PERSAMAAN MODEL CAMPURAN HENDERSON PADA MODEL SMALL AREA SEMIPARAMETRIK DENGAN SAMPLING INFORMATIF

Angela Nina R. C.[1,2], Sri Haryatmi[1], Danardono[1]

[1] **FMIPA, Universitas Gadjah Mada**
[2] **FKIP, Universitas Pancasakti Tegal**
angela.nina.r@ugm.ac.id

Abstrak: Estimasi small area dapat dilakukan dengan pendekatan berdasarkan model. Salah satu model small area adalah model regresi eror bersarang yang digunakan pada data terkluster . Dalam aplikasi, fungsi rata-rata model seringkali tidak terspesifikasi sebelumnya sehingga diperlukan pendekatan non-parametrik. Di sisi lain, data sampel yang digunakan seringkali dihasilkan dari desain sampling informatif. Efek keinformatifan sampel ini harus dipertimbangkan dalam proses inferensi. Tulisan ini memaparkan konsep penurunan estimator bagi parameter model dan prediktor bagi variabel random pada model regresi eror bersarang dimana efek keinformatifan sampel diperhitungkan dalam model. Model diperluas dengan menambahkan sebagai kovariat, fungsi dari peluang inklusi. Regresi spline terpenalti digunakan untuk mendekati fungsi rata-rata model dan fungsi peluang inklusi. Estimator dan prediktor diturunkan menggunakan persamaan model campuran Henderson. Dalam tulisan ini, dianalisa pula sifat estimator dan prediktor yang diperoleh.

Kata kunci: Estimasi *Small Area*, Pendekatan Berdasarkan Model, Sampling Informatif, Spline Terpenalti, Persamaan Model Campuran Henderson

PENDAHULUAN

Survei sampel telah digunakan secara luas untuk menghasilkan estimasi bagi kuantitas populasi maupun subpopulasi (domain) yang diamati. Domain dapat berupa wilayah geografis, kelompok sosio-demografi, ataupun subpopulasi lainnya. Domain, dimana banyaknya unit sampel domain tidak cukup besar untuk menghasilkan estimasi langsung dengan ketelitian yang memadai dalam statistik disebut sebagai s*mall area*.

Survei untuk tujuan analitis memerlukan pemodelan hubungan diantara variabel yang diamati. Analisis ini memerlukan antara lain asumsi model serta estimasi bagi parameter model. Salah satu model yang banyak digunakan dalam estimasi *small area* adalah model regresi eror bersarang. Pendekatan dengan model regresi eror bersarang mengasumsikan bahwa populasi tersusun dari unit-unit yang dikelompokkan dalam kluster (sebagai *small area*). Dalam setiap kluster yang terpilih sebagai sampel, dipilih unit-unit sampel dan observasi diasumsikan memiliki struktur korelasi yang terdefinisi oleh

$$y_{ij} = \mathbf{x}_{ij}^{*t} b + v_i + e_{ij}, \;\; i = 1, \cdots, m; \;\; j = 1, \cdots, N_i, \tag{1}$$

dengan $\mathbf{x}^*_{ij} = (1, x_{ij1}, ..., x_{ijp})$ adalah vektor variabel kovariat berordo 1x(p+1), b adalah koefisien regresi. Variabel random v_i menyatakan efek area, untuk menghitung variasi dalam y yang tidak dijelaskan oleh variabel kovariat dan e_{ij} menyatakan eror unit. m adalah banyaknya *small area* dan N_i adalah banyaknya unit populasi dalam *small area* i. Pada model (1) seringkali diasumsikan $v_i \sim_{iid} N(0,\sigma^2_v)$, $e_{ij} \sim_{iid} N(0,\sigma^2_e)$, v_i dan e_{ij} saling bebas . (Rao, 2003).

Estimator bagi efek tetap dan efek random dalam model campuran (1) dapat diperoleh secara simultan dengan menyelesaikan model campuran Henderson. Henderson, dalam Robinson (1991) dan Searle, dkk. (1992), mengasumsikan bahwa v dan e berdistribusi normal dan memaksimalkan densitas gabungan y dan v terhadap b dan v. Persamaan model campuran Henderson diperoleh dengan mendiferensialkan densitas gabungan y dan v terhadap b dan v kemudian menyamakannya dengan nol. Metode Henderson ini memberikan estimator tak bias linier dan terbaik bagi b dan prediktor tak bias linier terbaik bagi v.

Bentuk fungsional hubungan antara variabel yang diamati dengan variabel kovariat pada model terkadang tidak dapat dispesifikasi sebelumnya dalam praktek, sehingga diperlukan pendekatan nonparametrik bagi fungsi tersebut. Di sisi lain, data yang digunakan dalam aplikasi seringkali dihasilkan dari desain sampling kompleks dimana terdapat korelasi antara peluang inklusi dengan variabel hasil yang diamati. Desain sampling yang menghasilkan korelasi antara peluang inklusi dan nilai-nilai observasi disebut sebagai desain sampling informatif. Kegagalan dalam memperhitungkan atau pengabaian efek keinformatifan sampel dalam inferensi dapat membawa pada hasil inferensi yang buruk dengan biasnya estimasi.

Dalam pendekatan berdasarkan model, pemberian bobot untuk mengurangi efek keinformatifan sampling masih menjadi kontroversi. Alternatif lain adalah dengan penggunaan prinsip sederhana yakni, memasukkan semua variabel independen yang terkait dengan desain sampling ke dalam model. Namun, prinsip ini tidak langsung membawa pada cara yang tepat bagaimana memodelkan variabel observasi dengan variabel desain karena luasnya ragam struktur data (Reiter, dkk. , 2005). Lebih lanjut, Nathan (2005) menyatakan bahwa terdapat dua masalah utama dalam memasukkan semua variabel desain ke dalam model. Masalah tersebut adalah adanya kemungkinan tidak diketahuinya variabel mana yang tepat yang harus digunakan dan apabila diketahui, nilainya mungkin tidak tersedia; manakala variabel desain dimasukkan dalam model secara benar, estimasi yang dihasilkan dapat kurang bernilai karena variabel yang ditambahkan bukanlah materi subyek yang diamati.

Smith (1976, 1994), Kish (1995) dan Little (2004) menyatakan bahwa peluang inklusi mempunyai peran penting dalam inferensi berdasarkan desain maupun model dari sampel survei dengan peluang yang tidak sama. Dari pendapat tersebut, peluang inklusi dapat menjadi aspek relevan desain sampling yang dapat dipertimbangkan untuk digabungkan ke dalam model.

Tulisan ini memaparkan prosedur menentukan estimasi bagi efek tetap dan efek random model regresi eror bersarang yang diperluas menggunakan persamaan model campuran Henderson. Model regresi eror bersarang diperluas dengan cara menambahkan fungsi peluang inklusi sebagai kovariat. Dalam tulisan ini, fungsi peluang inklusi dan fungsi rata-rata dalam model dipandang sebagai fungsi yang tidak terspesifikasi sebelumnya dan diasumsikan didekati cukup baik oleh fungsi *spline* terpenalti (*p-spline*). Sifat estimator yang dihasilkan juga hendak diamati. Diharapkan tulisan ini memberi kontribusi konsep penggunaan model campuran linier Henderson pada model *small area* dimana efek sampling informatif dipertimbangkan dan pendekatan non-parametrik *p-spline* diterapkan.

HASIL DAN PEMBAHASAN

Dipandang (1) dengan $\mathbf{x}^*_{ij} = (1, x_{ij})$. Diasumsikan model populasi ini berlaku pada setiap area dalam populasi dan data kovariat unit $x_{ij} = (x_{ij1}, \cdot\cdot\cdot, x_{ijp})$ tersedia untuk setiap elemen populasi j dalam setiap *small area* i. Unit-unit sampel sebesar n_i diambil pada setiap kluster menggunakan skema sampling informatif dengan peluang inklusi $\pi_{j|i}$ yang diketahui. Model bagi sampel dibentuk dengan mempertimbangan efek keinformatifan sampling. Efek informatif sampling diperhitungkan dengan memasukkan fungsi $g(\pi_{j|i})$ sebagai kovariat ke dalam model, yakni

$$y_{ij} = \mathbf{x}^*_{ij} b + g(\pi_{j|i})\alpha + v_i + e_{ij},\ \ i = 1, \cdots, m;\ j = 1, \cdots, n_i \qquad (2)$$

dengan α koefisien yang tidak diketahui. Diasumsikan pula, $\mathbf{x}^*_{ij} b$ dan $g(\pi_{j|i})$ α dalam model merupakan fungsi mulus yang tidak diketahui dan dapat didekati dengan fungsi *p-spline*. Dalam bentuk model campuran *p-spline*, (2) dapat dinyatakan sebagai

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ij}^k + \sum_{k=1}^{K_1} t_k (x_{ij} - q_k)_+^p + \sum_{k=1}^{s} \delta_k \pi_{ij}^k + \sum_{k=1}^{K_2} r_k (\pi_{ij} - Q_k)_+^s + v_i + e_{ij} \qquad (3)$$

dengan p dan s berturut-turut adalah order *p-spline* untuk $\mathbf{x}^*_{ij} b$ dan $g(\pi_{j|i})$ α; β_k dan δ_k berturut-turut adalah vektor koefisien bagian parametrik bagi $\mathbf{x}^*_{ij} b$ dan $g(\pi_{j|i})$ α;

t_k dan r_k berturut-turut adalah vektor koefisien bagian *spline* bagi $\mathbf{x}^*_{ij} b$ dan $g(\pi_{j|i})$ α;

$q_1 < q_2 < \cdots < q_{K_1}$ dan $Q_1 < Q_2 < \cdots < Q_{K_2}$ berturut-turut adalah *knot* tertentu dalam x dan dalam $\pi_{j|i}$; $\beta_0, \beta_k, \delta_k, t_k$ dan r_k adalah koefisien yang diestimasi, $(u)_+^p = u$ jika t ≥ 0 dan 0 jika sebaliknya. *Knot* bagi X ditentukan dengan $K_k = kuantil\ sampel\ x_i\ yang\ tunggal\ ke\ (\frac{k+1}{K+2}),\ 1 \le k \le K,$ dengan K = min(n/4, 35). (Wand ,2003 dan Ruppert, Wand dan Carroll,2003)

Bentuk matrik bagi (3), dengan menggunakan seluruh data sampel, dinyatakan sebagai

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Ct} + \mathbf{\Pi\delta} + \mathbf{Dr} + \mathbf{Zv} + \mathbf{e}\ , \qquad (4)$$

dengan,

$\mathbf{y} = \underset{1\le i\le m}{col}(\mathbf{y_i})$, $\mathbf{y_i} = [y_{i1} \cdots y_{in_i}]^t$;

$\mathbf{X} = \underset{1\le i\le m}{col}(\mathbf{x_i})$, $\mathbf{x_i}$ matrik berordo n_ix(1+p) dengan baris $\mathbf{x_{ij}} = [1 \quad x_{ij} \quad \cdots \quad x_{ij}^p]^t$, j=1,…,$n_i$;

$\boldsymbol{\beta} = [\beta_0\ \beta_1 \cdots \beta_p]^t$; $\mathbf{t} = [t_1 \cdots t_{K_1}]^t$; $\boldsymbol{\delta} = [\delta_1 \cdots \delta_s]^t$; $\mathbf{r} = [r_1 \cdots r_{K_1}]^t$;

$\mathbf{C} = \underset{1\le i\le m}{col}(\mathbf{C_i})$, $\mathbf{C_i}$ matrik n_ixK_1 dengan baris $C_{ij} = [(x_{ij} - q_1)_+^p \cdots (x_{ij} - q_{K_1})_+^p]$, j = 1, …, n_i;

$\mathbf{\Pi} = \underset{1\le i\le m}{col}(\pi_i)$, π_i matrik n_i x s dengan baris $\pi_{ij} = [\pi_{ij} \cdots \pi_{ij}^s]$, j = 1, …, n_i ;

$\mathbf{D} = \underset{1\le i\le m}{col}(D_i)$, D_i matrik n_ixK_2 dengan baris $D_{ij} = [(\pi_{ij} - Q_1)_+^s \cdots (\pi_{ij} - Q_{K_2})_+^s]$;

$\mathbf{Z} = \underset{1\le i\le m}{diag}(\mathbf{W}_i)$, dengan $\mathbf{W}_i = \mathbf{1}_{n_i}$ vektor n_i x1 dengan elemen 1 ; $\mathbf{v} = [v_1 \cdots v_m]^t$;

$\mathbf{e} = \underset{1\le i\le m}{col}(\mathbf{e_i})$ dengan $e_i = [e_{i1} \cdots e_{in_i}]^t$.

Diasumsikan,

$$\mathbf{t} \overset{iid}{\sim} N(0,\ \mathbf{G}_1),\ \text{dengan } \mathbf{G}_1 = \sigma_t^2\,\mathbf{I}_{K_1}\ ;\ \mathbf{r} \overset{iid}{\sim} N(0,\ \mathbf{G}_2),\ \text{dengan } \mathbf{G}_2 = \sigma_r^2\,\mathbf{I}_{K_2}\,,$$

$$\mathbf{v} \overset{iid}{\sim} N(0,\ \mathbf{G}_3),\ \text{dengan } \mathbf{G}_3 = \sigma_v^2\,\mathbf{I}_m\ ;\ \mathbf{e} \overset{iid}{\sim} N(0,\ \mathbf{G}_4),\ \text{dengan } \mathbf{G}_4 = \sigma_e^2\,\mathbf{I}_n\,,\ n = \sum_{i=1}^{m} n_i\,,$$

dan **t**, **r**, **v** saling independen.

Model *p-spline* dapat dipandang sebagai model campuran dengan mengasumsikan koefisien **t** dan **r** sebagai variabel random. Dalam hal ini, penghalusan *p-spline* terkait dengan prediksi optimal dalam kerangka model campuran. Hal ini memungkinkan untuk menggunakan metodologi model campuran untuk regresi *p-spline*.

Persamaan model campuran Henderson bagi (4) dibentuk dengan memaksimalkan densitas gabungan bagi **y**, **t**, **r**, dan **v** dengan mengasumsikan $(\mathbf{y}\,|\ \mathbf{t},\ \mathbf{r},\ \mathbf{v}) \sim N(\mathbf{X\beta} + \mathbf{Ct} + \boldsymbol{\pi\delta} + \mathbf{Dr} + \mathbf{Zv},\ \mathbf{G}_4)$. Densitas gabungan bagi **y**, **t**, **r**, dan **v** dinyatakan sebagai:

$$f(\mathbf{y}, \mathbf{t}, \mathbf{r}, \mathbf{v}) = f(\mathbf{y}|\ \mathbf{t}, \mathbf{r}, \mathbf{v}).\ f(\mathbf{t}, \mathbf{r}, \mathbf{v})$$

$$= (2\pi)^{-(n+k_1+k_2+1)/2}\,|\,\mathbf{G}_1\,|^{-1/2}|\,\mathbf{G}_2\,|^{-1/2}|\,\mathbf{G}_3\,|^{-1/2}|\,\mathbf{G}_4\,|^{-1/2}\,.$$

$$exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Ct} - \boldsymbol{\pi\delta} - \mathbf{Dr} - \mathbf{Zv})'\}\mathbf{G}_4^{-1}.$$

$$(\mathbf{y} - \mathbf{X\beta} - \mathbf{Ct} - \boldsymbol{\pi\delta} - \mathbf{Dr} - \mathbf{Zv}) + \mathbf{t}'\mathbf{G}_1^{-1}\mathbf{t} + \mathbf{r}'\mathbf{G}_2^{-1}\mathbf{r} + \mathbf{v}'\mathbf{G}_1^{-1}\mathbf{v}\}$$

Pendiferensialan $\log f(\mathbf{y}, \mathbf{t}, \mathbf{r}, \mathbf{v})$ berturut-turut terhadap **β**, **δ**, **t**, **r**, dan **v**, kemudian menyamakan dengan nol masing-masing turunan parsial tersebut membawa pada persamaan

$$\begin{bmatrix} \mathbf{X}'\mathbf{G}_4^{-1}\mathbf{X} & \mathbf{X}'\mathbf{G}_4^{-1}\boldsymbol{\pi} & \mathbf{X}'\mathbf{G}_4^{-1}\mathbf{C} & \mathbf{X}'\mathbf{G}_4^{-1}\mathbf{D} & \mathbf{X}'\mathbf{G}_4^{-1}\mathbf{Z} \\ \boldsymbol{\pi}'\mathbf{G}_4^{-1}\mathbf{X} & \boldsymbol{\pi}'\mathbf{G}_4^{-1}\boldsymbol{\pi} & \boldsymbol{\pi}'\mathbf{G}_4^{-1}\mathbf{C} & \boldsymbol{\pi}'\mathbf{G}_4^{-1}\mathbf{D} & \boldsymbol{\pi}'\mathbf{G}_4^{-1}\mathbf{Z} \\ \mathbf{C}'\mathbf{G}_4^{-1}\mathbf{X} & \mathbf{C}'\mathbf{G}_4^{-1}\boldsymbol{\pi} & (\mathbf{C}'\mathbf{G}_4^{-1}\mathbf{C} + \mathbf{G}_1^{-1}) & \mathbf{C}'\mathbf{G}_4^{-1}\mathbf{D} & \mathbf{C}'\mathbf{G}_4^{-1}\mathbf{Z} \\ \mathbf{D}'\mathbf{G}_4^{-1}\mathbf{X} & \mathbf{D}'\mathbf{G}_4^{-1}\boldsymbol{\pi} & \mathbf{D}'\mathbf{G}_4^{-1}\mathbf{C} & (\mathbf{D}'\mathbf{G}_4^{-1}\mathbf{D} + \mathbf{G}_2^{-1}) & \mathbf{D}'\mathbf{G}_4^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{G}_4^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{G}_4^{-1}\boldsymbol{\pi} & \mathbf{Z}'\mathbf{G}_4^{-1}\mathbf{C} & \mathbf{Z}'\mathbf{G}_4^{-1}\mathbf{D} & (\mathbf{D}'\mathbf{G}_4^{-1}\mathbf{C} + \mathbf{G}_2^{-1}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{t} \\ \mathbf{r} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{G}_4^{-1}\mathbf{y} \\ \boldsymbol{\pi}'\mathbf{G}_4^{-1}\mathbf{y} \\ \mathbf{C}'\mathbf{G}_4^{-1}\mathbf{y} \\ \mathbf{D}'\mathbf{G}_4^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{G}_4^{-1}\mathbf{y} \end{bmatrix} \quad (5)$$

yang dinamakan persamaan model campuran Henderson bagi (4). Misal invers matrik partisi ruas kiri (5) dinyatakan sebagai

$$\mathbf{J}=\begin{bmatrix}\mathbf{J}_{11} & \mathbf{J}_{12} & \mathbf{J}_{13} & \mathbf{J}_{14} & \mathbf{J}_{15}\\ \mathbf{J}_{21} & \mathbf{J}_{22} & \mathbf{J}_{23} & \mathbf{J}_{24} & \mathbf{J}_{25}\\ \mathbf{J}_{31} & \mathbf{J}_{32} & \mathbf{J}_{33} & \mathbf{J}_{34} & \mathbf{J}_{35}\\ \mathbf{J}_{41} & \mathbf{J}_{42} & \mathbf{J}_{43} & \mathbf{J}_{44} & \mathbf{J}_{45}\\ \mathbf{J}_{51} & \mathbf{J}_{52} & \mathbf{J}_{53} & \mathbf{J}_{54} & \mathbf{J}_{55}\end{bmatrix}$$

Estimator bagi $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ dan prediktor bagi $\mathbf{t}$, $\mathbf{r}$, dan $\mathbf{v}$ diperoleh dengan menyelesaikan (5), yakni

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\delta}}\\ \hat{\mathbf{t}}\\ \hat{\mathbf{r}}\\ \hat{\mathbf{v}}\end{bmatrix}=\begin{bmatrix}\mathbf{J}_{11} & \mathbf{J}_{12} & \mathbf{J}_{13} & \mathbf{J}_{14} & \mathbf{J}_{15}\\ \mathbf{J}_{21} & \mathbf{J}_{22} & \mathbf{J}_{23} & \mathbf{J}_{24} & \mathbf{J}_{25}\\ \mathbf{J}_{31} & \mathbf{J}_{32} & \mathbf{J}_{33} & \mathbf{J}_{34} & \mathbf{J}_{35}\\ \mathbf{J}_{41} & \mathbf{J}_{42} & \mathbf{J}_{43} & \mathbf{J}_{44} & \mathbf{J}_{45}\\ \mathbf{J}_{51} & \mathbf{J}_{52} & \mathbf{J}_{53} & \mathbf{J}_{54} & \mathbf{J}_{55}\end{bmatrix}\begin{bmatrix}\mathbf{X}'\mathbf{G}_4^{-1}\\ \boldsymbol{\pi}'\mathbf{G}_4^{-1}\\ \mathbf{C}'\mathbf{G}_4^{-1}\\ \mathbf{D}'\mathbf{G}_4^{-1}\\ \mathbf{Z}'\mathbf{G}_4^{-1}\end{bmatrix}\mathbf{y}=\begin{bmatrix}\mathbf{U}_1'\\ \mathbf{U}_2'\\ \mathbf{U}_3'\\ \mathbf{U}_4'\\ \mathbf{U}_5'\end{bmatrix}\mathbf{y} \tag{6}$$

Persamaan model campuran (5) menyajikan prosedur penghitungan bagi $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\mathbf{t}}, \hat{\mathbf{r}}, \hat{\mathbf{v}}$. Selain itu, elemen-elemen dalam persamaan model campuran dapat digunakan untuk pengaturan prosedur iteratif bagi estimasi komponen variansi.

Selanjutnya, menggunakan (5) dapat dinyatakan,

$$\mathbf{G}_4'\left[\mathbf{X}\quad \boldsymbol{\pi}\quad \mathbf{C}\quad \mathbf{D}\quad \mathbf{Z}\right]'\left[\mathbf{X}\quad \boldsymbol{\pi}\quad \mathbf{C}\quad \mathbf{D}\quad \mathbf{Z}\right]=\mathbf{I}-\mathbf{J}\,diag\left[\mathbf{0}\quad \mathbf{0}\quad \mathbf{G}_1^{-1}\quad \mathbf{G}_2^{-1}\quad \mathbf{G}_3^{-1}\right] \tag{7}$$

Sehingga,

$$\begin{bmatrix}\mathbf{U}_1'\\ \mathbf{U}_2'\\ \mathbf{U}_3'\\ \mathbf{U}_4'\\ \mathbf{U}_5'\end{bmatrix}\left[\mathbf{X}\quad \boldsymbol{\pi}\quad \mathbf{C}\quad \mathbf{D}\quad \mathbf{Z}\right]=\mathbf{I}-\mathbf{J}.diag\left[\mathbf{0}\quad \mathbf{0}\quad \mathbf{G}_1^{-1}\quad \mathbf{G}_2^{-1}\quad \mathbf{G}_3^{-1}\right] \tag{8}$$

Menggunakan (7) dan (8) dapat diperoleh, $E(\hat{\boldsymbol{\beta}})=E(\mathbf{U}_1'\mathbf{y})=\mathbf{U}_1'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\pi}\boldsymbol{\delta})=\boldsymbol{\beta}$, $E(\hat{\boldsymbol{\delta}})=E(\mathbf{U}_2'\mathbf{y})=\mathbf{U}_2'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\pi}\boldsymbol{\delta})=\boldsymbol{\delta}$ yang menunjukkan bahwa estimator $\hat{\boldsymbol{\beta}}$ dan $\hat{\boldsymbol{\delta}}$ tak bias. Diperoleh pula, $E(\hat{\mathbf{t}}-\mathbf{t})=E(\mathbf{U}_3'\mathbf{y}-\mathbf{t})=\mathbf{0}$, $E(\hat{\mathbf{r}}-\mathbf{r})=E(\mathbf{U}_4'\mathbf{y}-\mathbf{r})=\mathbf{0}$, dan $E(\hat{\mathbf{v}}-\mathbf{v})=E(\mathbf{U}_5'\mathbf{y}-\mathbf{v})=\mathbf{0}$, yang menunjukkan bahwa prediktor $\hat{\mathbf{t}}, \hat{\mathbf{r}}, \hat{\mathbf{v}}$ tak bias.

Sifat terbaik prediktor $\hat{\mathbf{t}}$ dapat ditunjukkan dengan memandang $\mathbf{y}\,|\,\mathbf{r},\mathbf{v} \sim N(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\pi}\boldsymbol{\delta}+\mathbf{D}\mathbf{r}+\mathbf{C}\mathbf{t}+\mathbf{Z}\mathbf{v},\, \mathbf{C}'\mathbf{G}_1\mathbf{C}+\mathbf{G}_4)$, sehingga

$$\begin{aligned}E(\mathbf{t}\,|\,\mathbf{y}) &= E(\mathbf{t}) + \operatorname{cov}(\mathbf{t},\mathbf{y})\operatorname{var}^{-1}(\mathbf{y})(\mathbf{y}-E(\mathbf{y}))\\ &= \mathbf{G}_1\mathbf{C}'(\mathbf{C}'\mathbf{G}_1\mathbf{C}+\mathbf{G}_4)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\pi}\boldsymbol{\delta}-\mathbf{D}\mathbf{r}-\mathbf{Z}\mathbf{v})\end{aligned}$$

Menggunakan baris ketiga dari (5) dapat dinyatakan

$$\begin{aligned}\hat{\mathbf{t}} &= \left(\mathbf{C}'\mathbf{G}_4^{-1}\mathbf{C}+\mathbf{G}_1^{-1}\right)^{-1}\mathbf{C}'\mathbf{G}_4^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\boldsymbol{\pi}\hat{\boldsymbol{\delta}}-\mathbf{D}\hat{\mathbf{r}}-\mathbf{Z}\hat{\mathbf{v}}\right)\\ &= \left(\mathbf{C}'\mathbf{G}_4^{-1}\mathbf{C}+\mathbf{G}_1^{-1}\right)^{-1}\mathbf{C}'\mathbf{G}_4^{-1}\left(\mathbf{C}'\mathbf{G}_1^{-1}\mathbf{C}+\mathbf{G}_4\right)\left(\mathbf{C}'\mathbf{G}_1^{-1}\mathbf{C}+\mathbf{G}_4\right)^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\boldsymbol{\pi}\hat{\boldsymbol{\delta}}-\mathbf{D}\hat{\mathbf{r}}-\mathbf{Z}\hat{\mathbf{v}}\right)\\ &= \mathbf{G}_1\mathbf{C}'\left(\mathbf{C}'\mathbf{G}_1^{-1}\mathbf{C}+\mathbf{G}_4\right)^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\boldsymbol{\pi}\hat{\boldsymbol{\delta}}-\mathbf{D}\hat{\mathbf{r}}-\mathbf{Z}\hat{\mathbf{v}}\right)\end{aligned}$$

$\hat{\mathbf{t}} = E(\mathbf{t}\,|\,\mathbf{y})$ dengan $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{\delta}=\hat{\boldsymbol{\delta}}, \mathbf{r}=\hat{\mathbf{r}}, \mathbf{v}=\hat{\mathbf{v}}$, sehingga $\hat{\mathbf{t}}$ merupakan prediktor terbaik bagi **t**. Dengan cara yang similar, dapat pula dibuktikan bahwa $\hat{\mathbf{r}}$ dan $\hat{\mathbf{v}}$ juga merupakan prediktor terbaik bagi **r** dan **v**.

Komponen variansi $\mathbf{G}_1=\sigma_t^2\,\mathbf{I}_{K_1}$, $\mathbf{G}_2=\sigma_r^2\,\mathbf{I}_{K_2}$, $\mathbf{G}_3=\sigma_v^2\,\mathbf{I}_m$ dan $\mathbf{G}_4=\sigma_e^2\,\mathbf{I}_n$ seringkali tidak diketahui sehingga perlu diestimasi. Dalam tulisan ini, komponen variansi diestimasi secara iteratif menggunakan persamaan titik tetap. Mengingat $\operatorname{var}(\mathbf{y})=\mathbf{C}\mathbf{G}_1\mathbf{C}'+\mathbf{D}\mathbf{G}_2\mathbf{D}'+\mathbf{Z}\mathbf{G}_3\mathbf{Z}'+\mathbf{G}_4$, $\mathbf{J}\mathbf{J}^{-1}=\mathbf{I}$ serta dengan menggunakan (6) dapat diperoleh $\operatorname{var}(\hat{\boldsymbol{\beta}})=\mathbf{J}'_{11}$, $\operatorname{var}(\hat{\boldsymbol{\delta}})=\mathbf{J}'_{22}$, $\operatorname{var}(\hat{\mathbf{t}})=\mathbf{G}_1-\mathbf{J}_{33}$, $\operatorname{var}(\hat{\mathbf{r}})=\mathbf{G}_2-\mathbf{J}_{44}$ dan $\operatorname{var}(\hat{\mathbf{v}})=\mathbf{G}_3-\mathbf{J}_{55}$. Sehingga,

$$\begin{aligned}&E(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')=\mathbf{J}'_{11}+\boldsymbol{\beta}\boldsymbol{\beta}',\quad E(\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}')=\mathbf{J}'_{22}+\boldsymbol{\delta}\boldsymbol{\delta}',\quad E(\hat{\mathbf{t}}\hat{\mathbf{t}}')=\mathbf{G}_1-\mathbf{J}_{33},\\ &E(\hat{\mathbf{r}}\hat{\mathbf{r}}')=\mathbf{G}_2-\mathbf{J}_{44},\quad E(\hat{\mathbf{v}}\hat{\mathbf{v}}')=\mathbf{G}_3-\mathbf{J}_{55}\end{aligned}\tag{9}$$

Kemudian, oleh karena $\hat{\mathbf{e}}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\boldsymbol{\pi}\hat{\boldsymbol{\delta}}-\mathbf{C}\hat{\mathbf{t}}-\mathbf{D}\hat{\mathbf{r}}-\mathbf{Z}\hat{\mathbf{v}}=(\mathbf{I}_n-\sigma_e^{-2}\mathbf{M}\mathbf{J}\mathbf{M}')\mathbf{y}$ dengan $\mathbf{M}=\begin{bmatrix}\mathbf{X} & \boldsymbol{\pi} & \mathbf{C} & \mathbf{D} & \mathbf{Z}\end{bmatrix}$, dapat diperoleh

$$E(\hat{\mathbf{e}}\hat{\mathbf{e}}')=E(\hat{\mathbf{e}}\mathbf{e}')=(\sigma_e^2\mathbf{I}_n-\mathbf{M}\mathbf{J}\mathbf{M}')\ .\tag{10}$$

Dari (9) dapat diperoleh

$$E\left(\sum_{k=1}^{K_1}\hat{t}_k^2\right)=tr(E(\hat{\mathbf{t}}\hat{\mathbf{t}}'))=\sigma_t^2(K_1-d_1)\tag{11}$$

$$E\left(\sum_{k=1}^{K_2}\hat{r}_k^2\right)=tr(E(\hat{\mathbf{r}}\hat{\mathbf{r}}'))=\sigma_r^2(K_2-d_2)\tag{12}$$

$$E\left(\sum_{k=1}^{m}\hat{v}_k^2\right)=tr(E(\hat{\mathbf{v}}\hat{\mathbf{v}}'))=\sigma_v^2(\mathrm{m}-\mathrm{d}_3) \quad (13)$$

dengan $d_1=\sigma_t^{-2}tr(\mathbf{J}_{33})$, $d_2=\sigma_r^{-2}tr(\mathbf{J}_{44})$, $d_3=\sigma_v^{-2}tr(\mathbf{J}_{55})$, dan dari (10) diperoleh

$$E\left(\sum_{k=1}^{m}\hat{e}_k^2\right)=tr(E(\hat{\mathbf{e}}\hat{\mathbf{e}}'))=\sigma_e^2(n-\mathrm{d}_4)\ , \quad (14)$$

dengan $d_4=\sigma_e^{-2}tr(\mathbf{MJM}')$. Menggunakan (11), (12), (13), dan (14), estimasi bagi $\sigma_t^2,\sigma_r^2,\sigma_v^2,\sigma_e^2$ dapat diperoleh secara iteratif dari persamaan titik tetap,

$$\sigma_t^2=\sum_{k=1}^{K_1}\frac{\hat{t}_k^2}{K_1-d_1} \quad (15)$$

$$\sigma_r^2=\sum_{k=1}^{K_2}\frac{\hat{r}_k^2}{K_2-d_2} \quad (16)$$

$$\sigma_v^2=\sum_{k=1}^{K_1}\frac{\hat{v}_k^2}{m-d_3} \quad (17)$$

$$\sigma_e^2=\sum_{k=1}^{K_1}\frac{\hat{e}_k^2}{n-d_4} \quad (18)$$

Proses iteratif dilakukan sebagai berikut: (i) Menentukan nilai awal bagi $\boldsymbol{\theta}^{(0)}=(\sigma_t^2,\sigma_r^2,\sigma_v^2,\sigma_e^2)'$; (ii) menyelesaikan (6) dengan $\boldsymbol{\theta}^{(0)}$ untuk memperoleh $\hat{\mathbf{t}}$, $\hat{\mathbf{r}}$, $\hat{\mathbf{v}}$ serta $\hat{\mathbf{e}}$; (iii) menghitung suku ruas kanan (15) s.d. (18) menggunakan hasil (ii) untuk memperoleh $\boldsymbol{\theta}^{(1)}=(\hat{\sigma}_t^2,\hat{\sigma}_r^2,\hat{\sigma}_v^2,\hat{\sigma}_e^2)'$; (iv) mengulangi (i) s,d (iii) menggunakan hasil (iii) sampai diperoleh nilai $\boldsymbol{\theta}=(\sigma_t^2,\sigma_r^2,\sigma_v^2,\sigma_e^2)'$ yang konvergen. Nilai $\boldsymbol{\theta}=(\sigma_t^2,\sigma_r^2,\sigma_v^2,\sigma_e^2)'$ yang diperoleh kemudian dimasukkan ke (6) untuk memperoleh estimasi secara empiris bagi $\boldsymbol{\beta}$ dan $\boldsymbol{\delta}$, serta prediktor empiris bagi **t**, **r**, dan **v**.

SIMPULAN DAN SARAN

Tulisan ini merupakan konsep awal dimana efek desain sampling informatif disertakan dalam model regresi eror bersarang serta dengan melibatkan pendekatan *p-spline*. Persamaan model campuran Henderson pada model yang dikembangkan tersebut memberikan solusi estimator yang linier dan tak bias bagi parameter yang merupakan efek tetap model dan prediktor tak bias linier terbaik bagi efek random model. Model selanjutnya dapat digunakan untuk inferensi lebih lanjut seperti estimasi bagi mean area dan total area.

DAFTAR PUSTAKA

Kish, L. (1995). The Hundred Years wars of Survey Sampling. *Statistic in Transition*, 2, 813-830.

Little, R. J. (2004). To Model or Not to Model ? Competing Models of Inference for Finite Population Sampling. *Journal of the American Statistical Association*, 99(466), 546-556.

Reiter, J. P., Zanutto, E. L. dan Hunter, L. W. (2005). Analytical modeling in complex surveys of work practices. *ILR Review: the Journal of Work and Policy*, 59(1), 82-100.

Searle, S. R., Casella, G. dan Mc Culloh, C. E. (1992). *Variance Components*. New Jersey: John Wiley and Sons, Inc.

Smith, T. M. F. (1976). The Foundation of Survey Sampling: A Review (with discussion). *Journal of the Royal Statistical Society, Series A*, 139, 183-204.

Smith, T. M. F. (1994). Sample Surveys 1975-1990: An Age of Reconciliation? (with discussion). *International Statistical Review*, 62, 5-34.

Rao, J. N. K. (2003). *Small Area Estimation*. New Jersey: John Wiley & Sons Inc.

Robinson, G. K. (1991). That BLUP is a Good Thing: The Estimatin of Random Effects. *Statistical Science*, 6(1): 15-32.

Ruppert, D., Wand, M. P. dan Carroll, R. J. (2003). *Semiparametric Regression*. United Kingdom: Cambridge University Press.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistic*, 18: 223-249.

UJI PERUBAHAN STRUKTURAL PADA REGRESI KUANTIL DENGAN *LAGRANGE MULTIPLIER*

Triwik Jatu Parmaningsih [1, a)], Sri Haryatmi[2, b)], Danardono[3, c)]

[1] Pengajar Program Studi Matematika FMIPA UNS
[2,3]Pengajar Jurusan Matematika FMIPA UGM

[a)] triwik_jt@yahoo.com [b)]s_kartiko@yahoo.com
[c)]danardono@ugm.ac.id

Abstrak: Regresi kuantil merupakan teknik statistika yang digunakan untuk menduga hubungan antara variabel respon dengan variabel penjelas pada fungsi kuantil bersyarat tertentu. Regresi kuantil mengijinkan kita untuk memeriksa model regresi tidak hanya pada pusat sebaran data (seperti pada estimator kuadrat terkecil) tapi juga bagian atas atau bawah ekor sebaran. Inferensi pada model regresi kuantil merupakan topik penelitian yang sedang berkembang dan terdapat beberapa uji untuk memverifikasi eksistensi perubahan struktural, namun sebagian besar penelitian dalam literatur perubahan struktural hanya memfokuskan pada rata-rata bersyarat, sementara perubahan struktural yang terjadi pada distribusi bersyarat atau pada kuantil bersyarat adalah kunci penting. Artikel ini membahas uji perubahan struktural pada regresi kuantil dengan *Lagrange Multiplier.* Uji ini diimplementasikan dengan mengestimasi regresi *auxiliary.*

Kata kunci: Perubahan Struktural, Regresi Kuantil, Uji Hipotesis

PENDAHULUAN

Metode Ordinary Least Square (OLS) merupakan metode estimasi klasik yang digunakan untuk membangun model regresi. Dalam OLS disyaratkan untuk memenuhi asumsi- asumsi yang ada, antara lain asumsi residu berdistribusi normal (asumsi normalitas), variansi galat yang bersifat konstan (homokedastisitas), tidak adanya korelasi serial pada residual dan tidak adanya multikolinearitas antar variabel independen. Metode ini dikenal peka terhadap penyimpangan asumsi pada data, misalnya jika data tidak memenuhi salah satu asumsi regresi maka penduga OLS tidak lagi baik digunakan. Salah satu asumsi terpenting adalah asumsi sebaran normal (normalitas). Jika terjadi pelanggaran asumsi normalitas, solusi yang biasa dilakukan adalah melakukan transformas pada data, namun terkadang transformasi yang dilakukan belum bisa memenuhi asumsi sehingga menyebabkan nilai taksiran menjadi bias. Kemudian berkembanglah regresi median dengan pendekatan LAD (*Least Absolute Deviation*) yang dikembangkan dengan mengganti pendekatan *mean* pada OLS menjadi median. Suatu saat pendekatan median dirasa kurang tepat, karena hanya mendasarkan pada dua kelompok data yang dibagi pada nilai tengahnya saja. Padahal ada celah bahwa sebaran data terletak pada potongan kuantil tertentu saja. Berawal dari hal tersebut dikembangkan metode regresi kuantil. Regresi kuantil merupakan teknik statistika yang digunakan untuk menduga hubungan antara variabel respon dengan variabel penjelas pada fungsi kuantil bersyarat tertentu. Pendekatan metode regresi kuantil dengan memisahkan atau membagi data menjadi dua atau lebih kelompok dimana dicurigai adanya perbedaan nilai taksiran pada kuantil-kuantil tertentu. Metode regresi kuantil tidak membutuhkan asumsi parametrik (Buhai, 2005). Regresi kuantil mengijinkan kita untuk memeriksa model regresi tidak hanya pada pusat sebaran data (seperti pada estimator kuadrat terkecil) tapi juga bagian atas atau bawah ekor sebaran. Inferensi pada metode regresi kuantil merupakan topik penelitian yang sedang berkembang dan terdapat dua uji yang telah diajukan untuk memverifikasi eksistensi perubahan struktural pada regresi kuantil, yaitu uji berdasarkan estimasi fungsi objektif dan uji berdasarkan pada gradien. Uji ketidakstabilan parameter dan perubahan struktural merupakan isu yang relevan. Di bawah hipotesis nol parameter regresi diasumsikan konstan. Di bawah hipotesis alternatif parameter regresi diijinkan untuk merubah respon. Perubahan struktural merupakan perubahan pola yang terjadi pada deret waktu. Hansen (2001) menjelaskan bahwa perubahan struktural ditandai dengan perubahan parameter pada waktu tertentu. Namun oleh Qu (2008) istilah perubahan struktural ini digunakan dalam ranah regresi kuantil, sehingga perubahan struktural diartikan sebagai perubahan nilai parameter pada kuantil tertentu atau di beberapa kuantil. Dalam model regresi, misal Y adalah variabel dependen dan X adalah variabel independen, maka satu hal menarik yang menjadi perhatian adalah apakah hubungan antara Y dan X berubah atau tidak. Regresi kuantil mampu menjawab pertanyaan tersebut. Dengan melihat pada kuantil tertentu, kita dapat menjawab pertanyaan seperti ”Apakah hubungan y dan x pada kuantil ke 75 berubah?”. Dalam ha ini uji statistik dapat dikonstruksikan untuk mengevaluasi ketetapan hubungan persamaan Y dan X pada kuantil tertentu atau di beberapa kuantil. Regresi kuantil memungkinkan pengujian perubahan struktural untuk kuantil yang berbeda. Lebih lanjut, mungkin timbul masalah bahwa perubahan ekor sebaran tidak dapat dideteksi pada pusat sebaran data atau pada median. Untuk uji perubahan struktural pada regresi kuantil, Furno (2007) menggunakan uji *likelihood ratio* berdasar pada regresi kuantil, sebut C^1. Uji *likelihood ratio* ini membandingkan fungsi objektif pada *constrained* dan *unconstrained model*, yang diestimasi pada kuantil yang terpilih. *Constrained model* mengasumsikan koefisien konstan, sementara *uncon- strained model* mengijinkan parameter untuk merubah respon pada perubahan struktural. Qu (2008) memberikan uji yang berbeda untuk perubahan struktural, yang mendasarkan pada gradien regresi kuantil pada *constrained model*. Uji ini membandingkan sebagian dan jumlah total dari gradien tersebut. Perubahan struktural pada *constrained model* memberikan estimasi bahwa rata-rata nilai sebelum dan sesudah perubahan dan gradien pada sub sampel akan berbeda secara signifikan dengan gradien yang dihitung pada sampel keseluruhan. Uji ini tidak menganalisis tiap-tiap koefisien regresi. Namun, uji ini mengambil nilai maksimum dari koefisien- koefisien yang berbeda dan kuantil-kuantil yang berbeda. Furno (2011) memberikan uji tambahan untuk perubahan struktural, berdasar pada *Lagrange Multiplier* (LM). LM diimplementasikan dengan mengestimasi regresi *auxiliary* untuk memverifikasi apakah residual dari model koefisien konstan dapat dijelaskan lebih lanjut oleh data. Dengan memperhatikan uji yang sudah dibahas sebelumnya, LM *test* lebih mudah di implementasikan. Uji ini hanya membutuhkan estimasi model koefisien konstan dan regresi *auxiliary*. Keuntungan LM *test* terletak pada perhitungannya yang mudah. Dengan pertimbangan tersebut, maka uji perubahan struktural pada regresi kuantil dengan *Lagrange Multiplier* dijadikan pokok bahasan dalam penelitian ini.

METODE PENELITIAN

Penelitian ini mengkaji tentang model regresi kuantil, uji hipotesis untuk perubahan structural, dilanjutkan dengan uji perubahan struktural berdasarkan *Lagrange Multiplier test.*

HASIL DAN PEMBAHASAN

Model Regresi Kuantil

Koenker (2005) menjelaskan bahwa regresi dilakukan untuk memberikan ringkasan terhadap rata-rata dari distribusi yang sesuai dengan x. Kita bisa melangkah lebih jauh dan menghitung beberapa regresi yang berbeda sesuai dengan persentase berbagai poin dari distribusi dan dengan demikian mendapatkan gambaran yang lebih lengkap dari himpunan. Biasanya hal ini tidak dilakukan, sehingga regresi memberikan gam- baran yang tidak lengkap. Sama halnya pada mean, memberikan gambaran tidak lengkap dari distribusi tunggal, sehingga regresi memberikan gambaran tidak lengkap untuk distribusi. Regresi kuantil dimaksudkan untuk menawarkan strategi menyeluruh untuk melengkapi gambaran regresi.

Regresi kuantil merupakan teknik statistika yang digunakan untuk menduga hubungan antara variabel respon dengan variabel penjelas pada fungsi kuantil bersyarat tertentu. Pendekatan metode regresi kuantil dengan memisahkan atau membagi data menjadi dua atau lebih kelompok dimana dicurigai adanya perbedaan nilai taksiran pada kuantil-kuantil tertentu. Pendekatan ini memungkinkan menduga fungsi kuantil bersyarat pada berbagai nilai kuantil yang diinginkan. Secara umum regresi kuantil sangat bermanfaat ketika ingin menganalisis bagian tertentu dari suatu sebaran bersyarat. Contoh, adanya suatu kebijakan yang bertujuan untuk membantu siswa yang mempunyai nilai rendah. Dalam hal ini perhatian harus fokus pada kuantil yang lebih rendah dalam rangka untuk memahami dampak dari kebijakan tersebut. Misal Y adalah variabel dependen, X adalah variabel independen dan $F(y) = P(Y \leq y)$ adalah fungsi distribusi Y, maka untuk $0 < \theta < 1$, kuantil ke-θ dari Y adalah

$$Q_y(\theta) = F^{-1}(\theta) = \inf\{y : F_Y(y) \geq \theta\} \tag{1}$$

Saat mengevaluasi pada nilai kuantil yang berbeda, fungsi (1) memberikan gambaran yang lengkap atas pengaruh X terhadap Y. Dengan mempertimbangkan model regresi linier sederhana

$$y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$$

Maka regresi kuantil yang dirintis oleh Koenker dan Basset (1978), merumuskan fungsi kuantil bersyarat dari y dengan nilai x tertentu, $Q_y(\theta|x)$ sebagai fungsi linear dari parameter dan dipengaruhi leh m *structural break :*

$$Q_{y_i}(\theta|x_i) = \begin{cases} x_{i1}\beta_1(\theta) \\ x_{i2}\beta_2(\theta) \\ \vdots \\ x_{i(m+1)}\beta_{(m+1)}(\theta) \end{cases} \tag{2}$$

Dengan $\theta \in (0,1)$ merupakan kuantil terpilih, $\beta_j(\theta)$ merupakan parameter regresi dengan $j = 1,\ldots,m+1$ merupakan segmen index, $x_{i(m+1)}$ variabel independen pada segmen ke j dengan $i = n_{j-1}+1,\ldots,n_j$ dan $n_j\,(j = 1,\ldots,m)$ merupakan *breakpoint* yang tidak diketahui dengan $n_j = \lambda n,\ 0 < \lambda \leq 1.$ Vektor parameter $\beta(\theta)$ dapat diestimasi dengan meminimalkan masalah

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \rho_\theta \left(y_i - x_i \beta \right) \tag{3}$$

dengan $\rho_\theta(\varepsilon_i)$ merupakan *check function,* $\rho_\theta(\varepsilon_i) = \varepsilon_i \left(\theta - 1(\varepsilon_i < 0)\right)$ sehingga

$$\rho_\theta(\varepsilon_i) = \begin{cases} \theta \varepsilon_i, & \text{jika } \varepsilon_i \geq 0 \\ -(1-\theta)\varepsilon_i, & \text{jika } \varepsilon_i < 0 \end{cases}$$

dengan ε_i merupakan error regresi kuantil. Jika nilai β pada (3) diganti dengan estimatornya yaitu *b* maka persamaannya menjadi

$$\min_{b \in \mathbb{R}} \left[\sum_{i \in \{i: y \geq x_i b\}} \theta (y_i - x_i b) + \sum_{i \in \{i: y < x_i b\}} (1-\theta)(y_i - x_i b) \right] \tag{4}$$

Hal yang menarik adalah apakah hubungan antara *y* dan *x* berubah atau tidak. Regresi kuantil mampu menjawab pertanyaan tersebut. Untuk melihatnya, missal $\left((x_i, y_i),\ i = 1,\ldots,n\right)$ sampel berukuran *n* dan kuantil ke θ dari y_i diberikan oleh

$$Q_{y_i}\left(\theta \middle| x_i\right) = x_i \beta(\theta) \tag{5}$$

Pasangan sampel (x_{ik}, y_{ik}) berbeda dengan (x_{il}, y_{il}) (disebut terjadi perubahan struktural) jika

$$\beta_k(\theta) \neq \beta_l(\theta)$$

untuk suatu $\theta \in (0,1)$ dengan $k, l \in j,\ \ j = 1, 2, \ldots, m+1$.

Dengan kata lain, perubahan struktural terjadi jika ada perubahan nilai parameter pada tiap-tiap segmen atau tiap sub sampel.

Koenker (2005) mengatakan bahwa kuantil yang merupakan pengurutan pada sampel pengamatan dapat dinyatakan sebagai solusi dari masalah optimisasi sederhana, sehingga istilah pengurutan diganti dengan pengoptimalan. Masalah regresi kuantil pda (3) dapat diformulasikan sebagai sebuah program linier

$$\min_{(\beta,u,v) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \left\{ \theta 1_n^T u + (1-\theta) 1_n^T v \middle| X\beta + u - v = y \right\} \tag{6}$$

dengan 1_n^T merupakan vektor *n* berelemen 1, $\{u_i, v_i : 1, \ldots, n\}$ adalah vektor residu $y - X\beta$ yang bernilai positif dan negatif serta *X* pada (6) berukuran *nxp.*

Masalah ini sering disebut problem primal. Problem dual yang berkaitan adalah

$$\max\left\{ y^T a \middle| X^T a = (1-\theta) X^T 1_n,\ a \in [0,1]^n \right\} \tag{7}$$

Problem primal dapat dipandang seba- gai pengurutan sampel pengamatan, dan problem dual adalah pemberian ranking pada pengamatan tersebut (Koenker, 2005). Masalah dual pada persamaan (7) akan memiliki solusi sebagaimana yang diajukan oleh Gutenbrunner dan Jurecková (1992) sebagai $\hat{a}_i(\theta)$ yang didefinisikan

$$\hat{a}_i(\theta) = \begin{cases} 0, & \text{jika } \theta > \dfrac{R_i}{n}; \\ R_i - n\theta, & \text{jika } \dfrac{R_i - 1}{n} \le \theta \le \dfrac{R_i}{n}; \\ 1, & \text{jika } \theta < \dfrac{R_i - 1}{n}. \end{cases} \tag{8}$$

dengan R_i merupakan rang dari observasi ke *i*, Y_i dalam sampel $y_1, \ldots, y_n$.

Uji Hipotesis untuk Perubahan Struktural

Uji statistik dapat dikonstruksikan untuk mengevaluasi ketetapan hubungan pada persamaan (5) pada kuantil tertentu atau di beberapa kuantil (Qu, 2008). Dengan melihat pada kuantil tertentu, kita dapat menjawab pertanyaan seperti "Apakah hubungan y dan x pada kuantil ke 75 berubah?" Penelitian ini fokus pada model kuantil linear dan kuantil bersyarat dari y_i diasumsikan pada persamaan (5). Ada dua macam uji hipotesis untuk model (5). Uji perubahan struktural yang pertama untuk kuantil yang sudah diten- tukan sebelumnya dengan hipotesis nol (H_0) dan hipotesis altenatif (H_1) sebagai berikut

$$H_0 : \beta_i(\theta) = \beta_0(\theta) \ ,$$

untuk semua *I*, untuk suatu $\theta \in (0,1)$.

$$H_1 : \beta_i(\theta) = \begin{cases} \beta_1(\theta), & i = 1, 2, \ldots, n_1; \\ \beta_2(\theta), & i = n_1 + 1, \ldots, n. \end{cases}$$

untuk suatu $\theta \in (0,1)$.

Uji perubahan struktural yang kedua untuk beberapa kuantil, misalkan kuantil yang berada dalam himpunan £, dengan hipotesis nol (H_0^*) dan hipotesis alternatif (H_1^*) sebagai berikut

$$H_0^* : \beta_i(\theta) = \beta_0(\theta)$$

untuk semua *I*, dan untuk semua $\theta \in$ £.

$$H_1^* : \beta_i(\theta) = \begin{cases} \beta_1(\theta), & i = 1, 2, ..., n_1; \\ \beta_2(\theta), & i = n_1 + 1, ..., n. \end{cases}$$

untuk beberapa $\theta \in \pounds$.

Uji Perubahan Struktural

Hansen (2001) mengatakan bahwa perubahan struktural ditandai dengan perubahan parameter pada waktu tertentu. Perubahan struktural oleh Qu (2008) digunakan dalam ranah regresi kuantil, sehingga perubahan struktural diartikan sebagai perubahan nilai parameter pada kuantil tertentu atau di beberapa kuantil. Pada model regresi linear dengan error ε_i iid, uji untuk perubahan struktural di bawah hipotesis nol, persamaan *constrained* model

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, ..., n \qquad (9)$$

mengasumsikan koefisien konstan pada seluruh sampel. Di bawah hipotesis alternatif diberikan persamaan *unconstrained* model

$$y_i = \alpha + \beta x_i + \xi z_i + \varepsilon_i, \quad i = 1, 2, ..., n \qquad (10)$$

dimana $z_i = \delta_i x_i$ dan δ_i adalah variabel *dummy* dengan nilai unit setelah perubahan. Sehingga ξ mengukur perubahan, kenaikan atau pengurangan, dari koefisien regresi setelah perubahan. Model alternatif, dapat disajikan dalam bentuk lain. Sebagai ganti penambahan variabel *explanatory* (variabel bebas) z_i , sampel dipecah menjadi dua dan dua regresi diestimasi secara terpisah, satu sebelum dan satu setelah perubahan

$$\begin{aligned} y_i &= \alpha_1 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, ..., \lambda n \\ y_i &= \alpha_2 + \beta_2 x_i + \varepsilon_i, \quad i = \lambda n + 1, 2, ..., n \end{aligned} \qquad (11)$$

dengan $0 < \lambda \leq 1$.

Pada regresi kuantil, fungsi objektif untuk kuantil θ yang terpilih adalah

$$\begin{aligned} V(\theta) &= \sum_{y_i < \alpha + \beta x_i} (1-\theta)\left(y_i - \alpha(\theta) - \beta(\theta) x_i\right) \\ &\quad + \sum_{y_i \geq \alpha + \beta x_i} \theta\left(y_i - \alpha(\theta) + \beta(\theta) x_i\right) \\ &= \sum_i \rho_\theta\left(y_i - \alpha(\theta) + \beta(\theta) x_i\right) \\ &= \sum_i \rho_\theta(e_i) \end{aligned}$$

dengan ρ merupakan *check function* $\rho_\theta(e_i) = e_i(\theta - I(e_i < 0))$ (Koenker dan Basset (1978)) dan e_i adalah residu regresi kuantil. Uji perubahan struktural berdasar $V(\theta)$ diberikan oleh Furno (2007) :

$$C^1 = \frac{\left[\tilde{V}(\theta) - \hat{V}(\theta)\right]/d_1}{\hat{V}(\theta)/d_2} = \left(\frac{\tilde{V}(\theta)}{\hat{V}(\theta)} - 1\right)\frac{d_2}{d_1} \tag{12}$$

Fungsi C^1 berdistribusi F_{d_1,d_2} .

Notasi $\tilde{V}(\theta)$ merupakan estimasi fungsi objektif pada kuantil θ pada *constrined* model (9) dan $\hat{V}(\theta)$ adalah estimasi fungsi objektif pada kuantil θ pada *unconstrained* model (11).

Gradien dari fungsi objektif regresi kuantil untuk model dengan kendala $\tilde{V}(\theta)$ adalah dasar dari *Qu* (2008) *test* :

$$\begin{aligned} S_n(\theta) &= n^{-1/2}\sum_{i=1,n} x_i \psi_\theta\left(y_i - a(\theta) - b(\theta)x_i\right) \\ &= n^{-1/2}\sum_{i=1,n} x_i \psi_\theta(\tilde{e}_i) \end{aligned} \tag{13}$$

dengan $\psi_\theta(\tilde{e}_i) = \rho'_\theta(\tilde{e}_i) = \theta - 1(\tilde{e}_i)$ merupakan *sign function.* Selanjutnya dievaluasi gradien pada sub sampel berukuran λn, $S_n(\lambda, \theta)$ untuk $0 < \lambda \leq 1$, dan didefinisikan fungsi $H_{\lambda,n}(\theta)$ sebagai berikut :

$$\begin{aligned} S_n(\lambda,\theta) &= n^{-1/2}\sum_{i=1,n}^{\lambda n} x_i \psi_\theta\left(y_i - a(\theta) - b(\theta)x_i\right) \\ &= n^{-1/2}\sum_{i=1,n}^{\lambda n} x_i \psi_\theta(\tilde{e}_i) \end{aligned} \tag{14}$$

$$H_{\lambda,n}(\theta) = \left(X^T X\right)^{-1/2}\sum_{i=1}^{\lambda n} x_i \psi_\theta(\tilde{e}_i) \tag{15}$$

dimana X adalah (n, p) matriks variabel *explanatory*. Qu (2008) membandingkan $H_{\lambda,n}(\theta)$ dan $H_{1,n}(\theta)$ dalam pengujian berikut

$$Q_\theta = \sup_\lambda \left\| \left(\theta(1-\theta)\right)^{-1/2}\left[H_{\lambda,n}(\theta) - \lambda H_{1,n}(\theta)\right] \right\| \tag{16}$$

Uji *Lagrange Multiplier* untuk Perubahan Struktural

Uji *Lagrange Multiplier* merupakan uji yang cukup mudah dalam perhitungan karena ini hanya membutuhkan estimasi *constrained* model ditambah dengan regresi *artificial* yang sesuai Dengan regresi kuantil, diberikan persamaan *auxiliary* sebagai berikut :

$$\psi_\theta\left(\tilde{e}_i\right) = \xi z_i + \varepsilon_i \tag{17}$$

Dengan $\psi_\theta\left(\tilde{e}_i\right) = \rho'_\theta\left(\tilde{e}_i\right) = \theta - 1\left(\tilde{e}_i \leq 0\right)$ adalah *sign function* yang merupakan turunan pertama dari $\rho_\theta\left(\tilde{e}_i\right)$. Uji *Lagrange Multiplier* untuk perubahan struktural pada regresi kuantil didefinisikan dengan $LM^{1} = nR(\theta)$ yang berdistribusi χ^2 dengan derajat bebas 1 (Furno, 2011), dimana $R(\theta)$ adalah *goodness of fit index* pada (17). Adapun formula *goodness of fit index* pada regresi kuantil adalah

$$R\left(\theta\right) = 1 - \hat{V}\left(\theta\right) / \tilde{V}\left(\theta\right) \tag{18}$$

dengan $\tilde{V}$ (θ) dan $\hat{V}$ (θ) merupakan estimasi fungsi objektif *constrained* dan *unconstrained* model (Koenker dan Machado, 1999).

Pada persamaan (12) telah diasumsikan secara implisit bahwa titik perubahan (*breakpoint*) diketahui. Ketika hal tersebut tidak terjadi, maka pencarian disemua *breakpoint* λ yang mungkin harus diimplementasikan pada (16).

SIMPULAN DAN SARAN

Regresi kuantil merupakan teknik statistika yang digunakan untuk menduga hubungan antara variabel respon dengan variabel penjelas pada fungsi kuantil bersyarat tertentu. Uji *Lagrang Multiplier* yang berdasar pada residual regresi kuantil merupakan salah satu uji untuk memverifikasi eksistensi perubahan struktural.Uji perubahan struktural pada regresi kuantil mengijinkan untuk menguji di setiap nilai kuantil yang diiinginkan. Uji *Lagrange Multiplier* memerlukan estimasi *constrained model* dimana residu dari *constrained model* diregresikan pada variabel *dummy* yang mempunyai nilai unit setelah perubahan. Uji ini cukup mudah dalam pengerjaannya namun cenderung memiliki nilai yang hampir sama di semua kuantil karena ada satu elemen pada fungsi LM^{1}, yaitu ukuran sampel, yang tidak berubah pada kuantil yang berbeda. Uji perubahan struktur pada regresi kuantil dengan *Lagrange Multiplier* dapat juga dilakukan di beberapa kuantil.

DAFTAR PUSTAKA

Buhai, S (2005). Quantile Regression : Overview and Selected Application. *Ad Astra 4 , 1–16.*

Chow, G.C. (1960). Test of Equality between Sets of Coefficients in Two Linear Regressions. *Journal Econometrica 28 , no.3, 591-605.*

Furno, M. (2007). Parameter Instability in Quantile Regresion. *Statistical Modelling 7, 345–362.*

Furno, M. (2011). Tests for Structural Break in Quantile Regressions. *Springer-Verlag.*, AStA Adv Stat Anal, DOI 10.1007/s10182-012-0188-3.

Gutenbrunner, C. and J. Jureckova. Regression Rank Scores and Regression Quantiles. *The Annals of Statistics 1992, Vol. 20, no. 1, 305-330*.

Hansen, B.E. (2001). The New Econometrics of Struktural Change : Dating Breaks in U.S. Labor Productivity. *Journal of Eco- nomic Perspectives 15, no.4, 117-128.*

Koenker, R. (2005). *Quantile Regression,* Cambridge University Press, Cambridge.

Koenker, R. and Basset, G. (1978). Regression Quantiles. *Econometrica, Vol. 46, No.1, 33-50.*

Koenker, R. and Machado, J. (1999). Goodness of Fit and Related Inference for Quantile Regression. *Journal of the American Statistical Association, Vol. 94, No. 448, 1296-1310.*

Oka, T. and Qu, Z. (2011). Estimating Structural Changes in Regression Quantiles. *Journal of Econometrics, Vol. 162, 248-267.*

Qu, Z. (2008). Testing for Structural Change in Regression Quantiles. *Journal of Econometrics, Vol. 146, 170–184.*