum.

2.6 Koefisien Determinasi

Suatu nilai atau ukuran yang digunakan untuk mengukur seberapa baik garis regresi sesuai dengan data aktualnya disebut dengan koefisien determinasi (R^2). Koefisien determinasi ini mengukur presentase total variasi variabel dependen Y yang dijelaskan oleh variabel independen di dalam garis regresi[7]. Nilai dari koefisien determinasi dapat diperoleh dengan rumus sebagai berikut:

$$R^2 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(\hat{y}_{ij} - \bar{y}\right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}\right)^2}$$

2.7 Pemilihan Model Terbaik

Jika tujuan utama dari sebuah model adalah untuk meramalkan nilai yang akan datang, maka kriteria alternatif dalam tahap pemilihan model adalah berdasarkan nilai kesalahan terkecil. Besarnya kesalahan dapat dihitung dengan mengurangi data sebenarnya dengan besarnya hasil ramalan.

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

Salah satu cara yang digunakan untuk mengukur tingkat kesalahan yaitu *Mean Square Error* (MSE) yang didefinisikan sebagai berikut [5]:

$$\text{MSE}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} {e_{ij}}^2}{\sum_{i=1}^{n} n_i}$$

2.8 Ketepatan Metode Peramalan

Salah satu ukuran yang digunakan menyangkut galat presentase, yaitu *Mean Absolute Percentage Error* (MAPE) yang dirumuskan sebagai berikut [5]:

$$MAPE = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left|\left(\frac{y_{ij}-\hat{y}_{ij}}{y_{ij}}\right)\times 100\%\right|}{\sum_{i=1}^{n} n_i}$$

Berikut interpretasi dari nilai MAPE yang dihasilkan [1] :

1. MAPE $< 10\%$: peramalan sangat akurat
2. $10\% \leq$ MAPE $< 20\%$: peramalan tersebut baik
3. $20\% \leq$ MAPE $<50\%$: peramalan cukup baik
4. MAPE $\geq 50\%$: peramalan tidak akurat

3. METODE PENELITIAN

Data yang digunakan adalah data sekunder berupa data saham pada perusahaaan Astra Otoparts Tbk. yang merupakan kelompok kapitalisasi besar, perusahaan Astra Graphia Tbk. yang merupakan kelompok kapitalisasi sedang dan perusahaan Mahaka Media Tbk. yang merupakan kelompok kapitalisasi kecil. Data tersebut berupa harga penutupan (*closing price*) saham bulanan yang diambil pada bulan Januari 2012 sampai dengan bulan September 2014 sebagai data *in sample* dan data bulan Oktober 2014 sampai dengan bulan April 2015 sebagai data *out sample*. Data harga penutupan saham tersebut dapat diakses pada situs www.yahoo.finance.com

Tahapan analisis yang dilakukan, yaitu sebagai berikut:

1. Menentukan *bandwidth* optimum serta *arbited fix point* pada masing-masing perusahaan untuk setiap derajat polinomial dengan langkah sebagai berikut:
   a. Mendefinisikan variabel respon y_{ij} dan variabel prediktor t_{ij} pada data *in sample*.
   b. Menentukan fungsi kernel yang digunakan, yaitu kernel Gaussian, kernel Segitiga, kernel Epanechnikov dan kernel *Biweight*.
   c. Menentukan derajat polinomial (p) yang dicobakan, yaitu p=1, p=2, p=3 dan p=4.
   d. Pemilihan *bandwidth* optimum, derajat polinomial serta *arbited fix point* (titik waktu) menggunakan metode GCV.
2. Mengestimasi model regresi nonparametrik berdasarkan estimator polinomial lokal menggunakan titik waktu, *bandwidth* dan derajat polinomial optimal yang diperoleh pada langkah 1.
3. Menghitung nilai MSE dan R^2dari model yang dihasilkan.
4. Pemilihan model regresi nonparametrik pada data longitudinal terbaik dengan MSE minimum.
5. Menghitung ketepatan peramalan model terbaik menggunakan data *out sample*.

Pengolahan data dilakukan menggunakan bantuan Ms. Excel 2013 dan R 2.15.3

4. HASIL DAN PEMBAHASAN

Estimasi model nonparametrik polinomial lokal sangat bergantung pada pemilihan *banwidth* optimal dan derajat polinomial, dimana penentuan *bandwidth* optimal dan derajat polinomial dapat dilihat pada nilai GCV minimum. Dalam penelitian ini fungsi kernel yang

digunakan adalah fungsi kernel Gaussian, kernel Segitiga, kernel Epachenikov dan kernel *Biweight*. Setelah dicobakan pada berbagai fungsi kernel tersebut diperoleh model yang paling baik adalah model regresi menggunakan fungsi Kernel Gaussian dengan MSE sebesar 0,03053464 dan nilai koefisien determinasi sebesar 97,80174%. Pada kernel Gaussian diperoleh nilai *bandwidth* optimal, titik waktu derajat pada masing-masing subjek, dimana subjek 1 mempunyai bandwidth optimal= 3,10 , titik waktu= 7 dan derajat polinomial= 2. Pada subjek 2 mempunyai *bandwidth* optimal= 2,06 , titik waktu= 10 dan derajat polinomial= 1. Pada subjek 3 mempunyai *bandwidth* optimal= 2,27 , titik waktu= 9 dan derajat polinomial= 3. Berikut model untuk masing-masing subjek:

Subjek 1 (Astra Otoparts Tbk) :

$$\hat{y}_{1j} = 3{,}462309 + 0{,}04647756\left(t_{1j} - 7\right) - 0{,}001089802\left(t_{1j} - 7\right)^2$$

Subjek 2 (Astra Graphia Tbk.) :

$$\hat{y}_{2j} = 1{,}386388 + 0{,}03291576\left(t_{2j} - 10\right)$$

Subjek 3 (Mahaka Media Tbk.) :

$$\hat{y}_{3j} = 0{,}08947214 - 0{,}003179995\left(t_{3j} - 9\right) + 0{,}0002656927\left(t_{3j} - 9\right)^2 - (6.215146 \times 10^{-6})\left(t_{3j} - 9\right)^3$$

Dari model tersebut kemudian dilakukan uji ketepatan peramalan, diperoleh nilai MAPE untuk data *out sample* sebesar 11,74493% dengan nilai MSE sebesar 0,03970463

5. KESIMPULAN

Berdasarkan hasil penelitian diperoleh kesimpulan sebagai berikut:

1. Pada perbandingan keempat jenis fungsi kernel, diperoleh model regresi nonparametrik data longitudinal terbaik dengan menggunakan fungsi kernel Gaussian.
2. Diperoleh nilai MAPE untuk data *out sample* sebesar 11,74493%, dimana nilai tersebut terletak diantara 10% dan 20% sehingga dapat dikatakan bahwa model tersebut baik digunakan untuk melakukan prediksi yang akan datang

6. DAFTAR PUSTAKA

[1] Chen, R.J.C., Bloomfield, P., dan Cubbage, F.W. Comparing Forecasting Models in Tourism. *Journal of Hospitality & Tourism Research 2007*. DOI: 10.1177/1096348007309566.

[2] Eubank, R.L. 1999. *Nonparametric Regression and Spline Smoothing Second Edition*. Texas: Departmen of Statistics Southern Methodist Dallas University.

[3] Fan, J. dan Gijbels, I. 1997. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

[4] Hadi, N. 2013. *Pasar Modal: Acuan Teoritis dan Praktis Investasi di Keuangan Pasar Modal*. Yogyakarta : Graha Ilmu.

[5] Makridakis, S., Wheelewright, S.C. dan McGee, V.E. 1995. *Metode dan Aplikasi Peramalan*. Terjemahan Untung Sus Andriyanto dan Abdul Basith. Jakarta: Erlangga.

[6] Ogden, R.T. 1997. *Essential Wavelets for Statistical Applications and Data Analysis*. Boston: Birkhauser.

[7] Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. Yogyakarta: Unit Penerbit dan Percetakan STIM YKPN.

[8] Wu, H. dan Zhang, J. T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis.* New York: John Wiley and Sons, Inc.

[9] www.yahoo.finance.com