

UJI KOEFISIEN VARIANSI KONSTAN DALAM REGRESI NONPARAMETRIK

Oleh:

Asri Ode Samura

Tadris Matematika, IAIN Ternate
asri22samura@gmail.com

ABSTRAK

Tulisan ini membahas uji baru untuk hipotesis koefisien variansi konstan dalam model umum regresi nonparametrik. Uji ini didasarkan pada estimasi jarak antara kuadrat dari fungsi regresi dan fungsi variansi. Dalam tulisan ini telah dilakukan pengujian formal untuk hipotesis dari koefisien variansi konstan dalam regresi nonparametrik yaitu dengan melakukan sebuah uji simulasi kecil, yang mana hipotesis nol akan ditolak jika persamaannya terpenuhi, dan sebaliknya hipotesis nol akan ditolak jika hipotesisnya tidak terpenuhi. Dengan menggunakan teknik smoothing yakni fungsi kernel dalam hal ini dilakukan uji simulasi kecil untuk membuktikan hipotesis uji bootstrap

Kata Kunci : Analisis Regresi, Regresi Parametrik, Regresi Nonparametrik, Teknik Smoothing, Kernel, Bootstrap

ABSTRACT

This write propose a new test for the hypothesis of a constant coefficient of variance in the common nonparametric regression model. The test is based on the estimated distance between the square of the regression function and variance function. In this write have been done a formal test for the hypothesis of a constant coefficient of variance in nonparametric regression is to perform a small simulation test, in which the null hypothesis is rejected if the equation is fulfilled, and otherwise the null hypothesis is rejected if the hypothesis is not fulfilled. By using the smoothing technique that is the kernel function in this small simulation test to prove the hypothesis of the bootstrap test.

Keywords: Regression Analysis, Regression Nonparametric, Smoothing Technique, Kernel, Bootstrap

I. PENDAHULUAN

Analisis regresi berkembang pesat mengatasi permasalahan yang timbul dalam aplikasi, diantaranya transformasi dan pembobotan untuk mengatasi atau mengoreksi ketidaksesuaian model, perlakuan khusus dalam adanya observasi influensial, regresi polinomial, *piecewise polynomial fitting*, regresi dengan variabel indikator yang dapat digunakan sebagai pendekatan untuk analisis variansi, bagaimana mengatasi adanya multikolinearitas pada variabel independen, regresi robust, regresi nonlinear dan untuk variabel independen yang berupa waktu

perkembangan regresi mengarah ke runtun waktu (time series). Regresi di samping digunakan untuk mengetahui bentuk hubungan antar peubah, regresi juga dapat dipergunakan untuk maksud-maksud peramalan.

Tujuan analisis adalah menentukan hampiran yang beralasan untuk fungsi $m(\cdot)$ ini. Ada dua metode yang dapat digunakan, yaitu metode regresi parametrik dan metode regresi nonparametrik.

Pandang model umum regresi nonparametrik berikut:

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i; i = 1, 2, \dots, n \quad (1.2)$$

dimana m merupakan fungsi regresi dan σ fungsi variansi dan variabel random ε_i memenuhi $E[\varepsilon_i | X_i = x] = 0$ dan $E[\varepsilon_i^2 | X_i = x] = 1$. Dalam banyak aplikasi variansi dapat diasumsikan sepadan dengan rata-rata kuadrat (*squared mean*) yang berkorespondensi dengan asumsi dari koefisien variansi konstan.

Dalam regresi nonparametrik, tidak ada asumsi tentang bentuk fungsi regresi $m(\cdot)$. Umumnya fungsi regresi yang diasumsikan termuat dalam fungsi berdimensi tak hingga. Untuk mengkonstruksi model regresinya dipilih ruang fungsi yang sesuai yang mana fungsi regresi diasumsikan termasuk di dalamnya. Pemilihan ruang fungsi ini biasanya dimotivasi oleh sifat kelicinan (*smoothness*) yang diasumsikan dimiliki oleh fungsi regresi. Ada beberapa teknik untuk mengestimasi fungsi regresi $m(\cdot)$ dalam regresi nonparametrik, salah satunya adalah estimator kernel.

II. LANDASAN TEORI

1. Regresi Parametrik

Dengan menggunakan n pengamatan untuk suatu model linear sederhana:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.1)$$

dengan Y_i adalah peubah tidak bebas, X_i adalah peubah bebas dengan $i = 1, 2, \dots, n$, β_0 dan β_1 adalah parameter-parameter yang tidak diketahui, ε_i adalah *Disturbance error*. Diberlakukan asumsi-asumsi model ideal tertentu terhadap galat e yaitu bahwa galat menyebar $NID(0, \sigma^2)$. Dengan pemenuhan terhadap asumsi kenormalan dapat digunakan regresi parametrik untuk mengetahui bentuk hubungan antar peubah regresi pada data contoh yang diamati. Asumsi-asumsi tersebut biasanya disebut asumsi klasik regresi linear.

Diberikan n pengamatan $\{(X_i, Y_i) : i = 1, 2, 3, \dots, n\}; X_i \in R; Y_i \in R$. Pandang model regresi:

$$Y_i = m(X_i) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (2.2)$$

dengan Y_i variabel respon, X_i variabel prediktor dan ε_i sesatan random tidak terobservasi yang diasumsikan tidak berkorelasi dengan mean 0.

Salah satu analisis alternatif lain yang dapat digunakan adalah dengan regresi nonparametrik karena dalam regresi nonparametrik tidak diperlukan pemenuhan asumsi kenormalan.

2. Regresi Nonparametrik

Penggunaan regresi nonparametrik dilandasi pada asumsi:

- Contoh yang diambil bersifat acak dan kontinu;
- Regresi $(Y|X)$ bersifat linear;
- Semua nilai X_i saling bebas.

Sebagai bandingan dari metode nonparametrik tentunya adalah metode parametrik, yang mendominasi statistika klasik. Perbedaan penting antara metode regresi parametrik dan regresi nonparametrik adalah derajat ketergantungan terhadap informasi tentang m yang didapat dari data dan pelaku eksperimen. Untuk mengkonstruksi model regresi nonparametrik, dipilih ruang fungsi yang sesuai, dimana m termuat di dalamnya. Pemilihan ini tergantung pada sifat halus (*smoothness*) fungsi regresi yang harus dimiliki. Data kemudian digunakan untuk mengestimasi elemen ruang fungsi yang merupakan representatif fungsi yang diketahui.

Ada beberapa teknik untuk mengestimasi fungsi regresi $m(\cdot)$ dalam regresi nonparametrik, salah satunya adalah *kernel*. Sebagai contoh estimator kernel untuk fungsi regresi $m(\cdot)$ untuk variabel prediktor berdimensi satu dirumuskan dengan

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n K_h(x - X_i) Y_i}{\sum_{i=1}^n K_h(x - X_i)}$$

dengan $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ dengan $K(x)$ disebut fungsi kernel dan h disebut parameter *smoothing* atau *bandwidth*.

3. Teknik-teknik Smoothing

a. Ide Dasar Smoothing

Tujuan dari smoothing adalah untuk membuang variabilitas dari data yang tidak memiliki efek sampling sehingga ciri-ciri dari data akan tampak lebih jelas. Smoothing telah menjadi sinonim dengan metode-metode nonparametrik yang digunakan untuk mengestimasi fungsi-fungsi.

Jadi mungkin untuk menggunakan sesuatu seperti rata-rata lokal (*local averaging*) dari data yang mendekati x untuk membangun sebuah penduga (estimator) dari $m(x)$. Prosedur rata-rata lokal ini dapat dipandang sebagai dasar dari teknik-teknik *smoothing*. Secara formal prosedur ini didefinisikan sebagai

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^n W_{ni}(x) Y_i \quad (2.3)$$

dengan $\{W_{hi}(x): i = 1, 2, \dots, n\}$ adalah barisan pembobot dan $\hat{m}(x)$ menyatakan estimator dari $m(x)$. Dari (2.3.1) besarnya rata-rata dikontrol oleh $\{W_{hi}(x): i = 1, 2, \dots, n\}$ yang diukur oleh parameter *Smoothing*.

b. Kernel Smoothing

Fungsi kernel K yang umum dipakai adalah fungsi densitas dan biasanya dilengkapi dengan asumsi-asumsi tertentu. Diasumsikan bahwa fungsi kernel K kontinu, terbatas, simetri terhadap pusat interval yang terintegrasi ke satu,

$$\int K(u) du = 1 \quad (2.4)$$

Barisan pembobot untuk *kernel smoother* (untuk satu dimensi x) didefinisikan oleh

$$W_{ni}(x) = \frac{K_{h_n}(x - X_i)}{\hat{f}_{h_n}(x)} \quad (2.5)$$

dimana

$$\hat{f}_{h_n}(x) = n^{-1} \sum_{i=1}^n K_{h_n}(x - X_i) \quad (2.6)$$

dan dimana

$$K_{h_n}(u) = h_n^{-1} K\left(\frac{u}{h_n}\right)$$

adalah kernel dengan faktor skala h_n . Menekan ketergantungan $h = h_n$ pada ukuran sampel n , barisan pembobot kernel (2.5) dapat disingkat sebagai $\{W_{ni}(x)\}_{i=1}^n$. Fungsi $\hat{f}_h(\cdot)$ merupakan *estimator densitas kernel Rosenblatt-Parzen* (Rosenblatt (1956; Parzen (1962)) dari densitas (marginal) X . Bentuk (2.5) dari bobot kernel $W_{hi}(x)$ telah diusulkan oleh Nadaraya (1964) dan Watson (1964) dan oleh karena itu

$$\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{i=1}^n K_h(x - X_i)}$$

sering disebut *estimator Nadaraya-Watson*. Bentuk bobot kernel ditentukan oleh K , sedangkan ukuran bobot parameter ditentukan oleh h , yang disebut dengan *bandwidth* atau lebar pita.

c. Estimator Densitas Kernel

Definisi umum dari kernel K :

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

dimana $K(\cdot)$ merupakan *kernel* dan h adalah *bandwidth*. Fungsi K ini merupakan fungsi pembobot yang dinamakan fungsi kernel, selanjutnya estimator

$$\begin{aligned}\hat{f}_h(x) &= \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\end{aligned}\quad (2.7)$$

dinamakan estimator densitas kernel untuk fungsi densitas $f(x)$.

Terlihat dari rumus (2.7) bahwa $\hat{f}_h(x)$ tergantung pada kernel K dan parameter h . Parameter h ini dinamakan parameter smoothing atau sering dinamakan *bandwidth* atau lebar pita. Fungsi kernel K yang umum dipakai adalah fungsi densitas dan biasanya dilengkapi dengan asumsi-asumsi tertentu, sedangkan f minimal mempunyai kelicinan tingkat kedua atau mempunyai derivatif kedua ($f''(x)$). Diasumsikan bahwa fungsi kernel K kontinu, terbatas, simetris terhadap pusat interval, dan mempunyai support yang finit atau dengan kata lain $K(u)$ memenuhi kondisi:

$$\begin{aligned}\int K(u)du &= 1 \\ \int u K(u)du &= 0 \\ 0 < \int u^2 K(u)du &= \kappa_2 < \infty.\end{aligned}\quad (2.8)$$

Lemma 2.1: Jika $\int K(s)ds = 1$ maka $\int \hat{f}_h(x)dx = 1$

Diberikan beberapa contoh fungsi kernel yang lain:

1. Kernel Uniform: $\frac{1}{2}I(|u| \leq 1)$
2. Kernel Triangle: $(1 - |u|)I(|u| \leq 1)$
3. Kernel Epanechnikov: $\frac{3}{4}(1 - u^2)I(|u| \leq 1)$
4. Kernel Quartic: $\frac{15}{16}(1 - u^2)^2I(|u| \leq 1)$
5. Kernel Triweight: $\frac{35}{32}(1 - u^2)^3I(|u| \leq 1)$
6. Kernel Gaussian: $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$
7. Kernel Cosinus: $\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)I(|u| \leq 1)$

d. Statistik Densitas Kernel

Estimasi densitas kernel didasarkan pada dua parameter, bandwidth h dan fungsi densitas kernel K .

Teorema 2.1: Jika $\hat{f}_h(x)$ diberikan oleh (2.7), maka untuk $n \rightarrow \infty$. $E[\hat{f}_h(x)] \rightarrow f(x) \int K(s)ds = f(x)$, $h \rightarrow 0$.

Bukti:

$$\begin{aligned}E[\hat{f}_h(x)] &= \frac{1}{n} \sum_{i=1}^n E[K_h(x - X_i)] \\ &= E[K_h(x - X)] \\ &= \int K_h(x - u)f(u)du \\ &= \int K(s)f(x + sh)ds\end{aligned}$$

Jika dimisalkan $h \rightarrow 0$, dapat dilihat bahwa

$$E[\hat{f}_h(x)] \rightarrow f(x) \int K(s)ds = f(x), h \rightarrow 0.$$

Teorema 2.2: Jika $\hat{f}_h(x)$ diberikan oleh (2.7), maka:
 $Bias[\hat{f}_h(x)] = \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2)$ dengan $\mu_2(K) = \int s^2 K(s)ds$.

Bukti:

Dari Teorema (2.1) diperoleh:

$$\begin{aligned} Bias[\hat{f}_h(x)] &= \int K(s)f(x+sh)ds - f(x) \\ &= \int K(s) \left[f(x) + shf'(x) + \frac{h^2 s^2}{2} f''(x) + o(h^2) \right] ds - f(x) \\ &= f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) - f(x) \end{aligned} \quad (2.9)$$

Karena K simetris menuju 0, maka bentuk linear $\int sK(s)hf'(x)ds = 0$. Secara ringkas bias estimasi densitas kernel dapat ditulis:

$$Bias[\hat{f}_h(x)] = \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2), h \rightarrow 0 \quad (2.10)$$

Akibat dari Teorema (2.2) adalah:

$$\begin{aligned} [Bias[\hat{f}_h(x)]]^2 &= \left[\frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) \right]^2 \\ &= \frac{h^4}{4} [f''(x) \mu_2(K)]^2 + o(h^4) \end{aligned}$$

Selanjutnya variansi disajikan dalam Teorema berikut,

Teorema 2.3: Jika $\hat{f}_h(x)$ diberikan oleh (2.7), dengan $\|K\|_2^2 = \int K^2(s)ds$, maka
 $Var[\hat{f}_h(x)] = (nh)^{-1} \|K\|_2^2 f(x) + o((nh)^{-1})$, $nh \rightarrow \infty$.

Bukti:

Karena X_i adalah i.i.d,

$$\begin{aligned} Var[\hat{f}_h(x)] &= n^{-2} Var\left(\sum_{i=1}^n K_h(x - X_i)\right) \\ &= n^{-2} \sum_{i=1}^n Var[K_h(x - X_i)] \\ &= n^{-1} Var[K_h(x - X)] \\ &= n^{-1} \{E[K_h^2(x - X)] - (E[K_h(x - X)])^2\} \\ &= n^{-1} \left\{ h^{-2} \int K^2\left(\frac{x-u}{h}\right) f(u)du - (f(x) + o(h^2))^2 \right\} \\ &= n^{-1} \left\{ h^{-1} \int K^2(s)f(x+sh)ds - (f(x) + o(h^2))^2 \right\} \\ &= n^{-1} \left\{ h^{-1} \|K\|_2^2 (f(x) + o(h)) - (f(x) + o(h))^2 \right\} \end{aligned}$$

disini digunakan $E[K_h(x - X)] = f(x) + o(h)$ dari (2.10) dan

$$\int K^2(s)ds(f(x) + o(h)) = \|K\|_2^2(f(x) + o(h)).$$

Oleh karena itu, secara ringkas diperoleh *Variansi estimasi densitas kernel*

$$\text{Var}[\hat{f}_h(x)] = (nh)^{-1} \|K\|_2^2 f(x) + o((nh)^{-1}), nh \rightarrow \infty \quad (2.11)$$

Teorema 2.4: Jika $\hat{f}_h(x)$ diberikan oleh (2.7), maka

$$\text{MSE}[\hat{f}_h(x)] = \frac{1}{nh} f(x) \|K\|_2^2 + \frac{h^4}{4} (f'(x) \mu_2(K))^2 + o((nh)^{-1}) + o(h^4), \\ h \rightarrow 0, nh \rightarrow \infty.$$

e. Smoothing Regresi

Tujuan pemasangan (*fitting*) kurva regresi adalah mencari hubungan antara variabel-variabel $\{X_i\}_{i=1}^n$ dan $\{Y_i\}_{i=1}^n$ dimana X_i dipandang untuk menjelaskan nilai Y_i . Untuk pengepasan kurva regresi, bobot pengamatan Y bergantung pada jarak X ke- x , yakni digunakan estimator

$$\hat{m}_h(x) = n^{-1} \sum_{i=1}^n W_h(x; X_1, \dots, X_n) Y_i.$$

$W_h(\cdot)$ mendefinisikan fungsi bobot dependen atas parameter smoothing h dan sampel X_1, \dots, X_n sebagai variabel penjelas. Oleh karena itu, bentuk umum dari *Regresi Nonparametrik halus*

$$\hat{m}_h(x) = n^{-1} \sum_{i=1}^n W_{hi}(x) Y_i \quad (2.12)$$

f. Estimator Nadaraya-Watson

Estimator kernel untuk fungsi regresi $m(\cdot)$ dikonstruksikan sebagai berikut, perhatikan bahwa:

$$\begin{aligned} m(x) &= E(Y|X = x) \\ &= \int y f(y|x) dy \\ &= \int \frac{y f(x, y) dy}{f(x)} \end{aligned} \quad (2.13)$$

Untuk pembilang bisa diestimasi menggunakan densitas gabungan $f(x, y)$ kernel multiplikatif, (lihat Sesi 2.9, W. Hardle, (1990)).

$$\hat{f}_{h_1, h_2}(x, y) = n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) K_{h_2}(y - Y_i) \quad (2.14)$$

Oleh karena itu, dapat dikerjakan di luar dari estimasi pembilang, yang disajikan dalam Teorema berikut.

Teorema 2.5: Jika $\hat{f}_{h_1, h_2}(x, y)$ diberikan oleh (2.15), maka estimasi densitas dari

$$\int y \hat{f}_{h_1, h_2}(x, y) = n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) Y_i.$$

Bukti:

$$\begin{aligned} \int y \hat{f}_{h_1, h_2}(x, y) &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int y K_{h_2}(y - Y_i) dy \\ &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int \frac{y}{h_2} K\left(\frac{y - Y_i}{h_2}\right) dy \end{aligned}$$

$$= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int (sh_2 + Y_i) K(s) ds$$

$$= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) Y_i$$

Dengan mengambil $\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$ dan $\int y \hat{f}_{h_1, h_2}(x, y) = n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) Y_i$, diperoleh:

$$\hat{m}_h(x) = \frac{\int y \hat{f}_{h_1, h_2}(x, y) dy}{\hat{f}_h(x)}$$

$$= \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{j=1}^n K_h(x - X_j)}. \quad (2.15)$$

Dalam pengertian umum dari estimasi kurva regresi nonparametrik (lihat persamaan (2.12)), bahwa pengamatan dari bobot memiliki bentuk

$$W_{hi}(x) = \frac{h^{-1} K\left(\frac{x - X_i}{h}\right)}{\hat{f}_h(x)}$$

g. Statistik Nadaraya-Watson Estimator

Defenisi

$$r(x) = \int y f(x, y) dy = m(x) f(x)$$

$$\hat{r}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i$$

Estimasi kurva regresi diberikan dengan

$$\hat{m}_h(x) = \frac{\hat{r}_h(x)}{\hat{f}_h(x)}.$$

Untuk pembilang $\hat{r}_h(x)$ ekspektasi dan variansi secara rinci disajikan sebagai berikut:

Teorema 2.6: Jika $\hat{r}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i$, dengan

$$E\hat{r}_h(x) = \iint K_h(x - u) y f(y|u) dy du \quad \text{dan} \quad m(x) = E(Y|X = x) = \frac{\int y f(x, y) dy}{f(x)},$$

maka $E[\hat{r}_h(x)] = \int K_h(x - u) r(u) du.$ (2.16)

Teorema 2.7: Jika $E[\hat{r}_h(x)]$ diberikan oleh (2.16), dengan $\int s^2 K(s) = \mu_2(K)$ dan $\int K(s) shr'(x) ds = 0$, maka $E[\hat{r}_h(x)] = r(x) + \frac{h^2}{2} r''(x) \mu_2(K) + o(h^2)$ $h \rightarrow 0$.

Bukti:

$$E[\hat{r}_h(x)] = \int K_h(x - u) r(u) du$$

$$= \int \frac{1}{h} K\left(\frac{x - u}{h}\right) r(u) du$$

$$= h^{-1} \int K(s) r(x + sh) h ds$$

$$= \int K(s) r(x + sh) ds$$

$$= \int K(s) \left[r(x) + shr''(x) + \frac{h^2 s^2}{2} r''(x) + o(h^2) \right] ds$$

$$\begin{aligned} &= r(x) + \frac{h^2}{2} r''(x) \int s^2 K(s) ds + o(h^2) \\ &= r(x) + \frac{h^2}{2} r''(x) \mu_2(K) + o(h^2). \end{aligned}$$

Untuk menghitung variansi dari $\hat{r}_h(x)$ diberikan dalam Teorema berikut.

Teorema 2.8: Jika $\hat{r}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i$, dengan $\|K\|_2^2 = \int K^2(s) ds$ dan $s^2(x) = E[Y^2|X=x]$, maka $Var[\hat{r}_h(x)] = n^{-1} h^{-1} f(x) s^2(x) \|K\|_2^2 + o((nh)^{-1})$ ($nh \rightarrow h$). (2.17)

Selanjutnya, ketika persamaan (2.16) dan (2.17) digabungkan, maka akan diperoleh formula untuk *Mean Squared Error* dari $\hat{r}_h(x)$:

$$MSE[\hat{r}_h(x)] = \frac{1}{nh} f(x) s^2(x) \|K\|_2^2 + \frac{h^4}{4} [r''(x) \mu_2(K)]^2 + o(h^4) + o((nh)^{-1})$$

III. PEMBAHASAN

1. Uji Koefisien Variansi Konstan dalam Regresi Nonparametrik

Para ahli mempertimbangkan pengujian berbagai hipotesis yang berkaitan dengan rata-rata dan fungsi varians dalam model regresi nonparametrik (1.2), Dette dan Munk (2003). Hipotesis ini meliputi asumsi semi parametrik yang berkaitan dengan rata-rata dan fungsi varians, tapi banyak kelemahan upaya telah dihabiskan dalam menginvestigasi hubungan antara rata-rata dan varians pada model regresi nonparametrik (1.2). Penelitian ini mengidentifikasi hipotesis (1.3) dari koefisien variansi konstan menggunakan estimasi dari jarak L_2 antara varians dan fungsi regresi kuadrat.

Untuk sebarang c positif kita definisikan statistik $T_n(c)$ sebagai

$$\begin{aligned} T_n(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j) \{c^2 Y_i^2 - (c^2 + 1) \hat{m}^2(X_i)\} w(X_i) \\ &\quad \times \{c^2 Y_j^2 - (c^2 + 1) \hat{m}^2(X_j)\} w(X_j) \end{aligned} \quad (3.1)$$

dimana w merupakan *bobot fungsi*, $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$, $K(\cdot)$ merupakan *kernel* dan h adalah *bandwidth* menuju ke 0 dengan meningkatkan ukuran sampel. Jika estimasi \hat{m} konsisten, maka ukuran sampel yang besar

$$\begin{aligned} E[T_n(c)] &\approx E[K_h(X_1 - X_2) \{c^2 \sigma^2(X_1) \varepsilon_1^2 - 2c^2 m(X_1) \sigma(X_1) \varepsilon_1 - m^2(X_1)\} \times \\ &\quad \{c^2 \sigma^2(X_2) \varepsilon_2^2 - 2c^2 m(X_2) \sigma(X_2) \varepsilon_2 - m^2(X_2)\}] \\ &\approx E[f(X_i) \{c^2 \sigma^2(X_i) - m^2(X_i)\}^2 w^2(X_i)] \\ &= E[\Delta_c^2(X_i) f(X_i) w^2(X_i)], \end{aligned} \quad (3.2)$$

dimana f merupakan densitas (kepadatan) dari X dan

$$\Delta_c(x) = m^2(x) - c^2 \sigma^2(x). \quad (3.3)$$

Untuk tujuan ini, perhatikan masalah kuadrat terkecil berikut:

$$\begin{aligned}\hat{c}^2 &= \arg \min_{c \in R > 0} \sum_{i=1}^n (m^2(X_i) - c^2 \sigma^2(X_i))^2 w(X_i) \\ &= \frac{\sum_{i=1}^n m^2(X_i) \sigma^2(X_i) w(X_i)}{\sum_{i=1}^n \sigma^4(X_i) w(X_i)}\end{aligned}\quad (3.4)$$

dan estimasi jumlah yang tidak diketahui pada ruas kanan. Residualnya didefinisikan

$$\hat{r}(X_i) = Y_i - \hat{m}(X_i), (i = 1, 2, \dots, n) \quad (3.5)$$

dan estimasi

$$\hat{c}^2 = \frac{(1/n) \sum_{i=1}^n \hat{m}^2(X_i) \hat{r}^2(X_i) w(X_i)}{(1/n) \sum_{i=1}^n (\hat{\sigma}^2(X_i))^2 w(X_i)}, \quad (3.6)$$

dimana estimasi variansi $\hat{\sigma}$ didefinisikan dalam (3.10). Estimasi \hat{c}^2 dapat ditulis

$$c_0^2 = \frac{E[m^2(X) \sigma^2(X) w(X)]}{E[\sigma^4(X) w(X)]}, \quad (3.7)$$

yang bersamaan dengan konstanta c^2 jika hipotesis nol (1.3) dipenuhi.

2. Asimtotik Normalitas

Dimulai dengan definisi dari estimasi, untuk fungsi regresi kita menggunakan estimasi linear lokal, Fan dan Gijbels, (1996).

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i) s_{n,1}(x)] Y_i}{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i) s_{n,1}(x)]} \quad (3.8)$$

dimana $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$, $K(\cdot)$ adalah *kernel*, selanjutnya h merupakan *bandwidth* dan

$$s_{n,l}(x) = \sum_{i=1}^n K_h(X_i - x) (x - X_i)^l \quad l = 1, 2. \quad (3.9)$$

Dengan cara yang sama, estimasi dari fungsi variansi diperoleh dengan menggantikan Y_i dengan kuadrat residu $\hat{r}^2(X_i)$ yang didefinisikan pada (3.5) dan diberikan dengan

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i) s_{n,1}(x)] \hat{r}^2(X_i)}{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i) s_{n,1}(x)]}. \quad (3.10)$$

Untuk lebih jelasnya pertama diasumsikan bahwa $\{X_i, Y_i\}_{i=1}^n$ merupakan sampel dari pengamatan *independent identically distribution (i.i.d)*. *Bandwidth* (luas daerah) yang sama diasumsikan untuk menghitung estimasi regresi dan fungsi variansi agar bisa dinyatakan dalam notasi yang sederhana. Selanjutnya diasumsikan bahwa asumsi itu memenuhi:

- (A1) fungsi densitas f terdiferensial dua kali secara kontinu pada himpunan kompak,
- (A2) fungsi regresi m terdiferensial empat kali secara kontinu pada himpunan kompak,
- (A3) fungsi variansi σ^2 positif dan terdiferensial dua kali secara kontinu pada himpunan kompak,
- (A4) fungsi bobot w terdiferensial dua kali secara kontinu dan memiliki pendukung kompak yang terdapat pada $\{x | f(x) > 0\}$,

(A5) kernel K berorde 2, dan memenuhi kondisi Lipschitz,

(A6) jika $n \rightarrow \infty$ maka bandwidth g dan h memenuhi

$$h \sim n^{-1/5}, g = o(h^2), ng \rightarrow \infty$$

(3.11)

(A7) fungsi $m_k(x) = E[\varepsilon^k | X = x]$ kontinu untuk $k = 3, 4$ dan terbatas untuk

$$E[\varepsilon_t^k | X_t = x] \leq C < \infty, k \leq 8$$

(3.12)

(A8) regresi dan fungsi variansi memenuhi

$$E[m(X)]^k < \infty, \text{ untuk } k = 2, 4 \text{ dan } E[\sigma^2(X)]^k < \infty \text{ untuk } k = 1, 2.$$

Teorema 3.1. Diasumsikan bahwa asumsi (A1)-(A8) dipenuhi, maka

$$\begin{aligned} \hat{c}^2 - E[\hat{c}^2] = & \frac{1}{n} \sum_{i=1}^n \left\{ \tau_1(m^2(X_i)\sigma^2(X_i)w(X_i)\varepsilon_i^2 - E[m^2(X)\sigma^2(X)w(X)]) + \right. \\ & 2\tau_1m(X_i)\sigma^3(X_i)w(X_i)\varepsilon_i - \tau_2(\sigma^4(X_i)w(X_i) - E[\sigma^4(X)w(X)]) - \\ & \left. 2\tau_2\sigma^4(X_i)w(X_i)\{\varepsilon_i^2 - 1\} \right\} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (3.13)$$

Teorema 3.2. Diasumsikan bahwa asumsi-asumsi pada (A1)-(A8) dipenuhi. Dibawah null hypothesis (1.3) diperoleh

$$n\sqrt{g}T_n(\hat{c}) = n\sqrt{g}T_n(c) + o_p(1) \xrightarrow{D} N(0, \mu_0^2),$$

dimana konstanta μ_0^2 didefinisikan sebagai

$$\mu_0^2 = 2E\{[-1 + 4c^2 + 4cm_3(X) + m_4(X)]^2m^8(X)f(X)w^4(X)\} \int K^2(u)du$$

Bukti

Jika estimasi

$$T_n(c) - T_n(\hat{c}) = o_p\left(\frac{1}{n\sqrt{g}}\right) \quad (3.20)$$

dapat dibentuk, maka dekomposisi

$$\begin{aligned} T_n(c) = & (c^2 + 1)^2T_{1n} - 2(c^2 + 1)\{2c^2T_{2n} - T_{3n}(c)\} + T_{4n}(c) - \\ & 4c^2\{T_{5n}(c) - c^2T_{6n}\}, \end{aligned} \quad (3.21)$$

dengan

$$\begin{aligned} T_{1n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \delta_n(X_j) w(X_j), \\ T_{2n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) m(X_j) \sigma(X_j) w(X_j) \varepsilon_j, \\ T_{6n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \sigma(X_i) w(X_i) \varepsilon_i m(X_j) \sigma(X_j) w(X_j) \varepsilon_j, \\ T_{3n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j), \\ T_{4n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \Delta_c(X_i, \varepsilon_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j), \\ T_{5n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \Delta_c(X_i, \varepsilon_i) w(X_i) m(X_j) \sigma(X_j) w(X_j) \varepsilon_j, \end{aligned}$$

dimana

$$\begin{aligned}\Delta_c(X_i, \varepsilon_i) &= m^2(X_i) - c^2 \sigma^2(X_i) \varepsilon_i^2 \\ \delta_n(X_i) &= \hat{m}^2(X_i) - m^2(X_i).\end{aligned}$$

Selanjutnya diperoleh,

$$\begin{aligned}T_n(\hat{c}) &= (\hat{c}^2 + 1)^2 T_{1n} - 4\hat{c}^2(\hat{c}^2 + 1)T_{2n} + 2(\hat{c}^2 + 1)T_{3n}(\hat{c}) + T_{4n}(\hat{c}) - \\ &\quad 4\hat{c}^2 T_{5n}(\hat{c}) + 4\hat{c}^4 T_{6n} \\ &= 2(\hat{c}^2 + 1)T_{3n}(\hat{c}) + T_{4n}(\hat{c}) - 4\hat{c}^2 T_{5n}(\hat{c}) + 4\hat{c}^4 T_{6n} + o_p\left(\frac{1}{n\sqrt{g}}\right),\end{aligned}\quad (3.22)$$

Sebagai syarat ketiga dalam (3.21) diperoleh

$$T_{3n}(\hat{c}) = T_{3n}(c_0) - (\hat{c}^2 - c_0^2)T_{3n}^{(a)},$$

dimana

$$\begin{aligned}T_{3n}^{(a)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \sigma^2(X_j) w(X_j) \varepsilon_j^2 \\ &= O_p(h^2) + O_p\left(\frac{1}{nh}\right)\end{aligned}$$

Masing-masing diperoleh dari Teorema 3.1

$$T_{3n}(\hat{c}) - T_{3n}(c_0) = o_p\left(\frac{1}{n\sqrt{g}}\right). \quad (3.23)$$

Estimasi-estimasi yang bersesuaian

$$T_{4n}(\hat{c}) - T_{4n}(c_0) \underset{H_0}{=} o_p\left(\frac{1}{n\sqrt{g}}\right), \quad (3.24)$$

$$\hat{c}^2 T_{5n}(\hat{c}) = c_0^2 T_{5n}(c_0) + o_p\left(\frac{1}{n\sqrt{g}}\right), \quad (3.25)$$

$$(\hat{c}^2)^2 T_{6n} = c_0^4 T_{6n} + O_p\left(\frac{h^2}{n\sqrt{g}}\right). \quad (3.26)$$

Mengingat

$$c_0^2 \underset{H_0}{=} c^2$$

dengan penggabungan (3.21)-(3.24) dengan Teorema 3.1 menghasilkan

$$T_n(\hat{c}) \underset{H_0}{=} T_{4n}(\hat{c}) - 4\hat{c}^2 T_{5n}(\hat{c}) + 4\hat{c}^4 T_{6n} + o_p\left(\frac{1}{n\sqrt{g}}\right) \underset{H_0}{=} T_n(c) + o_p\left(\frac{1}{n\sqrt{g}}\right) \quad (3.27)$$

Pembuktian ini merupakan identitas pertama dalam Teorema 3.2.

Teorema 3.3. Diasumsikan bahwa asumsi-asumsi pada (A1)-(A8) dipenuhi. Dibawah fixed alternative

$$\varphi = E[\Delta_{c_0}(X) \sigma^2(X) f(X) w^2(x)] > 0$$

diperoleh

$$\sqrt{n}(T_n(\hat{c}) - E[T_n(\hat{c})]) \xrightarrow{D} N(0, \omega_1^2),$$

dimana

$$E[T_n(\hat{c})] = E[\Delta_{c_0}^2(X)f(X)w^2(X)] + h^2(B(c_0) - 2\phi I) + o(h^2),$$

dan $B(c_0)$ merupakan syarat bias dari statistik $T_n(c_0)$. Variansi asimtotik ω_1^2 dituliskan sebagai

$$\omega_1^2 = \mu_1^2(c_0) + 4\phi^2 v^2 - 4\phi v^2(c_0),$$

dimana $\mu_1^2(c_0)$ didefinisikan

$$\begin{aligned} \mu_1^2(c) = & 4var(\Delta_c^2(X)f(X)w^2(X)) + 16E[\Delta_c^2(X)m^2(X)\sigma^2(X)f^2(X)w^4(X)] \\ & + 4c^4E[\Delta_c^2(X)\sigma^4(X)f^2(X)\{m_4(X) - 1\}w^4(X)] \\ & - 16c^2E[\Delta_c^2(X)m(X)\sigma^3(X)f^2(X)m_3(X)w^4(X)], \end{aligned}$$

v^2 disesuaikan dengan variansi asimtotik dari \hat{c}^2 dalam Teorema 3.1 dan

$$\begin{aligned} v^2(c_0) = & 2\tau_1E[\Delta_{c_0}(X)(m^2(X) - c_0^2\sigma^2(X)m_4(X))m^2(X)\sigma^2(X)f(X)w^3(X)] - \\ & 2\tau_1E[\Delta_{c_0}^2(X)f(X)w^2(X)E[m^2(X)\sigma^2(X)w(X)]] - \\ & 4c_0^2\tau_1E[\Delta_{c_0}(X)m(X)\sigma^5(X)f(X)m_3(X)w^3(X)] - \\ & 2\tau_2Cov(\Delta_{c_0}^2(X)f(X)w^2(X), \sigma^4(X)w(X)) + \\ & 4c_0^2\tau_2E[\Delta_{c_0}(X)\sigma^6(X)f(X)\{m_4(X) - 1\}w^3(X)] + \\ & 4\tau_1E[\Delta_{c_0}(X)m^3(X)\sigma^3(X)f(X)m_3(X)w^3(X)] + \\ & 8\tau_1E[\Delta_{c_0}(X)m^2(X)\sigma^4(X)f(X)w^3(X)] - \\ & 8\tau_2E[\Delta_{c_0}(X)m(X)\sigma^5(X)f(X)m_3(X)w^3(X)]. \end{aligned}$$

IV. SIMULASI

Untuk mempelajari sifat sampel berhingga dari uji baru ini telah dilakukan sebuah uji simulasi yang sederhana, bootstrap merupakan alat bantu umum (*general tool*) yang biasa digunakan untuk mencari pendekatan dalam distribusi statistik yang kita inginkan. Bootstrap menggantikan atau bahkan seringkali memperbaiki hasil yang diperoleh berdasarkan analisa asimtotik secara klasik, terutama untuk data sampel yang berukuran kecil sampai menengah.

Algoritma bootstrap untuk data independent dilakukan dengan membangkitkan bilangan random secara diskrit uniform dengan *replacement*, misalnya, $I \sim d_{unif}(N, n)$, dimana N adalah jumlah data yang akan kita bangkitkan secara bootstrap. Data pada indeks inilah yang akan menjadi data bootstrap.

Sebagai contoh telah diterapkan prosedur smooth bootstrap untuk mendapatkan nilai-nilai kritis. Untuk tujuan ini fungsi regresi dan fungsi variansi diestimasi dengan estimasi linear lokal yang masing-masing didefinisikan dalam (3.8) dan (3.10) dan perhatikan standar residu berikut

$$\eta_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}, i = 1, 2, \dots, n, \quad (4.1)$$

dimana $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$; $i = 1, 2, \dots, n$, dengan $m(x) = c\sigma(x)$, yang berdistribusi normal dengan mean 0 dan variansi 1, yaitu:

$$\hat{\varepsilon}_i = \frac{\eta_i - \bar{\eta}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\eta_i - \bar{\eta})^2}}, \quad i = 1, 2, \dots, n. \quad (4.2)$$

Kemudian error bootstrap didefinisikan sebagai

$$\varepsilon_i^* = \hat{\varepsilon}_i^* + vN_i, \quad (4.3)$$

dimana $\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*$ diambil secara acak dengan pengembalian dari distribusi empiris pada standar residual $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ dan N_1, \dots, N_n yaitu standar distribusi normal variabel random independen dari sampel $y_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ dan $v=v_n$ merupakan parameter smoothing yang konvergen menuju 0 (nol) dengan peningkatan ukuran sampel.

Pada langkah berikutnya data bootstrap *digenerate* dengan model

$$Y_i^* = \hat{c}\hat{\sigma}(X_i) + \hat{\sigma}(X_i)\varepsilon_i^*, \quad i = 1, 2, \dots, n, \quad (4.4)$$

dimana \hat{c} merupakan estimasi least square (3.6) yang diperoleh dari data yang sesuai dengan interval $[X_{(0.05n)}, X_{(0.95n)}]$ dari prediktor, dimana $X_{(1)} \leq \dots \leq X_{(n)}$ menandakan urutan statistik dari X_1, \dots, X_n . Uji statistik T_n^* dihitung dari data bootstrap $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$. Jika replikasi bootstrap B telah dilakukan, maka hipotesis nol (1.3) akan ditolak jika

$$T_n > T_n^{*(\lfloor B(1-\alpha) \rfloor)}, \quad (4.5)$$

dimana $T_n^{*(1)} < \dots < T_n^{*(B)}$ menunjukkan urutan statistik dari sampel bootstrap. Pada simulasi ini, untuk ukuran replikasi bootstrap dipilih $B=100$, sedangkan 1000 simulasi dilakukan untuk menjalankan perhitungan tingkat empiris ini. Ukuran sampel diberikan dengan $n = 50, 75, 100$ dan parameter smoothing dalam uji statistik dan prosedur bootstrap dipilih masing-masing dengan $h = n^{-1/2}$ dan $v = 0.1$. Bandwidth untuk estimasi fungsi regresi dan fungsi variansi dipilih secara terpisah dengan validasi error kuadrat terkecil.

Sebagai contoh pandang model

$$m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x), \quad (4.6)$$

dimana $c = 0.5, 1.0, 1.5$. Prediktor X_1, \dots, X_n memiliki *independent identically distributed* mengikuti distribusi seragam pada interval $[0, 1]$, sementara error $\varepsilon_1, \dots, \varepsilon_n$ mempunyai standar distribusi normal. Dalam langkah kedua dipelajari uji power dan pandang model

$$m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x + \sqrt{n}), \quad (4.7)$$

$$m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x + 2\sqrt{x}). \quad (4.8)$$

Dimana hasil simulasi disajikan sebagai berikut:

n		50				75				100			
	c/a	2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
(4.6)	0.5	0.832	0.832	0.834	0.834	0.847	0.849	0.849	0.849	0.844	0.845	0.846	0.846
	1.0	0.476	0.478	0.480	0.481	0.521	0.525	0.526	0.527	0.501	0.502	0.503	0.503
	1.5	0.268	0.268	0.268	0.268	0.271	0.271	0.271	0.271	0.279	0.279	0.280	0.281
(4.7)	0.5	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000
	1.0	0.910	0.911	0.913	0.918	0.946	0.947	0.947	0.947	0.949	0.952	0.954	0.957
	1.5	0.728	0.734	0.738	0.740	0.789	0.792	0.793	0.793	0.774	0.775	0.778	0.779
(4.8)	0.5	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.0	0.985	0.986	0.987	0.988	0.998	0.998	0.998	0.999	1.000	1.000	1.000	1.000
	1.5	0.937	0.937	0.938	0.941	0.962	0.962	0.964	0.967	0.970	0.974	0.975	0.976

V. KESIMPULAN

Fungsi kernel K yang umum dipakai adalah fungsi densitas dan biasanya dilengkapi dengan asumsi-asumsi tertentu, dimana fungsi kernel K kontinu, terbatas, simetris terhadap pusat interval yang terintegrasi ke satu, $\int k(u)du = 1$.

Dalam makalah ini telah dilakukan pengujian formal untuk hipotesis dari koefisien variansi konstan dalam regresi nonparametrik model (1.2), yaitu dengan melakukan sebuah uji simulasi kecil, yang mana hipotesis nol (1.3) akan ditolak jika persamaan (4.5) terpenuhi, sebaliknya hipotesis nol (1.3) akan diterima jika persamaan (4.5) tidak terpenuhi. Kernel yang dipakai dalam simulasi itu adalah kernel Gaussian, $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$. Data yang dipakai merupakan data yang digenerate, kemudian sampel diambil secara acak dengan pengembalian, dengan replikasi bootstrap $B=100$ dan 1000 simulasi dijalankan.

DAFTAR PUSTAKA

- D. Holger, W. and Gabriele. (2008), "Testing for a constant coefficient of variation in nonparametric regression", *Journal Statistic Nonparametric*.
- Draper, N. Smith H. (1992), *Analisis Regresi Terapan*, Gramedia Pustaka Utama, Jakarta.
- Efron, B., and Tibshirani. (1993), R. J., *An Introduction to the Bootstrap*, Chapman & Hall.
- G.K. Eagleson, and H.G. Müller (1997). Transformations for smooth regression models with multiplicative errors. *J. R. Statist. Soc. B.*, 59(1), 173–189.
- Green, W. H. (2003), *Econometric Analysis 5th edition*, Prentice Hall.
- Gujarati, D. N. (2003), *Basic Econometrics*, McGraw-Hill.

- H. Dette, and A. Munk (2003). Some methodological aspects of validation of models in nonparametric regression. *Statistica Neerlandica*, 57(2), 207–244.
- Hardle, W. (1994), *Applied Nonparametric Regression*, Cambridge University Press, New York.
- Hardle, W. (1990), *Smoothing Techniques*, Springer-Verlag, New York.
- Montgomery, D. C. And Peck, A. E. (2006), *Introduction To Linear Regression Analysis*, John Wiley & Sons, New York.