PEMODELAN DATA INFLASI INDONESIA PADA SEKTOR TRANSPORTASI, KOMUNIKASI, DAN JASA KEUANGAN MENGGUNAKAN METODE KERNEL DAN SPLINE

Suparti[1] dan Tarno[2]
[1]Staf Jurusan Statistika FSM UNDIP, supartisudargo@yahoo.co.id
[2]Staf Jurusan Statistika FSM UNDIP, tarno.stat@gmail.com

Abstract

In this research, we study data modeling of Indonesian inflation in the transportation, communication and financial services sector using the kernel and spline models. Determination of the optimal models based on the smallest of GCV value and determination of the best model based on the smallest out sampels of Mean Square Error (MSE) value. By modeling the yoy (year on year) inflation data in Indonesia in the transportation, communication and financial services sector In January 2007 to January 2015, shows that the kernel model using Gaussian kernel function obtained optimal model with a bandwidth 0.24 and the optimal spline model with order 5 and 4 points knots. Based on out sampels data in February to August 2015, obtained out sampels MSE value of the spline model is smaller than the kernel model. So that the spline model is better than the kernel model to analyze the inflation data of transportation, communication and financial services sector.

Keywords: Inflation, Transportation, Communication and Financial Services Sector, Kernel, Spline, GCV, MSE.

1. Pendahuluan

Data inflasi merupakan salah satu data runtun waktu ekonomi yang mempunyai sifat volatilitas tinggi, sehingga jika data ini dimodelkan dengan model parametrik ARIMA Box-Jenkins sering mengalami kendala karena ada asumsi yang tidak dipenuhi. Kemudian dikembangkan metode nonparametrik yang tidak mengharuskan adanya asumsi yang ketat seperti halnya metode parametrik. Beberapa model nonparametrik yang dapat digunakan untuk memodelkan data inflasi antara lain model regresi kernel[8] dan model regresi spline[7].

Menurut Bunyamin dan Danila (2011), model inflasi Indonesia terbaik dengan Box - Jenkins menggunakan data inflasi umum tahunan 1998 - 2008 adalah model AR(2) dengan prediksi inflasi pada tahun 2009 sebesar 10,48%[1].Ternyata hasil prediksi ini sangat jauh dengan data aktual inflasi tahun 2009 yang besarnya hanya 2,78%. Suparti, dkk (2014) memodelkan data inflasi umum tahunan (*yoy*) menggunakan data bulan Desember 2006 – Desember 2013 dengan model regresi kernel diperoleh hasil prediksi inflasi tahun 2009 sebesar 2,56%. Prediksi ini lebih mendekati data aktualnya[8].

Suparti, dkk (2015) telah melakukan analisis data inflasi umum yoy menggunakan metode ARIMA Box-Jenkins, kernel dan spline. Metode terbaiknya adalah ARIMA Box-Jenkins yang mempunyai kinerja model sangat bagus. Namun model kernel dan spline juga layak digunakan sebagai model alternatif karena mempunyai MAPE antara 10% - 20% sehingga mempunyai kinerja yang bagus. Ketiga model menghasilkan hasil prediksi inflasi pada tahun 2015 hampir sama[9]. Keuntungan menggunakan metode kernel dan spline tidak mengharuskan datanya mempunyai sifat tertentu.

Dari pengamatan data selama 3 tahun (2011, 2012 dan 2013), telah terjadi kenaikan inflasi umum yang signifikan pada tahun 2013. Kenaikan ini disumbang oleh kenaikan

yang signifikan pada beberapa sektor/kelompok inflasi yaitu sektor transportasi, komunikasi dan jasa keuangan; sektor bahan makanan, sektor makanan jadi, minuman, rokok dan tembakau. Namun kenaikan yang sangat signifikan terjadi pada sektor transportasi, komunikasi dan jasa keuangan. Kenaikan inflasi yang signifikan pada tahun 2013 sebagai efek dari kebijakan pemerintah yang menaikkan harga Tarif Dasar Listrik (TDL) secara bertahap sepanjang tahun 2013 dan kenaikan Bahan Bakar Minyak (BBM) pada pertengahan tahun 2013[9] .

Dalam artikel ini penulis bertujuan melakukan analisis pemodelan data inflasi Indonesia pada Sektor Transportasi, Komunikasi dan Jasa Keuangan menggunakan model regresi kernel dan spline dengan optimasi modelnya menggunakan kriteria *Generalized Cross Validation* (GCV) minimum.

2. Tinjauan Pustaka

2.1. Pengertian dan Indikator Inflasi

Secara sederhana inflasi diartikan sebagai meningkatnya harga-harga secara umum dan terus menerus. Indikator yang sering digunakan untuk mengukur tingkat inflasi adalah Indeks Harga Konsumen (IHK)[4]. Inflasi yang diukur dengan IHK di Indonesia dikelompokkan ke dalam 7 kelompok/sektor pengeluaran yaitu sektor Bahan Makanan; Sektor Makanan Jadi, Minuman, Rokok dan Tembakau; Sektor Perumahan, Air, Listrik, Gas dan Bahan Bakar; Sektor Sandang; Sektor Kesehatan; Sektor Pendidikan, Rekreasi dan Olahraga; Sektor Transportasi, Komunikasi dan Jasa Keuangan[6]. Cara perhitungan inflasi berdasarkan IHK terbagi 3 yaitu inflasi bulanan (*month to month/mtm*) dan inflasi tahunan (*year on year/yoy*) dan inflasi tahun kalender (*year to date/ytd*)[5]. Inflasi bulanan (*mtm*) dihitung dari perubahan indeks bulan tertentu dan indeks bulan sebelumnya pada tahun yang sama yaitu:

$$Inflasi\ (mtm) bulan\ n\ tahun\ t = \frac{IHK\ bulan\ n\ tahun\ t - IHK\ bulan\ (n-1)\ tahun\ t}{IHK\ bulan\ \ (n-1)\ tahun\ t} x\ 100\% \quad (1)$$

Sedangkan inflasi tahunan (*yoy*) dihitung dari perubahan indeks bulan yang sama pada tahun tertentu dan tahun sebelumnya yaitu:

$$Inflasi\ (yoy)\ bulan\ n\ tahun\ t = \frac{IHK\ bulan\ n\ tahun\ t - IHK\ bulan\ n\ tahun\ (t-1)}{IHK\ bulan\ ke\ n\ tahun\ (t-1)} x\ 100\% \quad (2)$$

Inflasi tahun kalender bulan ke *n* dihitung dari perubahan indeks bulan tertentu dan indeks Bulan Desember tahun sebelumnya yaitu:

$$Inflasi\ (ytd)\ bulan\ n\ tahun\ t = \frac{IHK\ bulan\ n\ tahun\ t - IHK\ Bulan\ Desember\ \ tahun\ (t-1)}{IHK\ Bulan\ Desember\ \ tahun\ (t-1)} x\ 100\% \quad (3)$$

2.2. Pemodelan Regresi nonparametrik

Jika diberikan data pengamatan independen $\{(X_i, Y_i)\}_{i=1}^{n}$ maka model regresi nonparametrik adalah $Y_i = g(X_i) + \varepsilon_i,\ i = 1, 2, \ldots, n$ (4)
dengan g fungsi regresi yang tidak diketahui dan ε_i kesalahan pengamatan random independen dengan mean 0 dan varian σ^2 [3]. Pendekatan nonparametrik dilakukan jika asumsi bentuk g tidak diketahui. Untuk mengestimasi fungsi regresi g dapat dilakukan dengan metode kernel dan spline.

Regresi kernel adalah regresi terboboti menggunakan bobot fungsi kernel. Estimator regresi kernel model (4) adalah

$$\hat{g}(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i) Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)} \quad \text{dengan } K_h(x) = \frac{K(\frac{x}{h})}{h} \quad (5)$$

sehingga estimasi regresi di titik x = X_i adalah $\hat{g}(X_i) = \frac{\sum_{j=1}^{n} K_h(X_i - X_j)Y_j}{\sum_{k=1}^{n} K_h(X_i - X_k)} = \sum_{j=1}^{n} [H]_{ij} Y_j$.

Beberapa fungsi kernel diantaranya Kernel Seragam (uniform), Kernel Segitiga, Kernel Epanechnikov dan Kernel Gauss[3].

Model regresi spline keluarga polinomial *truncated* orde *m* dan k titik knot $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ adalah y = g(x) + ε dengan $g(x) = \sum_{i=0}^{m-1} \beta_i x^i + \sum_{j=1}^{k} \beta_{j+m-1} (x - \lambda_j)_+^{m-1}$, dan fungsi *truncated* $(x - \lambda_j)_+^{m-1} = \begin{cases} (x - \lambda_j)^{m-1} ; & x - \lambda_j \geq 0 \\ 0 ; & x - \lambda_j < < 0 \end{cases}$,

untuk $a < \lambda_1 < \lambda_2 < \ldots < \lambda_k < b$, dimana *a* dan *b* merupakan nilai minimum dan maksimum dari x[10] . Bentuk matrik dari model regresi splide adalah Y = X β + ε. Dengan menggunakan metode OLS, estimator parameter β dengan titik knot λ adalah $\hat{\beta}_\lambda = (X_\lambda^T X_\lambda)^{-1} X_\lambda^T Y$, sehingga estimator regresi spline adalah

$$\hat{g}_\lambda(x) = X_\lambda (X_\lambda^T X_\lambda)^{-1} X_\lambda^T Y = H_\lambda Y, \text{ dengan } H_\lambda = X_\lambda (X_\lambda^T X_\lambda)^{-1} X_\lambda^T \quad (6)$$

Keoptimalan estimasi regresi kernel sangat ditentukan oleh pemilihan *bandwidth h* sedangkan keoptimalan regresi spline sangat ditentukan oleh pemilihan titik knotnya. Salah satu metode untuk memperoleh bandwidth dan titik knot optimal dengan meminimalkan nilai *Generalized Cross Validation* (GCV) dengan

$$\text{GCV} = \frac{\text{MSE}}{(n^{-1}\text{trace}[I - H])^2} \text{ dan } MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \text{ [11]}.$$

Menurut Hardle (1990), pemodelan data runtun waktu $\{Z_i, i = 1, 2, ..., n\}$ dapat dimodifikasi menjadi data regresi (X_i, Y_i) dengan mengganti 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 dipandang sebagai masalah pemulusan regresi untuk $(X_i, Y_i) = (Z_{i-1}, Z_i)$, $i = 2, 3, ..., n$ yang dapat diselesaikan dengan model regresi nonparametrik[3].

2.3 Ukuran Kinerja Model

Menurut Lewis (1982) dalam Chen, *et al.* (2008), suatu model dikatakan mempunyai kinerja yang sangat bagus jika nilai MAPE kurang dari 10%, mempunyai kinerja yang bagus jika nilai MAPE besarnya 10% sampai 20%, mempunyai kinerja yang wajar jika MAPE besarnya 20% sampai 50% dan mempunyai kinerja yang tidak akurat jika MAPE besarnya 50% atau lebih[2], dengan rumus

$$\text{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| x100\% \quad (7)$$

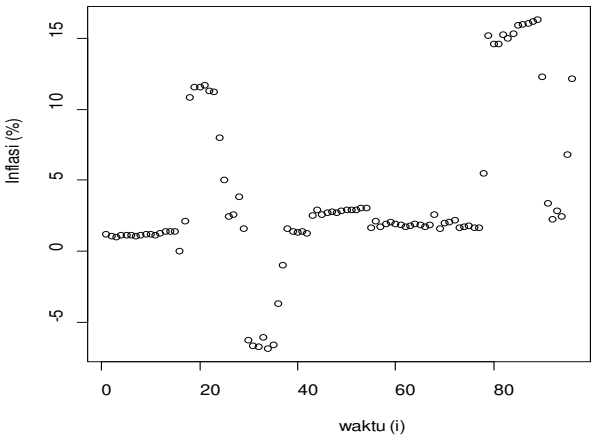
3. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data bulanan dari inflasi *yoy* (*year on year*) pada sektor transportasi, komunikasi dan jasa keuangan mulai bulan Januari 2007 – Januari 2015 sebagai data in sampel untuk membangun model dan data bulan Februari – Desember 2015 sebagai data out sampel digunakan untuk evaluasi kinerja model. Data diambil dari situs resmi Bank Indonesia yaitu data IHK bulanan pada sektor transportasi, komunikasi dan jasa keuangan[6]. Dari data mentah IHK kemudian diolah menjadi data inflasi *yoy* (Persamaan 2) dengan menyamakan tahun dasar perhitungan IHK. Data inflasi $\{Z_i, i = 1, 2, \ldots, 97\}$ merupakan data runtun waktu, selanjutnya data tersebut dimodifikasi

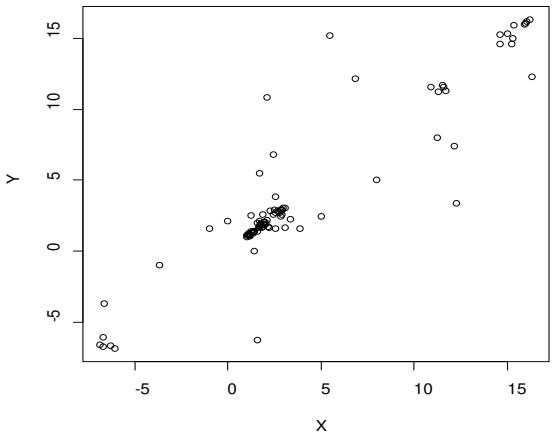
menjadi data $\{(X_i,Y_i),\ i = 2, 3, \ldots n\} = \{(Z_{i-1},Z_i),\ i = 2, 3, \ldots n\}$ dengan $n = 97$. Jadi masalah prediksi $\{Z_i\}$ ekivalen dengan mengestimasi $g(x) = E(Y|X = x)$ untuk runtun waktu dua dimensi $\{(X_i,Y_i),\ i = 2, 3, ..., 97\}$. Selanjutnya fungsi g diestimasi dengan model regresi kernel dengan fungsi kernel Gauss dan regresi spline truncated. Pada model regresi kernel pemilihan nilai *bandwidth h* dicobakan semua kemungkinan h positip dan setiap *h* yang dicobakan dihitung GCV-nya. Pada model regresi spline dicobakan beberapa orde (yaitu orde 2, 3, 4 dan 5) dan beberapa kombinasi titik knot (yaitu 1, 2, 3, 4 titik knot). Setiap model spline yang terbentuk dihitung nilai GCV-nya. Pemilihan model optimal untuk masing-masing metode menggunakan kriteria GCV minimum**.** Model optimal yang terbentuk selanjutnya digunakan untuk memprediksi data out sampel dan dihitung nilai MSE dan MAPE-nya untuk evaluasi kinerja model. Model terbaik berdasarkan nilai MSE out sampel terkecil.

4. Hasil Dan Pembahasan

Data inflasi pada Sektor Transportasi, Komunikasi dan Jasa Keuangan pada bulan Januari 2007 – Januari 2015 digambarkan dalam *scatter plot* Gambar 1. Dari Gambar 1 terlihat data memiliki volatilitas yang tinggi yang ditunjukkan adanya data pengamatan yang tiba-tiba naik dan turun secara tajam di beberapa titik. Kemudian data runtun waktu Gambar 1 dimodifikasi menjadi data (X_i,Y_i) yang diplotkan pada Gambar 2.



Gambar 1. Plot Data inflasi



Gambar 2. Plot Data Modifikasi

Data modifikasi pada Gambar 2 selanjutnya diolah dengan program R menggunakan model regresi kernel dengan fungsi pembobot kernel Gauss dan model regresi spline *truncated* beberapa orde dan titik knot. Beberapa orde dan titik knot yang dicobakan menghasilkan nilai optimal yang disajikan pada Tabel 1. Dari Tabel 1, untuk model kernel optimal diperoleh nilai GCV minimum sebesar 3,380 dicapai pada bandwidth *h* optimal sebesar 0,24 sedangkan pada regresi spline optimal diperoleh nilai GCV minimum sebesar 1,819 dicapai pada orde 5 dengan 4 titik knot (-5,652 ; 0,347; 2,348; 14,348). Sehingga model optimal yang terbentuk adalah sebagai berikut:

Untuk model kernel optimal sesuai Persamaan (5) adalah:

$$\hat{g}(x) = \frac{\sum_{j=1}^{n} K(\frac{x - X_j}{0,24}) Y_j}{\sum_{k=1}^{n} K(\frac{x - X_k}{0,24})} \quad \text{dengan} \quad K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{8}$$

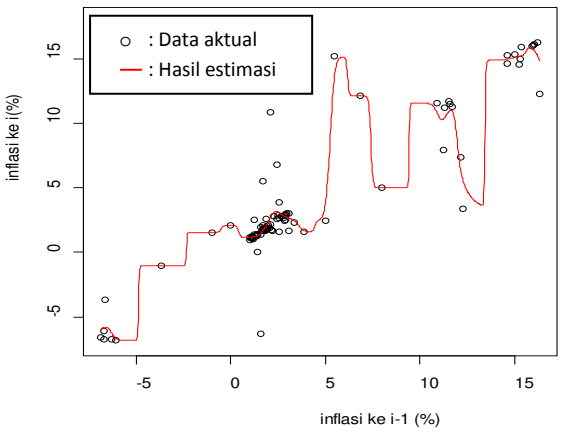
dan model spline optimal sesuai Persamaan (6) adalah:

$$\hat{g}(x) = \begin{cases} -42,961 - 31,730x - 8,355x^2 - 0,850x^3 - 0,028x^4, & \text{untuk } x < -5,652 \\ -42,961 - 31,730x - 8,355x^2 - 0,850x^3 - 0,028x^4 + 0,044(x+5,652)^4, & \text{untuk } -5,652 \le x < 0,348 \\ -42,961 - 31,730x - 8,355x^2 - 0,850x^3 - 0,028x^4 - 0,046(x-0,348)^4, & \text{untuk } 0,348 \le x < 2,348 \\ -42,961 - 31,730x - 8,355x^2 - 0,850x^3 - 0,028x^4 + 0,035(x-2,348)^4, & \text{untuk } 2,348 \le x < 14,348 \\ -42,961 - 31,730x - 8,355x^2 - 0,850x^3 - 0,028x^4 - 0,683(x-14,348)^4, & \text{untuk } x \ge 14,348 \end{cases} \tag{9}$$

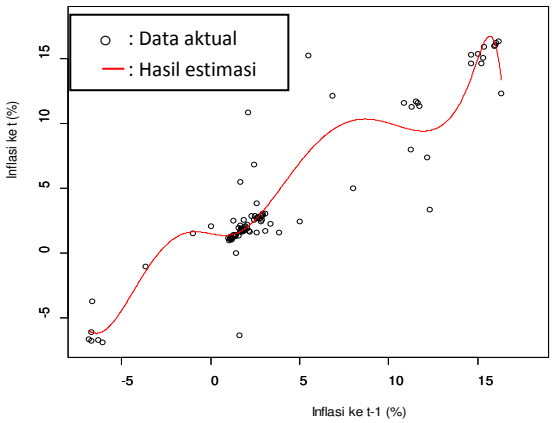
Estimasi model kernel (8) dan model spline (9) ditunjukkan pada Gambar 3 dan 4. Setelah data dikembalikan ke bentuk runtun waktu semula, estimasi model kernel (8) dan model spline (9) ditunjukkan pada Gambar 5. Secara visual, dari Gambar 3, 4 dan 5 terlihat estimasi kedua model mengikuti pola data yang ada.

Tabel 1. Perbandingan Nilai GCV Minimum Model Kernel dan Spline

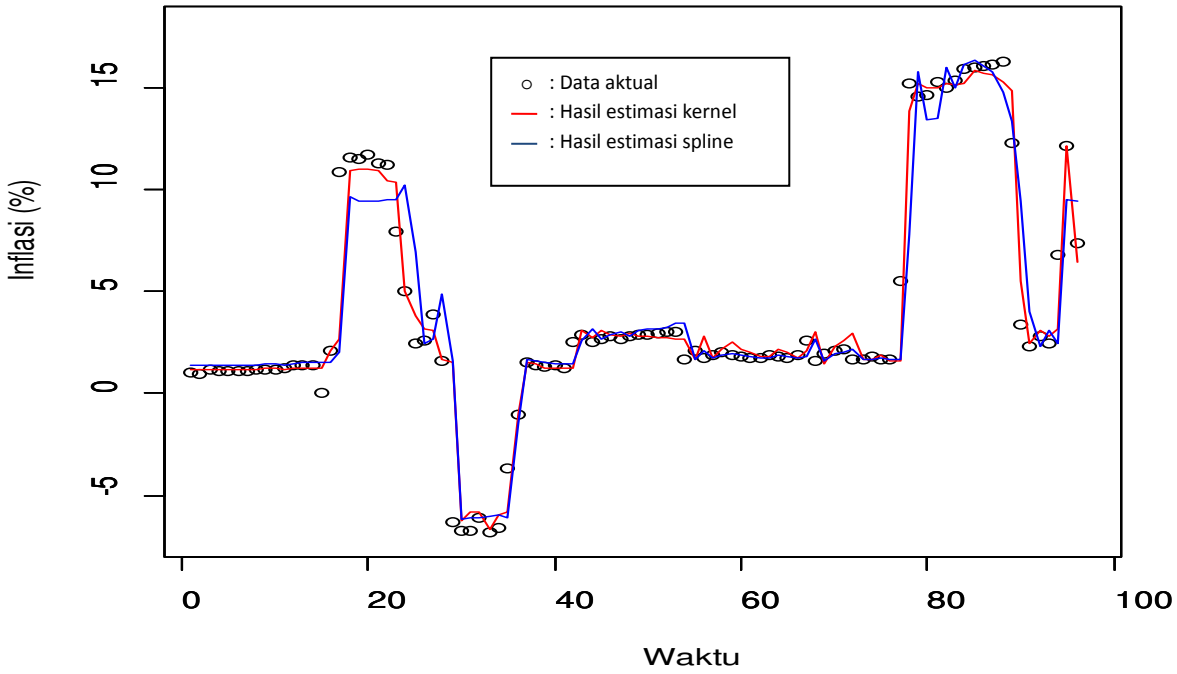
Metode Kernel			***h* optimal**	**GCV min**
			0,24	3,380
	Banyaknya knot	**Orde**	**Knot optimal**	**GCV min**
Metode Spline	1	2	16,194	5,087
		3	16,194	5,197
		4	16,194	5,227
		5	15,999	5,246
	2	2	(11,723; 12,293)	4,648
		3	(5,500; 6,839)	5,046
		4	(12,150; 12,293)	4,979
		5	(11,559 ; 11,613)	4,719
	3	2	(5,004 ; 5,500; 7,985)	3,791
		3	(11,723 ; 12,150; 12,293)	4,443
		4	(11,558; 11,613; 11,723)	4,586
		5	(-6,090; 11,311; 14,636)	3,585
	4	2	(3,348; 4,348; 6,348; 7,348)	3,990
		3	(4,348; 5,348; 6,348; 8,348)	4,072
		4	(-6,652; 2,348; 4,348; 7,348)	3,345
		5	(-5,652; 0,348; 2,348; 14,348)	1,819



Gambar 3. Plot Data Modifikasi dan Estimasi Model Kernel

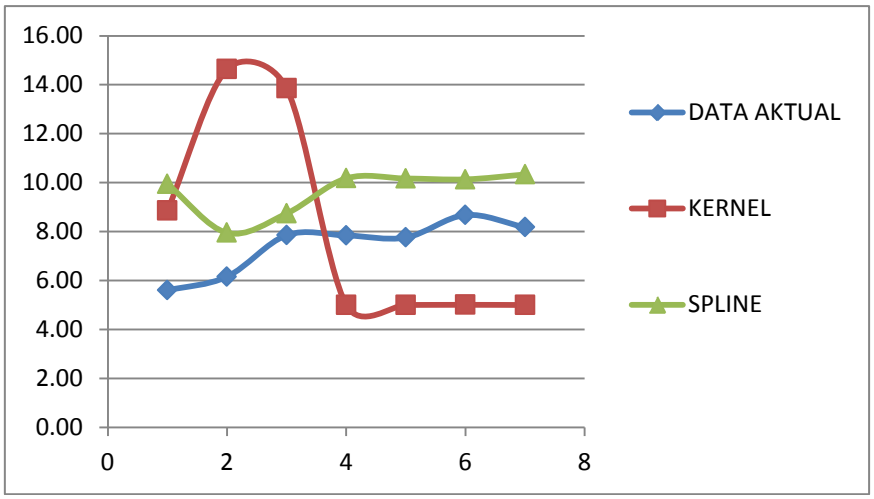


Gambar 4. Plot Data Modifikasi dan Estimasi Model Spline



Gambar 5. Plot Data Aktual dan Estimasi Model Kernel dan Spline

Dari kedua model optimal (8) dan (9) diperoleh nilai prediksi inflasi sektor transportasi, komunikasi dan jasa keuangan pada Bulan Februari – Agustus 2015 yang disajikan dalam Gambar 5, dan terlihat hasil prediksi yang lebih dekat dengan data aktual adalah model spline. Dari Tabel 2, dilihat dari besarnya MSE baik in sampel maupun out sampel dan MAPE out sampel, model spline optimal lebih kecil dari model kernel optimal. Ini menunjukkan bahwa metode spline lebih baik untuk memodelkan data inflasi sektor transportasi, komunikasi dan jasa keuangan dari metode kernel. Dilihat dari besarnya nilai MAPE, model spline mempunyai kinerja yang wajar sedangkan kinerja model kernel tergolong tidak akurat.



Gambar 5. Perbandingan Data Aktual dan Prediksi Kedua Model pada Bulan Februari – Agustus 2015

Tabel 2. Perbandingan Model Kernel dan Model Spline Optimal

	Metode	
	Kernel	**Spline**
Model	Kernel Gauss dengan h = 0,24	Spline orde 5 dengan 4 titik knot (-5,652; 0,348; 2,348; 14,348)
GCV in sampel	3,380	1,819
MSE in sampel	2,100	1,541
MSE out sampel	22,515	5,841
MAPE out sampel	60,682 %	33,526 %
Kinerja model	Tidak akurat	Wajar

5. Kesimpulan

Dari pemodelan menggunakan model nonparametrik kernel dan spline, model terbaik untuk inflasi Indonesia pada sektor transportasi, komunikasi dan jasa keuangan adalah model spline orde 5 dan 4 titik knot dengan kinerja model tergolong wajar.

DAFTAR PUSTAKA

1. Bunyamin dan Danila , N., Estimasi Inflasi di Indonesia dengan Menggunakan Metodologi Box-Jenkins, *National* Journals, 2011, Vol. 18, No. 2.
2. Chen, R.J.C., Blomfield, P., Cubbage, F.W., Comparing Forecasting Models in Touris. *Journal of Hospitality and Tourism Research*, 2008, Vol. 32, No. 1.

3. Hardle, W., *Applied Nonparametric Regression,* Cambridge University Press., New York, 1990.
4. http://www.bi.go.id/id/moneter/inflasi/pengenalan/Contents/Default.aspx.
5. http://www.bi.go.id/id/statistik/metadata/seki/Documents/12_Indeks_Harga_Konsumen_(IHK)_2014.pdf
6. http://www.bi.go.id/seki/tabel/TABEL8_1.pdf
7. Suparti, Analisis Data Inflasi di Indonesia Menggunakan Model Regresi Spline. *Jurnal Media Statistika*, 2013, Vol. 6, No. 1.
8. Suparti, Warsito, B. dan Mukid, M.A., Analisis Data Inflasi Di Indonesia Sebelum dan Sesudah Kenaikan TDL dan BBM Tahun 2013 Menggunakan Model Regresi Kernel*, Prosiding Konferensi Nasional Matematika ke XVII* ITS, Surabaya 11 – 14 Juni 2014.
9. Suparti, Warsito, B. dan Mukid, M.A., Analisis Data Inflasi Di Indonesia Menggunakan Model Arima Box-Jenkins, Kernel dan Spline, *Prosiding Seminar Nasional SAINTEKINFO* UNS, Surakarta, 2015.
10. Wu, H. and Zang, J. T., *Nonparametric Regression Methods for Longitudinal Data Analysis*, John Wiley and Sons, New Jersey, 2006.
11. Takezawa, K., *Introduction to Nonparametric Regression*, John Wiley & Sons, Inc., New Jersey, 2006.