MODEL MULTILEVEL *SPLINE LINEAR TRUNCATED* DALAM REGRESI NONPARAMETRIK MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

MULTILEVEL SPLINE LINEAR TRUNCATED MODEL IN NONPARAMETRIK REGRESSION USING THE METHOD RESTRICTED MAXIMUM LIKELIHOOD

HEDI KUSWANTO

PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2022

MODEL MULTILEVEL *SPLINE LINEAR TRUNCATED* DALAM REGRESI NONPARAMETRIK MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

Tesis

sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

HEDI KUSWANTO

H062202003

kepada

PROGRAM STUDI MAGISTER STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

TESIS

MODEL MULTILEVEL *SPLINE LINEAR TRUNCATED* DALAM REGRESI NONPARAMETRIK MENGGUNAKAN METODE *RESTRICTED MAXIMUM LIKELIHOOD*

HEDI KUSWANTO

H062202003

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 22 Juni 2022
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

Pembimbing Pendamping

Dr. Nirwan, M.Si.
NIP. 19630306 198702 1 002

Ketua Program Studi
Magister Statistika,

Dr. Dr. Georgina M. Tinungki, M.Si
NIP. 19620926 198702 2 001

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS HASANUDDIN
FAKULTAS MIPA

Dr. Eng. Amiruddin, M.Si
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS
DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul Model Multilevel *Spline Linear Truncated* Dalam Regresi Nonparametrik Menggunakan Metode *Restricted Maximum Likelihood* adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Anna Islamiyati, S.Si., M.Si dan Dr. Nirwan, M.Si). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (International Journal of Academic and Applied Research (IJAAR) ISSN: 2643-9603 Vol. 6 Issue 2, February - 2022, Pages:77-81) sebagai artikel dengan judul "Multilevel Nonparametrik Regression Model with Truncated Linear Spline Estimator on Students' National Examination Scores".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 22 Juni 2022

Yang Menyatakan,

Hedi Kuswanto

NIM. H062202003

UCAPAN TERIMA KASIH

Puji syukur atas kehadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyusun dan menyelesaikan tesis ini. Penulis menyadari sepenuhnya bahwa apa yang dikemukakan dalam tesis ini masih jauh dari kesempurnaan yang merupakan sebagai akibat dari keterbatasan kemampuan serta berbagai kesulitan yang penulis hadapi dalam penyusun tesis ini.

Penulis memanjatkan doa kepada Tuhan Yang Maha Esa agar memberikan rahmat-Nya kepada pihak yang banyak membantu dalam penyelesaian tesis ini. Penulis juga percaya tesis ini dapat selesai bukan hanya dengan kekuatan pikiran penulis semata akan tetapi karena bantuan dari berbagai pihak juga, baik selama proses perkuliahan bahkan sampai proses pengerjaan tesis di Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Namun demikian, penulis dengan senang hati menerima kritik dan saran yang bersifat membangun dari pembaca karya tulis ini demi sempurnanya tesis ini.

Terima kasih yang tak terhingga kepada kedua orang tua tercinta dan saudariku atas doa yang tak pernah putus, dukungan serta segala kebaikan mereka yang sampai kapan pun takkan pernah bisa terbalaskan atas kasih sayang yang tiada henti dalam penyelesaian tesis ini. Selanjutnya, saya ingin menyampaikan juga rasa hormat dan terima kasih kepada:

1. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. **Dr. Eng. Amiruddin, M.Si**. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.
3. **Dr. Nurtiti Sanusi, M.Si.** selaku Ketua Departemen Statistika yang menjadi salah satu tim penguji tesis sekaligus memberikan ilmu, dukungan, dan motivasi serta kemudahan kepada penulis dalam berbagai hal selama menjalani Pendidikan di Departemen Statistika.
4. **Dr. Dr. Georgina Maria Tinungki, M.Si**. selaku Ketua Program Studi Magister Statistika yang menjadi salah satu tim penguji tesis.

5. **Dr. Anna Islamiyati, S.Si., M.Si** selaku Pembimbing Utama yang telah bersabar dan bersedia meluangkan banyak waktunya untuk membimbing penulis dan memberikan masukan dalam penyelesaian tesis ini.
6. **Dr. Nirwan Ilyas, M.Si** selaku Pembimbing Pertama yang telah bersabar dan bersedia meluangkan banyak waktunya untuk membimbing penulis dan memberikan masukan dalam penyelesaian tesis ini.
7. **Dr. Erna Tri Herdiani, M.Si** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
8. Saudara tak sedarah **TRANSPOSE 2014**, terkhusus ketua angkatan **Fandy Heribet**, saudara-saudara di grup **SAHABAT LERR** terkhusus **Arfyan Saputra, Miftah Farid, Setiawan Ahmad, Nivel Saputra Wahid, Muh. Alfajar, Munawir Djamaluddin, Rachmat Darmawan, Ahmad Husain, Muhammad Mappanyompa, Muh. Sarwan, Muh. Ij'lal, Andi Muh. Fawzy, Syahrul** menjadi rumah ke dua bagi penulis.
9. Teman **Pondok Hisyam** yaitu **Muh Ashar, Samsir Aditya Ania, Muh Naim, Andi Arkan Alam Putra,** dan **Abu Hurairah** yang telah berkontribusi besar dengan saling mengingatkan dalam kebaikan, dan saling memotivasi,. Saya ucapkan Jazaakumullahu Khairon.
10. Teman **Besti Lab** yaitu **Agung Muhammad Takdir, Ratmila, Andi Isna Yunita, Alimatun Najiha, Ainun Utari,** dan **Asnidar** atas doa, semangat serta kebersamaannya selama ini yang banyak membantu penulis.
11. Teman – teman Mahasiswa Program Magister Statistika angkatan Angkatan 20202 terkhusus **Maktisen Ena, Trigarcia Randa,** dan **Muh. Qardawi Hamzah** terima kasih dukungan luar biasa kepada penulis.

Semoga Allah SWT memberikan pahala yang berlipat ganda atas segala kebaikan yang telah diberikan kepada penulis dan semoga penulisan tesis ini bermanfaat bagi perkembangan ilmu pengetahuan dan teknologi, khususnya dalam dunia statistika dan data sains.

Makassar, 22 Juni 2022

Hedi Kuswanto

ABSTRAK

HEDI KUSWANTO. **Model Multilevel *Spline Linear Truncated* Dalam Regresi Nonparametrik Menggunakan Metode *Restricted Maximum Likelihood*** (dibimbing oleh Dr. Anna Islamiyati, S.Si., M.Si dan Dr. Nirwan, M.Si)

Multilevel telah populer di industri dan penelitian karena sifat penskalaan resolusi pada level yang berbeda. Salah satu kegunaan model bertingkat yang telah dipelajari sebelumnya adalah dalam bidang pendidikan mengenai nilai Ujian Nasional sekolah menengah pertama. Besarnya sampel tiap kabupaten atau kota berbeda, sehingga metode estimasi *Restricted Maximum Likelihood* menjadi tepat. Selain itu, terdapat kecenderungan nilai Ujian Nasional Nasional tidak mengikuti pola parametrik, sehingga diperlukan pendekatan penduga spline terpotong. Hasil yang diperoleh adalah model terbaik dari model level-1 menggunakan tiga titik knot dengan nilai GCV 72.58384 dan model level-2 menggunakan dua titik knot dengan nilai GCV 12.4201.

Kata kunci: Model Multilevel, Spline Truncated, Restricted Maximum Likelihood, Ujian Nasional, GCV

ABSTRACT

HEDI KUSWANTO. **Multilevel Spline Linear Truncated Model in Nonparametrik Regression Using the Restricted Maximum Likelihood Method** (supervised by Dr. Anna Islamiyati, S.Si., M.Si and Dr. Nirwan, M.Si)

Multilevel has been popular in industry and research due to the nature of resolution scaling at different levels. One of the uses of the stratified model that has been studied previously is in the field of education regarding junior high school National Examination scores. The sample size for each district or city is different, so the Restricted Maximum Likelihood estimation method is appropriate. In addition, there is a tendency that the National National Examination scores do not follow a parametric pattern, so a truncated spline estimator approach is needed. The results obtained are the best model from the level-1 model using three knot points with a GCV value of 72.58384 and the level-2 model using two knot points with a GCV value of 12.4201.

Keywords: Multilevel Model, Spline Truncated, Restricted Maximum Likelihood, National Exam, GCV

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi linier merupakan metode yang digunakan sebagai pengambilan keputusan tentang hubungan antara peubah penjelas dengan peubah respon. Hubungan yang dimaksud dalam hal ini adalah hubungan sebab-akibat antar kedua peubah tersebut (Drapper & Smith, 1998). Secara umum, model regresi dibagi atas dua yaitu model regresi parametrik dan model regresi nonparametrik. Perbedaan kedua model regresi terletak pada asumsi bentuk pola datanya. Pola regresi parametrik diasumsikan diketahui dengan memperhatikan plot data awal atau informasi penelitian sebelumnya, sementara pola regresi nonparametrik tidak mengikuti bentuk pola parametrik dan diasumsikan tidak diketahui (Härdle, 1990).

Beberapa estimator yang telah dikembangkan oleh peneliti regresi nonparametrik diantaranya estimator Kernel (Aydin & Yilmaz, 2017), *Spline* (Budiantara, et al., 2009), Polinomial Lokal (Chamidah, et al., 2012), Deret *Fourier* (Mardianto, et al., 2019), dan MARS (Otok, 2009). Salah satu estimator yang telah dikembangkan dalam melakukan estimasi kurva regresi adalah *spline* karena metode *spline* dapat menyesuaikan pola data secara efektif (Wulandari, et al., 2014). Estimator *spline* merupakan salah satu estimator dalam regresi nonparametrik yang mendapatkan estimasi data berdasarkan pergerakan pola data, sehingga *spline* disebut estimator yang memiliki sifat estimasi yang fleksibel (Budiantara, 2011). Beberapa jenis *spline* yang telah dikembangkan diantaranya *Spline Truncated* (Kuswanto, et al., 2022), *Spline Smoothing* (Lestari, et al., 2010) dan *Penalized Spline* (Islamiyati, et al., 2020). *Spline Truncated* adalah salah satu bentuk *spline* yang bersifat piecewise polynomial, yaitu suatu potongan-potongan polinom yang memiliki sifat tersegmen pada selang yang terbentuk pada titik-titik knot (Wang & Lijian, 2009). Bentuk *Spline Linear Truncated* banyak digunakan dalam analisis data karena kemudahan interpretasi secara visual akibat keterlibatan titik knot dalam estimasi fungsi. Titik knot adalah titik terjadinya pola perubahan data. Kriteria yang dapat digunakan dalam pemilihan knot yang optimal yaitu *Generalized Cross Validation (GCV)* (Budiantara, 2011).

Regresi nonparametrik juga sudah dikembangkan pada kasus multilevel, diantaranya model multilevel *spline linear* (Howe, et al., 2016), model *spline* multilevel multivariat (Macdonald- Wallis, et al., 2012), dan pemodelan multilevel kernel (He, et al., 2017). Model multilevel adalah salah satu metode statistik yang berasal dari pengembangan analisis regresi sederhana. Model multilevel diperkenalkan oleh Goldstein (1995) yang menyebutkan bahwa model multilevel dikembangkan karena dapat mengatasi masalah data dengan kondisi struktur hierarki atau berjenjang. Pada pemodelan multilevel, peubah respon diukur pada level terendah, sedangkan peubah penjelas dapat didefinisikan pada setiap level. Apabila analisis yang dilakukan pada data yang berstruktur hierarki atau berjenjang dipaksakan menggunakan regresi sederhana, maka akan memberikan hasil yang tidak akurat, sehingga hasil yang diintrepretasikan juga tidak akurat (Laila, 2016). Model multilevel yang paling sederhana adalah model dua level dengan level pertama adalah data individu dan level kedua adalah data kelompok (West & Welch, 2007). Beberapa penelitian yang telah menggunakan model multilevel diantaranya Safitri et al. (2015) menggunakan model linear campuran pada data panel dengan mengabaikan tingkatan atau level dari sekolah terhadap kabupaten atau kota, Anggara et al. (2015) menggunakan model linear campuran pada data panel dengan memperhitungkan tingkatan atau level dari tingkat kemiskinan terhadap kabupaten atau kota, dan Zulvia (2017) menggunakan model linear campuran pada data panel dengan memperhitungkan tingkatan atau level dari sekolah terhadap kabupaten atau kota.

Beberapa metode estimasi parameter yang dapat digunakan dalam model multilevel diantaranya metode kuadrat terkecil dua tahap (*Two Stage* OLS) seperti yang dilakukan Ringdal (1992). Pada *Two Stage* OLS membutuhkan asumsi bahwa setiap kelompok pada level-2 mempunyai sampel yang sama sehingga penaksir ini bukan penaksir yang baik apabila ukuran sampel tiap kelompok pada level-2 tidak sama. Metode lain dapat dilakukan dengan menggunakan metode *Maximum Likelihood* (MLE) (Hox, 1995) dan metode *Restricted Maximum Likelihood* (REML). REML menghasilkan penaksir koefisien regresi dan komponen variansi yang efisien dan konsisten apabila sampel yang digunakan dalam jumlah besar sehingga pelanggaran asumsi dapat diabaikan (Cahyati, 2020). Penaksir REML adalah penaksir parameter dengan mengoptimalkan fungsi *likelihood* melalui proses iterasi sehingga mendapatkan nilai penaksir yang konvergen (West & Welch, 2007).

Salah satu aplikasi penggunaan model multilevel yang telah dikaji sebelumnya pada bidang pendidikan yaitu nilai Ujian Nasional (UN). UN merupakan sistem evaluasi standar pendidikan dasar dan menengah secara nasional dan Persamaan mutu tingkat pendidikan antar daerah yang dilakukan di Indonesia. Pada peraturan pemerintah Nomor 13 Tahun 2015 (Pusat Penilaian Pendidikan, 2015), hasil UN dapat digunakan sebagai salah satu bahan pertimbangan untuk pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upaya peningkatan mutu pendidikan. Hasil UN dapat mendorong peningkatan mutu pendidikan pada tingkat nasional, provinsi, kabupaten/kota, dan sekolah. Harapan secara khusus, hasil yang diperoleh dari setiap mata pelajaran dapat menjadi perbaikan proses pembelajaran pada tingkat sekolah. Rata-rata nilai UN yang diperoleh pada setiap sekolah dapat dipengaruhi oleh faktor internal dan eksternal. Faktor internal misalnya kualitas siswa, kompetensi guru, sarana dan prasarana sekolah. Selain itu, faktor eksternal dapat dilihat dari kondisi ekonomi daerah seperti Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Kelulusan (AL), Angka Melek Huruf (AMH), Rata-Rata Lama Sekolah (RLS) dan Angka Harapan Lama Lulus (AHLS) (Kemendikbud, 2013). Pemerintah selalu berupaya melakukan berbagai inovasi setiap tahun untuk mewujudkan harapan peningkatan mutu pendidikan. Salah satu cara yang dilakukan adalah menganalisis informasi untuk struktur data berjenjang pelaksanaan UN setiap tingkatan yang berasal dari kabupaten atau kota berbeda. Ukuran sampel tiap kabupaten atau kota yang berbeda menyebabkan metode estimasi *Restricted Maximum Likelihood* yang sesuai. Selain itu, ada kecenderungan nilai UN tidak mengikuti pola parametrik seperti pada penelitian Asmin (2010) yang menggunakan pendekatan estimator *spline* dalam mengestimasi nilai UN di Kota Makassar.

Berdasarkan uraian tersebut, penulis mengkaji estimasi model regresi multilevel *Spline Linear Truncated* menggunakan metode *Restricted Maximum Likelihood.* Metode diaplikasikan pada data rata-rata nilai UN SMP tiap-tiap sekolah dengan memperhitungkan adanya variansi antar kabupaten/kota di Provinsi Sulawesi Selatan. Pemilihan data rata-rata nilai UN pada penelitian ini disebabkan karena terdapat perubahan nilai yang bervariasi pada tingkat kabupaten/kota di Provinsi Sulawesi Selatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diperoleh, maka didapatkan rumusan masalah sebagai berikut:

1. Bagaimana mengestimasi parameter model regresi multilevel menggunakan metode REML dengan estimator *Spline Linear Truncated*?
2. Bagaimana model regresi multilevel pada nilai UN SMP antar sekolah dan kabupaten/kota di Provinsi Sulawesi Selatan?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Untuk memperoleh hasil estimasi parameter model regresi multilevel menggunakan metode REML dengan estimator *Spline Linear Truncated*
2. Untuk memperoleh model regresi multilevel pada nilai UN SMP antar sekolah dan kabupaten/kota di Provinsi Sulawesi Selatan

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Estimator *Spline Linear Truncated* yang digunakan dibatasi hingga menggunakan orde satu dan tiga titik knot.
2. Model multilevel yang digunakan dibatasi dua level
3. Pemilihan titik knot optimal menggunakan *Generalized Cross Validation*.
4. Data yang digunakan adalah nilai Ujian Nasional (UN) Sekolah Menengah Pertama (SMP) tahun 2019 di seluruh provinsi Sulawesi Selatan.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah wawasan dan pengetahuan mengenai estimasi parameter dan model multilevel *Spline Linear Truncated* dalam regresi nonparametrik menggunakan metode *Restricted Maximum Likelihood.*
2. Dapat dijadikan sebagai salah satu rujukan bagi pemerintah dalam usaha peningkatan kualitas pendidikan di Indonesia secara umum dan di Provinsi Sulawesi Selatan secara khusus.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Nonparametrik

Regresi nonparametrik merupakan suatu teknik analisis data dalam statistika yang dapat menjelaskan hubungan antara peubah prediktor dengan peubah respon yang tidak diketahui bentuk fungsinya karena sebelumnya tidak ada informasi tentang bentuk fungsi tersebut dan hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik sangat mempertahankan fleksibilitasnya. Secara umum, model regresi nonparametrik dapat disajikan sebagai berikut: (Eubank, 1999).

$$y_i = f(x_i) + \varepsilon_i \, , \qquad (2.1)$$

dengan y_i adalah peubah respon pada pengamatan ke-*i*, x_i adalah peubah prediktor pada pengamatan ke-*i*, $f(x_i)$ adalah adalah fungsi regresi nonparametrik yang memuat peubah prediktor. ε_i adalah faktor gangguan yang tidak dapat dijelaskan dengan model yang dapat disebut sebagai galat (*error),* yang diasumsikan sebagai peubah acak dengan rata-rata nol dan variansii σ^2, dan $i = 1,2, \dots , n$.

Fungsi regresi nonparametrik pada Persamaan (2.1) bentuk kurvanya tidak diketahui dan hanya diasumsikan termuat dalam suatu fungsi tertentu, pemilihan ruang fungsi didasarkan oleh sifat kemulusan *(smooth)* yang dimiliki oleh fungsi $f(x_i)$. Regresi *spline* merupakan salah satu bentuk pendekatan regresi nonparametrik. Pada data tertentu terkadang memiliki pola yang tidak beraturan, data tersebut dapat diatasi menggunakan regresi *spline* dengan bantuan titik knot sehingga kurva yang dihasilkannya relatif mulus (Wahba, 1990).

2.2 Estimator *Spline Linear Truncated*

Spline adalah salah satu jenis polinomial tersegmen yang memiliki sifat fleksibilitas. Sifat fleksibilitas ini yang membedakan *spline* dari polinomial, sehingga memungkinkan menyesuaikan diri lebih efektif terhadap karakteristik suatu fungsi atau data (Budiantara, 2006). Metode *spline* dalam regresi nonparametrik dapat ditemui dalam banyak bentuk, terdapat beberapa estimator *spline* diantaranya adalah *smoothing spline* dan *truncated spline*. Kedua metode *spline* tersebut masing-masing menggunakan parameter yang berbeda untuk membuat estimasi kurva regresi lebih fleksibel, yaitu parameter penghalus pada *smoothing spline* dan titik knot pada *truncated spline*. Perbedaan jenis parameter tersebut, menyebabkan optimasi untuk mendapatkan estimator pada kedua metode *spline* tersebut berbeda (Islamiyati & Chamidah, 2018).

Spline pada regresi nonparametrik mempunyai kemampuan untuk mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1988). Kemampuan estimasi pada perilaku data ini ditunjukkan oleh *truncated* (potongan-potongan) yang melekat pada estimator, potongan-potongan tersebut disebut dengan titik knot. Titik knot adalah titik perpaduan bersama yang menunjukkan perubahan pola perilaku fungsi pada selang yang berbeda. Misalkan terdapat satu peubah prediktor x_i, maka titik knot diambil pada selang interval $a < k_r < b$, dengan a adalah nilai minimum dan b adalah nilai maksimun pada prediktor x_i (Härdle, 1990).

Secara umum fungsi *Spline Linear Truncated* berorde q dengan titik-titik knot pada $k_1, k_2, \dots, k_r$ dapat dinyatakan dalam bentuk sebagai berikut: (Rodrıguez, 2001):

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - k_h)_+^q, \qquad (2.2)$$

dengan $\beta_0, \beta_1, \dots, \beta_q, \beta_{q+1}, \dots, \beta_{q+r}$ adalah parameter regresi, k_h adalah titik knot ke-h, $(h = 1,2, \dots, r)$. $(x_i - k_h)_+^q$ adalah fungsi polinomial *truncated* yang dijabarkan sebagai berikut:

$$(x_i - k_h)_+^q = \begin{cases} (x_i - k_h)^q, & \text{jika } x_i \geq k_h \\ 0, & \text{jika } x_i < k_h \end{cases}.$$

Jika fungsi yang menyatakan hubungan antara p prediktor dengan respon tunggal maka fungsi *spline* $f(x_i)$ dalam Persamaan (2.2) maka dapat dinyatakan dalam bentuk sebagai berikut:

$$y_i = f(x_{1i}) + f(x_{2i}) + \cdots + f(x_{pi}) + \varepsilon_i \qquad (2.3)$$

$$= \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i; i = 1,2, \dots, n,$$

dengan

$$f(x_{ji}) = \beta_{0j} + \sum_{l=1}^{q} \beta_{jl} x_{ji}^{l} + \sum_{h=1}^{r} \beta_{j(q+h)} (x_{ji} - k_{jh})_{+}^{q}$$

dengan

y_i : peubah respon pada pengamatan ke-i,

x_{ji} : peubah prediktor ke-j pada pengamatan ke-i,

β_{0j} : intersep prediktor ke-j,

β_{jl} : parameter polinomial pada prediktor ke-j dan orde ke-l,

k_{jh} : nilai titik knot pada prediktor ke-j dan titik knot ke-h,

r : banyaknya titik knot,

q : orde polinomial *Spline Linear Truncated* ,

p : banyaknya peubah prediktor,

$\beta_{j(q+h)}$: parameter *truncated* pada prediktor ke-j dan titik knot ke-$(q+h)$,

ε_i : galat *(error)* pada pengamatan ke-*i* yang diasumsikan saling bebas berdistribusi normal dengan mean 0 dan variansi σ^2 .

Persamaan (2.3) untuk n data pengamatan dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & {x_{21}}^2 & \cdots & x_{p1}^q & (x_{11} - k_{11})_+^q & \cdots & (x_{p1} - k_{pr})_+^q \\ 1 & x_{12} & {x_{22}}^2 & \cdots & x_{p2}^q & (x_{12} - k_{11})_+^q & \dots & (x_{p2} - k_{pr})_+^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & {x_{2n}}^2 & \dots & x_{pn}^q & (x_{1n} - k_{11})_+^q & \dots & (x_{pn} - k_{pr})_+^q \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \\ \beta_{(q+1)} \\ \vdots \\ \beta_{(q+r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan y adalah vektor yang memiliki ukuran $n \times 1$, matriks X yang memiliki ukuran $n \times (1 + q + r)$, vektor β ukuran $(1 + q + r) \times 1$, dan vektor ε ukuran $n \times 1$ (Budiantara, 2006).

2.3 Model Regresi Multilevel

Model multilevel merupakan salah satu pemodelan untuk menduga hubungan antar peubah yang diamati pada level-level yang berbeda dalam stuktur data berjenjang. Model regresi multilevel diasumsikan bahwa data yang digunakan memiliki struktur data yang berjenjang, terdiri dari atas amatan (individu) yang tersarang di dalam kelompok. Model regresi multilevel merupakan bagian dari model linear campuran karena terdapat dua parameter yaitu parameter efek tetap dan efek acak yang digabungkan kedalam satu Persamaan (Goldstein, 1995).

Suatu peubah bebas dikatakan memiliki efek tetap jika koefisien regresinya bernilai sama bagi seluruh anggota sampel dan suatu peubah bebas dikatakan memiliki efek acak jika nilai koefisien regresinya berbeda antar dua atau lebih grup anggota sampel (Harlan, 2016). Adapun Persamaan model linear campuran dalam bentuk sederhana adalah (Rencher & Schaalje, 2007):

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}u + \varepsilon \quad (2.4)$$

dengan:

y : vektor peubah respon

$\boldsymbol{X}$: matriks peubah penjelas untuk parameter tetap

β : vektor parameter efek tetep

$\boldsymbol{Z}$: matriks peubah penjelas untuk parameter acak

u : vektor efek acak

ε : vektor galat acak

dengan asumsi antar galat $\boldsymbol{\varepsilon}$ dan $\boldsymbol{u}$ tidak saling bebas dan berdistribusi normal dengan variansi θ yang dapat ditulis sebagai berikut:

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}\right) \quad (2.5)$$

dengan $\boldsymbol{G} = \boldsymbol{G}(\gamma)$ dan $\boldsymbol{R} = \boldsymbol{\Sigma}$ dengan $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\phi)$. γ dan ϕ adalah vektor parameter variansi yang terikat dengan u dan ε sedangkan σ^2 adalah parameter skala variansi. Parameter variansi yaitu adalah σ^2, ϕ dan γ. $\boldsymbol{G}, \boldsymbol{R}$ dan $\boldsymbol{\Sigma}$ merupakan matriks variansi yang diasumsikan definit positif.

$$\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \boldsymbol{H}) \tag{2.6}$$

dengan

$$\boldsymbol{H} = \boldsymbol{ZGZ}' + \boldsymbol{R} \tag{2.7}$$

2.4 Struktur Data Model 2-Level

Pada pemodelan multilevel dengan 2-level, terdapat sebanyak n individu yang berasal dari m kelompok. $Y_{1j}, Y_{2j}, \dots, Y_{n_ij}$ adalah peubah respon masing-masing n_j individu pada kelompok ke-j, $j = 1,2,\dots,m$ dan jika $X_{1j}, X_{2j}, \dots, X_{kj}$ adalah peubah penjelas pada level 1 untuk kelompok ke-j, serta $Z_1, Z_2, \dots, Z_l$ adalah peubah penjelas pada level 2, maka struktur pemodelan multilevel dengan 2-level dapat disajikan dalam Tabel 2.1 berikut.

Tabel II.1 Struktur Data Model 2 Level

<table>
<tr><th rowspan="2">Klp.</th><th rowspan="2">Observasi</th><th>Peubah Respon</th><th colspan="4">Peubah Penjelas Level 1</th><th colspan="4">Peubah Penjelas Level 2</th></tr>
<tr><th>Y</th><th>X_1</th><th>X_2</th><th>...</th><th>X_k</th><th>Z_1</th><th>Z_2</th><th>...</th><th>Z_l</th></tr>
<tr><td rowspan="6">1</td><td>1</td><td>y_{11}</td><td>x_{111}</td><td>x_{211}</td><td rowspan="6">...</td><td>x_{k11}</td><td rowspan="6">z_{11}</td><td rowspan="6">z_{21}</td><td rowspan="6">...</td><td rowspan="6">z_{l1}</td></tr>
<tr><td>2</td><td>y_{21}</td><td>x_{121}</td><td>x_{221}</td><td>x_{k21}</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td></tr>
<tr><td>i</td><td>y_{i1}</td><td>x_{1i1}</td><td>x_{2i1}</td><td>x_{ki1}</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td></tr>
<tr><td>n_j</td><td>y_{n_j1}</td><td>x_{1n_j1}</td><td>x_{2n_j1}</td><td>x_{kn_j1}</td></tr>
<tr><td rowspan="2">2</td><td>1</td><td>y_{12}</td><td>x_{112}</td><td>x_{212}</td><td rowspan="2">...</td><td>x_{k12}</td><td rowspan="2">z_{12}</td><td rowspan="2">z_{22}</td><td rowspan="2">...</td><td rowspan="2">z_{l2}</td></tr>
<tr><td>2</td><td>y_{22}</td><td>x_{122}</td><td>x_{222}</td><td>x_{k22}</td></tr>
</table>

	⋮	⋮	⋮	⋮		⋮				
	i	y_{i2}	x_{1i2}	x_{2i2}		x_{ki2}				
	⋮	⋮	⋮	⋮		⋮				
	n_j	y_{n_j2}	x_{1n_j2}	x_{2n_j2}		x_{kn_j2}				
⋮	⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
	i	y_{1m}	x_{11m}	x_{21m}		x_{k1m}				
	2	y_{2m}	x_{12m}	x_{22m}		x_{k2m}				
	⋮	⋮	⋮	⋮		⋮				
M	*i*	y_{im}	x_{1im}	x_{2im}	…	x_{kim}	z_{1m}	z_{2m}	…	z_{lm}
	⋮	⋮	⋮	⋮		⋮				
	n_j	y_{n_jm}	x_{1n_jm}	x_{2n_jm}		x_{kn_jm}				

2.5 Model Regresi 2-Level

Model multilevel mempunyai struktur data berjenjang dengan satu peubah respon yang hanya diukur pada level terendah saja dan peubah penjelas yang diukur pada semua level (Goldstein, 2011). Model multilevel yang paling sederhana adalah model 2-level dengan level 1 adalah data individu dan level 2 adalah data kelompok dengan level 1 sebagai level rendah yang tersarang pada level yang tinggi yaitu level 2 (Hox, et al., 2017). Berdasarkan struktur data pada Tabel 2.1 pemodelan regresi multilevel dengan 2-level adalah sebagai berikut: (Goldstein, 2011):

2.5.1 Model level-1

Definisi model level 1 adalah model yang disusun tanpa memperhitungkan pengaruh dari level kelompok. Pemodelan multilevel untuk tiap kelompok dapat ditulis sebagai berikut: (Zulvia, 2017):

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} + \varepsilon_{ij}$$

$$y_{ij} = \beta_{0j} + \sum_{q=1}^{k} \beta_{qj}X_{qij} + \varepsilon_{ij} \tag{2.8}$$

dengan $q = 1,2, \dots, k$, individu-$i = 1,2, \dots, n_j$ dan kelompok-$j = 1,2, \dots, m$ atau dinyatakan dalam matriks berikut

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j,\ \text{dengan } \boldsymbol{\varepsilon}_j \sim N(0, \sigma^2 I) \tag{2.9}$$

$$\begin{bmatrix} y_{1j} \\ y_2 j \\ \vdots \\ y_{n_j j} \end{bmatrix} = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \dots & x_{k1j} \\ 1 & x_{12j} & x_{22j} & \dots & x_{k2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{kn_j j} & x_{kn_j j} & \dots & x_{kn_j j} \end{bmatrix} \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{kj} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{nj} \end{bmatrix}$$

dengan:

$\boldsymbol{y}_j$: vektor peubah respon

$\boldsymbol{X}_j$: matriks peubah penjelas untuk parameter tetap

$\boldsymbol{\beta}_j$: vektor parameter efek tetap

$\boldsymbol{\varepsilon}_j$: vektor galat

2.5.2 Model level-2

Koefisien regresi pada level-1, β_{pj} dengan $p = 0,1,2, \dots, k$ dalam model level-1 memiliki nilai yang berbeda antar kelompok. Variasi nilai β_{pj} akan dijelaskan dengan membentuk model level 2. Pembentukan model level 2 dilakukan untuk setiap koefisien regresi sebagai respon ke-*p* dengan menggunakan peubah penjelas pada level-2. Bentuk pemodelan pada level-2 dapat ditulis sebagai berikut: (Zulvia, 2017):

$$\beta_{pj} = \gamma_{p0} + \gamma_{1p}Z_{1j} + \gamma_{2p}Z_{2j} + \dots + \gamma_{lp}Z_{lj} + u_{pj}$$

$$\beta_{pj} = \gamma_{p0} + \textstyle\sum_{p=1}^{k} \gamma_{pl}Z_{lj} + u_{pj} \tag{2.10}$$

dengan $p = 1,2, \dots, k$, dan $l = 1,2, \dots, n_j$ atau dinyatakan dalam matriks berikut

$$\boldsymbol{\beta}_p = \boldsymbol{Z}\boldsymbol{\gamma}_p + \boldsymbol{u}_p,\ \text{dengan } \boldsymbol{u}_p \sim N(0, \sigma_\tau^2 I) \tag{2.11}$$

$$\begin{bmatrix} \beta_{p1} \\ \beta_{p2} \\ \vdots \\ \beta_{pm} \end{bmatrix} = \begin{bmatrix} 1 & Z_{11} & Z_{21} & \dots & Z_{l1} \\ 1 & Z_{12} & Z_{22} & \dots & Z_{l2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Z_{1m} & Z_{2m} & \dots & Z_{lm} \end{bmatrix} \begin{bmatrix} \gamma_{0p} \\ \gamma_{1p} \\ \vdots \\ \gamma_{lp} \end{bmatrix} + \begin{bmatrix} u_{p1} \\ u_{p2} \\ \vdots \\ u_{pm} \end{bmatrix}$$

Jika Persamaan 2.9 disubstitusikan pada Persamaan 2.11 maka diperoleh model sebagai berikut:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} + \varepsilon_{ij}$$

$$y_{ij} = \gamma_{00} + \gamma_{10}Z_{1j} + \gamma_{20}Z_{2j} + \cdots + \gamma_{l0}Z_{lj} + u_{0j} + \big(\gamma_{01} + \gamma_{11}Z_{1j} + \gamma_{21}Z_{2j} \ldots + \gamma_{l1}Z_{lj} + u_{1j}\big)X_{1ij} + \cdots + \big(\gamma_{02} + \gamma_{12}Z_{1j} + Z_{2j} + \cdots + \gamma_{l2}Z_{lj} + u_{2j}\big)X_{2ij} + \cdots + \big(\gamma_{0k} + \gamma_{1k}Z_{1j} + \gamma_{2k}Z_{2j} + \cdots + \gamma_{lk}Z_{lj} + u_{kj}\big)X_{kij} + \varepsilon_{ij}$$

$$y_{ij} = \gamma_{00} + \sum_{q=1}^{l} \gamma_{0q}Z_{qj} + \sum_{p=1}^{k} \gamma_{0p}Z_{pij} + \sum_{p=1}^{k}\sum_{q=1}^{l} \gamma_{qp}Z_{qj}X_{pij} + u_{0j} + \sum_{p=1}^{k} u_{pj}X_{pij} + \varepsilon_{ij} \quad (2.12)$$

atau dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j \quad (2.13)$$

dengan $\boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma}$ adalah efek tetap dan $\left[\boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j\right]$ adalah efek acak.

Hox (2010) matriks $\boldsymbol{Z}$ dalam Persamaan 2.13 merupakan peubah perantara untuk menghubungkan $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$. Oleh karena itu, variasi hubungan antara $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$ tergantung pada $\boldsymbol{R}$. Interpretasi dari koefisien regresi pada model level 1 dan koefisien regresi model level 2 terhadap y tergantung pada tanda positif dan negatif dari kedua koefisien regresi tersebut. Jika koefisien regresi model level 1 dan koefisien regresi model level 2 bertanda sama berarti kedua koefisien regresi tersebut memiliki pengaruh berbanding lurus terhadap y. Pada $\boldsymbol{X}_j\boldsymbol{Z}_j$ merupakan faktor interaksi dalam model sebagai variansi *slope* peubah $\boldsymbol{X}$.

2.6 *Intraclass Correlation*

Model regresi multilevel dapat menghasilkan nilai dugaan bagi *intraclass correlation* (Hox & Maas, 2002). Nilai dugaan tersebut dapat diperoleh menggunakan model tanpa peubah penjelas atau dikenal *intercept-only model* atau *null model*. *Null model* adalah model regresi yang hanya terdiri dari intersep saja tanpa memasukkan pengaruh peubah bebas. Null model pada level-1 dapat dituliskan sebagai berikut:

$$\boldsymbol{Y}_{ij} = \boldsymbol{\beta}_{oj} + \boldsymbol{\varepsilon}_{ij} \quad (2.14)$$

Sedangkan pada level-2 Persamaannya sebagai berikut:

$$\boldsymbol{\beta}_{oj} = \boldsymbol{\gamma}_{00} + \boldsymbol{u}_{0j} \quad (2.15)$$

Pada Persamaan di atas $u_{0j} \sim N(0, \sigma_{\mathrm{u0}}^2)$ adalah variansi galat pada level kelompok dan $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ adalah variansi galat dari level individu. Sehingga dari *null model* inilah diperoleh *intraclass correlation* (ρ) dapat dirumuskan sebagai berikut:

$$\rho = \frac{\sigma_{u_{oj}}^2}{\sigma_{u_{oj}}^2 + \sigma_{\varepsilon_{ij}}^2} \qquad (2.16)$$

Dengan $\sigma_{u_{oj}}^2$ merupakan variansi level-2 dan $\sigma_{\varepsilon_{ij}}^2$ merupakan variansi level-1. Semakin besar nilai ρ menunjukkan semakin tinggi korelasi antar individu sehingga tidak bisa dilakukan analisis regresi biasa dan diperlukan analisis regresi multilevel.

2.7 Estimasi Parameter

Estimasi dalam model linear campuran melibatkan dua proses terkait yaitu akan diestimasi efek tetap dan efek acak ($\boldsymbol{\beta}$ dan $\boldsymbol{u}$) dan parameter variansi ($\boldsymbol{\theta}$ dan $\boldsymbol{\gamma}$). Estimasi ini dilakukan dengan menggunakan metode yang pertama kali ditetapkan oleh Henderson (1950) dan Henderson (1959) dengan metode yang kekar untuk estimasi parameter variansi dalam model linier campuran adalah estimasi REML (Thompson, 2008).

2.7.1 Estimasi Parameter Efek tetap dan Efek acak

Searle, et al., (2006) mengusulkan metode untuk mendapatkan solusi estimasi $\boldsymbol{\beta}$ dan $\boldsymbol{u}$ dengan memaksimalkan *likelihood* gabungan dari $(\boldsymbol{y}, \boldsymbol{u})$. Kepadatan peluang gabungan dari $\boldsymbol{y}$ dan $\boldsymbol{u}$ dapat ditulis dengan mengalikan kepadatan kondisional dari $\boldsymbol{y}$ terhadap $\boldsymbol{u}$ dan fungsi kepadan peluang dari **u**

$$f(\boldsymbol{y}, \boldsymbol{\beta}; \boldsymbol{u}) = f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}) f(\boldsymbol{u}|\boldsymbol{G}) \qquad (2.17)$$

dengan

$$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \boldsymbol{R}),$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta_0}, \boldsymbol{B}), \qquad \boldsymbol{u}|\boldsymbol{G} \sim N(\boldsymbol{0}, \boldsymbol{G})$$

Dengan fungsi *likelihood* sebagai berikut:

$$\ell(\beta,\sigma,y|u) = -\frac{1}{2}\left\{n\log 2\pi + \log|\boldsymbol{R}| + (\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zu})'\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zu})\right\} - \frac{1}{2}\left\{q\log 2\pi + \log|\boldsymbol{G}| + \boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\right\} \quad (2.18)$$

(Lee & Nelder, 2001)

2.7.2 Estimasi Parameter Variansi

Searle, et all (2006) mendefinisikan bahwa $\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \boldsymbol{V})$ dimana $\boldsymbol{K'X} = \boldsymbol{0}$,

$$\boldsymbol{K'y} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{K'VK})$$

dengan melakukan penggantian yang sesuai, $\boldsymbol{y}$ menjadi $\boldsymbol{K'y}$, $\boldsymbol{X}$ menjadi $\boldsymbol{K'X} = \boldsymbol{0}$, $\boldsymbol{Z}$ menjadi $\boldsymbol{K'Z}$, $\boldsymbol{V}$ menjadi $\boldsymbol{K'VK}$

dimana diketahui

$$\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'V}^{-1} = \boldsymbol{K}(\boldsymbol{K'VK})^{-1}\boldsymbol{K'} \quad (2.19)$$

dan

$$\boldsymbol{V} = \sum_{i=0}^{r} \boldsymbol{Z}_i\boldsymbol{Z}_i'\sigma_i^2 + \sigma_e^2\boldsymbol{I} \quad (2.20)$$

dengan fungsi *likelihood* sebagai berikut:

$$\ell(\boldsymbol{\sigma^2}, \boldsymbol{\gamma}; \boldsymbol{y}) = -\frac{1}{2}(n-r)log\ 2\pi - \frac{1}{2}\log|\boldsymbol{K'VK}| - \frac{1}{2}\boldsymbol{y'K}(\boldsymbol{K'VK})^{-1}\boldsymbol{K'y} \quad (2.21)$$

2.8 Pemilihan Titik Knot Optimal

Estimator yang digunakan pada *Spline Linear Truncated* didasarkan pada titik knot yang optimal. Titik knot inilah yang membuat suatu kurva *spline* menjadi tersegmen. Pemilihan lokasi titik knot yang berbeda akan menghasilkan fungsi yang berbeda pula (Budiantara, 2006). Model regresi *spline* terbaik tergantung pada titik knot optimal. Metode untuk memilih titik knot optimal salah satunya adalah dengan menggunakan *Generalized Cross Validation* (GCV) (Budiantara, 2006). Titik knot optimal untuk model regresi *spline* terbaik diperoleh dari nilai GCV yang terkecil. Metode GCV dapat dituliskan seperti berikut:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}trace[I - A(k)])^2}, \quad (2.22)$$

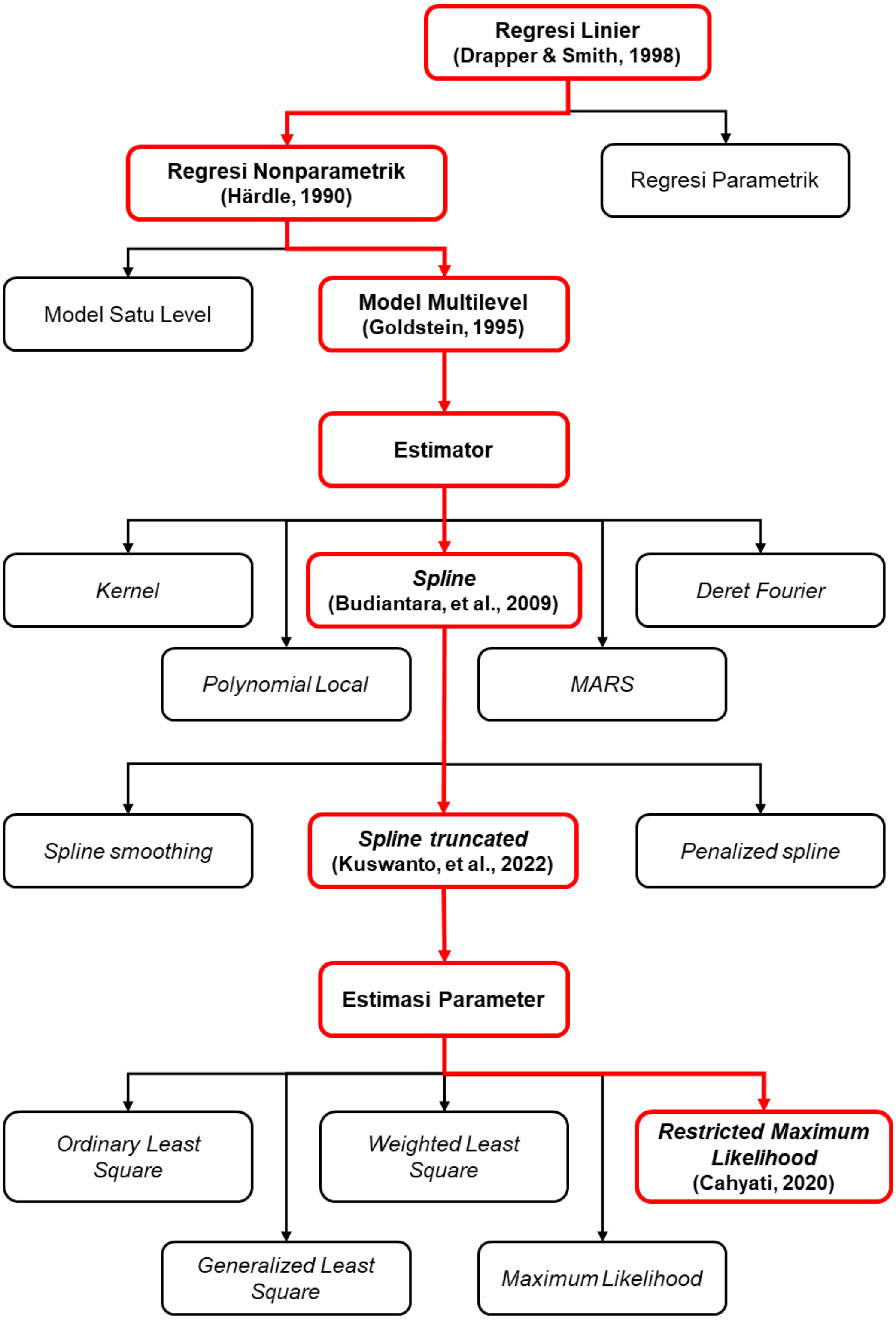
dengan $MSE(k) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y})^2$, $k = (k_1, k_2, \ldots k_r)$ adalah titik knot, *I* adalah matriks identitas, $\boldsymbol{A}[k] = \boldsymbol{X}[k](\boldsymbol{X}'[k]\boldsymbol{X}[k])^{-1}\boldsymbol{X}'[k]$, dan *n* adalah jumlah data (Budiantara, 2006).

2.9 Ujian Nasional

Ujian nasional adalah untuk melihat kondisi mutu pendidikan di Indonesia dan diharapkan terjadi pemerataan kualitas yang sama di seluruh daerah di Indonesia dengan memberikan standar nilai kelulusan yang sama di seluruh Indonesia (Sailan, 2016). Ujian Nasional adalah salah satu instrumen yang digunakan pemerintah untuk menilai pencapaian standar kompetensi lulusan siswa dari suatu mata pelajaran secara nasional (Hidayah & Astuti, 2016)

Ujian nasional merupakan salah satu bentuk dari kegiatan evaluasi pendidikan yang berupa evaluasi hasil belajar siswa. Melalui pengukuran dan penilaian pencapaian standar kompetensi lulusan secara nasional pada mata pelajaran tertentu yang ditetapkan oleh pemerintah. Selain itu ujian nasional juga dijadikan sebagai alat evaluasi pendidikan untuk pemetaan masalah mutu pendidikan dalam rangka menyusun kebijakan pendidikan nasional (Hidayat, 2016)

2.10 Kerangka Konseptual



Gambar II.1 Kerangka Konseptual