

Asimtotik Model Multivariate Adaptive Regression Spline

Bambang Widjanarko Otok¹⁾, Suryo Guritno ²⁾, Subanar ²⁾

¹⁾S3 Matematika UGM, Jurusan Statistika FMIPA - ITS Surabaya

²⁾Jurusan Matematika, FMIPA - UGM, Yogyakarta

Diterima 25-08-2007

Disetujui 25-03-2008

ABSTRACT

Parameter estimation in MARS model executed by minimizing penalized least-squares (PLS). Through some requirement, asymptotic estimator characteristic from MARS prediction model has been successfully proven. The research result shows that GCV can work properly to determine the best model that applied on MARS model. Solar's vehicles produce opacity that exceed the standard limit of emission quality which was adjusted in Kepmen LH No. 35 Year 1993, as large as 88 percent from 408 percent. Applying years, cylinder volume, type of machine, and vehicle's radius are the variables that influences the opacity.

Keywords: GCV, MARS, Opacity, penalized least-squares

PENDAHULUAN

Analisis regresi memperlihatkan hubungan dan pengaruh variabel prediktor terhadap variabel respon. Untuk mengestimasi kurva atau parameter regresi, ada dua pendekatan yaitu parametrik dan nonparametrik. Pendekatan parametrik digunakan jika ada informasi sebelumnya tentang bentuk kurva, yang diperoleh berdasarkan teori atau pengalaman masa lalu, sehingga dapat dikatakan bahwa mengestimasi kurva ekuivalen dengan mengestimasi parameter, dimana hasil estimasi mengikuti model tertentu. Sedang pendekatan nonparametrik digunakan jika tidak ada informasi tentang bentuk kurva $f(x_i)$, tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar (Eubank 1988; Emond & Steven 1997).

Hardle (1990), jika kurva regresi merupakan model parametrik maka disebut sebagai regresi parametrik dan apabila model yang diasumsikan ini benar, maka pendugaan parametrik sangat efisien, tetapi jika tidak, menyebabkan interpretasi data yang menyesatkan. Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain pendekatan histogram, estimator *spline*, estimator *kernel*, estimator deret orthogonal, analisis *wavelet* dan lain-lain. Pendekatan estimator *spline* ada bermacam-macam antara lain *spline original*, *spline type M*, *spline relaxed*, *spline* terbobot dan lain-lain. Pendekatan *spline* mempunyai suatu basis fungsi.

Basis fungsi yang biasa dipakai antara lain *spline truncated* dan *B-spline*.

Model MARS Friedman (1991), bertujuan untuk prediksi, yang merupakan kombinasi *recursive partitioning regression* (RPR) dengan *spline truncated* yang telah dimodifikasi. Tujuan paper ini mencari estimator dari model tersebut, dan mengkaji sifat-sifat asimtotik estimator. Sedangkan kajian terapan dalam rangka mengimplementasikan kajian teori, yaitu dengan menerapkan dibidang lingkungan.

Emisi Gas Buang. Emisi adalah makhluk hidup, zat energi, dan atau komponen lain yang dihasilkan dari kegiatan yang masuk atau dimasukkan ke dalam udara ambien yang mempunyai dan tidak mempunyai potensi sebagai pencemar (Bapedal 2002). Dalam Peraturan Pemerintah RI Nomor 41 Tahun 1999 pada Bab I Pasal 1 tentang pengendalian pencemaran udara, definisi dari pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan atau komponen lain ke dalam udara ambien oleh kegiatan manusia, sehingga mutu udara ambien turun ke tingkat tertentu yang menyebabkan udara ambien tidak memenuhi fungsinya.

Beberapa konsep dan istilah tentang emisi gas buang yang terdapat dalam Kepmen LH Nomor Kep.35/MenLH/10/1993 tentang ambang batas emisi gas buang kendaraan bermotor adalah sebagai berikut:

- kendaraan bermotor adalah kendaraan yang digerakkan oleh peralatan teknis yang berada pada kendaraan itu
- ambang batas emisi gas buang kendaraan bermotor adalah batas maksimum zat atau bahan pencemar yang boleh dikeluarkan langsung dari pipa gas buang kendaraan bermotor
- kendaraan bermotor selain sepeda motor 2 langkah dengan bahan bakar solar dengan bilangan setana ³ 45 ditentukan maksimum ekuivalen 50% Bosch untuk opasitas pada diameter 102 mm
- ketebalan asap gas buang diukur pada kondisi percepatan bebas (*idling*)

Gas karbon monoksida (CO) dikatakan paling beracun karena dapat menghalangi transportasi O₂ ke jaringan sel-sel tubuh. Dalam konsentrasi lebih dari 750 ppm dapat mengakibatkan terjadinya kematian dalam waktu singkat. Kombinasi gas CO dalam Hb (haemoglobin) darah dapat membentuk carboxyhaemoglobin (COHb), di mana daya ikat Hb terhadap CO adalah ± 210 kali lebih cepat dibandingkan terhadap O₂ (Seno 2005).

Baku mutu ambien untuk CO dalam udara bebas adalah 35 ppm (rata-rata 1 jam) dan 9 ppm (rata-rata 8 jam) yang dihitung berdasarkan rata-rata pergerakan yang tidak boleh terlampaui lebih dari satu kali dalam setahun. Nilai ini cenderung sulit dicapai di daerah perkotaan dan industri karena aktifitasnya yang menghasilkan CO sangat tinggi (Baumbach 1996)

Opasitas digunakan dalam berbagai konteks untuk menguraikan sesuatu yang tidak transparan. Dalam emisi gas buang khususnya untuk kendaraan bermotor bahan bakar solar, opasitas adalah kemampuan asap menyerap cahaya yang dinyatakan dalam satuan persen (%). Opasitas disebut juga sebagai kepekatan asap atau ketebalan asap (Wikipedia 2006)

Emisi gas buang kendaran bermotor terkandung zat-zat berbahaya yang sangat mempengaruhi kesehatan manusia. Bahaya atau efek samping yang ditimbulkan sebagai akibat dari terpenuhinya gas buang kendaraan bermotor tidak seketika dapat dirasakan dan tidak kasat mata, artinya tidak dapat dirasakan secara langsung. Beberapa akibat efek samping di antaranya adalah:

- Pb, dapat mengakibatkan tekanan darah tinggi serta dapat menurunkan tingkat kecerdasan dan perkembangan mental anak. Di samping itu dapat mengurangi fungsi ginjal dan mengurangi fungsi reproduksi pada kaum laki-laki.
- HC dan Nox dapat menyebabkan iritasi mata, memicu asma dan kanker paru-paru.
- CO menyebabkan mual-mual, kepala menjadi pusing, bahkan pada konsentrasi yang tinggi dapat menyebabkan kematian.
- Sox menyebabkan tenggorokan gatal dan batuk-batuk.

Tingginya konsentrasi dari emisi rumah kaca (CO, CO₂, HC, Nox dan Sox) menyebabkan peningkatan suhu bumi (efek rumah kaca) karena dapat menyerap sinar infra merah dari spektrum matahari. Fenomena ini menyebabkan masalah lain karena berakibat naiknya permukaan laut, perubahan iklim (musim panas atau kemarau lebih panjang), terperangkapnya polutan karena tidak dapat terdispersi secara vertikal akibat temperatur di lapisan atmosfer atas lebih tinggi (membentuk lapisan inversi) (Santawinata 2007).

BAHAN DAN METODE

Data yang digunakan adalah data sekunder hasil pengujian emisi opasitas gas buang pada kendaraan berbahan bakar solar yang dilakukan oleh Ikatan Peduli Otomotif Otopoint dalam beberapa kegiatan yang diselenggarakan di Jawa Timur pada tahun 2004-2007.

Variabel respon (y) adalah opasitas emisi gas buang kendaraan berbahan bakar solar. Sedangkan variabel Independen (x) terdiri atas tujuh variabel karakteristik kendaraan yang diduga berpengaruh terhadap opasitas (Sumargo 2005), yakni:

- a. Merk (X₁)
Terdapat empat kategori merk kendaraan berbahan bakar solar yang terobservasi yakni Daihatsu (kode 1), Mitsubishi (kode 2), Toyota (kode 3), dan Isuzu (kode 4).
- b. Model Kendaraan (X₂)
Model kendaraan merupakan variabel yang memiliki hubungan struktural dengan variabel merk kendaraan. Terdapat lima kategori model kendaraan yang terobservasi pada penelitian ini yakni Taft (kode 1), Kuda (kode 2), Elf (kode 3), Kijang (kode 4), dan Panther (kode 5).

c. Jenis (X3)

Kendaraan yang terobservasi dalam penelitian ini terdiri atas empat kategori jenis kendaraan yakni Pick up (kode 1), Jeep (kode 2), Minibus (kode 3), dan Station Wagon (kode 4).

d. Tipe mesin (X4)

e. Usia Pakai (X5)

Usia pakai merepresentasikan lama pakai kendaraan pada saat uji emisi dilaksanakan. Usia pakai dihitung melalui selisih antara tahun pelaksanaan uji emisi dengan tahun pembuatan kendaraan.

f. Isi silinder (cc) (X6)

Isi silinder yang kendaraan berbahan bakar solar yang terobservasi antara interval 2000 cc sampai 2800 cc.

g. Jarak tempuh (km) (X7)

Jarak tempuh merepresentasikan seberapa besar jarak yang ditempuh kendaraan yang tercatat pada saat uji emisi. Jarak tempuh ditunjukkan melalui odometer yang terpasang pada kendaraan.

Analisis data dilakukan dengan bantuan software SPSS versi 13.0 dan MARS versi 2.0 (Steinberg 2001) dengan tahapan sebagai berikut:

1. Tahapan yang dilakukan dalam pemodelan menggunakan metode MARS adalah sebagai berikut:
 - Menentukan titik-titik *knots* berdasarkan kriteria GCV minimum,
 - Menentukan maksimal jumlah fungsi basis. Friedman (1990), menyarankan maksimum fungsi basis yang digunakan adalah antara dua sampai empat kali ukuran variabel prediktor,
 - Menentukan maksimal jumlah interaksi, minimal jumlah pengamatan diantara *knots* atau minimum observasi (MO) dilakukan dengan *trial and error* sampai diperoleh model optimal dengan nilai GCV minimum,
 - Menduga koefisien model pada Model MARS.
2. Setelah diperoleh model terbaik, selanjutnya dilakukan perhitungan prediksi opasitas berdasarkan model yang diperoleh.

HASIL DAN PEMBAHASAN

Kajian Teori Model MARS. Misalkan \underline{Y} menunjukkan variabel respon tunggal tergantung pada M variabel prediktor \underline{x} , di mana $\underline{x} = (x_{1i}, \dots, x_{Mi})$, maka model regresi dapat ditulis sebagai berikut :

$$\underline{Y}_i = f(x_{1i}, \dots, x_{Mi}) + \varepsilon_i \tag{1}$$

atas beberapa domain. $(x_{1i}, \dots, x_{Mi}) \in D \subset R^M$ Variabel random ε diasumsikan mempunyai rata-rata nol dan variansi σ_ε^2 .

Estimasi dari kurva regresi $f(x)$ secara umum didapatkan melalui *penalized least squares* (PLS) yakni meminimumkan persamaan berikut:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + u^2 \int_a^b (f^{(m)}(x))^2 dx \tag{2}$$

Estimator Model MARS dapat ditulis sebagai berikut:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] \tag{3}$$

Di mana:

- a_0 = basis fungsi induk
- a_m = koefisien dari basis fungsi ke- m
- M = maksimum basis fungsi (*nonconstant basis function*)
- K_m = derajat interaksi
- s_{km} = nilainya ± 1
- $x_{v(k,m)}$ = variabel independen
- t_{km} = nilai *knots* dari variabel independen $x_{v(k,m)}$

Sehingga model MARS, dinyatakan dalam persamaan berikut:

$$y_i = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] + \varepsilon_i$$

dalam bentuk matrik dapat ditulis sebagai berikut:

$$\underline{Y} = \underline{B}\underline{a} + \underline{\varepsilon} \tag{4}$$

dengan $\underline{Y} = (y_1, \dots, y_n)^T$; $\underline{a} = (a_0, \dots, a_M)^T$

$$\mathbf{B} = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} s_{1m}(x_{1(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{K_M M}(x_{1(K_M, M)} - t_{K_M M}) \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{2(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{K_M M}(x_{2(K_M, M)} - t_{K_M M}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{n(1,m)} - t_{1m}) & \cdots & \prod_{k=1}^{K_M} s_{K_M M}(x_{n(K_M, M)} - t_{K_M M}) \end{pmatrix}$$

Estimasi Parameter . Model MARS tersebut dalam persamaan (4), $\underline{\mathbf{a}}$ merupakan parameter yang akan diestimasi dari data, menggunakan pendekatan PLS yang telah dimodifikasi. Hal ini dijabarkan dalam bukti teorema di bawah ini.

Teorema 1. Dengan menggunakan estimator MARS dalam persamaan (3), dan \mathbf{B} matrik non singular dan parameter smoothing $u^2 = 0$ maka penduga dari $\hat{\underline{\mathbf{a}}}$ adalah

$$\hat{\underline{\mathbf{a}}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T Y$$

dengan $\mathbf{B} = [1, (x_{v,(k,m)} - t_{km})_1^K]$

Bukti:

Perhatikan persamaan (2), yaitu:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + u^2 \int_a^b (f^{(m)}(u))^2 du$$

dengan $u^2 = 0$, persamaan (2) menjadi:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (5)$$

Dari persamaan (4), maka $\hat{f}(x_i) = \mathbf{B}\underline{\mathbf{a}}$, sehingga persamaan (5) menjadi:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{B}\underline{\mathbf{a}})^2 = (\underline{Y} - \mathbf{B}\underline{\mathbf{a}})^T (\underline{Y} - \mathbf{B}\underline{\mathbf{a}}) = Z$$

Untuk memperoleh estimator $\hat{\underline{\mathbf{a}}}$ digunakan metode kuadrat terkecil, yang pada prinsipnya meminimumkan Z, dinyatakan sebagai berikut:

$$Z = (\underline{Y} - \mathbf{B}\underline{\mathbf{a}})^T (\underline{Y} - \mathbf{B}\underline{\mathbf{a}})$$

$$Z = (\underline{Y}^T \underline{Y} - \underline{\mathbf{a}}^T \mathbf{B}^T \underline{Y} - \underline{Y}^T \mathbf{B} \underline{\mathbf{a}} + \underline{\mathbf{a}}^T \mathbf{B}^T \mathbf{B} \underline{\mathbf{a}})$$

$$Z = (\underline{Y}^T \underline{Y} - 2\underline{\mathbf{a}}^T \mathbf{B}^T \underline{Y} + \underline{\mathbf{a}}^T \mathbf{B}^T \mathbf{B} \underline{\mathbf{a}})$$

Untuk memperoleh persamaan normal dilakukan dengan menurunkan secara parsial terhadap $\underline{\mathbf{a}}$ dengan hasil sebagai berikut:

$$\frac{\partial Z}{\partial \underline{\mathbf{a}}} = -2\mathbf{B}^T \underline{Y} + 2\mathbf{B}^T \mathbf{B} \underline{\mathbf{a}} = \mathbf{0}$$

$$-\mathbf{B}^T \underline{Y} + \mathbf{B}^T \mathbf{B} \hat{\underline{\mathbf{a}}} = \mathbf{0} \quad (6)$$

$$\mathbf{B}^T \mathbf{B} \hat{\underline{\mathbf{a}}} = \mathbf{B}^T \underline{Y}$$

karena \mathbf{B} matrik non singular, maka

$$\hat{\underline{\mathbf{a}}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \underline{Y} \quad (7)$$

sedangkan turunan kedua terhadap $\underline{\mathbf{a}}$ adalah

$$\frac{\partial^2 Z}{\partial \underline{\mathbf{a}}^2} = 2\mathbf{B}^T \mathbf{B} > \mathbf{0}$$

Jadi teorema 1 terbukti.

Sifat Asimtotik Model MARS. Untuk menyelidiki sifat-sifat asimtotik $\hat{\underline{\mathbf{a}}}$ diperlukan beberapa asumsi:

Asumsi 1: Jika (X_i, t_i) adalah titik-titik pada barisan yang tetap, dan ada fungsi kontinu $h_j(\cdot)$ pada $[0,1]$ maka masing-masing elemen X_i memenuhi, $X_i = h_j(t_i) + u_{ij}$, $1 \leq i \leq n, 1 \leq j \leq p$ dengan $\{u_{ij}\}$ adalah barisan real yang memenuhi,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n u_i^T u_i = \mathbf{C} \text{ dan, } \limsup_{n \rightarrow \infty} \frac{1}{a_n} \max_{1 \leq k \leq n} \left| \sum_{i=1}^n u_{ij} \right| < \infty$$

Di mana, $u_{ij} = (u_{i1}, \dots, u_{ip})^T$, $a_n = n^{1/2} \log n$,

\mathbf{C} = matrik definit positif dengan orde p

Asumsi 2: $g(\cdot)$ dan $h(\cdot)$ memenuhi kondisi Lipschitz order 1.

Asumsi 3: Fungsi bobot $W_{ni}(\cdot)$ memenuhi:

$$(i) \quad \text{Max}_{1 \leq i \leq n} \sum_{j=1}^n W_{ni}(t_j) = O(1), \quad \text{Max}_{1 \leq j \leq n} \sum_{i=1}^n W_{ni}(t_j) = O(1)$$

$$(ii) \quad \text{Max}_{1 \leq i, j \leq n} W_{ni}(t_j) = O(b_n)$$

$$(iii) \quad \text{Max}_{1 \leq i \leq n} \sum_{j=1}^n W_{ni}(t_j) I(|t_i - t_j| > c_n) = O(c_n)$$

dengan b_n dan c_n adalah dua barisan yang memenuhi

$$\text{untuk } n \rightarrow \infty, \quad nb_n^2 \log^4 n < \infty, \quad nc_n^2 > 0,$$

$$nc_n^4 \log n < \infty, \quad nb_n^2 c_n^2 < \infty$$

Sifat asimtotik kenormalan estimator $\hat{\mathbf{a}}$ digunakan lemma berikut.

Lemma 1: Jika asumsi 1, 2 dan 3 terpenuhi, maka

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{X} = \mathbf{C} \text{ dengan } \mathbf{C} \text{ matriks definit positif.}$$

Bukti: Jika $\bar{h}_{ns}(t_i) = h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{ks}$ dan $X_{is} = h_s(t_i) + u_{is}$,

maka $\mathbf{X}^T \mathbf{X}$ dengan $(s, m = 1, \dots, p)$ dapat didekomposisikan sebagai berikut:

$$\begin{aligned} \sum_{i=1}^n X_{is} X_{im} &= \sum_{i=1}^n \left[X_{is} - \sum_{k=1}^n W_{nk}(t_i) X_{ks} \right] \left[X_{im} - \sum_{k=1}^n W_{nk}(t_i) X_{km} \right] \\ &= \sum_{i=1}^n \left[(h_s(t_i) + u_{is}) - \sum_{k=1}^n W_{nk}(t_i) X_{ks} \right] \left[(h_m(t_i) + u_{im}) - \sum_{k=1}^n W_{nk}(t_i) X_{km} \right] \\ &= \sum_{i=1}^n \left[\left(h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{ks} \right) + u_{is} \right] \left[\left(h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{km} \right) + u_{im} \right] \\ &= \sum_{i=1}^n \left[\bar{h}_{ns}(t_i) + u_{is} \right] \left[\bar{h}_{nm}(t_i) + u_{im} \right] \\ &= \sum_{i=1}^n u_{is} u_{im} + \sum_{i=1}^n \bar{h}_{ns}(t_i) \bar{h}_{nm}(t_i) + \sum_{i=1}^n \left[\bar{h}_{ns}(t_i) u_{im} \right] + \sum_{i=1}^n \left[\bar{h}_{nm}(t_i) u_{is} \right] \\ &= \sum_{i=1}^n u_{is} u_{im} + T_n^1 + T_n^2 + T_n^3 \end{aligned}$$

$$\begin{aligned} T_n^1 &= \sum_{i=1}^n \bar{h}_{ns}(t_i) \bar{h}_{nm}(t_i) \\ &= \sum_{i=1}^n \left[h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{ks} \right] \left[h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{km} \right] \\ &= \sum_{i=1}^n \left[h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_s(t_k) + u_{ks}) \right] \left[h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_m(t_k) + u_{km}) \right] \\ &= \sum_{i=1}^n \left[h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_s(t_k)) - \sum_{k=1}^n W_{nk}(t_i) u_{ks} \right] \left[h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_m(t_k)) - \sum_{k=1}^n W_{nk}(t_i) u_{km} \right] \\ &= \sum_{i=1}^n A_i A_2 \\ &= O(c_n) + O(a_n b_n) \end{aligned}$$

$$\begin{aligned} T_n^2 &= \sum_{i=1}^n \left[\bar{h}_{ns}(t_i) u_{im} \right] \\ &= \sum_{i=1}^n \left[h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{ks} \right] u_{im} \\ &= \sum_{i=1}^n \left[h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_s(t_k) + u_{ks}) \right] u_{im} \\ &= \sum_{i=1}^n \left[\left(h_s(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_s(t_k)) \right) u_{im} - \left(\sum_{k=1}^n W_{nk}(t_i) u_{ks} \right) u_{im} \right] \\ &= O(c_n a_n) + O(a_n^2 b_n) \end{aligned}$$

$$\begin{aligned} T_n^3 &= \sum_{i=1}^n \left[\bar{h}_{nm}(t_i) u_{is} \right] \\ &= \sum_{i=1}^n \left[h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) X_{km} \right] u_{is} \\ &= \sum_{i=1}^n \left[h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_m(t_k) + u_{km}) \right] u_{is} \\ &= \sum_{i=1}^n \left[\left(h_m(t_i) - \sum_{k=1}^n W_{nk}(t_i) (h_m(t_k)) \right) u_{is} - \left(\sum_{k=1}^n W_{nk}(t_i) u_{km} \right) u_{is} \right] \\ &= O(c_n a_n) + O(a_n^2 b_n) \end{aligned}$$

sehingga,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_{is} X_{im} &= \lim_{n \rightarrow \infty} \frac{1}{n} \left(\sum_{i=1}^n u_{ij} u_{ik} + T_n^1 + T_n^2 + T_n^3 \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \left(\sum_{i=1}^n u_{ij} u_{ik} + [O(c_n) + O(a_n b_n)] + [O(a_n c_n) + O(a_n^2 b_n)] + [O(a_n c_n) + O(a_n^2 b_n)] \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n u_{ij} u_{ik} + \\ &\quad \lim_{n \rightarrow \infty} \frac{1}{n} \left([O(c_n) + O(a_n b_n)] + [O(a_n c_n) + O(a_n^2 b_n)] + [O(a_n c_n) + O(a_n^2 b_n)] \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n u_{ij} u_{ik} + 0 \\ &= c_{sm} \end{aligned}$$

Jadi Lemma 1 terbukti.

Dari Lemma 1, maka $\mathbf{X}^T \mathbf{X} = O(n)$ artinya terdapat bilangan real M_1 , sedemikian hingga

$$\frac{1}{n} (\mathbf{X}^T \mathbf{X})_i \leq M_1, \text{ dan}$$

$$E(\mathbf{X}^T \boldsymbol{\xi}) = \mathbf{0}; \text{ Var}(\mathbf{X}^T \boldsymbol{\xi}) = \sigma^2 (\mathbf{X}^T \mathbf{X})$$

Dengan menggunakan teorema 1, jika $k = n^{1/2} M_2$

dan $\frac{1}{n} (\mathbf{X}^T \mathbf{X})_i \leq M_1$ terhadap variabel random $\mathbf{X}^T \boldsymbol{\xi}$

maka setiap $M_2 > 0$ berlaku

$$\begin{aligned} P[|\mathbf{X}^T \boldsymbol{\xi} - \mathbf{0}| \geq n^{1/2} M_2] &\leq \frac{\sigma^2 (\mathbf{X}^T \mathbf{X})}{(n^{1/2} M_2)^2} \\ &\leq \sigma^2 \frac{(\mathbf{X}^T \mathbf{X})}{n M_2^2} \\ &\leq \sigma^2 \frac{1}{n} \mathbf{X}^T \mathbf{X} \frac{1}{M_2^2} \\ &\leq \sigma^2 \frac{M_1}{M_2^2} \end{aligned}$$

atau,

$$P[|n^{1/2}\mathbf{X}^T\varepsilon|\geq M_2]\leq\sigma^2\frac{M_1}{M_2^2} \quad (8)$$

jika untuk sembarang $\varepsilon_1 > 0$ dipilih M_1 sehingga

$$\sigma^2\frac{M_1}{M_2^2}\leq\varepsilon_1 \text{ maka,}$$

$$M_2\geq\sqrt{\sigma^2\frac{M_1}{\varepsilon_1}}=\sigma\sqrt{\frac{M_1}{\varepsilon_1}}$$

sehingga persamaan (8) dapat ditulis sebagai,

$$P[|n^{1/2}\mathbf{X}^T\varepsilon|\geq M_2]\leq\varepsilon_1 \quad (9)$$

Misalkan X_i vektor yang berukuran $p \times 1$ yaitu baris ke- i dari matrik \mathbf{X} dan $Z_i = X_i\varepsilon_i$ maka $V_i = \sigma^2 X_i X_i^T$ dan

$$\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n V_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma^2 X_i X_i^T = \sigma^2 \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i X_i^T = \sigma^2 \mathbf{C}$$

Jadi,

$$n^{-1/2}\mathbf{X}^T\varepsilon \xrightarrow{d} N(0, \sigma^2\mathbf{C}) \quad (10)$$

Dari persamaan (4) dan substitusikan ke persamaan (7), diperoleh:

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{a} + \varepsilon) \\ &= (\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{B})\mathbf{a} + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon \\ &= \mathbf{a} + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon \end{aligned} \quad (11)$$

sehingga, $\hat{\mathbf{a}} - \mathbf{a} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon$

$$\begin{aligned} \sqrt{n}(\hat{\mathbf{a}} - \mathbf{a}) &= \sqrt{n}[(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon] \\ &= n\frac{1}{\sqrt{n}}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon \\ &= n(\mathbf{B}^T\mathbf{B})^{-1}\frac{1}{\sqrt{n}}\mathbf{B}^T\varepsilon \\ &= \left(\frac{1}{n}\mathbf{B}^T\mathbf{B}\right)^{-1}\left(\frac{\mathbf{B}^T\varepsilon}{\sqrt{n}}\right) \end{aligned} \quad (12)$$

Dengan Lemma 1, dan persamaan (10) maka:

$$\begin{aligned} \sqrt{n}(\hat{\mathbf{a}} - \mathbf{a}) &\xrightarrow{d} \mathbf{C}^{-1}\mathbf{Z} \text{ dengan } \mathbf{Z} \sim N(0, \sigma^2\mathbf{C}) \text{ adalah} \\ \mathbf{C}^{-1}\mathbf{Z} &\sim N(0, \sigma^2\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}). \end{aligned}$$

Sehingga

$$\sqrt{n}(\hat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, \sigma^2\mathbf{C}^{-1})$$

Selanjutnya mengkaji estimator $\hat{\mathbf{a}}$ konsisten untuk

\mathbf{a} yang artinya $\lim_{n \rightarrow \infty} P(|\hat{\mathbf{a}}_n - \mathbf{a}| < \varepsilon) = 1$ atau $p \lim \hat{\mathbf{a}} = \mathbf{a}$.

Pandang persamaan (4) dan substitusikan ke persamaan (7), diperoleh:

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{a} + \varepsilon) \\ &= (\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{B})\mathbf{a} + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon \\ &= \mathbf{a} + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon \\ &= \mathbf{a} + \left(\frac{1}{n}\mathbf{B}^T\mathbf{B}\right)^{-1}\frac{1}{n}\mathbf{B}^T\varepsilon \end{aligned} \quad (13)$$

maka,

$$\begin{aligned} p \lim \hat{\mathbf{a}} &= \mathbf{a} + p \lim \left[\left(\frac{1}{n}\mathbf{B}^T\mathbf{B}\right)^{-1}\frac{1}{n}\mathbf{B}^T\varepsilon \right] \\ p \lim \hat{\mathbf{a}} &= \mathbf{a} + \lim_{n \rightarrow \infty} \left(\frac{1}{n}\mathbf{B}^T\mathbf{B}\right)^{-1} p \lim \left[\frac{1}{n}\mathbf{B}^T\varepsilon \right] \end{aligned} \quad (14)$$

menurut lemma 1, persamaan (14) menjadi;

$$\begin{aligned} p \lim \hat{\mathbf{a}} &= \mathbf{a} + [(\mathbf{C})^{-1} \times 0] \\ p \lim \hat{\mathbf{a}} &= \mathbf{a} \end{aligned}$$

sedangkan varians estimator adalah,

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Var}(\hat{\mathbf{a}}_n) &= \lim_{n \rightarrow \infty} \text{Var}[\mathbf{a}_n + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon] \\ &= \lim_{n \rightarrow \infty} \text{Var}[(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon] \\ &= \lim_{n \rightarrow \infty} (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\text{Var}(\varepsilon) \\ &= \lim_{n \rightarrow \infty} (\mathbf{B}^T\mathbf{B})^{-1}\sigma^2 = \lim_{n \rightarrow \infty} \left(\frac{1}{n}\mathbf{B}^T\mathbf{B}\right)^{-1}\left(\frac{1}{n}\sigma^2\right) \\ &= \mathbf{C}^{-1} \times 0 = 0 \end{aligned}$$

Berdasarkan hasil di atas, ternyata estimator $\hat{\mathbf{a}}$ secara asimtotik mempunyai sifat kenormalan, konsisten dan varians minimum.

Penerapan MARS pada Opasitas. Pemodelan MARS dilakukan pada berbagai kombinasi nilai maksimum basis fungsi (BF), maksimum interaksi (MI), dan minimum observasi (MO). Model terbaik yang dihasilkan adalah model yang menghasilkan nilai GCV minimum yakni model MARS pada BF = 14, MI = 3, dan MO = 1. Persamaan (4.1) merupakan fungsi prediksi yang dihasilkan untuk pemodelan antara variabel merk (x_1), model (x_2), jenis (x_3), tipe mesin (x_4), usia pakai (x_5), isi silinder (x_6), dan jarak tempuh (x_7) kendaraan terhadap opasitas gas buang yang diukur dalam satuan persen (y).

$$\begin{aligned} \hat{f}(x) &= 87.690 - 2.056BF_2 - 0.050BF_3 \\ &\quad - 0.160BF_4 + 0.748734 \times 10^{-7}BF_7 \end{aligned} \quad (15)$$

dengan

$$\begin{aligned} BF_2 &= \max(0, 11 - x_5) \\ BF_3 &= \max(0, x_6 - 2200) \\ BF_4 &= \max(0, 2200 - x_6) \end{aligned}$$

$$BF_6 = \max(0, 611022 - x_7) * BF_3$$

$$BF_7 = (x_4 = 2 \text{ atau } x_4 = 4 \text{ atau } x_4 = 5 \text{ atau } x_4 = 6 \text{ atau } x_4 = 7 \text{ atau } x_4 = 9 \text{ atau } x_4 = 10 \text{ atau } x_4 = 11) * BF_6$$

Persamaan (15) menunjukkan bahwa :

1. Setiap kenaikan satu satuan $(11 - x_5)$, akan mengurangi opasitas sebesar 2,056 dengan menganggap BF_3 , BF_4 , dan BF_7 konstan. Perubahan tersebut terjadi pada kendaraan yang memiliki usia pakai kurang dari 11 tahun.
2. Setiap kenaikan satu satuan $(x_6 - 2200)$, akan mengurangi opasitas sebesar 0,050 dengan menganggap BF_2 , BF_4 , dan BF_7 konstan. Perubahan tersebut terjadi pada kendaraan dengan isi silinder di atas 2200 cc.
3. Setiap kenaikan satu satuan $(2200 - x_6)$, akan mengurangi opasitas sebesar 0,160 dengan menganggap BF_2 , BF_3 , dan BF_7 konstan. Perubahan tersebut terjadi pada kendaraan dengan isi silinder kurang dari 2200 cc.
4. Setiap kenaikan satu satuan $(611022 - x_7)$, akan menambah opasitas sebesar 0.748734×10^{-7} dengan menganggap BF_2 , BF_3 , dan BF_4 konstan. Perubahan tersebut terjadi pada kendaraan dengan tipe mesin F 70 atau LF 80 atau VB 5 W atau TBR 541 LV 25 atau TBR 541 LM 25 atau LF 82 atau TBR 541 LS 25 atau TBR 54 PRLC dengan isi silinder di atas 2200 cc dan apabila kendaraan memiliki jarak tempuh lebih dari 611022 km.

Berdasarkan beberapa variabel yang dilibatkan dalam model tersebut, dapat diketahui bahwa variabel yang berpengaruh terhadap opasitas adalah tipe mesin, usia pakai, jarak tempuh, dan isi silinder kendaraan. Hal ini didukung dengan nilai skor variabel penting seperti tercantum dalam Tabel 1.

Tabel 1. menunjukkan besarnya pengaruh variabel yang masuk dalam model. Variabel usia pakai dan isi silinder memiliki pengaruh kuat terhadap besarnya

Tabel 1. Skor variabel penting dalam model

Variabel	Skor
Usia Pakai	100,000
Isi Silinder	67,901
Tipe mesin	53,670
Jarak Tempuh Kendaraan	53,670
Merk Kendaraan	0,000
Model Kendaraan	0,000
Jenis Kendaraan	0,000

opasitas. Hal ini terlihat dari skor variabel pentingnya yang cukup besar, masing-masing bernilai 100 dan 67,901. Variabel tipe mesin dan jarak tempuh kendaraan juga memberikan pengaruh terhadap opasitas dengan skor variabel penting yang sama yakni sebesar 53,67. Untuk variabel merk, model, dan jenis kendaraan memiliki skor yang sangat kecil dan tidak memberikan sumbangan yang berarti dalam pembentukan model. Berdasarkan informasi tersebut, dapat diketahui bahwa secara statistik pengaruh variabel merk, model, dan jenis kendaraan terhadap nilai opasitas dapat diabaikan karena tertutupi (*masked*) oleh variabel yang paling berpengaruh (usia pakai, isi silinder, tipe mesin, dan jarak tempuh kendaraan).

KESIMPULAN

Model regresi $y_i = \hat{f}(x) + \varepsilon_i$ dengan kurva regresi f yang didekati dengan model MARS, mempunyai bentuk umum secara matematis seperti pada persamaan (3.1.2). Estimasi parameter pada kurva regresi $\hat{f}(x)$ dilakukan dengan meminimumkan *penalized least-squares* (PLS) pada persamaan (2) yang dimodifikasi bertujuan menemukan suatu penyelesaian \hat{a}_0 dan \hat{a}_m , $m = 1, \dots, M$. Sebagaimana telah dijabarkan dalam teorema 3.1.1 dan teorema 3.1.2 diperoleh estimasi untuk model MARS. Melalui beberapa persyaratan, sifat asimtotik estimator dari model MARS prediksi telah berhasil dibuktikan.

Hasil kajian empiris menunjukkan bahwa GCV dapat bekerja dengan baik dalam menentukan pemilihan model terbaik yang diterapkan pada model MARS. Sehingga dapat diambil kesimpulan bahwa 88 persen dari 408 kendaraan yang diamati tidak lulus uji emisi sehingga dapat dikatakan bahwa sebagian besar kendaraan berbahan bakar solar menghasilkan opasitas yang melebihi batas baku mutu emisi yang ditetapkan dalam Kepmen LH Nomor 35 Tahun 1993. Usia pakai, isi silinder, tipe mesin, dan jarak tempuh kendaraan merupakan variabel yang berpengaruh terhadap opasitas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih pada Ikatan Peduli Otomotif Otopoint, khususnya Pak Suwito atas diskusi dan bantuan memperoleh data Emisi Gas Buang (Opasitas), sehingga penelitian dapat diselesaikan.

DAFTAR PUSTAKA

- Emond, M. & Steven, G.** 1997. An efficient estimator for the generalized semi linear model. *Journal of the American Statistical Association* **92**, 1033-1040.
- Eubank, R.** 1988. Spline smoothing and nonparametric regression. New York: Marcel Dekker.
- Friedman, J.H.** 1991. Multivariate Adaptive Regression Splines (with discussion). *The Annals of Statistics* **19**: 1–141
- Hardle, W.** 1990. Smoothing Techniques With Implementation in S. Springer-Verlag: New York
- Menteri Negara Lingkungan Hidup.** 1993. *Keputusan Menteri Negara Lingkungan Hidup nomor 35 tahun 1993 tentang ambang batas emisi gas buang kendaraan bermotor, Kementrian Lingkungan Hidup, Republik Indonesia.*
- Santawinata, S.** 2007. Polusi Karena Mesin Bermasalah. Pikiran Rakyat Bandung. <http://www.pikiranrakyatbandung.com/> (13 Maret 2007)
- Seno, P.A.** 2005. Motor “2 tak” penyebab polusi (2-habis) emisi gas buangnya sama-sama berbahaya, Suara Merdeka Online. <http://www.suaramerdeka.com> (7 Pebruari 2007)
- Sholekhudin, M.** 2006. Tanpa asap belum tentu oke emisinya. Intisari Online. <http://www.intisari-online.com> (7 Pebruari 2007)
- Steinberg, D., et al.** 2001. MARS User Guide. Salford Systems.
- Sumargo S.** 2005. Ambang emisi gas buang tidak realistis (Evaluasi Terhadap Kepmen LH Nomor35 Tahun 1993). <http://www.suaramerdeka.com> (13 Maret 2007)