



TESIS - KS185411

**ESTIMASI KURVA REGRESI SEMIPARAMETRIK UNTUK
DATA LONGITUDINAL MENGGUNAKAN ESTIMATOR
CAMPURAN SPLINE TRUNCATED DAN DERET FOURIER**

ANJAR WICITRA WENING
06211750010029

Dosen Pembimbing
Prof. Dr. Drs. I Nyoman Budiantara, M.Si
Dr. Ismaini Zain, M.Si

Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
2020



TESIS - KS185411

**ESTIMASI KURVA REGRESI SEMIPARAMETRIK UNTUK
DATA LONGITUDINAL MENGGUNAKAN ESTIMATOR
CAMPURAN SPLINE TRUNCATED DAN DERET FOURIER**

**ANJAR WICITRA WENING
06211750010029**

**Dosen Pembimbing
Prof. Dr. Drs. I Nyoman Budiantara, M.Si
Dr. Ismaini Zain, M.Si**

**Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
2020**



THESIS - KS185411

**SEMIPARAMETRIC REGRESSION CURVE ESTIMATION FOR
LONGITUDINAL DATA USING MIXED SPLINE TRUNCATED
AND FOURIER SERIES ESTIMATOR**

**ANJAR WICITRA WENING
06211750010029**

Supervisor
Prof. Dr. Drs. I Nyoman Budiantara, M.Si
Dr. Ismaini Zain, M.Si

Department of Statistics
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
2020

LEMBAR PENGESAHAN TESIS

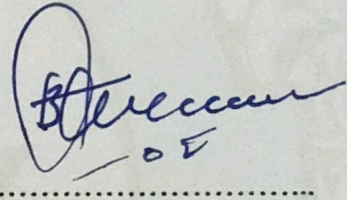
Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Statistika (M.Stat)
di
Institut Teknologi Sepuluh Nopember

Oleh:
ANJAR WICITRA WENING
NRP: 06211750010029

Tanggal Ujian : 20 Januari 2020
Periode Wisuda : Maret 2020

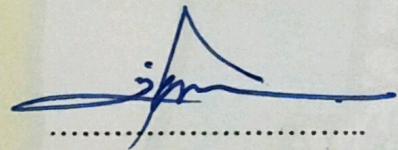
Disetujui oleh:
Pembimbing:

1. Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
NIP: 19650603 198903 1 003



.....

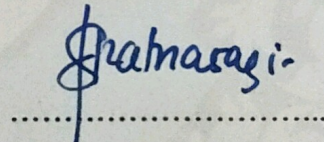
2. Dr. Ismaini Zain, M.Si.
NIP: 19600525 198803 2 001



.....

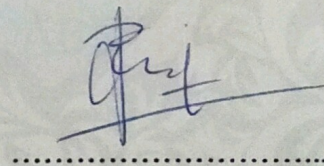
Penguji :

1. Dr. Vita Ratnasari, M.Si.
NIP: 19700910 199702 2 001



.....

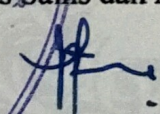
2. Santi Puteri Rahayu, M.Si., Ph.D.
NIP: 19750115 199903 2 003



.....



Kepala Departemen Statistika
Fakultas Sains dan Analitika Data


Dr. Kartika Fithriasari, M.Si.
NIP: 19691212 199303 2 002

(Halaman ini sengaja dikosongkan)

ESTIMASI KURVA REGRESI SEMIPARAMETRIK UNTUK DATA LONGITUDINAL MENGGUNAKAN ESTIMATOR CAMPURAN SPLINE TRUNCATED DAN DERET FOURIER

Nama Mahasiswa : Anjar Wicitra Wening
NRP : 06211750010029
Pembimbing : Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Co. Pembimbing : Dr. Ismaini Zain, M.Si

ABSTRAK

Regresi semiparametrik adalah pendekatan regresi yang digunakan ketika bentuk kurva regresi diasumsikan sebagian diketahui dan sebagian tidak diketahui. Jenis metode estimasi yang sama umumnya digunakan untuk beberapa atau semua variabel prediktor pada model regresi semiparametrik. Padahal ditemukan banyak contoh kasus data yang memiliki pola berbeda pada setiap variabel prediktornya, sehingga jika data dipaksa untuk menggunakan hanya satu bentuk estimator untuk mengestimasi kurva regresi, akan menghasilkan estimasi yang tidak sesuai dengan pola data. Oleh karena itu, estimator campuran perlu dikembangkan. Namun, estimator campuran dalam penelitian sebelumnya hanya mampu menangani data *cross-section*. Maka penelitian ini menggunakan data longitudinal yang memiliki kelebihan, salah satunya adalah estimasi masing-masing karakteristik individu dan waktu diperoleh secara terpisah. Oleh karena itu, penelitian ini mengembangkan suatu model estimator campuran Spline Truncated dan Deret Fourier dalam regresi semiparametrik untuk data longitudinal dengan menggunakan metode *Weighted Least Square* dalam penyelesaian estimasinya. Pemilihan model terbaik didasarkan pada titik knot dan parameter osilasi yang dapat dipilih secara optimal menggunakan nilai *Generalized Cross Validation* (GCV) minimum. Model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier diterapkan pada data persentase kejadian pneumonia di Indonesia. Model terbaik yang diperoleh adalah model regresi dengan satu titik knot dan osilasi $k = 1$, dengan nilai GCV sebesar 17,52002.

Kata Kunci : Semiparametrik, Spline Truncated, Deret Fourier, Data Longitudinal, *Weighted Least Square*

(Halaman ini sengaja dikosongkan)

SEMIPARAMETRIC REGRESSION CURVE ESTIMATION FOR LONGITUDINAL DATA USING MIXED SPLINE TRUNCATED AND FOURIER SERIES ESTIMATOR

Name : Anjar Wicitra Wening
NRP : 06211750010029
Supervisors : Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Dr. Ismaini Zain, M.Si

ABSTRACT

Semiparametric regression is a regression approach that is used when the form of regression curve is assumed to be partly known and partly unknown. In the semiparametric regression model, the same type of estimation method is generally used for some or all of the predictor variables. There are many examples of data cases that have different patterns on each predictor variable, so if the data is forced to use only one form of estimator to estimate the regression curve, it will produce estimation that will not be appropriate to the data pattern. Therefore, a mixed estimator needs to be developed. However, mixed estimator in previous studies was only able to handle cross-sectional data. So this study uses longitudinal data that has advantages, one of which is estimation of each individual characteristic and time characteristic are obtained separately in longitudinal data. Therefore, this study develops a mixed Spline Truncated and Fourier Series estimator in semiparametric regression for longitudinal data using the Weighted Least Square method. The selection of the best model is based on knot points and optimum oscillation parameters that should be selected optimally using minimum Generalized Cross Validation (GCV) value. Semiparametric regression model for longitudinal data with mixed Spline Truncated and Fourier Series estimator is used for percentage of pneumonia case data in Indonesia. The best model obtained is a regression model with one knot point and oscillation $k = 1$, with GCV value 17,52002.

Keyword : Semiparametric, Spline Truncated, Fourier Series, Longitudinal Data, *Weighted Least Square*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya serta shalawat dan salam selalu tercurah pada Nabi Muhammad SAW atas suri tauladannya dalam kehidupan ini sehingga penulis dapat menyelesaikan penyusunan Tesis yang berjudul **“Estimasi Kurva Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline Truncated dan Deret Fourier”**. terselesaikannya Tesis ini tidak lepas dari bantuan berbagai pihak. Penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada Bapak Mustofa dan Ibu Kusniah sebagai orang tua atas dorongan moral dan materiil yang sangat berarti mulai awal perkuliahan hingga akhir penyusunan tesis. Selanjutnya, penulis ingin menyampaikan rasa terima kasih juga kepada :

1. Prof. Dr. Drs. I Nyoman Budiantara M.Si. dan Dr. Ismaini Zain, M.Si. selaku dosen pembimbing yang telah banyak meluangkan waktunya untuk memberikan ilmu, bimbingan, arahan serta dorongan semangat kepada penulis dalam penyelesaian tesis ini.
2. Dr. Vita Ratnasari, M.Si dan Santi Puteri Rahayu, M.Si., Ph.D. selaku dosen penguji yang memberi saran dan evaluasi demi menjadikan tesis ini lebih baik.
3. Dr. Kartika Fithriasari, M.Si sebagai Ketua Departemen Statistika FSAD ITS
4. Dr.rer.pol. Dedy Dwi Prastyo, M.Si sebagai Ketua Program Studi Pascasarjana Departemen Statistika ITS beserta jajarannya atas kemudahan dan fasilitas yang diberikan selama studi.
5. Santi Puteri Rahayu, M.Si., Ph.D..selaku dosen wali yang telah memberikan semangat, motivasi serta masukan demi kelancaran dan terselesaikannya studi.
6. Segenap dosen pengajar dan para staff Departemen Statistika FSAD-ITS yang memberikan bekal ilmu, memfasilitasi dan membantu penulis selama masa perkuliahan.
7. Kedua adik tersayang Anggoro Krido Laksono dan Airlangga Tri Atmojo yang selalu memberikan doa dan dukungan sehingga menjadi dukungan moral bagi penulis.

8. Teman seperjuangan tesis, Miftahul Jannah Maulidia dan Dasty Dewi Prawanti yang selalu menemani diskusi, berbagi ilmu dan saling menyemangati dalam penyelesaian tesis.
9. Sahabat-sahabat : Nurma Istifara, Intan Diyora, Silvia Tri Putri, Daniatus Syahr Hajj, Din Amalia, Meirinda Fauziah, Aisyah Abdullah dan Adawiyah Asti serta keluarga Archimedes tercinta yang selalu memberikan semangat, support dan doa.
10. Teman-teman seperjuangan S2 2017 Ganjil dan Genap atas segala bantuan dan persaudaraan semasa perkuliahan.
11. Seluruh pihak yang telah banyak membantu penulis dan tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam penulisan tesis ini masih terdapat kekurangan baik dari segi penyusunan, bahasa atau penulisan, maka penulis mengharapkan kritik dan saran yang bersifat membangun guna menjadi acuan dalam bekal pengalaman bagi penulis untuk lebih baik di masa yang akan datang. Semoga tesis ini bermanfaat bagi pembaca.

Surabaya, 31 Januari 2020

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK	iii
ABSTRACT	v
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	6
BAB 2 TINJAUAN PUSTAKA	7
2.1 Regresi Parametrik, Nonparametrik dan Semiparametrik	7
2.2 Regresi Semiparametrik Spline Truncated	8
2.3 Regresi Semiparametrik Deret Fourier	11
2.4 Regresi Semiparametrik Campuran Spline Truncated dan Deret Fourier pada data Cross Section	13
2.5 <i>Weighted Least Square</i>	14
2.6 Definisi dan Teorema Dasar Aljabar Matriks	15
2.7 Data Longitudinal	16
2.8 <i>Generalized Cross Validation</i>	17
2.9 Penerapan Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline <i>Truncated</i> dan Deret Fourier pada Data Persentase Kejadian Pneumonia	18
BAB 3 METODOLOGI PENELITIAN	21
3.1 Sumber Data	21
3.2 Variabel Penelitian	21

3.3 Langkah-Langkah Penelitian.....	23
BAB 4 HASIL DAN PEMBAHASAN	29
4.1 Estimasi Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline <i>Truncated</i> dan Deret Fourier	29
4.2 Pemodelan Persentase Kejadian Pneumonia di Indonesia dengan Pendekatan Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline <i>Truncated</i> dan Deret Fourier	39
4.2.1 Deskripsi Variabel Penelitian.....	39
4.2.2 Pemodelan Persentase Kejadian Pneumonia.....	47
BAB 5 KESIMPULAN.....	57
5.1 Kesimpulan	57
5.2 Saran.....	58
DAFTAR PUSTAKA	59
LAMPIRAN	63

DAFTAR TABEL

Tabel 3.1	Variabel-Variabel dalam Penelitian	22
Tabel 3.2	Struktur Data Penelitian.....	23
Tabel 4.1	Perhitungan Nilai GCV untuk Semua Kemungkinan Model.....	46
Tabel 4.2	Komponen Parametrik dan Nonparametrik	46
Tabel 4.3	Perbandingan Estimator Tunggal dan Estimator Campuran.....	47
Tabel 4.4	Perbandingan Nilai GCV pada Masing-Masing Pembobot.....	47
Tabel 4.5	Nilai GCV dengan Satu Titik Knot dan Osilasi $k = 1$	49
Tabel 4.6	Nilai GCV dengan Satu Titik Knot dan Osilasi $k = 2$	49
Tabel 4.7	Nilai GCV dengan Satu Titik Knot dan Osilasi $k = 3$	50
Tabel 4.8	Hasil Perhitungan GCV antara Tiga Model.....	50

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

Gambar 1.1	Posisi Penelitian Ini Dibandingkan Penelitian Sebelumnya	5
Gambar 2.1	Kerangka Konsep Pemilihan Variabel	19
Gambar 3.1	Diagram Alir Estimasi Kurva Model Regresi Semiparametrik untuk Data Longitudinal dengan Estimator Campuran Spline Truncated dan Deret Fourier	26
Gambar 3.2	Diagram Alir Pemodelan Persentase Kejadian Pneumonia dengan Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline Truncated dan Deret Fourier	28
Gambar 4.1	Persentase Kejadian Pneumonia di Indonesia Tahun 2010-2017	40
Gambar 4.2	<i>Scatterplot</i> data <i>cross-section</i> antara persentase kejadian pneumonia dan persentase cakupan imunisasi dasar pada bayi di Indonesia tahun 2010-2017	41
Gambar 4.3	<i>Scatterplot</i> data <i>cross-section</i> antara persentase kejadian pneumonia dan persentase penduduk miskin di Indonesia tahun 2010-2017	41
Gambar 4.4	<i>Scatterplot</i> data <i>cross-section</i> antara persentase kejadian pneumonia dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Indonesia tahun 2010-2017	42
Gambar 4.5	<i>Scatterplot</i> data <i>time series</i> antara persentase kejadian pneumonia dan persentase cakupan imunisasi dasar pada bayi di Indonesia tahun 2010-2017	43
Gambar 4.6	<i>Scatterplot</i> data <i>time series</i> antara persentase kejadian pneumonia dan persentase penduduk miskin di Indonesia tahun 2010-2017	44
Gambar 4.7	<i>Scatterplot</i> data <i>time series</i> antara persentase kejadian pneumonia dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Indonesia tahun 2010-2017	45
Gambar 4.8	Plot antara nilai y pengamatan dan nilai estimasi y di provinsi Sulawesi Utara	54

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1	Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017.....	63
Lampiran 2	<i>Syntax</i> R Estimasi Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline <i>Truncated</i> dan Deret Fourier	70
Lampiran 3	Titik Knot Optimal	77
Lampiran 4	<i>Output</i> Estimasi Parameter	78
Lampiran 5	<i>Output</i> Estimasi Variabel Respon (y).....	82
Lampiran 6	Plot antara y dan \hat{y} di Masing-Masing Provinsi di Indonesia	86
Lampiran 7	Estimasi Model Regresi Semiparametrik dengan Estimator Campuran Spline <i>Truncated</i> dan Deret Fourier Berdasarkan Data Series	90

(Halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Menyelidiki pola hubungan fungsional antara variabel respon dan variabel prediktor melalui estimasi kurva dapat dilakukan dengan analisis regresi. Terdapat tiga pendekatan dalam mengestimasi kurva regresi, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Apabila bentuk kurva regresi diketahui, maka digunakan pendekatan regresi parametrik. Apabila hubungan variabel respon dan variabel prediktor tidak diketahui bentuk polanya, maka digunakan pendekatan regresi nonparametrik. Apabila bentuk kurva regresi sebagian diasumsikan diketahui dan sebagian lagi tidak diketahui, maka digunakan pendekatan regresi semiparametrik (Budiantara, 2014). Terdapat beberapa metode untuk mengestimasi kurva regresi semiparametrik, diantaranya yaitu Spline, Kernel, dan Deret Fourier. Penelitian yang telah dilakukan sebelumnya diantaranya adalah Bilodeau (1992) serta Tripena dan Budiantara (2006) yang menggunakan estimator Deret Fourier, serta Setiawan, dkk (2017) yang menggunakan estimator Spline Truncated.

Pada model regresi nonparametrik atau semiparametrik umumnya digunakan jenis metode estimasi yang sama untuk sebagian atau semua variabel prediktornya. Pada kenyataannya banyak kasus yang memiliki pola data berbeda pada masing-masing variabel prediktornya. Apabila dipaksakan hanya digunakan satu bentuk estimator untuk mengestimasi kurva regresi akan menghasilkan estimasi yang tidak sesuai dengan pola data dan cenderung menghasilkan *error* yang besar. Untuk mengatasi masalah tersebut beberapa peneliti telah mengembangkan estimator kurva regresi campuran dimana masing-masing pola data dalam model regresi dihipotesiskan dengan estimator kurva yang sesuai. Penelitian yang melibatkan estimator campuran pernah dilakukan oleh Ratnasari, dkk (2016), Rory (2016) dan Purnomo (2016) mengenai estimator campuran Spline Truncated dan Kernel dalam regresi nonparametrik. Selain itu, Afifah, Budiantara dan Latra (2017) melakukan penelitian tentang estimator campuran

Kernel dan Deret Fourier dalam regresi nonparametrik. Nisa, Budiantara dan Rumiati (2017) melakukan penelitian tentang model regresi semiparametrik multivariabel dengan estimator campuran Deret Fourier dan Kernel. Penggunaan estimator campuran Deret Fourier dan Spline Truncated juga telah dilakukan oleh Sudiarsa, dkk (2015) dalam regresi nonparametrik serta Nisa (2017) dalam regresi semiparametrik.

Salah satu estimator campuran yang penting untuk diteliti adalah estimator campuran Spline Truncated dan Deret Fourier. Estimator campuran Spline Truncated dan Deret Fourier digunakan apabila teridentifikasi pola hubungan antara variabel respon dengan variabel-variabel prediktor pada pemodelan regresi sebagian mengikuti pola tertentu, sebagian berubah pada sub-sub interval tertentu, dan sebagian yang lainnya mempunyai pola berulang (Nisa, 2017). Spline memiliki interpretasi statistik dan interpretasi visual yang sangat baik, serta memiliki fleksibilitas yang tinggi (Eubank, 1999). Spline juga memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Budiantara, 2014). Estimator Deret Fourier merupakan polynomial trigonometri yang mempunyai fleksibilitas, sehingga dapat menyesuaikan diri secara efektif terhadap sifat lokal data (Asrini dan Budiantara, 2014). Estimator ini umum digunakan apabila data tidak diketahui polanya dan ada kecenderungan berulang (Bilodeau, 1992). Pola data berulang yang dimaksud adalah pengulangan terhadap nilai variabel dependen untuk variabel independen yang berbeda-beda (Prahutama, 2013). Estimator Spline Truncated dan Deret Fourier pernah dilakukan pada penelitian Ratnasari, dkk (2015) yang membandingkan kedua estimator tersebut dalam model regresi nonparametrik multivariabel yang menyimpulkan bahwa Spline Truncated dan Deret Fourier memberikan hasil koefisien determinasi (R^2) yang tinggi.

Penelitian dengan estimator campuran masih banyak dilakukan pada data *cross-section*, dan belum banyak dilakukan pada data longitudinal. Data longitudinal merupakan data yang didapat dari pengamatan berulang dari setiap subjek pada interval waktu yang berbeda. Data ini berkorelasi pada subjek yang sama dan independen antar subjek yang berbeda (Wu dan Zhang, 2006). Data longitudinal lebih mampu mengidentifikasi dan mengukur efek yang tidak dapat

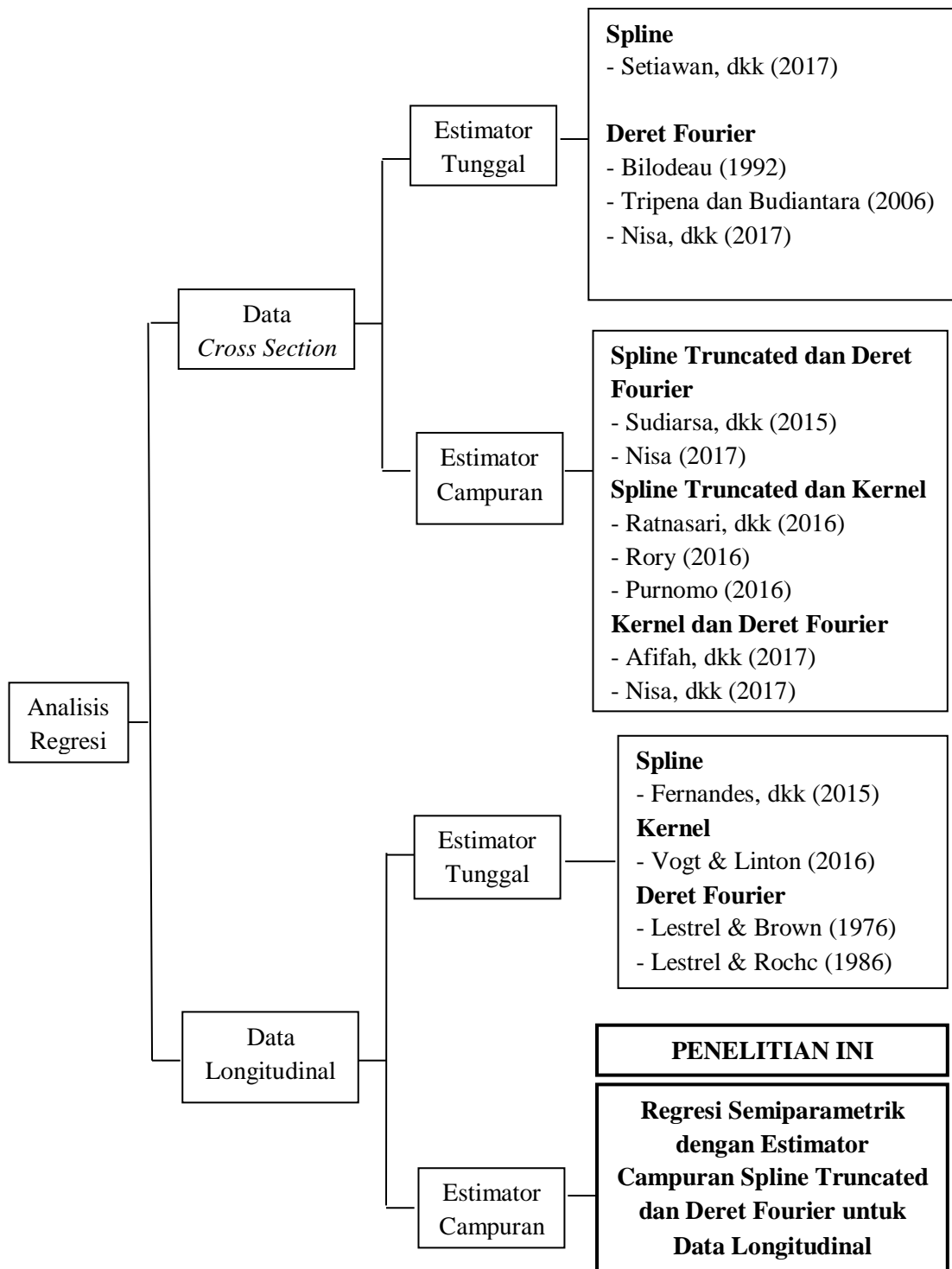
dideteksi dalam data *cross-section* atau *time series* (Baltagi, 2005). Dengan menerapkan proses estimasi pada data longitudinal, maka secara bersamaan dapat mengestimasi karakteristik individu dengan memperhatikan adanya dinamika antar waktu dan masing-masing variabel dalam penelitian (Umam, 2018). Penelitian dengan data longitudinal pada regresi nonparametrik telah dilakukan sebelumnya oleh Lestrel dan Brown (1976) serta Lestrel dan Roche (1986) menggunakan estimator deret fourier, Fernandes, dkk (2015) menggunakan estimator Spline serta Vogt dan Linton (2016) menggunakan estimator Kernel. Penelitian-penelitian pada data longitudinal tersebut masih dilakukan dengan estimator tunggal, dan belum dilakukan dengan estimator campuran.

Cakupan data longitudinal sangat luas, salah satunya adalah pada bidang kesehatan. Salah satu fokus kajian di bidang kesehatan adalah penyebaran penyakit menular maupun tidak menular serta penanganannya, salah satunya adalah pneumonia. Pneumonia merupakan salah satu penyakit yang tetap menjadi penyebab utama morbiditas dan mortalitas meskipun terdapat kemajuan dalam pengobatan dan terapi (Mattila *et al*, 2014). Penelitian tentang pneumonia pernah dilakukan sebelumnya dengan pendekatan *Geographically Weighted Regression* untuk menganalisis faktor-faktor eksternal pneumonia pada balita di Jawa Timur (Santoso dkk, 2012). Penelitian tersebut menggunakan metode yang berkaitan dengan spasial. Padahal data tentang kejadian pneumonia tercatat setiap tahun, sehingga dimungkinkan untuk digunakan analisis pada data longitudinal. Pada penelitian ini dilakukan studi awal mengenai persentase kejadian pneumonia dengan faktor-faktor yang diduga mempengaruhinya adalah persentase cakupan imunisasi dasar pada bayi, persentase penduduk miskin, dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak. Berdasarkan paparan di atas, maka akan diteliti estimasi kurva regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier yang diterapkan pada data kejadian pneumonia.

1.2 Rumusan Masalah

Banyak kasus di dunia nyata yang datanya memiliki pola berbeda pada masing-masing variabel prediktornya, sehingga apabila dipaksakan hanya digunakan satu bentuk estimator untuk mengestimasi kurva regresi akan menghasilkan estimasi yang tidak sesuai dengan pola data dan cenderung menghasilkan *error* yang besar. Oleh karena itu perlu dikembangkan suatu estimator campuran untuk memperoleh estimasi model kurva regresi yang sesuai dengan pola data. Pada penelitian regresi nonparametrik dan semiparametrik sebelumnya telah banyak digunakan macam-macam estimator, beberapa diantaranya adalah Spline Truncated dan Deret Fourier. Apabila pola hubungan antara variabel-variabel prediktor terhadap variabel respon memiliki pola yang sebagian mengikuti pola tertentu, sebagian berubah pada sub-sub interval tertentu, dan sebagian yang lainnya mempunyai pola berulang, maka digunakan regresi semiparametrik dengan estimator campuran Spline Truncated dan Deret Fourier.

Akan tetapi, estimator campuran ini masih mempunyai beberapa kelemahan, salah satu diantaranya adalah hanya mampu untuk menangani data *cross section* saja. Padahal banyak data yang tergolong data longitudinal, sehingga dengan analisis pada data longitudinal dapat diperoleh estimasi masing-masing karakteristik individu maupun karakteristik waktu (periode) secara terpisah. Berdasarkan uraian yang telah diberikan di atas, maka akan diteliti suatu model estimator campuran Spline *Truncated* dan Deret Fourier dalam regresi semiparametrik untuk data longitudinal yang merupakan pengembangan dari penelitian sebelumnya yaitu model estimator campuran Spline *Truncated* dan Deret Fourier dalam regresi semiparametrik untuk data *cross-section*. Penelitian ini diharapkan mampu mengatasi kelemahan dari penelitian sebelumnya, yaitu diperoleh model untuk masing-masing subjek. Gambaran posisi penelitian ini dibandingkan penelitian sebelumnya disajikan dalam gambar berikut.



Gambar 1.1 Posisi Penelitian Ini Dibandingkan Penelitian Sebelumnya

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah pada uraian sebelumnya, maka tujuan penulisan penelitian ini adalah sebagai berikut.

1. Memperoleh hasil kajian estimasi model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier.
2. Mendapatkan hasil estimasi model persentase kejadian pneumonia di Indonesia dengan pendekatan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier.

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Memberikan wawasan mengenai pemodelan kurva regresi semiparametrik campuran Spline Truncated dan Deret Fourier pada data longitudinal, sehingga dapat menjadi model alternatif untuk menganalisis persentase kejadian pneumonia.
2. Penelitian ini dapat menjadi masukan untuk pemerintah Republik Indonesia agar upaya pencegahan dan penanganan pneumonia akan lebih tepat.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan adalah persentase kejadian pneumonia di seluruh provinsi di Indonesia sejak tahun 2010-2017, kecuali Provinsi Kalimantan Utara, Papua, dan Papua Barat.
2. Metode yang digunakan dalam memilih titik knot optimal dan parameter osilasi dalam penelitian ini adalah metode *Generalized Cross Validation*.
3. Spline Truncated yang digunakan adalah Spline Truncated linier dengan satu titik knot.
4. Parameter osilasi yang digunakan dibatasi sampai tiga parameter osilasi.
5. Estimasi parameter diperoleh menggunakan *Weighted Least Square*.

BAB 2

TINJAUAN PUSTAKA

Pada bab ini akan dibahas mengenai beberapa tinjauan pustaka yang akan digunakan untuk membantu pengerjaan penelitian ini. Berdasarkan paparan pada bab 1 sebelumnya, maka tinjauan pustaka yang digunakan dalam penelitian ini adalah sebagai berikut.

2.1 Regresi Parametrik, Nonparametrik dan Semiparametrik

Analisis regresi digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon dan variabel prediktor melalui estimasi kurva. Terdapat tiga pendekatan dalam mengestimasi kurva regresi, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Menurut Gujarati dan Porter (2015) bentuk regresi parametrik secara umum adalah:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i; i = 1, 2, \dots, n$$

dimana n menunjukkan banyaknya observasi, y_i adalah variabel respon, $\beta_1, \beta_2, \dots, \beta_p$ sebagai parameter yang tidak diketahui dan ε_i adalah *error* random yang identik, independen, berdistribusi normal dengan mean nol dan variansi σ^2 .

Regresi nonparametrik merupakan pendekatan regresi yang sesuai untuk pola data yang tidak diketahui bentuk kurva regresinya, atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1999). Bentuk kurva regresi nonparametrik diasumsikan *smooth*, dalam arti kontinu dan diferensiabel. Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi, karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektivitas peneliti (Budiantara, 2009). Menurut Eubank (1999), jika diberikan data (t_i, y_i) , $i = 1, 2, \dots, n$ dan pola hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuknya, maka mengikuti model regresi nonparametrik sebagai berikut:

$$y_i = g(t_i) + \varepsilon_i, i = 1, 2, \dots, n$$

dengan y_i merupakan variabel respon, t_i merupakan variabel prediktor, $g(t_i)$ sebagai fungsi regresi yang bentuk kurvanya tidak diketahui dan ε_i adalah *error* random berdistribusi normal dengan mean nol dan varian σ^2 .

Regresi semiparametrik merupakan gabungan antara komponen parametrik dan komponen nonparametrik. Pendekatan regresi semiparametrik digunakan apabila bentuk kurva regresi sebagian diasumsikan diketahui dan sebagian lagi tidak diketahui. Misalkan terdapat data berpasangan (y_i, x_i, t_i) dengan $i = 1, 2, \dots, n$ dan hubungan antara y_i , x_i , dan t_i diasumsikan mengikuti model regresi semiparametrik sebagai berikut :

$$y_i = f(x_i) + g(t_i) + \varepsilon_i$$

Dengan y_i adalah variabel respon pada pengamatan ke- i , $f(x_i)$ adalah fungsi regresi parametrik, $g(t_i)$ adalah fungsi regresi nonparametrik dan ε_i adalah *error* random yang diasumsikan independen dan berdistribusi normal dengan mean nol dan varians σ^2 .

2.2 Regresi Semiparametrik Spline Truncated

Spline merupakan salah satu metode estimasi model regresi nonparametrik dan semiparametrik yang sering digunakan. *Spline* mempunyai sifat fleksibilitas yang tinggi dan memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Budiantara, 2014). Kelebihan ini dapat dilihat pada fungsi *truncated* (potongan-potongan) yang disebut dengan titik knot. Titik knot merupakan titik perpaduan bersama dimana fungsi tersebut berubah polanya pada sub-sub interval yang berlainan. Jika jumlah knot terlalu kecil, maka bias dalam estimator menjadi besar (Mohaisen dan Abdulhussein, 2015). Untuk knot yang sangat besar akan menghasilkan bentuk kurva regresi yang sangat halus. Sebaliknya untuk knot yang kecil akan memberikan bentuk kurva regresi yang sangat kasar (Wahba, 1990)

Misalkan diberikan data berpasangan (y_i, x_i, t_i) dengan $i = 1, 2, \dots, n$ dan hubungan antara y_i , x_i , dan t_i dapat dinyatakan dalam model regresi :

$$y_i = f(x_i) + g(t_i) + \varepsilon_i \quad (2.1)$$

dengan y_i adalah variabel respon pada pengamatan ke- i , $f(x_i)$ adalah fungsi regresi parametrik, $g(t_i)$ adalah fungsi regresi nonparametrik dan ε_i adalah *error* random yang diasumsikan independen dan berdistribusi normal dengan mean nol dan varians σ^2 . Fungsi regresi parametrik $f(x_i)$ dengan p merupakan banyaknya variabel prediktor dapat dinyatakan dalam bentuk persamaan sebagai berikut:

$$f(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad (2.2)$$

Sehingga, untuk $i = 1$; $y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_p x_{p1} + \varepsilon_1$

$$i = 2 ; y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_p x_{p2} + \varepsilon_2$$

⋮

$$i = n ; y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_p x_{pn} + \varepsilon_n$$

Persamaan di atas dapat dinyatakan dalam bentuk matriks sebagai berikut :

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

atau

$$\tilde{f}(x) = \mathbf{X} \tilde{\beta}, \quad (2.3)$$

dengan $\tilde{f}(x)$ adalah vektor dari fungsi f berukuran $n \times 1$, \mathbf{X} merupakan matriks berukuran $n \times (p+1)$ dan $\tilde{\beta}$ adalah vektor parameter yang akan diestimasi berukuran $(p+1) \times 1$.

Selanjutnya, kurva regresi $g(t_i)$ pada persamaan (2.1) dihampiri dengan fungsi Spline Truncated orde v dengan knot K_1, K_2, \dots, K_r . Secara umum, fungsi Spline Truncated dapat disajikan dalam bentuk :

$$g(t_i) = \sum_{k=1}^v \gamma_k t_i^k + \sum_{u=1}^r \gamma_{v+u} (t_i - K_u)_+^v \quad (2.4)$$

$\sum_{k=1}^v \gamma_k t_i^k$ merupakan komponen polinomial dan $\sum_{u=1}^r \gamma_{v+u} (t_i - K_u)_+^v$ merupakan komponen *truncated* dengan fungsi *truncated* sebagai berikut :

$$(t_i - K_u)_+^v = \begin{cases} (t_i - K_u)^v, & t_i \geq K_u \\ 0 & t_i < K_u \end{cases}$$

dimana γ_k dengan $k = 1, 2, \dots, v$ dan γ_{v+u} dengan $u = 1, 2, \dots, r$ merupakan parameter-parameter yang tidak diketahui, serta K_1, K_2, \dots, K_r merupakan titik-titik knot.

Berdasarkan model regresi komponen Spline pada persamaan (2.4), maka dapat dikembangkan bentuk model regresi semiparametrik Spline Truncated multivariabel. Jika variabel t_1, t_2, \dots, t_q bersifat aditif, maka fungsi multivariabel g adalah:

$$g(t_1, t_2, \dots, t_q) = \sum_{l=1}^q g_l(t_l) \quad (2.5)$$

Apabila fungsi $g_l(t_l)$ dihampiri dengan fungsi Spline Truncated orde v dengan titik-titik knot $K_{1l}, K_{2l}, \dots, K_{rl}$ dengan $l = 1, 2, \dots, q$ maka fungsi multivariabel $g(t_1, t_2, \dots, t_q)$ disajikan dalam bentuk sebagai berikut.

$$g(t_1, t_2, \dots, t_q) = \sum_{l=1}^q \left(\sum_{k=1}^v \gamma_{kl} t_{il}^k + \sum_{u=1}^r \gamma_{(v+u)l} (t_{il} - K_{ul})_+^v \right) \quad (2.6)$$

Dengan titik-titik knot K_1, K_2, \dots, K_r dan q merupakan banyaknya variabel prediktor pada persamaan, sehingga fungsi g merupakan model regresi nonparametrik Spline Truncated multivariabel seperti pada persamaan (2.6) sebelumnya.

Apabila dimisalkan nilai $v=1$ pada persamaan (2.6) maka disebut fungsi Spline Truncated linier multivariabel yang didefinisikan sebagai berikut :

$$g(t_1, t_2, \dots, t_q) = \sum_{l=1}^q \left(\gamma_l t_{il} + \sum_{u=1}^r \gamma_{(1+u)l} (t_{il} - K_{ul})_+ \right) \quad (2.7)$$

Dari persamaan (2.7) di atas, untuk $i=1, 2, \dots, n$ dapat diubah ke dalam bentuk persamaan vektor dan matriks sebagai berikut :

$$\tilde{g}(t) = \mathbf{T} \tilde{\gamma} \quad (2.8)$$

Dengan $\tilde{\gamma} = (\gamma_{11}, \gamma_{21}, \dots, \gamma_{(1+r)1}, \dots, \gamma_{1q}, \gamma_{2q}, \dots, \gamma_{(1+r)q})^T$

$$\mathbf{T} = \begin{pmatrix} t_{11} & (t_{11} - K_{11})_+ & \cdots & (t_{11} - K_{r1})_+ & \vdots & \cdots & \vdots & t_{1q} & (t_{1q} - K_{1q})_+ & \cdots & (t_{1q} - K_{rq})_+ \\ t_{21} & (t_{21} - K_{11})_+ & \cdots & (t_{21} - K_{r1})_+ & \vdots & \cdots & \vdots & t_{2q} & (t_{2q} - K_{1q})_+ & \cdots & (t_{2q} - K_{rq})_+ \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{n1} & (t_{n1} - K_{11})_+ & \cdots & (t_{n1} - K_{r1})_+ & \vdots & \cdots & \vdots & t_{nq} & (t_{nq} - K_{1q})_+ & \cdots & (t_{nq} - K_{rq})_+ \end{pmatrix}$$

Akibatnya didapat model regresi semiparametrik Spline Truncated :

$$\tilde{y} = \mathbf{X} \tilde{\beta} + \mathbf{T} \tilde{\gamma} + \tilde{\varepsilon} \quad (2.9)$$

Komponen $\mathbf{X} \tilde{\beta}$ disebut komponen parametrik, komponen $\mathbf{T} \tilde{\gamma}$ disebut komponen nonparametrik Spline Truncated dan *error* random $\tilde{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$.

2.3 Regresi Semiparametrik Deret Fourier

Deret Fourier merupakan polinomial trigonometri yang mempunyai fleksibilitas, sehingga dapat menyesuaikan diri secara efektif terhadap sifat lokal data. Estimator Deret Fourier dalam regresi nonparametrik maupun semiparametrik umumnya digunakan apabila data yang diselidiki polanya tidak diketahui dan ada kecenderungan pola berulang (Tripena dan Budiantara, 2006). Kemudian estimator Deret Fourier ini juga baik digunakan untuk menjelaskan kurva yang menunjukkan gelombang sinus dan cosinus (Asrini dan Budiantara, 2014). Estimator deret Fourier dalam regresi semiparametrik adalah estimator bias dan linier dalam pengamatan y (Pane dkk, 2014).

Misalkan diberikan data berpasangan (y_i, x_i, z_i) dengan $i = 1, 2, \dots, n$ dan hubungan antara y_i , x_i , dan z_i dapat dinyatakan dalam model regresi semiparametrik sebagai berikut :

$$y_i = f(x_i) + h(z_i) + \varepsilon_i \quad (2.10)$$

dengan y_i adalah variabel respon pada pengamatan ke- i dan $f(x_i)$ adalah fungsi regresi parametrik yang dapat dinyatakan seperti persamaan (2.2) dan dapat dinyatakan dalam bentuk matriks seperti persamaan (2.3). *Error* random ε_i diasumsikan independen dan berdistribusi normal dengan mean nol dan varians σ^2 . Sedangkan $h(z_i)$ adalah fungsi regresi nonparametrik yang bentuk kurva regresinya diasumsikan tidak diketahui dan termuat dalam ruang fungsi kontinu $C(0, \pi)$. Karena $h(z_i)$ kontinu pada interval $(0, \pi)$, maka dapat dihampiri oleh fungsi Deret Fourier T , dimana :

$$T(z_i) = \sum_{d=1}^s \left(b_d z_{id} + \frac{1}{2} a_0 + \sum_{m=1}^M a_{dm} \cos m z_{id} \right) \quad (2.11)$$

dimana b , a_0 , dan a_m dengan $m = 1, 2, \dots, M$ merupakan parameter-parameter model. Persamaan (2.11) di atas dapat dinyatakan dalam notasi matriks sebagai berikut:

$$\tilde{h}(z) = \mathbf{Z}\tilde{\alpha} \quad (2.12)$$

dengan $\tilde{\alpha} = (b_1 \ a_0 \ a_{11} \ \dots \ a_{1m} \ \dots \ b_s \ a_0 \ a_{s1} \ \dots \ a_{sM})^T$

$$\mathbf{Z} = \begin{pmatrix} z_{11} & \frac{1}{2} & \cos z_{11} & \dots & \cos m z_{11} & \vdots & \dots & \vdots & z_{1s} & \frac{1}{2} & \cos z_{1s} & \dots & \cos M z_{1s} \\ z_{21} & \frac{1}{2} & \cos z_{21} & \dots & \cos m z_{21} & \vdots & \dots & \vdots & z_{2s} & \frac{1}{2} & \cos z_{2s} & \dots & \cos M z_{2s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{n1} & \frac{1}{2} & \cos z_{n1} & \dots & \cos m z_{n1} & \vdots & \dots & \vdots & z_{ns} & \frac{1}{2} & \cos z_{ns} & \dots & \cos M z_{ns} \end{pmatrix}$$

Akibatnya diperoleh model regresi semiparametrik Deret Fourier :

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\alpha} + \tilde{\varepsilon} \quad (2.13)$$

Komponen $X\tilde{\beta}$ disebut komponen parametrik, komponen $Z\tilde{a}$ disebut komponen nonparametrik Deret Fourier dan *error* random $\tilde{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$.

2.4 Regresi Semiparametrik Campuran Spline Truncated dan Deret Fourier pada Data Cross Section

Regresi semiparametrik merupakan gabungan antara komponen parametrik dan komponen nonparametrik. Pada model regresi nonparametrik atau semiparametrik umumnya digunakan jenis metode estimasi yang sama untuk sebagian atau semua variabel prediktornya. Pada kenyataannya banyak kasus yang memiliki pola data berbeda pada masing-masing variabel prediktornya. Apabila hanya digunakan satu bentuk estimator untuk mengestimasi kurva regresi, akan mengakibatkan estimator yang diperoleh kurang sesuai dan menghasilkan estimasi model regresi yang tidak tepat serta cenderung memiliki error yang besar, sehingga untuk mengatasi masalah tersebut dikembangkan estimator kurva regresi campuran dimana masing-masing pola data dalam model regresi dihipotesiskan dengan estimator kurva yang sesuai.

Estimator Spline Truncated digunakan apabila pola data yang berubah pada sub-sub interval tertentu, sedangkan Deret Fourier digunakan apabila terdapat pola yang berulang, sehingga apabila pola hubungan antara variabel-variabel prediktor terhadap variabel respon memiliki pola yang berbeda-beda yaitu sebagian mengikuti pola tertentu, sebagian berubah pada sub-sub interval tertentu, dan sebagian yang lainnya mempunyai pola berulang, maka digunakan regresi semiparametrik dengan estimator campuran Spline Truncated dan Deret Fourier.

Menurut Nisa (2017), diberikan data berpasangan $(x_{i1}, x_{i2}, \dots, x_{ip}, t_{i1}, t_{i2}, \dots, t_{iq}, z_{i1}, z_{i2}, \dots, z_{is}, y_i)$ yang mengikuti model regresi semiparametrik sebagaimana persamaan (2.14).

$$y_i = \sum_{h=1}^p f(x_{ih}) + \sum_{l=1}^q g(t_{il}) + \sum_{d=1}^s h(z_{id}) + \varepsilon_i \quad (2.14)$$

dimana p adalah jumlah variabel komponen parametrik yang dihampiri dengan fungsi linier, q adalah jumlah variabel prediktor yang dihampiri dengan fungsi Spline *Truncated*, dan s adalah jumlah variabel prediktor yang dihampiri dengan fungsi Deret Fourier. Model (2.14) memuat n subjek yang dapat diuraikan sebagai berikut:

$$y_i = \sum_{h=1}^p \beta_h x_{ih} + \sum_{l=1}^q \left(\gamma_l t_{il} + \sum_{u=1}^r \gamma_{l(1+u)} (t_{il} - K_{ul})_+ \right) + \sum_{d=1}^s \left(b_d z_{id} + \frac{1}{2} a_0 + \sum_{m=1}^M a_{md} \cos m z_{id} \right) + \varepsilon_i$$

dimana y_i adalah variabel respon, $\sum_{h=1}^p f(x_{ih})$ adalah komponen parametrik yang

dihampiri dengan fungsi linier (2.3), $\sum_{l=1}^q g(t_{il})$ adalah komponen nonparametrik

Spline *Truncated* yang dihampiri dengan fungsi spline (2.8), serta $\sum_{d=1}^s h(z_{id})$

adalah komponen nonparametrik Deret Fourier yang dihampiri dengan fungsi Deret Fourier (2.12), dan ε_i adalah *error* random yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian σ^2 , sehingga diperoleh model regresi semiparametrik campuran Spline *Truncated* dan Deret Fourier pada data cross section sebagai berikut.

$$\tilde{y} = \mathbf{X} \tilde{\beta} + \mathbf{T} \tilde{\gamma} + \mathbf{Z} \tilde{a} + \tilde{\varepsilon} \quad (2.15)$$

Dengan komponen $\mathbf{X} \tilde{\beta}$ disebut komponen parametrik, komponen $\mathbf{T} \tilde{\gamma}$ disebut komponen nonparametrik Spline *Truncated*, komponen $\mathbf{Z} \tilde{a}$ disebut komponen nonparametrik Deret Fourier dan *error* random $\tilde{\varepsilon} \sim N(\tilde{0}, \sigma^2 \mathbf{I})$.

2.5 Weighted Least Square

Perbedaan metode *least square* antara data *cross-section* dan data longitudinal adalah adanya bobot (*weight*). Estimasi parameter pada data *cross-section* dapat diperoleh dengan menggunakan *Ordinary Least Square*. Sementara itu, data longitudinal menggunakan *Weighted Least Square* (WLS) untuk memperoleh estimasi parameter, dikarenakan metode WLS dapat mengatasi

korelasi dalam subjek pengamatan yang sama (Umam, 2018). Matriks \mathbf{W} merupakan matriks diagonal yang berisi pembobot untuk estimasi parameter.

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{W}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{W}_n \end{bmatrix} \text{ dimana } \mathbf{W}_i = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_i^2 \end{bmatrix} \quad (2.16)$$

dengan n adalah banyaknya subjek, dan t adalah banyaknya ulangan dalam subjek. Menurut Wu dan Zhang (2007), terdapat tiga metode dalam menentukan matriks pembobot yaitu:

1. $\mathbf{W}_i = N^{-1}\mathbf{I}$, $i = 1, 2, \dots, n$; dimana bobot ini memberikan perlakuan yang sama pada setiap pengamatan.
2. $\mathbf{W}_i = n^{-1}\mathbf{I}$, $i = 1, 2, \dots, n$; dimana bobot ini memberikan perlakuan yang sama pada setiap pengamatan dalam subjek.
3. $\mathbf{W}_i = \mathbf{V}_i^{-1}$ dimana $\mathbf{V}_i = \text{cov}(\mathbf{y}_i)$, $i = 1, 2, \dots, n$; ; bobot ini memperhitungkan korelasi dalam subjek pengamatan.

2.6 Definisi dan Teorema Dasar Aljabar Matriks

Berikut merupakan definisi dan teorema dasar terkait dengan aljabar matriks yang digunakan dalam penyelesaian estimasi parameter berdasarkan Rencher dan Scaalje (2007):

1. Transpose

$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ dengan syarat matriks \mathbf{A} dan \mathbf{B} masing-masing merupakan matriks yang memenuhi sifat perkalian..

2. Invers

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \text{ dan } (\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

3. Idempoten

Matriks \mathbf{A} dikatakan idempoten jika $\mathbf{A}^2 = \mathbf{A}$

4. Simetris

Matriks \mathbf{A} dikatakan simetris jika $\mathbf{A}^T = \mathbf{A}$

Teorema 1

Jika A adalah matriks berukuran $n \times p$ dan B adalah matriks berukuran $p \times n$ maka $tr(AB) = tr(BA)$

Teorema 2

Jika $u = \tilde{a}^T \tilde{x} = \tilde{x}^T \tilde{a}$, dengan $\tilde{a}^T = (a_1, a_2, \dots, a_p)$ adalah vektor konstanta, maka $\frac{\partial u}{\partial \tilde{x}} = \frac{\partial(\tilde{a}^T \tilde{x})}{\partial \tilde{x}} = \frac{\partial(\tilde{x}^T \tilde{a})}{\partial \tilde{x}} = \tilde{a}$.

Teorema 3

Jika $u = \tilde{x}^T A \tilde{x}$, dengan A adalah matrik simetri dari suatu konstanta, maka

$$\frac{\partial u}{\partial \tilde{x}} = \frac{\partial(\tilde{x}^T A \tilde{x})}{\partial \tilde{x}} = 2A\tilde{x}.$$

2.7 Data Longitudinal

Terdapat dua jenis data dalam analisis regresi, yaitu data longitudinal dan data *cross-section*. Data *cross-section* merupakan data yang hanya dilakukan satu kali pengamatan pada setiap subjek dan saling independen. Sedangkan data longitudinal merupakan data yang didapat dari pengamatan berulang dari setiap subjek pada interval waktu yang berbeda. Beberapa keuntungan studi mengenai data longitudinal antara lain dapat mengetahui perubahan individu, dan membutuhkan subjek yang tidak terlalu banyak karena pengamatan berulang (Wu & Zhang (2006). Data longitudinal juga lebih andal dalam mencari jawaban tentang dinamika perubahan (Nurdini, 2006). Inferensi pada model data longitudinal didasarkan pada data individu dengan asumsi masing-masing individu independen, tetapi dengan memperhatikan bahwa observasi berulang untuk tiap-tiap individu tidak independen (Wu & Zhang, 2006; Danardono, 2015)

Pada bidang ilmu ekonomi, dikenal istilah data panel. Menurut Danardono (2015), pada awalnya istilah data panel merujuk pada satu kelompok individu yang disurvei beberapa kali, sehingga interval pengamatan biasanya sama untuk setiap individu (*regular time*). Hal ini agak sedikit berbeda dengan data longitudinal yang waktu pengamatan berulangnya tidak harus sama. Pada data longitudinal, jumlah titik waktu ini dapat mencakup interval waktu yang

sama (*regularly spaced measurements*) ataupun tak sama (*irregularly spaced measurements*) (Harlan, 2018).

2.8 Generalized Cross Validation

Estimator spline sangat tergantung pada pemilihan titik knot K . Budiantara (2014) menjelaskan bahwa titik knot sendiri adalah titik perpaduan bersama dimana terdapat perubahan perilaku pola data pada interval yang berlainan. Sementara itu, Deret Fourier bergantung pada osilasi m . Osilasi m merupakan banyaknya osilasi dari gelombang cosinus pada model (Nisa, 2017). Salah satu metode yang dapat digunakan dalam memilih titik knot, dan osilasi m yang optimal adalah metode *Generalized Cross Validation* (GCV). Menurut Wahba (1990), metode yang sangat baik untuk memilih parameter penghalus optimal dalam estimator spline adalah *Generalized Cross Validation* (GCV). Hal ini dikarenakan metode GCV mempunyai sifat optimal asimtotik yang tidak dimiliki metode lain. Metode GCV juga memiliki kelebihan tidak membutuhkan pengetahuan terhadap variansi populasi serta metode GCV invarians terhadap transformasi (Nisa, 2017). Nilai knot yang paling optimum yaitu nilai knot dengan nilai GCV paling minimum. Metode GCV secara umum didefinisikan sebagai berikut.

$$GCV(K, m) = \frac{MSE(K, m)}{\left[n^{-1} \text{trace}(I - S(K, m)) \right]^2} \quad (2.17)$$

dengan $MSE(K, m) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

Estimator campuran Spline Truncated dan Deret Fourier untuk data longitudinal sangat bergantung pada banyaknya titik knot, lokasi titik knot, serta parameter osilasi. Menurut Prawanti (2019), metode *Generalized Cross Validation* (GCV) untuk data longitudinal adalah sebagaimana persamaan (2.18).

$$GCV(K, m) = \frac{(nJ)^{-1} \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \hat{y}_{ij})^2}{\left((nJ)^{-1} \text{tr}(\mathbf{I} - \mathbf{D}(K, m)) \right)^2} \quad (2.18)$$

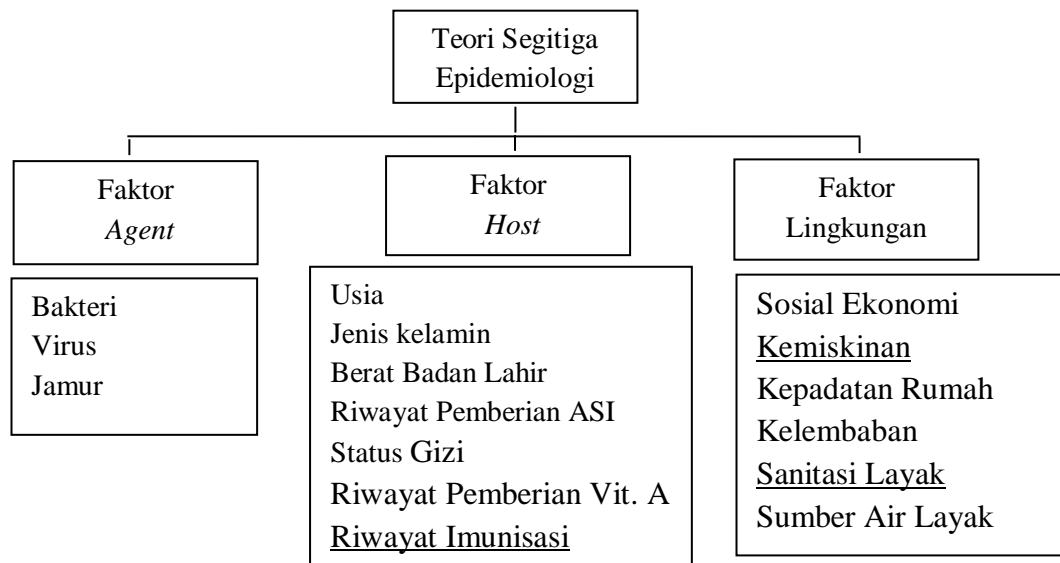
digunakan untuk mendapatkan titik knot dan parameter osilasi optimal

2.9 Penerapan Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline *Truncated* dan Deret Fourier pada Data Persentase Kejadian Pneumonia

Pneumonia adalah infeksi akut yang mengenai jaringan paru-paru (alveoli) yang dapat disebabkan oleh berbagai mikroorganisme seperti virus, jamur dan bakteri. Gejala penyakit pneumonia yaitu menggigil, demam, sakit kepala, batuk, mengeluarkan dahak dan sesak napas (Kementerian Kesehatan RI, 2017). Pemilihan faktor yang berpengaruh terhadap pneumonia didasarkan pada konsep model segitiga epidemiologi atau *triad* epidemiologi atau model rantai penyakit infeksi yang menggambarkan interaksi tiga komponen penyakit yaitu manusia (*Host*), penyebab (*Agent*) dan lingkungan (*Environment*). Sebuah penyakit dapat terjadi karena adanya ketidakseimbangan antara faktor *agent*, *host*, dan *environment* (Hockenberry dan Wilson, 2009). Faktor *agent* adalah faktor yang bertanggung jawab terhadap penyebab penyakit meliputi organisme penyebab penyakit atau *infectious agent*, *physical agent*, dan *chemical agent*. Faktor *host* adalah individu atau populasi yang berisiko menjadi penyebab penyakit meliputi faktor genetik atau gaya hidup. Faktor *environment* adalah tempat dimana *host* hidup termasuk kondisi cuaca dan faktor-faktor yang berhubungan dengan rumah, tetangga dan sekolah.

Hartati (2011) memaparkan hubungan tiga komponen yang terdapat dalam segitiga epidemiologi dengan faktor resiko terjadinya infeksi pneumonia pada balita. Berdasarkan faktor *agent* pneumonia dibedakan menjadi pneumonia bacterial, pneumonia atipikal, pneumonia virus, dan pneumonia jamur. Kemudian faktor risiko infeksi pneumonia pada pasien (*host*) meliputi usia, jenis kelamin, berat badan lahir, riwayat pemberian ASI, status gizi, riwayat pemberian vitamin A, riwayat imunisasi. Faktor lingkungan yang dapat menjadi risiko terjadinya pneumonia pada balita meliputi sosial ekonomi, kemiskinan, kepadatan rumah, kelembaban, cuaca, polusi udara, sanitasi layak dan sumber air layak. Dari beberapa faktor yang telah dipaparkan, dipilih tiga faktor yang diasumsikan mempengaruhi kejadian pneumonia pada penelitian ini.

Berikut adalah kerangka konsep pemilihan variabel dalam penelitian ini.



Gambar 2.1 Kerangka Konsep Pemilihan Variabel

Didefinisikan tiga faktor yang diasumsikan mempengaruhi kejadian pneumonia pada penelitian ini yaitu persentase cakupan imunisasi dasar pada bayi, persentase penduduk miskin, dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak.

- a. Faktor pertama yaitu persentase cakupan imunisasi dasar pada bayi. Imunisasi wajib untuk bayi di Indonesia dan diberikan secara gratis adalah Hepatitis B, Polio, BCG, campak dan pentavalen (DPT-HB-HiB). Imunisasi diberikan pada bayi dengan tujuan salah satunya adalah mencegah anak terserang penyakit infeksi menular. Balita yang telah mendapatkan imunisasi campak diharapkan dapat terhindar dari penyakit campak dan pneumonia yang merupakan komplikasi yang paling sering terjadi pada anak yang mengalami campak. Selain itu, imunisasi DPT juga dapat mencegah terjadinya penyakit difteri, pertusis, tetanus dan pneumonia yang merupakan komplikasi penyakit pertusis (Hartati, Nurhaeni dan Gayatri, 2012)
- b. Faktor kedua yaitu persentase penduduk miskin. Kemiskinan menyebabkan beberapa permasalahan yang cukup kompleks, misalnya pendidikan kurang, pemenuhan kebutuhan di bawah garis rata-rata, dan pola kehidupan yang masih rendah. Faktor sosio-ekonomi berkontribusi besar terhadap penyakit

saluran pernapasan akibat hubungan terbalik antara status sosial ekonomi dan morbiditas infeksi saluran pernafasan akut. Pada status sosio-ekonomi yang menurun, semua penyebab kematian dan seluruh angka kesakitan akan meningkat. Kemiskinan merupakan pangkal penyebab risiko pneumonia balita pada level rumah tangga yang lebih besar. Balita bergizi baik dan buruk jika berada dalam rumah tangga miskin berisiko lebih besar untuk menderita pneumonia (Machmud, 2009).

- c. Faktor ketiga yaitu persentase rumah tangga yang memiliki akses terhadap sanitasi layak. Akses terhadap sanitasi layak merupakan salah satu fondasi inti dari masyarakat yang sehat. Sanitasi yang baik merupakan elemen penting yang menunjang kesehatan manusia, sebab sanitasi berhubungan dengan kesehatan lingkungan yang mempengaruhi derajat kesehatan masyarakat. Buruknya kondisi sanitasi akan berdampak negatif di banyak aspek kehidupan, mulai dari turunnya kualitas lingkungan hidup masyarakat, tercemarnya sumber air minum bagi masyarakat, meningkatnya jumlah kejadian diare dan munculnya beberapa penyakit lainnya, salah satunya pneumonia (Kementerian Kesehatan RI, 2013).

BAB 3

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Kementerian Kesehatan Republik Indonesia (Kemenkes RI) yaitu Profil Kesehatan Indonesia dari tahun 2010 sampai tahun 2017. Data yang diambil dari Profil Kesehatan Indonesia meliputi persentase kejadian pneumonia pada 31 provinsi di Indonesia dan variabel-variabel yang diduga berpengaruh terhadap persentase kejadian pneumonia. Profil Kesehatan Indonesia merupakan salah satu media publikasi data dan informasi yang berisi situasi dan kondisi kesehatan di Indonesia yang cukup komprehensif. Profil Kesehatan Indonesia merupakan publikasi dari Kementerian kesehatan yang diterbitkan setiap tahun.

Menurut Kementerian Kesehatan RI (2014), metode pengumpulan data yang digunakan adalah metode rutin dan metode non-rutin. Pengumpulan data metode rutin dilakukan secara berkala, yang dikumpulkan dari catatan kegiatan harian atau rekam medik pasien. Pengumpulan data metode non-rutin adalah pengumpulan data sewaktu yang dilakukan melalui survey, dengan lingkup kabupaten/kota, provinsi atau nasional yang periodenya bisa tahunan, tiga tahunan atau lebih.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon (y) dan variabel prediktor (x). Penelitian ini menggunakan satu variabel respon dan tiga variabel prediktor. Variabel-variabel yang akan digunakan dalam penelitian ini tertera pada tabel 3.1 berikut :

Tabel 3.1 Variabel-Variabel dalam Penelitian

Komponen	Variabel	Keterangan
Faktor <i>Host</i>	x_1	Persentase cakupan imunisasi dasar pada bayi
Faktor Lingkungan	x_2	Persentase penduduk miskin
	x_3	Persentase rumah tangga yang memiliki akses terhadap sanitasi layak
-	y	Persentase kejadian pneumonia

Berikut merupakan definisi operasional dari variabel-variabel yang digunakan dalam penelitian ini :

- a. Variabel respon yang digunakan dalam penelitian ini adalah persentase kejadian pneumonia di masing-masing provinsi di Indonesia tahun 2010-2017. Persentase kejadian pneumonia yaitu jumlah penderita pneumonia balita usia 0-4 tahun dibagi dengan total populasi balita di wilayah tersebut dikali seratus persen (Kementerian Kesehatan RI, 2010).
- b. Variabel prediktor pertama yaitu persentase cakupan imunisasi dasar pada bayi. Persentase cakupan imunisasi dasar pada bayi yaitu jumlah balita yang memperoleh cakupan imunisasi dasar dibagi jumlah sasaran balita yang ditetapkan memperoleh imunisasi dasar dikali seratus persen.
- c. Variabel prediktor kedua yaitu persentase penduduk miskin. Kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Jadi Penduduk Miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan. Garis kemiskinan adalah penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM).
- d. Variabel prediktor ketiga yaitu persentase rumah tangga yang memiliki akses terhadap sanitasi layak. Persentase rumah tangga yang memiliki akses terhadap sanitasi layak yaitu jumlah rumah tangga yang memiliki akses sanitasi layak dibagi jumlah rumah tangga dikali seratus persen.

Unit observasi yang digunakan dalam penelitian ini adalah provinsi-provinsi di Indonesia dengan satu variabel respon, tiga variabel prediktor, dan lama pengamatan adalah 8 tahun yaitu sejak tahun 2010 hingga tahun 2017. Data penelitian tersebut disusun dalam sebuah struktur data yang disajikan pada tabel berikut.

Tabel 3.2 Struktur Data Penelitian

Subjek (<i>i</i>)	Waktu (<i>j</i>)	<i>y</i>	Variabel Komponen Parametrik				Variabel Komponen Spline <i>Truncated</i>				Variabel Komponen Deret Fourier			
			x_1	x_2	...	x_p	t_1	t_2	...	t_q	z_1	z_2	...	z_s
1	1	y_{11}	x_{111}	x_{121}	...	x_{1p1}	t_{111}	t_{121}	...	t_{1q1}	z_{111}	z_{121}	...	z_{1s1}
	2	y_{12}	x_{112}	x_{122}	...	x_{1p2}	t_{112}	t_{122}	...	t_{1q2}	z_{112}	z_{122}	...	z_{1s2}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
	<i>J</i>	y_{1J}	x_{11J}	x_{12J}	...	x_{1pJ}	t_{11J}	t_{12J}	...	t_{1qJ}	z_{11J}	z_{12J}	...	z_{1sJ}
⋮	⋮	⋮	⋮	⋮	⋮					⋮	⋮	⋮	⋮	
<i>N</i>	1	y_{n1}	x_{n11}	x_{n21}	...	x_{np1}	t_{n11}	t_{n21}	...	t_{nq1}	z_{n11}	z_{n21}	...	z_{ns1}
	2	y_{n2}	x_{n12}	x_{n22}	...	x_{np2}	t_{n12}	t_{n22}	...	t_{nq2}	z_{n12}	z_{n22}	...	z_{ns2}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
	<i>J</i>	y_{nJ}	x_{n1J}	x_{n2J}	...	x_{npJ}	t_{n1J}	t_{n2J}	...	t_{nqJ}	z_{n1J}	z_{n2J}	...	z_{nsJ}

3.3 Langkah-Langkah Penelitian

Untuk menyelesaikan tujuan penelitian, dibuat langkah-langkah yang disusun mengikuti tahapan-tahapan berikut.

A. Mengkaji estimasi kurva model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier dengan langkah-langkah sebagai berikut.

- (1). Diberikan data $(x_{i1j}, x_{i2j}, \dots, x_{ipj}, t_{i1j}, t_{i2j}, \dots, t_{iqj}, z_{i1j}, z_{i2j}, \dots, z_{isj}, y_{ij})$ yang mengikuti model regresi semiparametrik multivariabel untuk data longitudinal sebagai berikut.

$$y_{ij} = \sum_{h=1}^p f_{ihj}(x_{ihj}) + \sum_{l=1}^q g_{ilj}(t_{ilj}) + \sum_{d=1}^s h_{idj}(z_{idj}) + \varepsilon_{ij}$$

Dimana p adalah jumlah variabel komponen parametrik yang dihipotesiskan dengan fungsi linier, q adalah jumlah variabel prediktor yang dihipotesiskan dengan fungsi Spline Truncated, dan s adalah jumlah variabel prediktor yang dihipotesiskan dengan fungsi Deret Fourier. Kemudian, $i = 1, 2, \dots, n$ merupakan subjek yang diamati sebanyak n dan $j = 1, 2, \dots, J$ merupakan pengulangan sebanyak J .

- (2). Menghampiri kurva regresi komponen parametrik $f_{ihj}(x_{ihj})$ dimana $h = 1, 2, \dots, p$ dengan fungsi linier:

$$f_{ihj}(x_{ihj}) = \beta_0 + \sum_{h=1}^p \beta_{ih} x_{ihj}$$

- (3). Menghampiri kurva regresi komponen nonparametrik $g_{ilj}(t_{ilj})$ dimana $l = 1, 2, \dots, q$ menggunakan Spline Truncated linier dengan knot sebanyak r untuk masing-masing x_l yaitu,

$$g_{ilj}(t_{ilj}) = \gamma_{il} t_{ilj} + \sum_{u=1}^r \gamma_{(1+u)il} (t_{ilj} - K_{uil})_+$$

- (4). Menghampiri kurva regresi komponen nonparametrik $h_{idj}(z_{idj})$ dimana $d = 1, 2, \dots, s$ menggunakan Deret Fourier,

$$h_{idj}(z_{idj}) = b_{id} z_{idj} + \frac{1}{2} a_{0id} + \sum_{m=1}^M a_{mid} \cos m z_{idj}$$

- (5). Menuliskan model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier,

$$\begin{aligned} y_{ij} &= \sum_{h=1}^p f_{ihj}(x_{ihj}) + \sum_{l=1}^q g_{ilj}(t_{ilj}) + \sum_{d=1}^s h_{idj}(z_{idj}) + \varepsilon_{ij} \\ &= \sum_{h=1}^p \beta_{ih} x_{ihj} + \sum_{l=1}^q \left(\gamma_{il} t_{ilj} + \sum_{u=1}^r \gamma_{(1+u)il} (t_{ilj} - K_{uil})_+ \right) + \\ &\quad \sum_{d=1}^s \left(b_{id} z_{idj} + \frac{1}{2} a_{0id} + \sum_{m=1}^M a_{mid} \cos m z_{idj} \right) + \varepsilon_{ij} \end{aligned}$$

- (6). Menuliskan model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier dalam bentuk matriks,

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{T}\tilde{\gamma} + \mathbf{Z}\tilde{\alpha} + \tilde{\varepsilon}$$

- (7). Menulis *error* dari model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier dalam bentuk matriks,

$$\tilde{\varepsilon} = \tilde{y} - (\mathbf{X}\tilde{\beta} + \mathbf{T}\tilde{\gamma} + \mathbf{Z}\tilde{\alpha})$$

- (8). Mendapatkan estimasi untuk parameter melalui optimasi Weighted Least Square (WLS).

$$Q(\tilde{\beta}, \tilde{\gamma}, \tilde{\alpha}) = \underset{\tilde{\beta}, \tilde{\gamma}, \tilde{\alpha}}{\text{Min}} \left\{ (\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{T}\tilde{\gamma} - \mathbf{Z}\tilde{\alpha})^T \mathbf{W} (\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{T}\tilde{\gamma} - \mathbf{Z}\tilde{\alpha}) \right\}$$

- (9). Menyelesaikan optimasi WLS menggunakan derivatif parsial.

$$\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{\alpha})}{\partial \tilde{\beta}} = 0, \quad \frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{\alpha})}{\partial \tilde{\gamma}} = 0 \quad \text{dan} \quad \frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{\alpha})}{\partial \tilde{\alpha}} = 0$$

- (10). Mendapatkan estimator campuran Spline Truncated dan Deret Fourier.

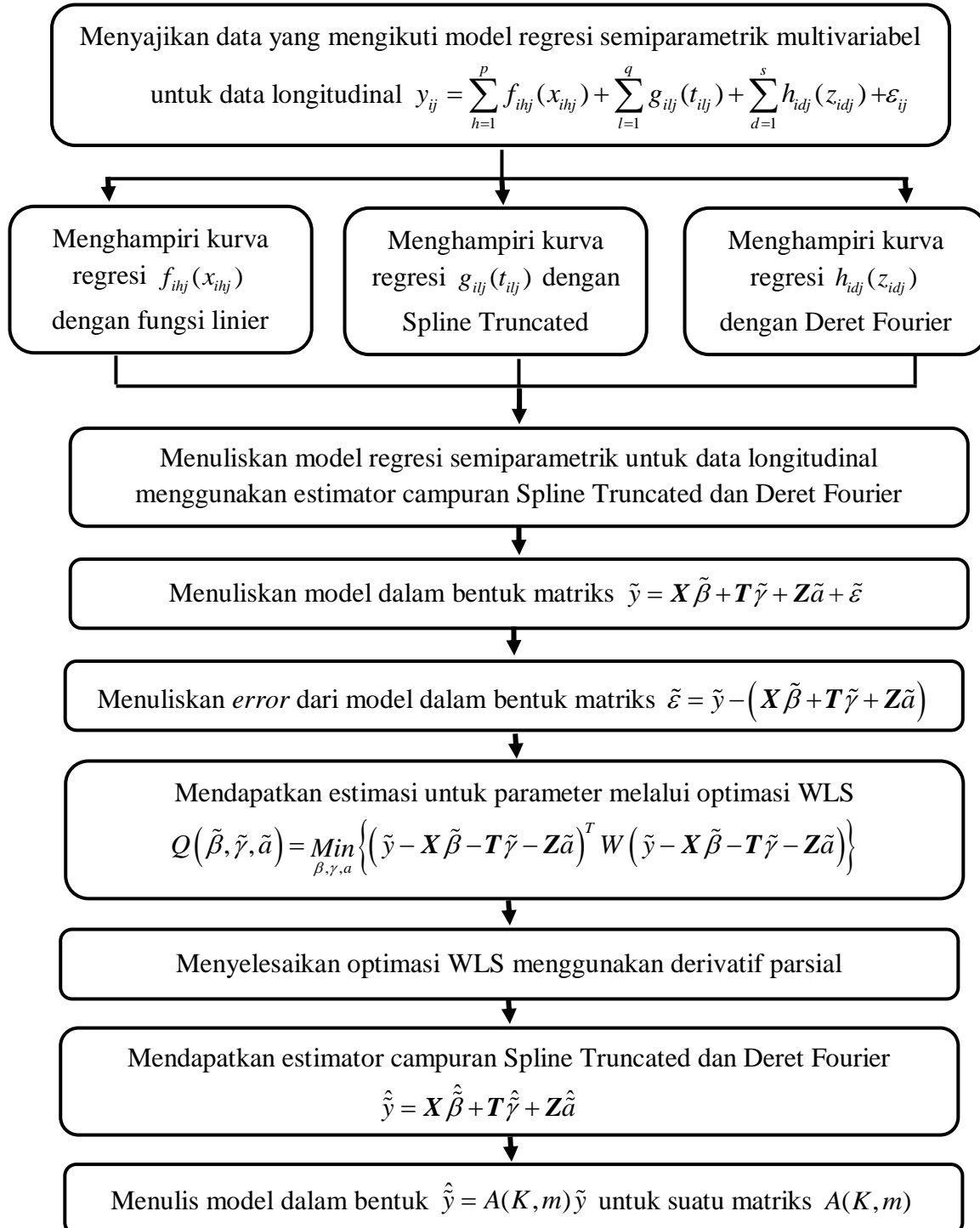
$$\hat{y} = \mathbf{X}\hat{\beta} + \mathbf{T}\hat{\gamma} + \mathbf{Z}\hat{\alpha}$$

- (11). Menulis model pada langkah (10) dalam bentuk,

$$\hat{y} = A(K, m)\tilde{y}$$

untuk suatu matriks $A(K, m)$.

Langkah-langkah analisis untuk tujuan penelitian yang pertama dapat digambarkan pada diagram alir berikut.

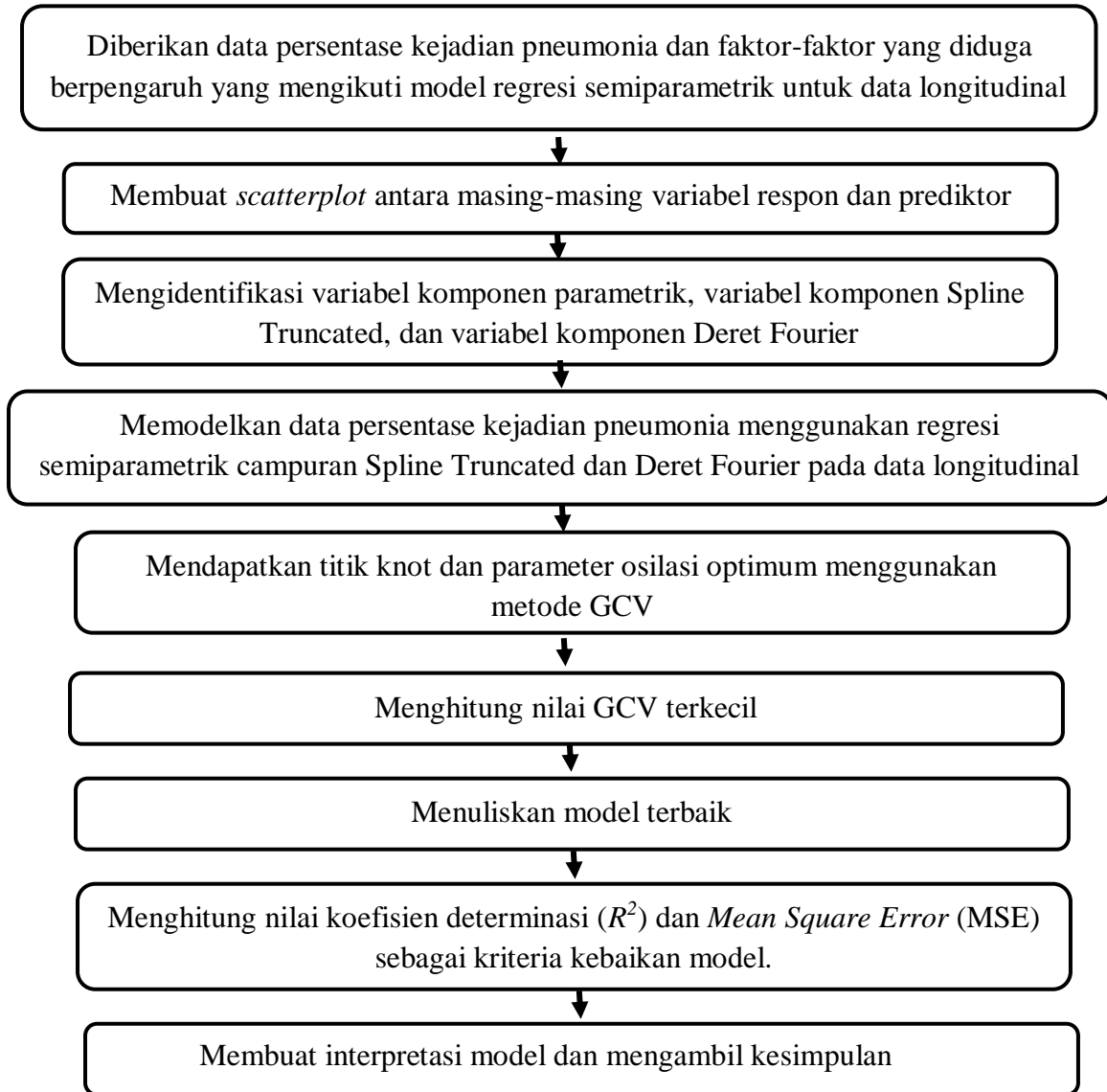


Gambar 3.1. Diagram Alir Estimasi Kurva Model Regresi Semiparametrik untuk Data Longitudinal dengan Estimator Campuran Spline Truncated dan Deret Fourier

B. Memodelkan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 dengan pendekatan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier dengan langkah-langkah sebagai berikut.

- (1). Melakukan eksplorasi data untuk mengetahui gambaran umum tentang data persentase kejadian pneumonia di Indonesia.
- (2). Membuat *scatterplot* antara masing-masing variabel respon dan variabel prediktor.
- (3). Mengidentifikasi variabel prediktor komponen parametrik, komponen nonparametrik Spline Truncated dan komponen nonparametrik Deret Fourier.
- (4). Memodelkan data persentase kejadian pneumonia dengan menggunakan estimator campuran Spline Truncated dan Deret Fourier dalam regresi semiparametrik pada data longitudinal.
- (5). Mendapatkan titik knot dan parameter osilasi optimum menggunakan metode *Generalized Cross Validation (GCV)*.
- (6). Menghitung nilai *Generalized Cross Validation GCV* terkecil dari langkah (4).
- (7). Menuliskan model terbaik dari regresi semiparametrik dengan estimator campuran Spline Truncated dan Deret Fourier pada data longitudinal.
- (8). Menghitung nilai koefisien determinasi (R^2) dan *Mean Square Error*.
- (9). Melakukan interpretasi model dengan pendekatan regresi semiparametrik menggunakan estimator campuran Spline Truncated dan Deret Fourier pada data longitudinal, serta mengambil kesimpulan.

Langkah-langkah analisis untuk tujuan penelitian yang kedua dapat digambarkan pada diagram alir berikut.



Gambar 3.2 Diagram Alir Pemodelan Persentase Kejadian Pneumonia dengan Regresi Semiparametrik Data Longitudinal Menggunakan Estimator Campuran Spline Truncated dan Deret Fourier

BAB 4

HASIL DAN PEMBAHASAN

Pada bab ini dilakukan kajian mengenai estimasi model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier. Selain itu, dikaji juga metode pemilihan titik knot optimal pada estimator Spline *Truncated* dan parameter osilasi optimal pada estimator Deret Fourier. Metode yang digunakan untuk mendapatkan estimasi fungsi regresi adalah dengan metode *Weighted Least Square* (WLS). Pemilihan parameter penghalus optimal menggunakan metode *Generalized Cross Validation* (GCV). Kemudian hasilnya akan diaplikasikan pada data kejadian pneumonia di Indonesia tahun 2010-2017.

4.1 Estimasi Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline *Truncated* dan Deret Fourier

Berikut ini dilakukan kajian mengenai estimasi model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier. Diberikan data berpasangan untuk pengamatan $i = 1, 2, \dots, n$ yaitu $(x_{i1j}, x_{i2j}, \dots, x_{ipj}, t_{i1j}, t_{i2j}, \dots, t_{iqj}, z_{i1j}, z_{i2j}, \dots, z_{isj}, y_{ij})$ yang mengikuti model regresi semiparametrik multivariabel untuk data longitudinal.

$$y_{ij} = \sum_{h=1}^p f_{ihj}(x_{ihj}) + \sum_{l=1}^q g_{ilj}(t_{ilj}) + \sum_{d=1}^s h_{idj}(z_{idj}) + \varepsilon_{ij} \quad (4.1)$$

dimana p adalah jumlah variabel komponen parametrik yang dihipotesiskan dengan fungsi linier, q adalah jumlah variabel prediktor yang dihipotesiskan dengan fungsi Spline *Truncated*, dan s adalah jumlah variabel prediktor yang dihipotesiskan dengan fungsi Deret Fourier. Model (4.1) memuat n subjek dengan subjek ke- i mempunyai J pengamatan, yang dapat diuraikan sebagai berikut.

$$y_{i1} = f(x_{i11}) + f(x_{i21}) + \dots + f(x_{ip1}) + g(t_{i11}) + g(t_{i21}) + \dots + g(t_{iq1}) + h(z_{i11}) + h(z_{i21}) + \dots + h(z_{is1}) + \varepsilon_{i1}$$

$$y_{i2} = f(x_{i12}) + f(x_{i22}) + \dots + f(x_{ip2}) + g(t_{i12}) + g(t_{i22}) + \dots + g(t_{iq2}) + h(z_{i12}) + h(z_{i22}) + \dots + h(z_{is2}) + \varepsilon_{i2}$$

⋮

$$y_{iJ} = f(x_{i1J}) + f(x_{i2J}) + \dots + f(x_{ipJ}) + g(t_{i1J}) + g(t_{i2J}) + \dots + g(t_{iqJ}) + h(z_{i1J}) + h(z_{i2J}) + \dots + h(z_{isJ}) + \varepsilon_{iJ}$$

Bentuk pola hubungan antara variabel respon y_{ij} dengan variabel prediktor $(x_{i1j}, x_{i2j}, \dots, x_{ipj})$ diasumsikan mengikuti pola linear. Kemudian pola hubungan antara variabel respon y_{ij} dengan variabel prediktor $(t_{i1j}, t_{i2j}, \dots, t_{iqj})$ diasumsikan berubah-ubah pada sub-sub interval tertentu, serta pola hubungan antara variabel respon y_{ij} dengan variabel prediktor $(z_{i1j}, z_{i2j}, \dots, z_{isj})$ diasumsikan memiliki pola berulang. Model persamaan (4.1) secara keseluruhan dapat didekati dengan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier. Model pada persamaan (4.1) dapat dijabarkan sebagai berikut.

$$y_{ij} = \beta_{0i} + \sum_{h=1}^p \beta_{ih} x_{ihj} + \sum_{l=1}^q \left(\gamma_{il} t_{lij} + \sum_{u=1}^r \gamma_{il(1+u)} (t_{lij} - K_{uil})_+ \right) + \sum_{d=1}^s \left(b_{id} z_{idj} + \frac{1}{2} a_{0id} + \sum_{m=1}^M a_{mid} \cos m z_{idj} \right) + \varepsilon_{ij} \quad (4.2)$$

Selanjutnya dapat dijabarkan sebagai berikut.

$$y_{11} = \beta_{01} + \beta_{11} x_{111} + \dots + \beta_{1p} x_{1p1} + \gamma_{111} t_{111} + \sum_{u=1}^r \gamma_{11(1+u)} (t_{111} - K_{u11})_+ + \dots + \gamma_{1q1} t_{1q1} + \sum_{u=1}^r \gamma_{1q(1+u)} (t_{1q1} - K_{u1q})_+ + b_{11} z_{111} + \frac{1}{2} a_{011} + \sum_{m=1}^M a_{m11} \cos m z_{111} + \dots + \sum_{m=1}^M a_{m1s} \cos m z_{1s1} + \varepsilon_{11}$$

$$y_{12} = \beta_{01} + \beta_{11} x_{112} + \dots + \beta_{1p} x_{1p2} + \gamma_{111} t_{112} + \sum_{u=1}^r \gamma_{11(1+u)} (t_{112} - K_{u11})_+ + \dots + \gamma_{1q1} t_{1q2} + \sum_{u=1}^r \gamma_{1q(1+u)} (t_{1q2} - K_{u1q})_+ + b_{11} z_{112} + \frac{1}{2} a_{011} + \sum_{m=1}^M a_{m11} \cos m z_{112} + \dots + \sum_{m=1}^M a_{m1s} \cos m z_{1s2} + \varepsilon_{12}$$

⋮

$$\begin{aligned}
y_{1J} &= \beta_{01} + \beta_{11}x_{11J} + \dots + \beta_{1p}x_{1pJ} + \gamma_{111}t_{11J} + \sum_{u=1}^r \gamma_{11(1+u)} (t_{11J} - K_{u11})_+ + \dots + \gamma_{1q1}t_{1qJ} + \\
&\quad \sum_{u=1}^r \gamma_{1q(1+u)} (t_{1qJ} - K_{u1q})_+ + b_{11}z_{11J} + \frac{1}{2}a_{011} + \sum_{m=1}^M a_{m11} \cos mz_{11J} + \dots + \sum_{m=1}^M a_{m1s} \cos mz_{1sJ} + \varepsilon_{1J} \\
&\quad \vdots \\
y_{n1} &= \beta_{0n} + \beta_{n1}x_{n11} + \dots + \beta_{np}x_{np1} + \gamma_{n11}t_{n11} + \sum_{u=1}^r \gamma_{n1(1+u)} (t_{n11} - K_{un1})_+ + \dots + \gamma_{nq1}t_{nq1} + \\
&\quad \sum_{u=1}^r \gamma_{nq(1+u)} (t_{nq1} - K_{unq})_+ + b_{n1}z_{n11} + \frac{1}{2}a_{0n1} + \sum_{m=1}^M a_{mn1} \cos mz_{n11} + \dots + \sum_{m=1}^M a_{mns} \cos mz_{ns1} + \varepsilon_{n1} \\
y_{n2} &= \beta_{0n} + \beta_{n1}x_{n12} + \dots + \beta_{np}x_{np2} + \gamma_{n11}t_{n12} + \sum_{u=1}^r \gamma_{n1(1+u)} (t_{n12} - K_{un1})_+ + \dots + \gamma_{nq1}t_{nq2} + \\
&\quad \sum_{u=1}^r \gamma_{nq(1+u)} (t_{nq2} - K_{unq})_+ + b_{n2}z_{n21} + \frac{1}{2}a_{0n2} + \sum_{m=1}^M a_{mn2} \cos mz_{n22} + \dots + \sum_{m=1}^M a_{mns} \cos mz_{ns2} + \varepsilon_{n2} \\
&\quad \vdots \\
y_{nJ} &= \beta_{0n} + \beta_{n1}x_{n1J} + \dots + \beta_{np}x_{npJ} + \gamma_{n11}t_{n1J} + \sum_{u=1}^r \gamma_{n1(1+u)} (t_{n1J} - K_{un1})_+ + \dots + \gamma_{nq1}t_{nqJ} + \\
&\quad \sum_{u=1}^r \gamma_{nq(1+u)} (t_{nqJ} - K_{unq})_+ + b_{n1}z_{n1J} + \frac{1}{2}a_{0n1} + \sum_{m=1}^M a_{mn1} \cos mz_{n1J} + \dots + \sum_{m=1}^M a_{mns} \cos mz_{nsJ} + \varepsilon_{nJ}
\end{aligned}$$

Model pada persamaan (4.2) dapat dinyatakan dalam bentuk matriks sebagaimana persamaan (4.3).

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{T}\tilde{\gamma} + \mathbf{Z}\tilde{a} + \tilde{\varepsilon} \quad (4.3)$$

dimana

$$\begin{aligned}
\tilde{y} &= \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix}_{nJ \times 1} ; \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_n \end{bmatrix}_{nJ \times n(p+1)} ; \tilde{\beta} = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_n \end{bmatrix}_{n(p+1) \times 1} ; \\
\mathbf{T} &= \begin{bmatrix} \mathbf{T}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{T}_n \end{bmatrix}_{nJ \times nq(1+r)} ; \tilde{\gamma} = \begin{bmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \\ \vdots \\ \tilde{\gamma}_n \end{bmatrix}_{nq(1+r) \times 1} ;
\end{aligned}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix}_{nJ \times ns(M+2)} ; \tilde{\mathbf{a}} = \begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_n \end{bmatrix}_{ns(M+2) \times 1} ; \tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{bmatrix}_{nJ \times 1}$$

Vektor dan matriks pada persamaan (4.3) diperoleh dari vektor dan matriks berikut untuk $i = 1, 2, \dots, n$

$$\tilde{\mathbf{y}}_i = [y_{i1} \quad y_{i2} \quad \cdots \quad y_{iJ}]^T$$

$$\tilde{\boldsymbol{\varepsilon}}_i = [\varepsilon_{i1} \quad \varepsilon_{i2} \quad \cdots \quad \varepsilon_{iJ}]^T$$

Berikut vektor-vektor penyusun vektor parameter $\tilde{\boldsymbol{\beta}}$ yaitu

$$\tilde{\boldsymbol{\beta}}_1 = [\beta_{01} \quad \beta_{11} \quad \beta_{12} \quad \cdots \quad \beta_{1p}]^T$$

$$\tilde{\boldsymbol{\beta}}_2 = [\beta_{02} \quad \beta_{21} \quad \beta_{22} \quad \cdots \quad \beta_{2p}]^T$$

\vdots

$$\tilde{\boldsymbol{\beta}}_n = [\beta_{0n} \quad \beta_{n1} \quad \beta_{n2} \quad \cdots \quad \beta_{np}]^T$$

dimana masing-masing $\tilde{\boldsymbol{\beta}}_i$ dengan $i = 1, 2, \dots, n$ berukuran $(p+1) \times 1$

Selanjutnya matriks-matriks penyusun matriks prediktor \mathbf{X} adalah sebagai berikut

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_{111} & x_{121} & \cdots & x_{1p1} \\ 1 & x_{112} & x_{122} & \cdots & x_{1p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11J} & x_{12J} & \cdots & x_{1pJ} \end{bmatrix}$$

$$\mathbf{X}_2 = \begin{bmatrix} 1 & x_{211} & x_{221} & \cdots & x_{2p1} \\ 1 & x_{212} & x_{222} & \cdots & x_{2p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{21J} & x_{22J} & \cdots & x_{2pJ} \end{bmatrix}$$

\vdots

$$\mathbf{X}_n = \begin{bmatrix} 1 & x_{n11} & x_{n21} & \cdots & x_{np1} \\ 1 & x_{n12} & x_{n22} & \cdots & x_{np2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1J} & x_{n2J} & \cdots & x_{npJ} \end{bmatrix}$$

dimana masing-masing \mathbf{X}_i dengan $i = 1, 2, \dots, n$ berukuran $J \times (p+1)$

Kemudian vektor-vektor penyusun vektor parameter $\tilde{\gamma}$ adalah

$$\begin{aligned}\tilde{\gamma}_1 &= [\gamma_{111} \quad \gamma_{112} \quad \cdots \quad \gamma_{1q(1+r)}]^T \\ \tilde{\gamma}_2 &= [\gamma_{211} \quad \gamma_{212} \quad \cdots \quad \gamma_{2q(1+r)}]^T \\ &\quad \vdots \\ \tilde{\gamma}_n &= [\gamma_{n11} \quad \gamma_{n12} \quad \cdots \quad \gamma_{nq(1+r)}]^T\end{aligned}$$

dimana masing-masing $\tilde{\gamma}_i$ dengan $i = 1, 2, \dots, n$ berukuran $q(1+r) \times 1$.

Matriks-matriks penyusun matriks prediktor T adalah sebagai berikut:

$$\begin{aligned}\mathbf{T}_1 &= \begin{bmatrix} t_{111} & (t_{111} - K_{111})_+ & (t_{111} - K_{211})_+ & \cdots & (t_{1q1} - K_{r1q})_+ \\ t_{112} & (t_{112} - K_{111})_+ & (t_{112} - K_{211})_+ & \cdots & (t_{1q2} - K_{r1q})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{11J} & (t_{11J} - K_{111})_+ & (t_{11J} - K_{211})_+ & \cdots & (t_{1qJ} - K_{r1q})_+ \end{bmatrix} \\ \mathbf{T}_2 &= \begin{bmatrix} t_{211} & (t_{211} - K_{121})_+ & (t_{211} - K_{221})_+ & \cdots & (t_{2q1} - K_{r2q})_+ \\ t_{212} & (t_{212} - K_{121})_+ & (t_{212} - K_{221})_+ & \cdots & (t_{2q2} - K_{r2q})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{21J} & (t_{21J} - K_{121})_+ & (t_{21J} - K_{221})_+ & \cdots & (t_{2qJ} - K_{r2q})_+ \end{bmatrix} \\ &\quad \vdots \\ \mathbf{T}_n &= \begin{bmatrix} t_{n11} & (t_{n11} - K_{1n1})_+ & (t_{n11} - K_{2n1})_+ & \cdots & (t_{nq1} - K_{rnq})_+ \\ t_{n12} & (t_{n12} - K_{1n1})_+ & (t_{n12} - K_{2n1})_+ & \cdots & (t_{nq2} - K_{rnq})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1J} & (t_{n1J} - K_{1n1})_+ & (t_{n1J} - K_{2n1})_+ & \cdots & (t_{nqJ} - K_{rnq})_+ \end{bmatrix}\end{aligned}$$

dimana masing-masing T_i dengan $i = 1, 2, \dots, n$ berukuran $J \times q(1+r)$

Vektor-vektor penyusun vektor parameter \tilde{a} adalah:

$$\begin{aligned}\tilde{a}_1 &= [b_{11} \quad a_{011} \quad a_{111} \quad \cdots \quad a_{M1s}]^T \\ \tilde{a}_2 &= [b_{21} \quad a_{021} \quad a_{121} \quad \cdots \quad a_{M2s}]^T \\ &\quad \vdots \\ \tilde{a}_n &= [b_{n1} \quad a_{0n1} \quad a_{1n1} \quad \cdots \quad a_{Mns}]^T\end{aligned}$$

dimana masing-masing \tilde{a}_i dengan $i = 1, 2, \dots, n$ berukuran $s(M+2) \times 1$

Dan matriks-matriks penyusun matriks prediktor \mathbf{Z} adalah sebagai berikut:

$$\mathbf{Z}_1 = \begin{bmatrix} z_{111} & \frac{1}{2} & \cos z_{111} & \cdots & \cos Mz_{1s1} \\ z_{112} & \frac{1}{2} & \cos z_{112} & \cdots & \cos Mz_{1s2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{11J} & \frac{1}{2} & \cos z_{11J} & \cdots & \cos Mz_{1sJ} \end{bmatrix}$$

$$\mathbf{Z}_2 = \begin{bmatrix} z_{211} & \frac{1}{2} & \cos z_{211} & \cdots & \cos Mz_{2s1} \\ z_{212} & \frac{1}{2} & \cos z_{212} & \cdots & \cos Mz_{2s2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{21J} & \frac{1}{2} & \cos z_{21J} & \cdots & \cos Mz_{2sJ} \\ \vdots & & & & \end{bmatrix}$$

$$\mathbf{Z}_n = \begin{bmatrix} z_{n11} & \frac{1}{2} & \cos z_{n11} & \cdots & \cos Mz_{ns1} \\ z_{n12} & \frac{1}{2} & \cos z_{n12} & \cdots & \cos Mz_{ns2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1J} & \frac{1}{2} & \cos z_{n1J} & \cdots & \cos Mz_{nsJ} \end{bmatrix}$$

dimana masing-masing \mathbf{Z}_i dengan $i = 1, 2, \dots, n$ berukuran $J \times s(M + 2)$

Selanjutnya dilakukan penyelesaian optimasi *Weighted Least Square* (WLS) untuk memperoleh estimasi parameter $\tilde{\beta}$, $\tilde{\gamma}$, dan \tilde{a} pada persamaan (4.3) dengan menggunakan matriks \mathbf{W} . Matriks \mathbf{W} merupakan matriks pembobot (matriks varians kovarians) berukuran $N \times N$ dimana $N = nJ \times nJ$ dan berisi diagonal $(\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n)$ sebagaimana persamaan (2.16).

Maka berdasarkan model pada persamaan (4.3) diperoleh :

$$\tilde{\varepsilon} = \tilde{\mathbf{y}} - \mathbf{X}\tilde{\beta} - \mathbf{T}\tilde{\gamma} - \mathbf{Z}\tilde{a}$$

Kemudian melalui optimasi *Weighted Least Square* (WLS) didapat:

$$\begin{aligned}
Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a}) &= \tilde{\varepsilon}^T \mathbf{W} \tilde{\varepsilon} \\
&= (\tilde{y} - \mathbf{X} \tilde{\beta} - \mathbf{T} \tilde{\gamma} - \mathbf{Z} \tilde{a})^T \mathbf{W} (\tilde{y} - \mathbf{X} \tilde{\beta} - \mathbf{T} \tilde{\gamma} - \mathbf{Z} \tilde{a}) \\
&= \tilde{y}^T \mathbf{W} \tilde{y} - 2 \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \tilde{y} - 2 \tilde{\gamma}^T \mathbf{T}^T \mathbf{W} \tilde{y} - 2 \tilde{a}^T \mathbf{Z}^T \mathbf{W} \tilde{y} + 2 \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + 2 \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a} + 2 \tilde{\gamma}^T \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a} \\
&\quad + \tilde{\beta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + \tilde{a}^T \mathbf{Z}^T \mathbf{W} \mathbf{Z} \tilde{a}
\end{aligned}$$

Selanjutnya derivatif parsial terhadap $\tilde{\beta}$ didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{\beta}} = -2 \mathbf{X}^T \mathbf{W} \tilde{y} + 2 \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + 2 \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a} + 2 \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} \quad (4.4)$$

Derivatif parsial terhadap $\tilde{\gamma}$ didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{\gamma}} = -2 \mathbf{T}^T \mathbf{W} \tilde{y} + 2 \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} + 2 \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a} + 2 \mathbf{T}^T \mathbf{W} \mathbf{T} \tilde{\gamma} \quad (4.5)$$

Derivatif parsial terhadap \tilde{a} didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{a}} = -2 \mathbf{Z}^T \mathbf{W} \tilde{y} + 2 \mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} + 2 \mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + 2 \mathbf{Z}^T \mathbf{W} \mathbf{Z} \tilde{a} \quad (4.6)$$

Hasil derivatif-derivatif yang diperoleh pada persamaan (4.4), (4.5) dan (4.6) disamadengankan nol, sebagai berikut

$$\begin{aligned}
\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{\beta}} &= 0 \\
-2 \mathbf{X}^T \mathbf{W} \tilde{y} + 2 \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + 2 \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a} + 2 \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} &= 0 \\
-\mathbf{X}^T \mathbf{W} \tilde{y} + \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a} + \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} &= 0 \\
\mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\beta} &= \mathbf{X}^T \mathbf{W} \tilde{y} - \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} - \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a} \\
\hat{\tilde{\beta}} &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \tilde{y} - \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} - \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a}) \\
&= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \tilde{y} - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{T} \tilde{\gamma} - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z} \tilde{a}
\end{aligned} \quad (4.7)$$

Selanjutnya untuk $\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{\gamma}} = 0$ dapat dijabarkan sebagai berikut :

$$\begin{aligned}
-2 \mathbf{T}^T \mathbf{W} \tilde{y} + 2 \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} + 2 \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a} + 2 \mathbf{T}^T \mathbf{W} \mathbf{T} \tilde{\gamma} &= 0 \\
-\mathbf{T}^T \mathbf{W} \tilde{y} + \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} + \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a} + \mathbf{T}^T \mathbf{W} \mathbf{T} \tilde{\gamma} &= 0 \\
\mathbf{T}^T \mathbf{W} \mathbf{T} \tilde{\gamma} &= \mathbf{T}^T \mathbf{W} \tilde{y} - \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} - \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a}
\end{aligned}$$

$$\begin{aligned}\hat{\gamma} &= (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} (\mathbf{T}^T \mathbf{W} \tilde{\mathbf{y}} - \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} - \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a}) \\ &= (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{W} \tilde{\mathbf{y}} - (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{W} \mathbf{X} \tilde{\beta} - (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{W} \mathbf{Z} \tilde{a}\end{aligned}\quad (4.8)$$

Kemudian untuk $\frac{\partial Q(\tilde{\beta}, \tilde{\gamma}, \tilde{a})}{\partial \tilde{a}} = 0$ dapat dijabarkan sebagai berikut:

$$\begin{aligned}-2\mathbf{Z}^T \mathbf{W} \tilde{\mathbf{y}} + 2\mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} + 2\mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + 2\mathbf{Z}^T \mathbf{W} \mathbf{Z} \tilde{a} &= 0 \\ -\mathbf{Z}^T \mathbf{W} \tilde{\mathbf{y}} + \mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} + \mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma} + \mathbf{Z}^T \mathbf{W} \mathbf{Z} \tilde{a} &= 0 \\ \mathbf{Z}^T \mathbf{W} \mathbf{Z} \tilde{a} &= \mathbf{Z}^T \mathbf{W} \tilde{\mathbf{y}} - \mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} - \mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma} \\ \hat{a} &= (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{W} \tilde{\mathbf{y}} + \mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} + \mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma}) \\ &= (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \tilde{\mathbf{y}} - (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{X} \tilde{\beta} - (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{T} \tilde{\gamma}\end{aligned}\quad (4.9)$$

Berdasarkan hasil yang diperoleh pada persamaan (4.7), (4.8), dan (4.9) terlihat bahwa $\tilde{\beta}$, $\tilde{\gamma}$, dan \tilde{a} masih mengandung parameter, maka untuk menyelesaikan permasalahan tersebut digunakan metode eliminasi dan substitusi. Untuk memudahkan perhitungan, maka persamaan (4.7), (4.8), dan (4.9) dapat dituliskan sebagai berikut :

$$\hat{\beta} = \mathbf{P} \tilde{\mathbf{y}} - \mathbf{P} \mathbf{T} \tilde{\gamma} - \mathbf{P} \mathbf{Z} \tilde{a} \quad (4.10)$$

dengan $\mathbf{P} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$

$$\hat{\gamma} = \mathbf{R} \tilde{\mathbf{y}} - \mathbf{R} \mathbf{X} \tilde{\beta} - \mathbf{R} \mathbf{Z} \tilde{a} \quad (4.11)$$

dengan $\mathbf{R} = (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{W}$

$$\hat{a} = \mathbf{U} \tilde{\mathbf{y}} - \mathbf{U} \mathbf{X} \tilde{\beta} - \mathbf{U} \mathbf{T} \tilde{\gamma} \quad (4.12)$$

dengan $\mathbf{U} = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W}$

Langkah pertama adalah menggunakan metode eliminasi untuk mendapatkan persamaan yang memuat dua parameter. Melakukan metode eliminasi pada persamaan (4.10) dan (4.12) dengan cara mengalikan kedua ruas pada persamaan (4.12) dengan $\mathbf{P} \mathbf{Z}$. Selanjutnya persamaan (4.10) dikurangi dengan hasil perkalian $\mathbf{P} \mathbf{Z}$ dengan persamaan (4.12), sehingga diperoleh:

$$\begin{aligned}\hat{\beta} - \mathbf{P} \mathbf{Z} \mathbf{U} \mathbf{X} \hat{\beta} + \mathbf{P} \mathbf{T} \hat{\gamma} - \mathbf{P} \mathbf{Z} \mathbf{U} \mathbf{T} \hat{\gamma} &= \mathbf{P} \tilde{\mathbf{y}} - \mathbf{P} \mathbf{Z} \mathbf{U} \tilde{\mathbf{y}} \\ (\mathbf{I} - \mathbf{P} \mathbf{Z} \mathbf{U} \mathbf{X}) \hat{\beta} + (\mathbf{P} \mathbf{T} - \mathbf{P} \mathbf{Z} \mathbf{U} \mathbf{T}) \hat{\gamma} &= (\mathbf{P} - \mathbf{P} \mathbf{Z} \mathbf{U}) \tilde{\mathbf{y}}\end{aligned}\quad (4.13)$$

Langkah kedua adalah melakukan metode eliminasi pada persamaan (4.11) dan (4.12) dengan cara mengalikan persamaan (4.12) dengan \mathbf{RZ} . Selanjutnya persamaan (4.11) dikurangi dengan hasil perkalian \mathbf{RZ} dengan persamaan (4.12), sehingga diperoleh:

$$\begin{aligned} \hat{\gamma} - \mathbf{RZUT}\hat{\gamma} + \mathbf{RX}\tilde{\beta} - \mathbf{RZUX}\tilde{\beta} &= \mathbf{R}\tilde{y} - \mathbf{RZU}\tilde{y} \\ (\mathbf{I} - \mathbf{RZUT})\hat{\gamma} + (\mathbf{RX} - \mathbf{RZUX})\tilde{\beta} &= (\mathbf{R} - \mathbf{RZU})\tilde{y} \end{aligned} \quad (4.14)$$

Berdasarkan hasil dari dua metode eliminasi yang telah dilakukan, didapatkan persamaan (4.13) dan (4.14) yang mana hanya terdapat dua parameter pada kedua persamaan tersebut. Selanjutnya melakukan metode eliminasi pada persamaan (4.13) dan (4.14) dengan cara mengalikan persamaan (4.14) dengan $(\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}$. Kemudian persamaan (4.13) dikurangi dengan hasil perkalian $(\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}$ dengan persamaan (4.14), sehingga diperoleh,

$$\begin{aligned} (\mathbf{I} - \mathbf{PZUX})\hat{\beta} - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{RX} - \mathbf{RZUX})\hat{\beta} &= \\ (\mathbf{P} - \mathbf{PZU})\tilde{y} - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{R} - \mathbf{RZU})\tilde{y} \end{aligned}$$

Sehingga dari persamaan diatas, diperoleh:

$$\begin{aligned} \hat{\beta} &= \left[(\mathbf{I} - \mathbf{PZUX}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{RX} - \mathbf{RZUX}) \right]^{-1} \\ &\quad \left[(\mathbf{P} - \mathbf{PZU}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{R} - \mathbf{RZU}) \right] \tilde{y} \\ &= \mathbf{A}(K, m)\tilde{y} \end{aligned} \quad (4.15)$$

dengan

$$\mathbf{A}(K, m) = \left[(\mathbf{I} - \mathbf{PZUX}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{RX} - \mathbf{RZUX}) \right]^{-1} \left[(\mathbf{P} - \mathbf{PZU}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{R} - \mathbf{RZU}) \right]$$

Kemudian mensubstitusi persamaan (4.15) pada persamaan (4.13), sehingga diperoleh:

$$\begin{aligned} (\mathbf{I} - \mathbf{PZUX})\mathbf{A}(K, m)\tilde{y} + (\mathbf{PT} - \mathbf{PZUT})\hat{\gamma} &= (\mathbf{P} - \mathbf{PZU})\tilde{y} \\ (\mathbf{PT} - \mathbf{PZUT})\hat{\gamma} &= (\mathbf{P} - \mathbf{PZU})\tilde{y} - (\mathbf{I} - \mathbf{PZUX})\mathbf{A}(K, m)\tilde{y} \end{aligned}$$

$$\begin{aligned}\hat{\gamma} &= (\mathbf{PT} - \mathbf{PZUT})^{-1} [(\mathbf{P} - \mathbf{PZU}) - (\mathbf{I} - \mathbf{PZUX})\mathbf{A}(K, m)] \tilde{y} \\ &= \mathbf{B}(K, m) \tilde{y}\end{aligned}\quad (4.16)$$

dengan

$$\mathbf{B}(K, m) = (\mathbf{PT} - \mathbf{PZUT})^{-1} [(\mathbf{P} - \mathbf{PZU}) - (\mathbf{I} - \mathbf{PZUX})\mathbf{A}(K, m)]$$

Selanjutnya mensubstitusikan persamaan (4.15) dan (4.16) pada persamaan (4.12), sehingga diperoleh:

$$\begin{aligned}\hat{a} &= \mathbf{U}\tilde{y} - \mathbf{UXA}\tilde{y} - \mathbf{UT} [(\mathbf{PT} - \mathbf{PZUT})^{-1} [(\mathbf{P} - \mathbf{PZU}) - (\mathbf{I} - \mathbf{PZUX})\mathbf{A}] \tilde{y}] \\ &= [\mathbf{U} - \mathbf{UXA}(K, m) - \mathbf{UT}(\mathbf{B}(K, m))] \tilde{y} \\ &= \mathbf{C}(K, m) \tilde{y}\end{aligned}\quad (4.17)$$

dengan

$$\mathbf{C}(K, m) = [\mathbf{U} - \mathbf{UXA}(K, m) - \mathbf{UT}(\mathbf{B}(K, m))]$$

Berdasarkan persamaan (4.15), (4.16) dan (4.17), diperoleh estimasi model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier sebagai berikut:

$$\begin{aligned}\hat{y} &= \mathbf{X}\hat{\beta} + \mathbf{T}\hat{\gamma} + \mathbf{Z}\hat{a} \\ &= \mathbf{XA}(K, m) \tilde{y} + \mathbf{TB}(K, m) \tilde{y} + \mathbf{ZC}(K, m) \tilde{y} \\ &= \mathbf{D}(K, m) \tilde{y}\end{aligned}\quad (4.18)$$

dengan $\mathbf{D}(K, m) = \mathbf{XA}(K, m) + \mathbf{TB}(K, m) + \mathbf{ZC}(K, m)$

Berdasarkan ulasan yang telah diberikan, tujuan pertama pada penelitian ini telah tercapai.

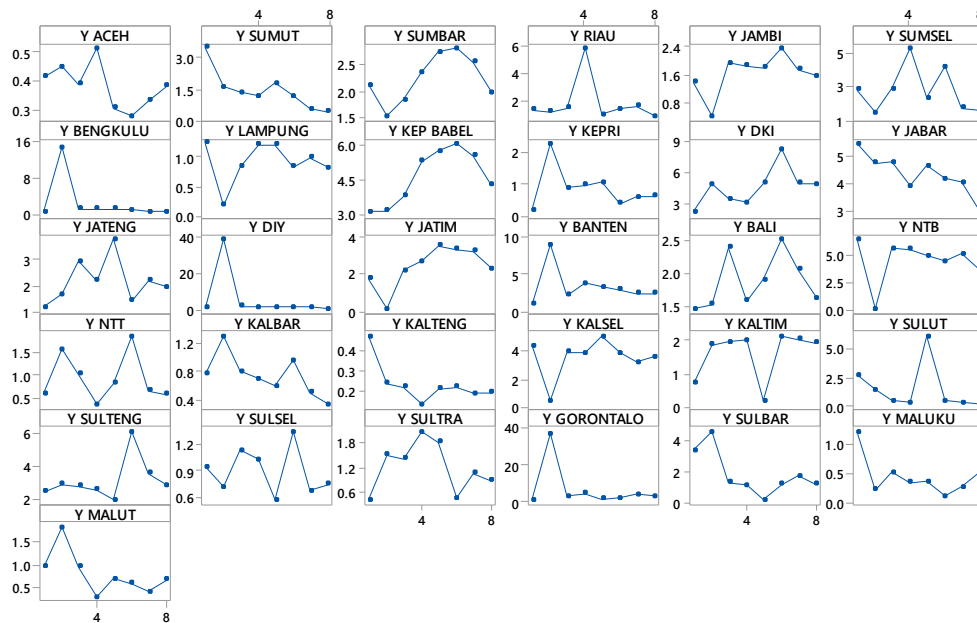
4.2 Pemodelan Persentase Kejadian Pneumonia di Indonesia dengan Pendekatan Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline Truncated dan Deret Fourier

Pada subbab ini akan dibahas mengenai penerapan estimasi model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier pada data persentase kejadian pneumonia tahun 2010-2017 di Indonesia. Unit observasi yang digunakan dalam penelitian ini adalah 31 provinsi di Indonesia. Variabel respon yang digunakan adalah persentase kejadian pneumonia di masing-masing provinsi di Indonesia tahun 2010-2017 (y). sedangkan variabel prediktor yang digunakan sebanyak tiga variabel yaitu persentase cakupan imunisasi dasar pada bayi (x_1), persentase penduduk miskin (x_2), dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3).

4.2.1 Deskripsi Variabel Penelitian

Pneumonia adalah infeksi akut yang mengenai jaringan paru-paru (alveoli) yang dapat disebabkan oleh berbagai mikroorganisme seperti virus, jamur, dan bakteri (Kementerian Kesehatan RI, 2017). Pada tahun 2018, angka kematian akibat pneumonia pada balita sebesar 0.08%. Angka kematian akibat pneumonia pada kelompok bayi lebih tinggi yaitu sebesar 0.16% dibandingkan pada kelompok anak umur 1-4 tahun sebesar 0.05% (Kementerian Kesehatan, 2018).

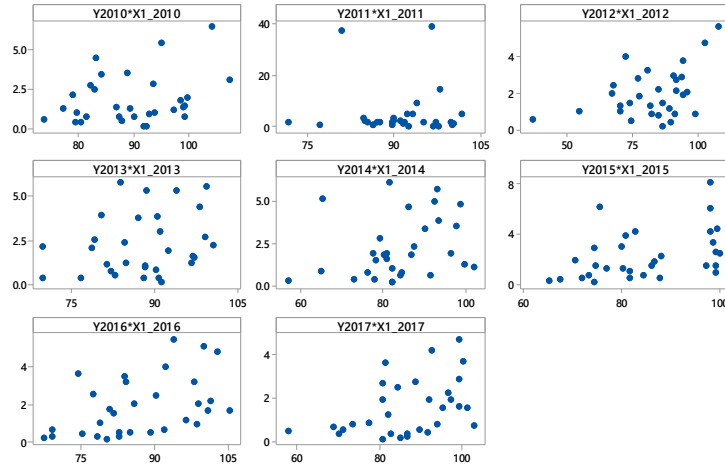
Selengkapnya data persentase kejadian pneumonia dan variabel-variabel yang diduga mempengaruhinya selama periode tahun 2010-2017 di masing-masing provinsi di Indonesia dapat dilihat pada Lampiran 1. Gambaran persentase kejadian pneumonia di masing-masing provinsi di Indonesia tahun 2010-2017 disajikan pada Gambar 4.1. Berdasarkan Gambar 4.1, terlihat masing-masing provinsi di Indonesia mengalami kenaikan atau penurunan yang tidak menentu. Akan tetapi, apabila dilihat dua sampai tiga tahun terakhir, beberapa provinsi mengalami penurunan persentasi kejadian pneumonia. Beberapa provinsi yang signifikan mengalami penurunan selama periode tahun 2010-2017 adalah provinsi Sumatera Utara, Jawa Barat, dan Kalimantan Barat.



Gambar 4.1 Persentase Kejadian Pneumonia di Indonesia Tahun 2010-2017

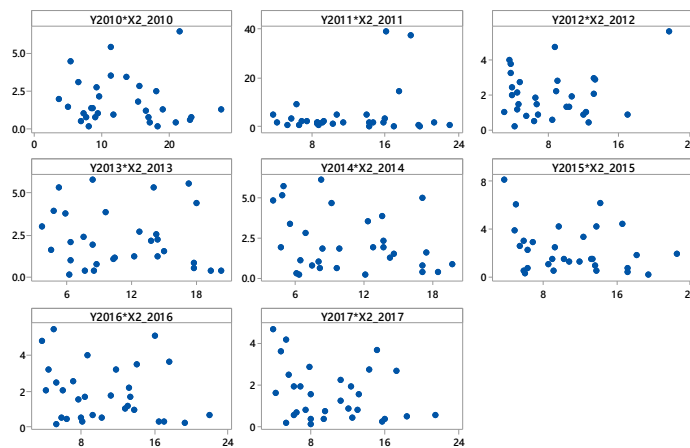
Tahap awal dalam pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 dengan pendekatan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier adalah mengetahui bentuk pola hubungan antara variabel respon dan masing-masing variabel prediktornya. Informasi mengenai bentuk pola hubungan variabel respon dan setiap variabel prediktor dapat diperoleh dengan membuat *scatterplot* antara variabel respon dan masing-masing variabel prediktornya.

Hasil *scatterplot* digunakan untuk mengidentifikasi variabel prediktor komponen parametrik, komponen nonparametrik Spline Truncated dan komponen nonparametrik Deret Fourier. Penelitian ini menggunakan data longitudinal, sehingga dibuat *scatterplot* dengan dua jenis data yaitu *scatterplot* berdasarkan data *cross-section* dan *scatterplot* berdasarkan data *time series*. Berikut merupakan hasil *scatterplot* berdasarkan data *cross-section* antara variabel persentase kejadian pneumonia (y) dan persentase cakupan imunisasi dasar pada bayi (x_1).



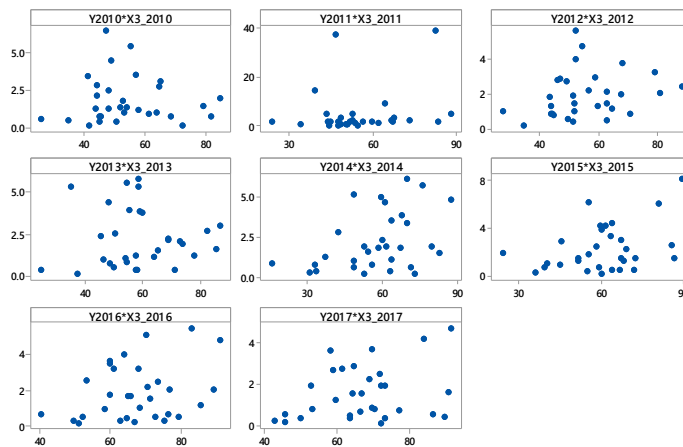
Gambar 4.2 *Scatterplot* data *cross-section* antara persentase kejadian pneumonia dan persentase cakupan imunisasi dasar pada bayi di Indonesia tahun 2010-2017

Gambar 4.2 menggambarkan pola hubungan antara persentase kejadian pneumonia (y) dan persentase cakupan imunisasi dasar pada bayi (x_1) membentuk pola linier. Berdasarkan hasil *scatterplot* pada Gambar 4.6, maka variabel persentase cakupan imunisasi dasar pada bayi (x_1) dapat didekati dengan fungsi linier. Selanjutnya dibuat pula *scatterplot* antara variabel persentase kejadian pneumonia (y) dan persentase penduduk miskin (x_2) berdasarkan data *cross-section* yang disajikan pada Gambar 4.3 berikut.



Gambar 4.3 *Scatterplot* data *cross-section* antara persentase kejadian pneumonia dan persentase penduduk miskin di Indonesia tahun 2010-2017

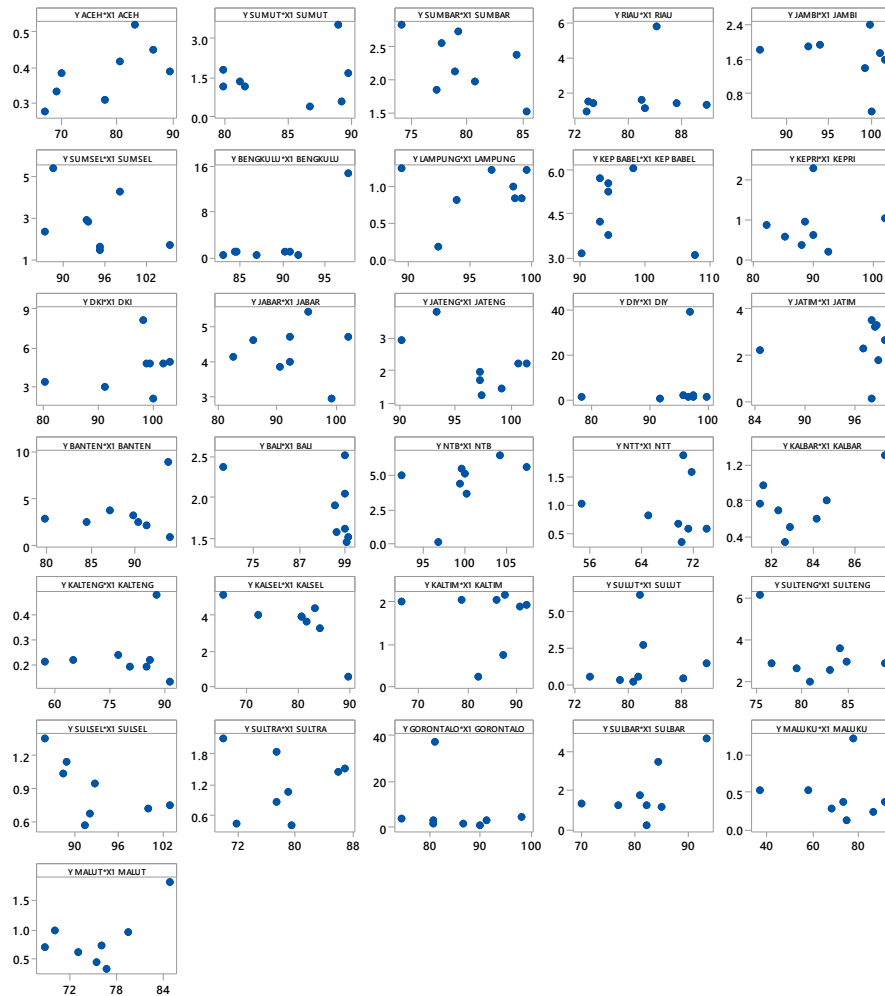
Gambar 4.3 menunjukkan bahwa pola hubungan antara persentase kejadian pneumonia (y) dan persentase penduduk miskin (x_2) tidak diketahui bentuk pola hubungannya, sehingga dapat dimodelkan secara nonparametrik. Bentuk pola hubungan antara persentase kejadian pneumonia (y) dan persentase penduduk miskin (x_2) cenderung mengalami perubahan perilaku pola data yang berulang, sehingga dapat didekati dengan fungsi Deret Fourier. Selanjutnya dibuat pula *scatterplot* antara variabel persentase kejadian pneumonia (y) dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3). Berikut merupakan hasil *scatterplot* berdasarkan data *cross-section* antara variabel persentase kejadian pneumonia (y) dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3).



Gambar 4.4 *Scatterplot* data *cross-section* antara persentase kejadian pneumonia dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Indonesia tahun 2010-2017

Gambar 4.4 menunjukkan bahwa pola hubungan antara persentase kejadian pneumonia (y) dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3) tidak diketahui bentuk pola hubungannya, sehingga dapat dimodelkan secara nonparametrik. Bentuk pola hubungan antara persentase kejadian pneumonia (y) dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3) cenderung mengalami perubahan perilaku pada sub interval tertentu, sehingga dapat didekati dengan fungsi Spline Truncated.

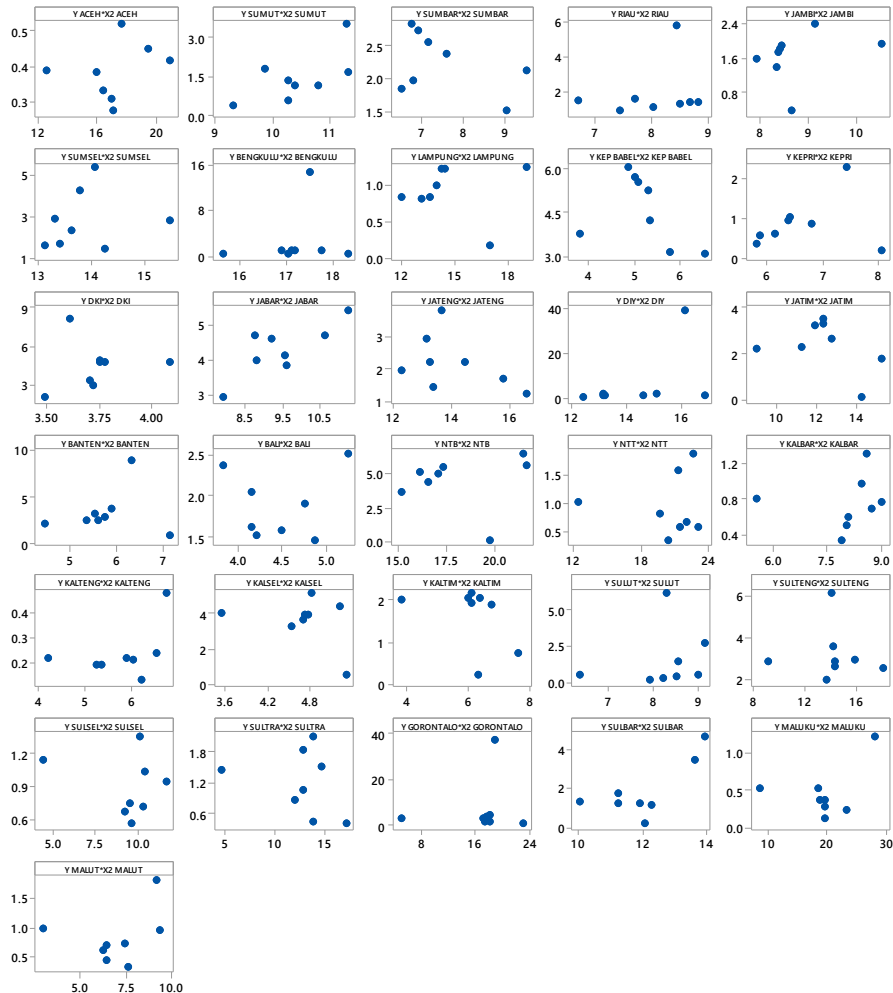
Selanjutnya dibuat pula *scatterplot* berdasarkan data *time series* antara variabel respon dan masing-masing variabel prediktor di 31 provinsi di Indonesia. Berikut merupakan hasil *scatterplot* berdasarkan data *time series* antara variabel persentase kejadian pneumonia (y) dan persentase cakupan imunisasi dasar pada bayi (x_1).



Gambar 4.5 *Scatterplot* data *time series* antara persentase kejadian pneumonia dan persentase cakupan imunisasi dasar pada bayi di Indonesia tahun 2010-2017

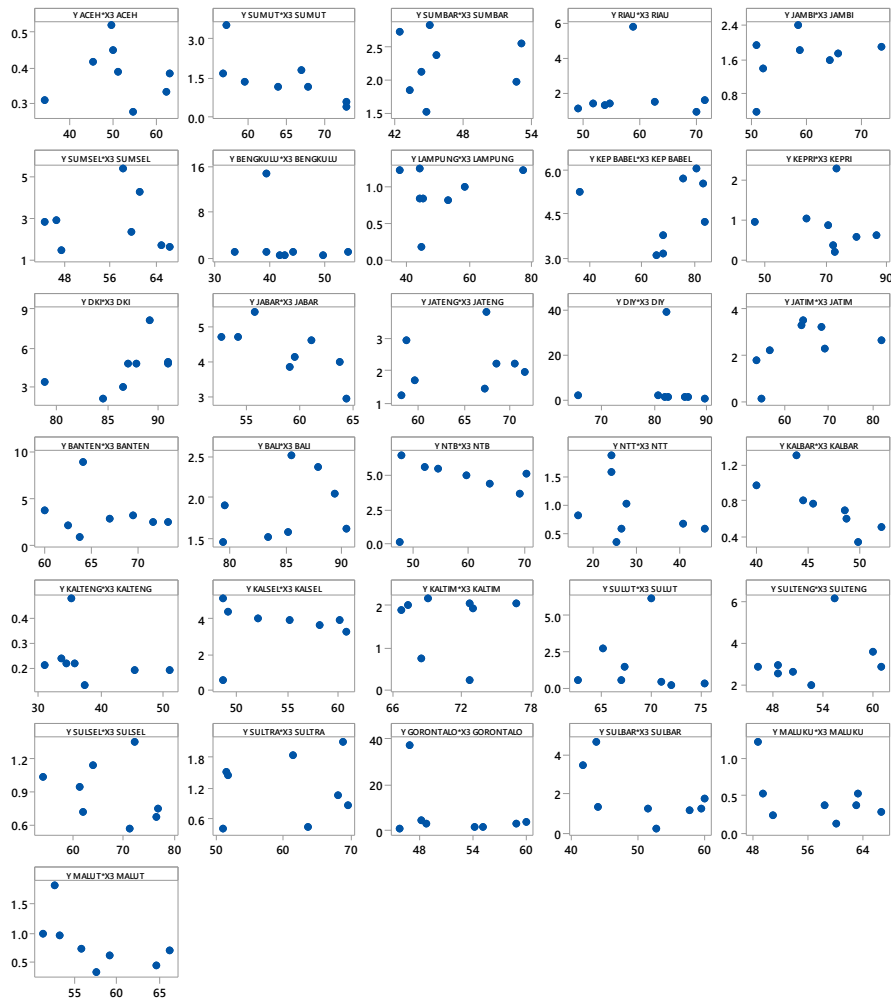
Gambar 4.5 menunjukkan variabel persentase cakupan imunisasi dasar pada bayi (x_1) cenderung memiliki pola linier pada sebagian besar provinsi di Indonesia, sehingga berdasarkan hal tersebut variabel persentase cakupan imunisasi dasar pada bayi (x_1) dapat diasumsikan ke dalam komponen parametrik. Selanjutnya, gambaran hasil *scatterplot* dari segi data *time series* antara variabel

persentase kejadian pneumonia (y) dan variabel persentase penduduk miskin (x_3) tersaji pada Gambar 4.6 berikut.



Gambar 4.6 *Scatterplot* data *time series* antara persentase kejadian pneumonia dan persentase penduduk miskin di Indonesia tahun 2010-2017

Gambar 4.6 menunjukkan variabel persentase penduduk miskin (x_2) cenderung memiliki pola yang tidak beraturan dan cenderung mengalami perubahan perilaku pola data yang berulang, sehingga berdasarkan hal tersebut variabel persentase penduduk miskin (x_3) dapat diasumsikan ke dalam komponen nonparametrik Deret Fourier. Selanjutnya, gambaran hasil *scatterplot* dari segi data *time series* antara variabel persentase kejadian pneumonia (y) dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3), tersaji pada Gambar 4.7 berikut.



Gambar 4.7 *Scatterplot* data *time series* antara persentase kejadian pneumonia dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Indonesia tahun 2010-2017

Gambar 4.7 menunjukkan variabel persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3) cenderung memiliki pola yang tidak beraturan dan cenderung mengalami perubahan perilaku pada sub interval tertentu, sehingga berdasarkan hal tersebut variabel persentase rumah tangga yang memiliki akses terhadap sanitasi layak (x_3) dapat diasumsikan ke dalam komponen nonparametrik Spline Truncated. Hasil *scatterplot* baik berdasarkan data *cross-section* maupun *time-series* menunjukkan hasil kesimpulan yang sama.

Mengidentifikasi variabel prediktor komponen parametrik, komponen nonparametrik Spline Truncated dan komponen nonparametrik Deret Fourier selain menggunakan *scatterplot* juga dapat dilakukan dengan memeriksa nilai *Generalized Cross Validation* (GCV) dari masing-masing variabel prediktor. Berikut ini adalah hasil perhitungan nilai GCV untuk semua kemungkinan model campuran Spline Truncated dan Deret Fourier, masing-masing dengan satu titik knot dan parameter osilasi $k = 1$.

Tabel 4.1 Perhitungan Nilai GCV untuk Semua Kemungkinan Model

No	Variabel			GCV
	Spline Truncated	Parametrik	Deret Fourier	
1	x_3	x_1	x_2	17.52002
2	x_2	x_1	x_3	41.41452
3	x_2	x_3	x_1	35.48621
4	x_1	x_2	x_3	41.415
5	x_3	x_2	x_1	35.48621

Berdasarkan Tabel 4.1 didapatkan hasil kesimpulan yang sama dengan identifikasi *scatterplot* sebelumnya. Ringkasan hasil penentuan komponen parametrik dan komponen nonparametrik disajikan pada Tabel 4.2 berikut.

Tabel 4.2 Komponen Parametrik dan Nonparametrik

Notasi	Variabel	Komponen	
x_1	Persentase cakupan imunisasi dasar pada bayi	Parametrik	(x_{ij})
x_2	Persentase penduduk miskin	Nonparametrik Deret Fourier	(z_{ij})
x_3	Persentase rumah tangga yang memiliki akses terhadap sanitasi layak	Nonparametrik Spline Truncated	(t_{ij})

Kemudian dilakukan perbandingan antara pemodelan dengan estimator campuran dan pemodelan dengan estimator tunggal. Hasil perbandingan nilai GCV untuk semua kemungkinan model dengan estimator tunggal dan model dengan estimator campuran Spline Truncated dan Deret Fourier disajikan dalam Tabel 4.3

Tabel 4.3 Perbandingan Estimator Tunggal dan Estimator Campuran

Estimator	GCV
Parametrik	- ($R^2 = 51,956$; MSE = 6,548079)
Spline Truncated	26,19232 ($R^2 = 51,956$; MSE = 6,548079)
Deret Fourier	71,9697 ($R^2 = 91,20833$; MSE = 1,198247)
Campuran	17,52002 ($R^2 = 81,92317$; MSE = 2,463752)

Berdasarkan Tabel 4.3, terlihat bahwa pemodelan dengan estimator campuran Spline Truncated dan Deret Fourier lebih baik dibandingkan pemodelan dengan estimator tunggal. Hal tersebut ditandai dengan nilai GCV estimator campuran Spline Truncated dan Deret Fourier yang memiliki nilai paling minimum, sehingga dengan ini dapat disimpulkan bahwa pemodelan untuk persentase kejadian pneumonia di Indonesia tahun 2010-2017 lebih baik menggunakan estimator campuran Spline Truncated dan Deret Fourier.

4.2.2 Pemodelan Persentase Kejadian Pneumonia

Pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 dengan pendekatan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier dengan dibatasi satu titik knot dan parameter osilasi optimum maksimum $k = 3$. Tahapan setelah identifikasi pola data yaitu memilih pembobot terbaik di antara tiga pembobot yang disebut oleh Wu dan Zhang. Hasil perbandingan nilai GCV pada masing-masing pembobot dalam pemodelan regresi semiparametrik dengan estimator campuran Spline Truncated dan Deret Fourier disajikan pada Tabel 4.4.

Tabel 4.4 Perbandingan Nilai GCV pada Masing-Masing Pembobot

Pembobot	Keterangan	GCV
W_1	$W_i = N^{-1}I, i = 1, 2, \dots, n$	17,520017
W_2	$W_i = n^{-1}I, i = 1, 2, \dots, n$	17,520017
W_3	$W_i = V_i^{-1}; V_i = cov(y_i), i = 1, 2, \dots, n ;$	71,891201

Pemilihan pembobot terbaik dilakukan dengan cara membandingkan nilai GCV terkecil. Berdasarkan Tabel 4.4 terlihat bahwa nilai GCV minimum terletak pada pembobot \mathbf{W}_1 dan \mathbf{W}_2 . Dikarenakan nilai GCV yang sama pada pembobot \mathbf{W}_1 dan \mathbf{W}_2 , maka dipilih salah satu pembobot diantara keduanya sebagai pembobot terbaik. Pembobot terbaik yang dipilih adalah \mathbf{W}_i dengan $\mathbf{W}_i = \mathbf{N}^{-1}\mathbf{I}$, untuk $i = 1, 2, \dots, n$ karena bobot tersebut memberikan perlakuan yang sama pada setiap pengamatan.

Langkah selanjutnya adalah menentukan titik knot dan parameter osilasi optimum menggunakan metode *Generalized Cross Validation* (GCV). Kriteria pemilihan titik knot dan parameter osilasi optimum adalah berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil sebagaimana persamaan (2.18). Model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier diterapkan pada data persentase kejadian pneumonia dengan bantuan *software R*.

Syntax pada *software R* diperoleh dari Prawanti (2019) dengan dilakukan modifikasi pada estimator Deret Fourier. Selengkapnya, *syntax* model regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier dapat dilihat pada Lampiran 2. Berikut ini model regresi semiparametrik campuran Spline Truncated dan Deret Fourier untuk data longitudinal dengan satu titik knot dan osilasi $k=1$ untuk $i=1, 2, \dots, 31$ dan $j=1, 2, \dots, 8$.

$$\hat{y}_{ij} = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_{ij} + \hat{\gamma}_{1i}t_{ij} + \hat{\gamma}_{2i}(t_{ij} - K_{li})_+ + \hat{b}_i z_{ij} + \frac{1}{2}\hat{a}_{0i} + \hat{a}_{1i} \cos z_{ij}$$

Mendapatkan estimasi parameter $\tilde{\beta}$ dengan persamaan (4.15), estimasi parameter $\tilde{\gamma}$ dengan persamaan (4.16), estimasi parameter \tilde{a} dengan persamaan (4.17), dan estimasi model dengan persamaan (4.18). Lokasi titik-titik knot dan nilai GCV yang diperoleh dari model dengan satu titik knot dan osilasi $k=1$ dapat dilihat pada Tabel 4.5. Berdasarkan Tabel 4.5 diperoleh nilai GCV minimum adalah 17,52002 yang terdapat pada ulangan ke-18.

Tabel 4.5 Nilai GCV dengan Satu Titik Knot dan Osilasi $k = 1$

Ulangan	Knot ($K_{ui}, u = 1 ; i = 1, 2, \dots, 31$)				GCV
	K_{11}	K_{12}	...	$K_{1,31}$	
1	35.462	57.4618	...	52.2304	3466.316
2	36.056	57.7924	...	52.5272	2690.643
3	39.62	59.776	...	54.308	44.3882
4	40.214	60.1066	...	54.6048	42.06262
⋮	⋮	⋮	⋮	⋮	⋮
18	33.68	56.47	...	51.34	17.52002
19	34.274	56.8006	...	51.6368	3077.739
20	34.868	57.1312	...	51.9336	3517.839
⋮	⋮	⋮	⋮	⋮	⋮
48	61.598	72.0082	...	65.2896	35.57239
49	62.192	72.3388	...	65.5864	65.17372
50	62.786	72.6694	...	65.8832	164.921

Selanjutnya, akan diperiksa model regresi regresi semiparametrik campuran Spline Truncated dan Deret Fourier untuk data longitudinal dengan satu titik knot dan osilasi $k = 2$, yang dapat diuraikan dalam model berikut untuk $i = 1, 2, \dots, 31$ dan $j = 1, 2, \dots, 8$.

$$\hat{y}_{ij} = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_{ij} + \hat{\gamma}_{i1}t_{ij} + \hat{\gamma}_{i2}(t_{ij} - K_{1i})_+ + \hat{b}_i z_{ij} + \frac{1}{2}\hat{a}_{0i} + \hat{a}_{1i} \cos z_{ij} + \hat{a}_{2i} \cos 2z_{ij}$$

Lokasi titik-titik knot dan nilai GCV yang diperoleh dari model dengan satu titik knot dan osilasi $k = 2$ dapat dilihat pada Tabel 4.6.

Tabel 4.6 Nilai GCV dengan Satu Titik Knot dan Osilasi $k = 2$

Ulangan	Knot ($K_{ui}, u = 1 ; i = 1, 2, \dots, 31$)				GCV
	K_{11}	K_{12}	...	$K_{1,31}$	
1	34.868	57.1312	...	51.9336	63.47365
2	35.462	57.4618	...	52.2304	295.4559
3	36.056	57.7924	...	52.5272	72.30761
4	56.252	69.0328	...	62.6184	36.80001
⋮	⋮	⋮	⋮	⋮	⋮
17	33.68	56.47	...	51.34	20.03952
18	34.274	56.8006	...	51.6368	59.14678
19	38.432	59.1148	...	53.7144	48.82499
⋮	⋮	⋮	⋮	⋮	⋮
48	55.658	68.7022	...	62.3216	50.31494
49	59.222	70.6858	...	64.1024	37.01889
50	59.816	71.0164	...	64.3992	40.46558

Berdasarkan Tabel 4.6 diperoleh GCV minimum adalah 20,03952 yang terdapat pada ulangan ke-17. Selanjutnya, akan diperiksa model regresi regresi semiparametrik campuran Spline Truncated dan Deret Fourier untuk data longitudinal dengan satu titik knot dan osilasi $k = 3$, yang dapat diuraikan dalam model berikut untuk $i = 1, 2, \dots, 31$ dan $j = 1, 2, \dots, 8$.

$$\hat{y}_{ij} = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_{ij} + \hat{\gamma}_{1i}t_{ij} + \hat{\gamma}_{i2} \left(t_{ij} - K_{1i} \right)_+ + \hat{b}_i z_{ij} + \frac{1}{2} \hat{a}_{0i} + \hat{a}_{1i} \cos z_{ij} + \hat{a}_{2i} \cos 2z_{ij} + \hat{a}_{3i} \cos 3z_{ij}$$

Lokasi titik-titik knot dan nilai GCV yang diperoleh dari model dengan satu titik knot dan osilasi $k = 3$ dapat dilihat pada Tabel 4.7.

Tabel 4.7 Nilai GCV dengan satu titik knot dan osilasi $k = 3$

Ulangan	Knot ($K_{ui}, u = 1 ; i = 1, 2, \dots, 31$)				GCV
	K_{11}	K_{12}	...	$K_{1,31}$	
1	36.056	57.7924	...	52.5272	49948.5
2	36.65	58.123	...	52.824	176822.7
3	59.222	70.6858	...	64.1024	1751.27
4	59.816	71.0164	...	64.3992	1679.106
⋮	⋮	⋮	⋮	⋮	⋮
32	47.342	64.0738	...	58.1664	214.1479
33	33.68	56.47	...	51.34	18.14955
34	34.274	56.8006	...	51.6368	489876
⋮	⋮	⋮	⋮	⋮	⋮
48	54.47	68.041	...	61.728	9868.678
49	55.064	68.3716	...	62.0248	435.7441
50	55.658	68.7022	...	62.3216	255.2882

Berdasarkan Tabel 4.7 diperoleh nilai GCV minimum sebesar 18,14955 pada ulangan ke-33. Kemudian ketiga hasil GCV minimum di atas dibandingkan nilai GCV-nya dan diringkas dalam Tabel 4.8.

Tabel 4.8 Hasil Perhitungan GCV antara Tiga Model

Titik Knot				Osilasi	GCV
K_{11}	K_{12}	...	$K_{1,31}$		
33.68	56.47	...	51.34	1	17.52002
33.68	56.47	...	51.34	2	20.03952
33.68	56.47	...	51.34	3	18.14955

Berdasarkan Tabel 4.8 terlihat bahwa model regresi yang memiliki nilai *Generalized Cross Validation* minimum adalah model dengan satu titik knot dan osilasi $k=1$, sehingga dipilihlah model tersebut sebagai model terbaik untuk pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 dengan pendekatan regresi semiparametrik untuk data longitudinal menggunakan estimator campuran Spline Truncated dan Deret Fourier. Titik-titik knot optimal untuk masing-masing provinsi di Indonesia tertera pada Lampiran 3.

Model terbaik regresi semiparametrik untuk pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 menggunakan estimator campuran Spline Truncated dan Deret Fourier adalah model regresi dengan satu titik knot dan osilasi $k=1$, di mana model tersebut dapat dituliskan dalam persamaan model sebagai berikut.

$$\hat{y}_{ij} = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_{ij} + \hat{\gamma}_{i1}t_{ij} + \hat{\gamma}_{i2}(t_{ij} - K_{1i})_+ + \hat{b}_i z_{ij} + \frac{1}{2}\hat{a}_{0i} + \hat{a}_{1i} \cos z_{ij}$$

di mana $i = 1, 2, \dots, 31$ dan $j = 1, 2, \dots, 8$.

Persamaan di atas dapat dijabarkan pada setiap provinsi yaitu:

1. Provinsi Aceh

$$\hat{y}_{1j} = -4.16 \times 10^{-16} + 0.01062x_{1j} + 0.00237t_{1j} + 0.00237(t_{1j} - 33.68)_+ + 0.0124z_{1j} - \frac{1}{2} \times 1.65 - 0.02326 \cos z_{1j}$$

di mana $j = 1, 2, \dots, 8$.

2. Provinsi Sumatera Utara

$$\hat{y}_{2j} = -1.66 \times 10^{-15} - 0.01996x_{2j} - 0.03768t_{2j} - 0.03768(t_{2j} - 56.47)_+ - 0.3893z_{2j} + \frac{1}{2} \times 20.7527 + 1.12185 \cos z_{2j}$$

di mana $j = 1, 2, \dots, 8$.

3. Provinsi Sumatera Barat

$$\hat{y}_{3j} = -3.15 \times 10^{-16} - 0.00668x_{3j} - 0.01519t_{3j} - 0.01519(t_{3j} - 42.34)_+ + 2.5836z_{3j} - \frac{1}{2} \times 34.2766 + 3.8603 \cos z_{3j}$$

di mana $j = 1, 2, \dots, 8$.

⋮
⋮
⋮

31. Maluku Utara

$$\hat{y}_{31,j} = -1.16 \times 10^{-15} + 0.05152x_{31,j} + 0.01353t_{31,j} + 0.01353(t_{31,j} - 51.34)_+ - 0.05823z_{31,j} - \frac{1}{2} \times 6.9992 - 0.3714 \cos z_{31,j}$$

di mana $j = 1, 2, \dots, 8$.

Estimasi parameter dari model regresi semiparametrik untuk pemodelan persentase kejadian pneumonia pada tahun 2010-2017 di masing-masing provinsi di Indonesia menggunakan estimator campuran Spline Truncated dan Deret Fourier dengan satu titik knot dan osilasi $k = 1$ selengkapnya tertera pada Lampiran 4. Selanjutnya hasil estimasi persentase kejadian pneumonia (\hat{y}) tertera pada Lampiran 5.

Pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 menggunakan regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier menghasilkan koefisien determinasi (R^2) sebesar 81,92%, yang berarti bahwa variabel-variabel prediktor yang digunakan mampu menjelaskan variasi dalam variabel respon persentase kejadian pneumonia sebesar 81,92%. Kemudian akan dibahas interpretasi model regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier. Provinsi yang dijadikan contoh interpretasi adalah provinsi Sulawesi Utara. Berikut merupakan estimasi model regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier pada provinsi Sulawesi Utara.

$$\hat{y}_{24,j} = -2.42 \times 10^{-15} - 0.0896x_{24,j} - 0.0388t_{24,j} - 0.0388(t_{24,j} - 62.48)_+ - 6.0349z_{24,j} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24,j}$$

Estimasi model di atas jika dijabarkan terhadap pengulangan pengamatan (tahun pengamatan) adalah sebagai berikut.

Tahun 2010 ($j = 1$) maka,

$$\hat{y}_{24(2010)} = -2.42 \times 10^{-15} - 0.0896x_{24(2010)} - 0.0388t_{24(2010)} - 0.0388(t_{24(2010)} - 62.48)_+ - 6.0349z_{24(2010)} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24(2010)}$$

Tahun 2011 ($j = 2$) maka,

$$\hat{y}_{24(2011)} = -2.42 \times 10^{-15} - 0.0896x_{24(2011)} - 0.0388t_{24(2011)} - 0.0388(t_{24(2011)} - 62.48)_+ - 6.0349z_{24(2011)} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24(2011)}$$

⋮

Tahun 2017 ($j = 8$) maka,

$$\hat{y}_{24(2017)} = -2.42 \times 10^{-15} - 0.0896x_{24(2017)} - 0.0388t_{24(2017)} - 0.0388(t_{24(2017)} - 62.48)_+ - 6.0349z_{24(2017)} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24(2017)}$$

Berikut merupakan interpretasi dari estimasi model di atas.

- a. Estimasi model untuk variabel persentase cakupan imunisasi dasar ($x_{24,j}$)

Mengasumsikan bahwa variabel prediktor lainnya selain persentase cakupan imunisasi dasar adalah konstan, sehingga diperoleh model:

$$\hat{y}_{24,j} = -0.0896x_{24,j} + c_1$$

dengan

$$c_1 = -2.42 \times 10^{-15} - 0.0388t_{24,j} - 0.0388(t_{24,j} - 62.48)_+ - 6.0349z_{24,j} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24,j}$$

sehingga berdasarkan model di atas dapat diartikan bahwa apabila persentase cakupan imunisasi dasar di Sulawesi Utara pada tahun ke- j meningkat satu satuan, maka persentase kejadian pneumonia di Sulawesi Utara akan turun sebesar 0.0896%.

- b. Estimasi model untuk variabel persentase rumah tangga yang memiliki akses terhadap sanitasi layak ($t_{24,j}$)

Mengasumsikan bahwa variabel prediktor lainnya selain persentase rumah tangga yang memiliki akses terhadap sanitasi layak adalah konstan, sehingga diperoleh model:

$$\hat{y}_{24,j} = -0.0388t_{24,j} - 0.0388(t_{24,j} - 62.48)_+ + c_2$$

dengan

$$c_2 = -2.42 \times 10^{-15} - 0.0896x_{24,j} - 6.0349z_{24,j} + \frac{1}{2} \times 116.013 - 9.5448 \cos z_{24,j}$$

Model tersebut dapat diinterpretasikan dengan menggunakan fungsi *truncated* berikut:

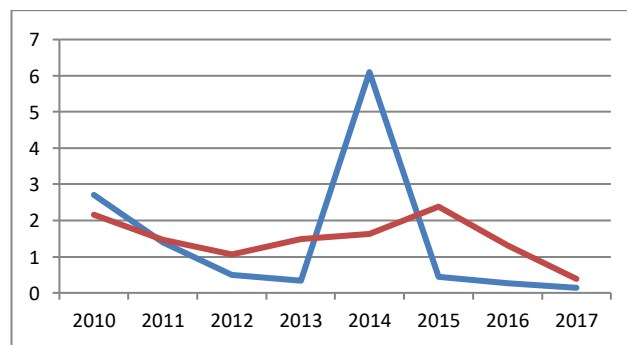
$$\hat{y}_{24,j} = \begin{cases} -0.0388t_{24,j} + c_2, & t_{24,j} < 62.48 \\ -0.0776t_{24,j} + c_3, & t_{24,j} \geq 62.48 \end{cases}$$

dengan $c_3 = 2.424224 + c_2$

sehingga berdasarkan model di atas dapat didefinisikan dalam dua kejadian, yaitu:

1. Ketika persentase rumah tangga yang memiliki akses terhadap sanitasi layak kurang dari 62,48%, maka apabila persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Sulawesi Utara pada tahun ke- j meningkat satu satuan akan menyebabkan persentase kejadian pneumonia di Sulawesi Utara akan turun sebesar 0,0388%.
2. Ketika persentase rumah tangga yang memiliki akses terhadap sanitasi layak lebih dari atau sama dengan 62,48%, maka apabila persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Sulawesi Utara pada tahun ke- j meningkat satu satuan akan menyebabkan persentase kejadian pneumonia di Sulawesi Utara akan turun sebesar 0,0776%.

Berikut merupakan gambaran plot antara nilai y pengamatan dan nilai estimasi y di provinsi Sulawesi Utara.



Gambar 4.8 Plot antara nilai y pengamatan dan nilai estimasi y di provinsi Sulawesi Utara

Gambar 4.8 merupakan gambaran plot antara nilai y pengamatan dan nilai estimasi y di provinsi Sulawesi Utara dimana garis biru menunjukkan nilai y pengamatan, dan garis merah menunjukkan nilai estimasi y . Sulawesi Utara mengalami kenaikan yang cukup signifikan pada tahun 2014, yang juga digambarkan pada nilai estimasinya. Namun kenaikan pada nilai pengamatan dan nilai estimasi tidak persis sama. Pada tahun 2015-2017, garis biru dan garis merah sama-sama menunjukkan penurunan. Sulawesi Utara terus mengalami penurunan persentase kejadian pneumonia pada 3 tahun terakhir, yang juga digambarkan sama pada nilai estimasinya. Plot antara y dan \hat{y} di masing-masing provinsi di Indonesia secara lengkap tersaji pada Lmpiran 6.

Selanjutnya, dengan pemodelan regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier ini, diperoleh pula interpretasi secara *series* atau berdasarkan ulangan (tahun). Berikut merupakan estimasi model regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier dengan satu titik knot dan osilasi $k = 1$ berdasarkan *series* pada tahun 2017.

$$\hat{y}_i = -2,31 \times 10^{-16} + 0,042358x_i + 0,00051t_i + 0,000411(t_i - 58,98286)_+ - 0,04118z_i - \frac{1}{2} \times 3,40824 - 0,09032 \cos z_i$$

Berikut ini merupakan interpretasi dari estimasi model di atas.

- a. Estimasi model untuk variabel persentase cakupan imunisasi dasar (x_i)

Mengasumsikan bahwa variabel prediktor lainnya selain persentase cakupan imunisasi dasar adalah konstan, sehingga diperoleh model:

$$\hat{y}_i = 0,042358x_i + c_1$$

dengan

$$c_1 = -2,31 \times 10^{-16} + 0,00051t_i + 0,000411(t_i - 58,98286)_+ - 0,04118z_i - \frac{1}{2} \times 3,40824 - 0,09032 \cos z_i$$

sehingga berdasarkan model di atas dapat diartikan bahwa apabila persentase cakupan imunisasi dasar meningkat satu satuan, maka persentase kejadian pneumonia akan naik sebesar 0,042%.

- b. Estimasi model untuk variabel persentase rumah tangga yang memiliki akses terhadap sanitasi layak (t_i)

Mengasumsikan bahwa variabel prediktor lainnya selain persentase rumah tangga yang memiliki akses terhadap sanitasi layak adalah konstan, sehingga diperoleh model:

$$\hat{y}_i = 0,00051t_i + 0,000411(t_i - 58,98286)_+ + c_2$$

dengan

$$c_2 = -2,31 \times 10^{-16} + 0,042358x_i + -0,04118z_i - \frac{1}{2} \times 3,40824 - 0,09032 \cos z_i$$

Model tersebut dapat diinterpretasikan dengan menggunakan fungsi *truncated* berikut:

$$\hat{y}_i = \begin{cases} 0,00051t_i + c_2, & t_i < 58,98 \\ 0,00092t_i + c_3, & t_i \geq 58,98 \end{cases}$$

dengan $c_3 = 0,0242 + c_2$

sehingga berdasarkan model di atas dapat didefinisikan dalam dua kejadian, yaitu:

1. Ketika persentase rumah tangga yang memiliki akses terhadap sanitasi layak kurang dari 58,98%, maka apabila persentase rumah tangga yang memiliki akses terhadap sanitasi layak di suatu provinsi meningkat satu satuan akan menyebabkan persentase kejadian pneumonia meningkat sebesar 0,00051%.
2. Ketika persentase rumah tangga yang memiliki akses terhadap sanitasi layak lebih dari atau sama dengan 58,98%, maka apabila persentase rumah tangga yang memiliki akses terhadap sanitasi layak di suatu provinsi meningkat satu satuan akan menyebabkan persentase kejadian pneumonia meningkat sebesar 0,00092%.

Selengkapnya, hasil estimasi model regresi semiparametrik untuk data longitudinal dengan estimator campuran Spline Truncated dan Deret Fourier pada data persentase kejadian pneumonia setiap tahunnya atau secara *series* tersaji pada Lampiran 7.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut.

1. Estimasi model regresi untuk data longitudinal menggunakan estimator campuran Spline *Truncated* dan Deret Fourier adalah

$$\begin{aligned}\hat{y} &= \mathbf{X}\hat{\beta} + \mathbf{T}\hat{\gamma} + \mathbf{Z}\hat{a} \\ &= \mathbf{XA}(K, m)\tilde{y} + \mathbf{TB}(K, m)\tilde{y} + \mathbf{ZC}(K, m)\tilde{y} \\ &= \mathbf{D}(K, m)\tilde{y}\end{aligned}$$

di mana $\mathbf{D}(K, m) = \mathbf{XA}(K, m) + \mathbf{TB}(K, m) + \mathbf{ZC}(K, m)$;

$$\hat{\beta} = \left[(\mathbf{I} - \mathbf{PZUX}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{RX} - \mathbf{RZUX}) \right]^{-1} \left[(\mathbf{P} - \mathbf{PZU}) - (\mathbf{PT} - \mathbf{PZUT})(\mathbf{I} - \mathbf{RZUT})^{-1}(\mathbf{R} - \mathbf{RZU}) \right] \tilde{y} ;$$

$$\hat{\gamma} = (\mathbf{PT} - \mathbf{PZUT})^{-1} \left[(\mathbf{P} - \mathbf{PZU}) - (\mathbf{I} - \mathbf{PZUX})\mathbf{A}(K, m) \right] \tilde{y} ; \text{ dan}$$

$$\hat{a} = \mathbf{U}\tilde{y} - \mathbf{UXA}\tilde{y} - \mathbf{UT} \left[(\mathbf{PT} - \mathbf{PZUT})^{-1} \left[(\mathbf{P} - \mathbf{PZU}) - (\mathbf{I} - \mathbf{PZUX})\mathbf{A} \right] \tilde{y} \right]$$

Estimasi dari parameter-parameter yang tidak diketahui yaitu $\tilde{\beta}$, $\tilde{\gamma}$, dan \tilde{a} diperoleh melalui optimasi *Weighted Least Square* (WLS).

2. Model terbaik regresi semiparametrik untuk pemodelan persentase kejadian pneumonia di Indonesia pada tahun 2010-2017 menggunakan estimator campuran Spline *Truncated* dan Deret Fourier adalah model regresi dengan satu titik knot dan osilasi $k = 1$. Koefisien determinasi (R^2) yang diperoleh adalah sebesar 81,92%, yang berarti bahwa variabel-variabel prediktor yang digunakan mampu menjelaskan variasi dalam variabel respon persentase kejadian pneumonia sebesar 81,92%.

5.2 Saran

Berdasarkan hasil dan pembahasan estimasi model regresi semiparametrik campuran Spline Truncated dan Deret Fourier, saran yang dapat diberikan adalah:

1. Penelitian ini terbatas pada estimasi titik, sehingga untuk penelitian selanjutnya dapat dilakukan estimasi interval dan uji statistik untuk mengidentifikasi variabel prediktor yang memberi pengaruh signifikan terhadap variabel respon.
2. Penelitian ini terbatas menggunakan satu variabel prediktor untuk setiap komponen parametrik atau nonparametrik, sehingga pada penelitian selanjutnya dapat digunakan variabel yang lebih banyak untuk aplikasi pada kasus data lainnya.

DAFTAR PUSTAKA

- Afifah, N., Budiantara, I. N. dan Latra, I.N. (2017). Estimator of Kernel and Fourier Series in Semiparametric Regression. *International Conference on Mathematics : Education, Theory and Application*.
- Ardyanto, F. (2018). *Estimasi Interval Parameter Komponen Parametrik dan Spline dalam Model Regresi Semiparametrik Campuran Spline dan Kernel (Aplikasi: Tingkat Pengangguran Terbuka di Indonesia)*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Asrini, L. J. and Budiantara, I N. (2014). Fourier Series Semiparametric Regression Models (Case Study: The Production of Lowland Rice Irrigation in Central Java). *ARPJ Journal of Engineering and Applied Sciences*. **9**. No. 9. 1501-1506.
- Baltagi, B.H. (2005). *Econometric Analysis of Panel Data, 3rd Edition*, New Jersey: John Wiley & Sons Inc.
- Bilodeau, Martin. (1992). Fourier Smoother and Additive Models. *The Canadian Journal of Statistics*. **20**. No 3. 257-269
- Budiantara, I. N. (2014). Pemodelan Regresi Nonparametrik dan Semiparametrik Spline (Konsep, Metode, dan Aplikasinya). *Prosiding Seminar Nasional Matematika*. Universitas Udayana. Denpasar. hal 1-16
- Danardono. (2015). *Analisis Data Longitudinal*. Yogyakarta: Gadjah Mada University Press
- Eubank, Randall L. (1999). *Nonparametric Regression and Spline Smoothing, 2nd Edition*. New York: Marcel Dekker, Inc.
- Fadhilah, K.N., Suparti, dan Tarno (2016). Pemodelan Regresi Spline Truncated untuk Data Longitudinal (Studi Kasus: Harga Saham Bulanan pada Kelompok Saham Perbankan Periode Januari 2009-Desember 2015). *Jurnal Gaussian*. **5**. No. 3. 447-454.
- Fernandes, A.A.R., Janssen, P., Sa'adah, U., Solimun, Effendi, A., Nurjannah, dan Amalina, L. (2017). Comparison of Spline Estimator at Various Levels of Autocorrelation in Smoothing Spline Nonparametric Regression for Longitudinal Data. *Communications in Statistics-Theory and Method*, **47**, No. 21, 5265-5285.
- Gujarati, D. N., dan Porter, D. C. (2015). *Dasar-Dasar Ekonometrika Volume 1, 2nd Edition*. Jakarta: Penerbit Salemba Empat
- Harlan, Johan. (2018). *Analisis Data Longitudinal*. Depok: Penerbit Gunadarma

- Hartati, Susi. (2011). *Analisis Faktor Risiko yang Berhubungan dengan Kejadian Pneumonia pada Anak Balita di RSUD Pasar Rebo Jakarta*. Tesis. Depok: Universitas Indonesia
- Hartati, Susi., Nurhaeni N., dan Gayatri D. (2012). Faktor Resiko Terjadinya Pneumonia pada Anak Balita. *Jurnal Keperawatan Indonesia*. **5**. No 1. 13-20
- Hockenberry, M.J., dan Wilson, D. (2009). *Wong's Essentials of Pediatric Nursing, 8th Edition*. St. Louis Missouri: Elsevier Mosby
- Kementerian Kesehatan RI. (2011). Profil Kesehatan Indonesia 2010. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2012). Profil Kesehatan Indonesia 2011. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2013). Profil Kesehatan Indonesia 2012. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2014a). Profil Kesehatan Indonesia 2013. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2014b). Petunjuk Teknis Penyusunan Profil Kesehatan kabupaten/Kota 2013. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2015). Profil Kesehatan Indonesia 2014. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2016). Profil Kesehatan Indonesia 2015. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2017). Profil Kesehatan Indonesia 2016. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2018). Profil Kesehatan Indonesia 2017. Jakarta : Kementerian Kesehatan RI
- Kementerian Kesehatan RI. (2019). Profil Kesehatan Indonesia 2018. Jakarta : Kementerian Kesehatan RI
- Lestrel, P.E. dan Brown, H.D. (1976). Fourier Analysis of Adolescent Growth of the Cranial Vault: A Longitudinal Study. *Human Biology*. **48**, No.33, 517-528
- Lestrel, P.E. dan Roche, A.F. (1986). Cranial Base Shape Variation with Age: A Longitudinal Study of Shape Using Fourier Analysis. *Human Biology*. **58**, No.4, 527-540

- Machmud, Rizanda. (2009). Pengaruh Kemiskinan Keluarga pada Kejadian Pneumonia Balita di Indonesia. *Jurnal Kesehatan Masyarakat Nasional*. **4**. No.1. 36-41
- Mattila, J.T., Fine, M.J., Limper, A.H., Murray, P.R., Chen, B.B. and Lin, P.L. (2014). Pneumonia Treatment and Diagnosis. *AnnalsATS*, **11**, 189-192
- Mohaisen, A.J. and Abdulhusein A.M. (2015). Spline Semiparametric Regression Models. *Journal of Kufa for Mathematics and Computer*. **2**. No 3. 1-10.
- Nisa, Khaerun. (2017). *Model Regresi Semiparametrik Campuran Spline Truncated dan Deret Fourier (Studi Kasus : Angka Harapan Hidup Provinsi Jawa Timur)*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Nisa, K., Budiantara, I. N. dan Rumiati, A.T. (2017). Multivariable Semiparametric Regression Model with Combined Estimator of Fourier Series and Kernel. *International Conference Series : Earth and Environmental Science*.
- Nurdini, A. (2006). "Cross Sectional versus Longitudinal": Pilihan Rancangan Waktu dalam Penelitian Perumahan Pemukiman. *Dimensi Teknik Arsitektur*. **34**. No 1. 52-58.
- Pane, R., Budiantara, I.N., Zain, I., and Otok, B.W. (2014). Parametric and Nonparametric Estimators in Fourier Series Semiparametric Regression and Their Characteristics. *Applied Mathematical Sciences*. **8**. No. 102. 5053-5064
- Prahotama, A. (2013). Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur. *Prosiding Seminar Nasional Statistika*. Universitas Diponegoro. Semarang. hal 69-76.
- Prawanti, D. D. (2019). Estimasi Interval Parameter Model Regresi Semiparametrik Spline Truncated untuk data Longitudinal. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Purnomo, A.A.S.I. (2016). *Estimator Campuran Kernel dan Spline Truncated Linier Multivariabel dalam Regresi Nonparametrik*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Ratnasari, V., Budiantara, I.N., Zain, I., Ratna, M., and Mariati N.P.A.M. (2015). Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models (Applications: Data of Poverty in Papua Indonesia). *International Journal of Basic & Applied Sciences IJBAS-IJENS*. **15**. No. 4. 9-12.

- Ratnasari, V., Budiantara, I.N., Ratna, M., and Zain, I. (2016). Estimation of Nonparametric Regression Curve using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel. *Global Journal of Pure and Applied Mathematics*. **12**. No. 6. 5047-5057.
- Rencher, A. C. and Schaalje, G. B. (2007), *Linear Model in Statistics, 2nd Edition*. New Jersey : John Wiley & Sons.
- Rory. (2016). *Regresi Campuran Nonparametrik Spline Linier Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Santoso, F.P., Wulandari, S.P. dan Purnadi. (2012). Faktor-Faktor Eksternal Pneumonia pada Balita di Jawa Timur dengan Pendekatan Geographically Weighted Regression. *Jurnal Sains dan Seni ITS*, **1**, No. 1,37-42
- Setiawan, R.N.S., Budiantara, I.N., dan Ratnasari, V. (2017). Application of Confidence Intervals for Parameters of Nonparametric Spline Truncated Regression on Index Development Gender in East Java. *IPTEK Journal of Science*, **2**, No. 3, 49-55.
- Sudiarsa, I.W., Budiantara, I.N., Suhartono and Purnami, S.W. (2015). Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression. *Applied Mathematical Sciences*, **9**, No. 100, 4997-5010.
- Tripena, A., dan Budiantara, I.N. 2006. Fourier Estimator in Nonparametric Regression. *International Conference On Natural and Applied Natural Scienses*. Ahmad Dahlan University. Yogyakarta
- Umam, K. (2018). *Interval Konfidensi Parameter Regresi Nonparametrik Spline Truncated untuk Data Longitudinal*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- Vogt, M., dan Linton, O. (2016). Classification of Non-Parametric Regression Functions in Longitudinal Data Models. *Journal of The Royal Statistical Society. Series B (Statistical Methodology)*, **79**, No. 1, 5-27.
- Wahba, G. (1990). Smoothing Spline ANNOVA with Component-Wise Bayesian “Confidence Intervals”. *Journal of Computational and Graphical Statistics.*, Vol. 2, No. 1, pp. 9-117.
- Wu, H., dan Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. New Jersey: John Willey & Sons Inc.

LAMPIRAN

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017

Provinsi	Tahun	y	x_1	x_2	x_3
Aceh	2010	0.41	80.41	20.98	45.17
	2011	0.45	86.40	19.57	50.10
	2012	0.39	89.20	12.47	51.11
	2013	0.51	83.00	17.72	49.81
	2014	0.30	77.60	16.98	33.68
	2015	0.27	67.10	17.11	54.68
	2016	0.33	69.10	16.43	62.68
	2017	0.38	70.00	15.92	63.38
Sumatera Utara	2010	3.47	88.98	11.31	57.10
	2011	1.59	89.70	11.33	56.47
	2012	1.32	81.10	10.28	59.43
	2013	1.12	81.50	10.39	63.77
	2014	1.75	79.90	9.85	66.92
	2015	1.13	79.90	10.79	67.89
	2016	0.52	89.20	10.27	72.86
	2017	0.35	86.70	9.28	73.00
Sumatera Barat	2010	2.10	78.93	9.50	44.26
	2011	1.50	85.30	9.04	44.67
	2012	1.81	77.20	6.45	43.12
	2013	2.35	84.50	7.56	45.59
	2014	2.72	79.10	6.89	42.34
	2015	2.81	74.10	6.71	45.02
	2016	2.54	77.60	7.14	53.24
	2017	1.95	80.70	6.75	52.77
Riau	2010	1.32	86.98	8.65	54.27
	2011	1.20	91.50	8.47	53.29
	2012	1.44	73.70	6.68	62.58
	2013	5.77	83.90	8.42	58.44
	2014	0.98	82.20	7.99	48.74
	2015	1.35	74.50	8.82	51.30
	2016	1.55	81.70	7.67	71.36
	2017	0.84	73.30	7.41	70.04
Jambi	2010	1.35	99.03	8.34	51.98
	2011	0.35	99.90	8.65	50.65
	2012	1.91	93.90	10.53	50.89
	2013	1.85	92.40	8.42	73.61
	2014	1.79	86.80	8.39	58.58
	2015	2.36	99.80	9.12	58.21
	2016	1.72	100.80	8.37	65.65
	2017	1.54	101.40	7.90	64.20

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
Sumatera Selatan	2010	2.77	93.46	15.47	44.36
	2011	1.37	95.10	14.24	47.36
	2012	2.85	93.30	13.29	46.65
	2013	5.31	88.50	14.06	58.29
	2014	2.28	87.30	13.62	59.79
	2015	4.17	98.10	13.77	61.30
	2016	1.68	105.30	13.39	65.05
	2017	1.54	95.30	13.10	66.36
Bengkulu	2010	0.17	91.92	18.30	41.64
	2011	14.46	97.80	17.50	39.22
	2012	0.89	90.90	16.89	44.14
	2013	0.81	90.20	17.75	54.48
	2014	0.76	84.50	17.09	33.18
	2015	0.68	84.20	17.16	39.22
	2016	0.29	82.90	17.03	49.75
	2017	0.27	86.80	15.59	42.71
Lampung	2010	1.23	89.40	18.94	43.85
	2011	0.16	92.40	16.93	44.33
	2012	0.83	98.70	11.88	43.35
	2013	1.20	96.80	14.39	77.55
	2014	1.21	99.60	14.21	37.27
	2015	0.82	99.20	13.53	44.83
	2016	0.97	98.60	13.86	58.58
	2017	0.81	93.90	13.04	52.89
Kep. Bangka Belitung	2010	3.05	107.45	6.51	65.06
	2011	3.09	89.90	5.75	67.64
	2012	3.73	94.00	3.73	67.92
	2013	5.26	94.10	5.25	35.06
	2014	5.71	92.80	4.97	75.67
	2015	6.03	98.00	4.83	80.80
	2016	5.50	94.00	5.04	83.16
	2017	4.21	92.80	5.30	83.56
Kep. Riau	2010	0.16	92.41	8.05	72.37
	2011	2.26	89.80	7.40	73.01
	2012	0.85	81.90	6.77	70.36
	2013	0.94	88.40	6.35	46.37
	2014	1.02	101.80	6.40	63.45
	2015	0.36	87.70	5.78	71.97
	2016	0.55	85.00	5.84	79.55
	2017	0.58	89.70	6.13	86.33

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
DKI Jakarta	2010	1.96	99.75	3.48	84.57
	2011	4.67	101.70	3.75	87.83
	2012	3.23	80.10	3.70	78.72
	2013	2.94	91.00	3.72	86.60
	2014	4.77	98.70	4.09	87.05
	2015	8.13	97.90	3.61	89.28
	2016	4.81	102.80	3.75	91.13
	2017	4.72	99.20	3.78	91.13
Jawa Barat	2010	5.41	95.18	11.27	55.57
	2011	4.69	92.20	10.65	52.50
	2012	4.72	102.10	8.71	54.07
	2013	3.85	90.50	9.61	58.92
	2014	4.59	86.00	9.18	61.00
	2015	4.11	82.50	9.57	59.43
	2016	3.99	92.20	8.77	63.79
	2017	2.91	99.20	7.83	64.40
Jawa Tengah	2010	1.19	97.41	16.56	57.76
	2011	1.67	97.20	15.76	59.42
	2012	2.94	90.20	13.11	58.48
	2013	2.20	100.70	14.44	68.62
	2014	3.79	93.30	13.58	67.43
	2015	1.44	99.20	13.32	67.20
	2016	2.20	101.40	13.19	70.66
	2017	1.94	97.30	12.23	71.84
DI Yogyakarta	2010	0.72	99.31	16.83	81.85
	2011	38.76	96.50	16.08	82.15
	2012	2.01	95.50	13.10	80.37
	2013	1.53	97.20	15.03	65.11
	2014	1.44	78.20	14.55	82.50
	2015	1.37	97.20	13.16	86.31
	2016	1.15	96.40	13.10	85.78
	2017	0.42	91.60	12.36	89.40
Jawa Timur	2010	1.74	98.63	15.26	52.96
	2011	0.09	97.70	14.23	54.21
	2012	2.19	84.50	8.90	56.14
	2013	2.65	99.30	12.73	82.03
	2014	3.50	97.70	12.28	63.70
	2015	3.28	98.40	12.28	63.48
	2016	3.22	98.10	11.85	68.15
	2017	2.27	96.70	11.20	68.83

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
Banten	2010	0.97	93.95	7.16	63.78
	2011	8.86	93.80	6.32	64.15
	2012	2.15	91.30	4.41	62.43
	2013	3.72	87.20	5.89	59.81
	2014	3.29	89.90	5.51	69.51
	2015	2.96	79.70	5.75	67.04
	2016	2.50	90.30	5.36	73.42
	2017	2.49	84.40	5.59	71.68
Bali	2010	1.46	99.38	4.88	79.13
	2011	1.52	99.80	4.20	83.26
	2012	2.38	67.00	3.81	87.86
	2013	1.58	97.00	4.49	85.15
	2014	1.91	96.40	4.76	79.38
	2015	2.51	99.10	5.25	85.46
	2016	2.06	99.00	4.15	89.33
	2017	1.63	99.20	4.14	90.51
Nusa Tenggara Barat	2010	6.46	104.17	21.55	47.43
	2011	0.12	96.70	19.73	47.34
	2012	5.62	107.40	21.65	51.66
	2013	5.52	99.50	17.25	54.35
	2014	4.96	92.40	17.05	59.41
	2015	4.42	99.30	16.54	63.72
	2016	5.15	100.10	16.02	70.31
	2017	3.71	100.20	15.05	69.25
Nusa Tenggara Timur	2010	0.60	73.75	23.03	26.23
	2011	1.59	71.60	21.23	23.82
	2012	1.02	54.40	12.21	27.33
	2013	0.35	69.90	20.24	24.91
	2014	0.82	64.80	19.60	16.12
	2015	1.88	70.10	22.58	23.90
	2016	0.67	69.30	22.01	40.46
	2017	0.58	70.90	21.38	45.31
Kalimantan Barat	2010	0.77	81.38	9.02	45.32
	2011	1.31	87.40	8.60	43.81
	2012	0.81	84.60	5.49	44.39
	2013	0.70	82.20	8.74	48.50
	2014	0.59	84.10	8.07	48.59
	2015	0.97	81.50	8.44	39.78
	2016	0.51	82.80	8.00	52.06
	2017	0.34	82.60	7.86	49.65

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
Kalimantan Tengah	2010	0.47	87.87	6.77	35.14
	2011	0.24	77.00	6.56	33.72
	2012	0.21	86.20	4.21	34.31
	2013	0.13	91.40	6.23	37.50
	2014	0.21	57.00	6.07	30.85
	2015	0.22	64.90	5.91	35.88
	2016	0.18	80.30	5.36	50.97
	2017	0.19	85.00	5.26	45.46
Kalimantan Selatan	2010	4.42	83.20	5.21	48.95
	2011	0.49	89.50	5.29	48.38
	2012	3.99	72.00	3.56	51.96
	2013	3.90	80.40	4.76	55.16
	2014	5.13	65.30	4.81	48.44
	2015	3.87	80.60	4.72	60.13
	2016	3.26	84.20	4.52	60.89
	2017	3.64	81.50	4.70	58.09
Kalimantan Timur	2010	0.72	87.26	7.66	68.37
	2011	1.90	90.90	6.77	66.56
	2012	1.98	66.50	3.82	67.09
	2013	2.03	78.80	6.38	72.61
	2014	0.20	82.10	6.31	72.65
	2015	2.16	87.80	6.10	68.83
	2016	2.05	85.80	6.00	76.76
	2017	1.94	92.00	6.08	72.83
Sulawesi Utara	2010	2.70	82.14	9.10	64.87
	2011	1.40	91.60	8.51	67.23
	2012	0.50	73.80	6.36	62.48
	2013	0.34	88.00	8.50	70.99
	2014	6.10	81.40	8.26	69.82
	2015	0.45	81.30	8.98	66.79
	2016	0.27	78.40	8.20	75.27
	2017	0.14	80.60	7.90	71.93
Sulawesi Tengah	2010	2.44	82.89	18.07	48.25
	2011	2.87	84.70	15.83	48.39
	2012	2.76	76.50	9.02	45.74
	2013	2.52	79.20	14.32	50.25
	2014	1.90	80.70	13.61	52.47
	2015	6.10	75.30	14.07	55.37
	2016	3.53	83.90	14.09	59.94
	2017	2.78	88.80	14.22	61.12

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
Sulawesi Selatan	2010	0.94	92.79	11.60	61.45
	2011	0.71	100.10	10.29	62.02
	2012	1.14	88.80	4.44	64.06
	2013	1.03	88.30	10.32	54.12
	2014	0.56	91.20	9.54	71.07
	2015	1.35	85.90	10.12	72.36
	2016	0.66	91.90	9.24	76.51
	2017	0.74	102.90	9.48	76.73
Sulawesi Tenggara	2010	0.40	79.44	17.05	50.87
	2011	1.52	87.00	14.56	51.43
	2012	1.43	86.00	4.62	51.63
	2013	2.09	69.90	13.73	68.87
	2014	1.83	77.30	12.77	61.26
	2015	0.44	71.60	13.74	63.62
	2016	1.05	79.00	12.77	68.26
	2017	0.87	77.30	11.97	69.52
Gorontalo	2010	0.69	90.02	23.19	45.66
	2011	37.16	80.90	18.75	46.68
	2012	2.69	91.10	4.80	48.65
	2013	4.36	98.30	18.01	48.02
	2014	1.52	80.80	17.41	54.07
	2015	1.73	86.60	18.16	54.96
	2016	3.68	74.50	17.63	59.85
	2017	2.67	80.60	17.14	58.75
Sulawesi Barat	2010	3.41	84.22	13.58	41.30
	2011	4.62	93.20	13.89	43.40
	2012	1.30	69.80	10.03	43.57
	2013	1.15	84.80	12.23	57.59
	2014	0.20	81.90	12.05	52.45
	2015	1.17	76.70	11.90	51.21
	2016	1.74	80.80	11.19	59.81
	2017	1.22	81.90	11.18	59.48
Maluku	2010	1.22	77.16	27.74	48.28
	2011	0.22	86.30	23.00	50.75
	2012	0.53	36.50	8.39	49.04
	2013	0.35	90.90	19.27	58.17
	2014	0.37	72.70	18.44	62.87
	2015	0.12	74.20	19.36	60.02
	2016	0.26	67.60	19.26	66.81
	2017	0.52	57.80	18.29	63.29

Lampiran 1. Data Persentase Kejadian Pneumonia dan Variabel yang Mempengaruhinya tahun 2010-2017 (Lanjutan)

Provinsi	Tahun	y	x_1	x_2	x_3
Maluku Utara	2010	0.96	79.58	9.42	53.26
	2011	1.82	84.80	9.18	52.53
	2012	0.97	69.90	2.92	51.34
	2013	0.32	76.70	7.64	57.49
	2014	0.71	76.10	7.41	55.75
	2015	0.62	73.00	6.22	59.17
	2016	0.44	75.30	6.41	64.71
	2017	0.70	68.80	6.44	66.18
	2016	0.96	79.58	9.42	53.26
	2017	1.82	84.80	9.18	52.53

Keterangan :

Simbol	Variabel	Komponen
x_1	Persentase cakupan imunisasi dasar pada bayi	Parametrik
x_2	Persentase penduduk miskin	Nonparametrik Deret Fourier
x_3	Persentase rumah tangga yang memiliki akses terhadap sanitasi layak	Nonparametrik Spline Truncated
y	Persentase kejadian pneumonia	-

Lampiran 2. *Syntax* R Estimasi Model Regresi Semiparametrik untuk Data Longitudinal Menggunakan Estimator Campuran Spline *Truncated* dan Deret Fourier

```
library(pracma)
library(MASS)
data0<-read.table("D://DATA COBA//Data 2010-2017.txt",
header=TRUE)
data<-cbind(data0$Tahun,data0$Y,data0$X4,data0$X2,
data0$X3)

estimasi<-function(data,M)
{
  y<-as.matrix(data[,2])
  x<-as.matrix(data[,c(3:5)])
  nt<-nrow(x)
  t<-max(data[,1])-min(data[,1])+1
  n<-nt/t
  p<-1 #banyak variabel parametrik
  q<-1 #banyak variabel nonparametrik spline
  s<-1 #banyak variabel nonparametrik deret fourier
  r<-1

  #=====Pembobot W=====
  cat("=====PEMILIHAN MATRIX PEMBOBOT W=====\\n")
  pembobot<-as.numeric(readline("Pembobot yang digunakan:
"))
  if(pembobot==1)
  {
    W<-diag(1/nt,nt,nt)
  } else if (pembobot==2) {
    W=diag(1/n,nt,nt)
  } else {
    ww=matrix(0,n,t)
    d=n
    a=1
    for (i in 1:t)
    {
      me=mean(y[a:d])
      ww[,i]=y[a:d]-me
      a=a+n
      d=d+n
    }
    W=matrix(0,nt,nt)
    mI=diag(1,t,t)
    a=1
    c=1
    for (j in 1:n)
    {
      w=as.vector(ww[j,])%*%t(as.vector(ww[j,]))
    }
  }
}
```



```

w1=mI*diag(w)
for (k in 1:t)
{
  b=c
  for (l in 1:t)
  {
    W[a,b]=w1[k,l]
    b=b+1
  }
  a=a+1
}
c=a
}
}#=====Matriks Respon per Provinsi=====
Y<-matrix(0,n,t)
d<-n
a<-1
for (i in 1:t)
{
  Y[,i]<-y[a:d]
  a<-a+n
  d<-d+n
}
Yk<-0
for (i in 1:n)
{
  Yk<-cbind(Yk,t(Y[i,]))
}
Yk<-as.vector(Yk[1,2:ncol(Yk)])

#=====Matriks Prediktor per Provinsi=====
Xt<-matrix(0,n*t,p+q+s)
a<-1
Xt<-matrix(0,1,p+q+s)
for (i in 1:n)
{
  aa<-matrix(0,1,p+q+s)
  b<-i
for (j in 1:t)
{
  aa<-rbind(aa,x[b,])
  b<-b+n
}
aa<-aa[2:nrow(aa),]
Xt<-rbind(Xt,aa)
}
Xt<-Xt[2:nrow(Xt),]

#=====Penentuan Titik Knot=====
nknot=51
nkomb=nknot
knot1=matrix(0,nkomb,1)

```

```

a=1
for (l in 1:n)
{
  xx=Xt[a:(l*t),]
  kn=matrix(0,nknot,q)
  for (o in 1:q)
  {
    kn[,o]=seq(min(xx[,o]),max(xx[,o]),length.out=nknot)
  }
  knot=matrix(0,nkomb,1)
  v=1
  for (i in 1:nknot)
  {
    knot[v,]=c(kn[i,1])
    v=v+1
  }
  knot1=cbind(knot1,knot)
  a=a+t
}
knot1=knot1[1:(nrow(knot1)-1),2:ncol(knot1)]
nknot=nrow(knot1)

#====Penentuan K optimal=====
#Inisiasi Matriks Prediktor Deret Fourier
N=length(y)
af<-s*(M+2)
Zk<-matrix(0,N,af)
for(i in 1:N)
{
  for(j in 1:M)
  {
    Zk[i,1]=Xt[i,3]
    Zk[i,2]=0.5
    Zk[i,2+j]<-cos(j*Xt[i,3])
  }
}
Z=matrix(0,nt,(n*s*(M+2)))
a=1
b=1
c=s*(M+2)
for (j in 1:n)
{
  Z[a:(j*t),b:(j*c)]=Zk[a:(j*t),]
  a=a+t
  b=b+c
}

#Inisiasi Matriks Prediktor Parametrik
betanol<-matrix(1,nt,1)
XX<-cbind(betanol,Xt[,2])
X=matrix(0,nt,n*(p+1))
a=1

```

```

b=1
c=p+1
for (j in 1:n)
{
  X[a:(j*t),b:(j*c)]=XX[a:(j*t),]
  a=a+t
  b=b+c
}

#=====GCV MINIMUM=====
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
T=matrix(0,nt,(n*q*(1+r)))
for (i in 1:nknot)
{
  b=1
  a=1
  Xk=matrix(0,1,r*q)
  for (j in 1:n)
  {
    xa=matrix(0,t)
    for (k in 1:q)
    {
      aa=cbind(pmax(0,Xt[b:(j*t),k]-knot1[i,a]))
      xa=cbind(xa,aa)
      a=a+1
    }
    xa=matrix(xa[,2:ncol(xa)])
    Xk=rbind(Xk,xa)
    b=b+t
  }
  Xk=Xk[2:nrow(Xk),]
  Xk=cbind(Xt[,1],Xk)
}

#Inisiasi Matriks Prediktor Spline
a=1
b=1
rq=(q*(1+r))
for (j in 1:n)
{
  T[a:(j*t),b:(j*rq)]=Xk[a:(j*t),]
  a=a+t
  b=b+rq
}

P=pinv(t(X)**ginv(W)**X)**t(X)**ginv(W)
R=pinv(t(T)**ginv(W)**T)**t(T)**ginv(W)
U=pinv(t(Z)**ginv(W)**Z)**t(Z)**ginv(W)

##Estimasi Beta (Parameter Reg. Parametrik)
I=diag(rep(1,(n*(p+1))))
A1=(I-(P**Z**U**X))

```

```

A2=( (P**T) - (P**Z**U**T) )
A3=pinv(I - (R**Z**U**T) )
A4=(R**X - (R**Z**U**X) )
AA=pinv(A1 - (A2**A3**A4) )
Aa=(P - (P**Z**U) )
Ab=(R - (R**Z**U) )
AB=(Aa - (A2**A3**Ab) )
AKm=AA**AB
Beta=AKm**Yk

##Estimasi Gamma (Parameter Spline Truncated)
B1=pinv(A2)
B2=(Aa - (A1**AKm) )
BKm=B1**B2
Gamma=BKm**Yk

##Estimasi Alpha (Parameter Deret Fourier)
C1=(U - (U**X**AKm) )
C2=(U**T**BKm)
CKm=C1 - C2
Alpha=CKm**Yk

DKm=(X**AKm) + (T**BKm) + (Z**CKm)
yhat=DKm**Yk
error=Yk - yhat
MSE[i]=(t(error)**error) / nt
db=(sum(diag(DKm)) / nt)
GCV[i]=MSE[i] / ((1 - db)^2)
}
optimum=cbind(knot1, MSE, GCV)
GCVmin=optimum[order(optimum[, ncol(optimum)]), ]

#=====ESTIMASI=====
T=matrix(0, nt, (n*q*(1+r)))
b=1
a=1
Xk=matrix(0, 1, r*q)
for (j in 1:n)
{
  xa=matrix(0, t)
  for (k in 1:q)
  {
    aa=cbind(pmax(0, Xt[b:(j*t), k] - GCVmin[1, a]))
    xa=cbind(xa, aa)
    a=a+1
  }
  xa=matrix(xa[, 2:ncol(xa)])
  Xk=rbind(Xk, xa)
  b=b+t
}
Xk=Xk[2:nrow(Xk), ]
Xk=cbind(Xt[, 1], Xk)

```

```

a=1
b=1
rq=(q*(1+r))
for (j in 1:n)
{
  T[a:(j*t),b:(j*rq)]=Xk[a:(j*t),]
  a=a+t
  b=b+rq
}

P=pinv(t(X)%*%ginv(W)%*%X)%*%t(X)%*%ginv(W)
R=pinv(t(T)%*%ginv(W)%*%T)%*%t(T)%*%ginv(W)
U=pinv(t(Z)%*%ginv(W)%*%Z)%*%t(Z)%*%ginv(W)

##Estimasi Beta (Parameter Reg. Parametrik)
I=diag(rep(1,(n*(p+1))))
A1=(I-(P%*%Z%*%U%*%X))
A2=((P%*%T)-(P%*%Z%*%U%*%T))
A3=pinv(I-(R%*%Z%*%U%*%T))
A4=(R%*%X-(R%*%Z%*%U%*%X))
AA=pinv(A1-(A2%*%A3%*%A4))
Aa=(P-(P%*%Z%*%U))
Ab=(R-(R%*%Z%*%U))
AB=(Aa-(A2%*%A3%*%Ab))
AKm=AA%*%AB
Beta=AKm%*%Yk

##Estimasi Gamma (Parameter Spline Truncated)
B1=pinv(A2)
B2=(Aa-(A1%*%AKm))
BKm=B1%*%B2
Gamma=BKm%*%Yk

##Estimasi Alpha (Parameter Deret Fourier)
C1=(U-(U%*%X%*%AKm))
C2=(U%*%T%*%BKm)
CKm=C1-C2
Alpha=CKm%*%Yk

DKm=(X%*%AKm)+(T%*%BKm)+(Z%*%CKm)
yhat=DKm%*%Yk
error<-yhat-Yk
MSE=(t(error)%*%error)/nt

SSE=sum((Yk-yhat)^2)
SSR=sum((yhat-mean(Yk))^2)
SST=SSR+SSE
Rsqr=(SSR/(SST))*100
cat("\n====R Square =====\n")
print(Rsqr)

```

```
#=====OUTPUT=====
write.csv(GCVmin,file = "D://OUTPUT//OUTPUT GCV.csv")
write.csv(cbind(Yk,yhat,error),file = "D://OUTPUT//OUTPUT
yhat.csv")
write.csv(Beta,file = "D://OUTPUT//OUTPUT Beta.csv")
write.csv(Gamma,file = "D://OUTPUT//OUTPUT Gamma.csv")
write.csv(Alpha,file = "D://OUTPUT//OUTPUT Alpha.csv")
}

##Untuk 1 Variabel Parametrik, 1 Variabel Spline, 1 Variabel
Deret Fourier
##Variabel X (Spline kolom 1, Parametrik kolom 2, Deret
Fourier kolom 3
```

Lampiran 3. Titik Knot Optimal

Provinsi	Titik Knot	Provinsi	Titik Knot
Aceh	33.68	Bali	79.13
Sumatera Utara	56.47	Nusa Tenggara Barat	47.34
Sumatera Barat	42.34	Nusa Tenggara Timur	16.12
Riau	48.74	Kalimantan Barat	39.78
Jambi	50.65	Kalimantan Tengah	30.85
Sumatera Selatan	44.36	Kalimantan Selatan	48.38
Bengkulu	33.18	Kalimantan Timur	66.56
Lampung	37.27	Sulawesi Utara	62.48
Kepulauan Bangka Belitung	35.06	Sulawesi Tengah	45.74
Kepulauan Riau	46.37	Sulawesi Selatan	54.12
DKI Jakarta	78.72	Sulawesi Tenggara	50.87
Jawa Barat	52.5	Gorontalo	45.66
Jawa Tengah	57.76	Sulawesi Barat	41.3
DI Yogyakarta	65.11	Maluku	48.28
Jawa Timur	52.96	Maluku Utara	51.34
Banten	59.81		

Lampiran 4. *Output* Estimasi Parameter

Provinsi	Parameter	Estimasi Parameter	Provinsi	Parameter	Estimasi Parameter
Aceh (y_{1j})	β_{01}	-4.16E-16	Jambi (y_{5j})	β_{05}	5.22E-16
	β_{11}	0.010621124		β_{51}	-0.01092553
	γ_{111}	0.002377176		γ_{511}	0.024030305
	γ_{112}	0.002377176		γ_{512}	0.024030305
	b_{11}	0.01239742		b_{51}	0.405626511
	a_{011}	-1.65000766		a_{051}	-5.18965465
	a_{111}	-0.02326262		a_{151}	-0.18519328
Sumatera Utara (y_{2j})	β_{02}	-1.66E-15	Sumatera Selatan (y_{6j})	β_{06}	-2.38E-15
	β_{21}	-0.0199649		β_{61}	-0.10175607
	γ_{211}	-0.03768324		γ_{611}	0.025205548
	γ_{212}	-0.03768324		γ_{612}	0.025205548
	b_{21}	-0.38929774		b_{61}	1.975366053
	a_{021}	20.75274015		a_{061}	-34.4290982
	a_{121}	1.121854427		a_{161}	1.773481138
Sumatera Barat (y_{3j})	β_{03}	-3.15E-16	Bengkulu (y_{7j})	β_{07}	-1.27E-13
	β_{31}	-0.00668811		β_{71}	0.861872785
	γ_{311}	-0.01519335		γ_{711}	-0.06815387
	γ_{312}	-0.01519335		γ_{712}	-0.06815387
	b_{31}	2.583596115		b_{71}	4.987545561
	a_{031}	-34.2766459		a_{071}	-313.080943
	a_{131}	3.860298895		a_{171}	-8.54268438
Riau (y_{4j})	β_{04}	3.02E-15	Lampung (y_{8j})	β_{08}	-6.99E-14
	β_{41}	-0.08220808		β_{81}	0.119948303
	γ_{411}	0.091563462		γ_{811}	0.007935432
	γ_{412}	0.091563462		γ_{812}	0.007935432
	b_{41}	-26.9612722		b_{81}	0.179630774
	a_{041}	425.7276978		a_{081}	-27.8727849
	a_{141}	-34.2693084		a_{181}	0.537959218

Lampiran 4. *Output* Estimasi Parameter (Lanjutan)

Provinsi	Parameter	Estimasi Parameter	Provinsi	Parameter	Estimasi Parameter
Kep. Bangka Belitung ($y_{9,j}$)	β_{09}	2.83E-15	Jawa Tengah ($y_{13,j}$)	$\beta_{0,13}$	-6.41E-14
	β_{91}	0.346852423		$\beta_{13,1}$	-0.28815203
	γ_{911}	0.001790756		$\gamma_{13,11}$	0.088839694
	γ_{912}	0.001790756		$\gamma_{13,12}$	0.088839694
	b_{91}	-10.1244806		$b_{13,1}$	0.329704468
	a_{091}	38.72661418		$a_{0,13,1}$	38.13435296
	a_{191}	12.46676585		$a_{1,13,1}$	0.031784295
Kep. Riau ($y_{10,j}$)	$\beta_{0,10}$	1.09E-16	DI Yogyakarta ($y_{14,j}$)	$\beta_{0,14}$	-1.40E-13
	$\beta_{10,1}$	0.002183989		$\beta_{14,1}$	1.539560742
	$\gamma_{10,11}$	0.00249845		$\gamma_{14,11}$	0.982221594
	$\gamma_{10,12}$	0.00249845		$\gamma_{14,12}$	0.982221594
	$b_{10,1}$	1.118465559		$b_{14,1}$	-12.721709
	$a_{0,10,1}$	-16.830073		$a_{0,14,1}$	-97.1130808
	$a_{1,10,1}$	1.974251012		$a_{1,14,1}$	-41.5568058
DKI Jakarta ($y_{11,j}$)	$\beta_{0,11}$	-2.52E-14	Jawa Timur ($y_{15,j}$)	$\beta_{0,15}$	6.95E-14
	$\beta_{11,1}$	-0.13055314		$\beta_{15,1}$	0.123824196
	$\gamma_{11,11}$	0.225305763		$\gamma_{15,11}$	-0.00584872
	$\gamma_{11,12}$	0.225305763		$\gamma_{15,12}$	-0.00584872
	$b_{11,1}$	-6.87886793		$b_{15,1}$	-0.47580343
	$a_{0,11,1}$	61.90525725		$a_{0,15,1}$	-6.76908093
	$a_{1,11,1}$	11.91957404		$a_{1,15,1}$	0.500310577
Jawa Barat ($y_{12,j}$)	$\beta_{0,12}$	1.77E-14	Banten ($y_{16,j}$)	$\beta_{0,16}$	-1.94E-15
	$\beta_{12,1}$	0.039569765		$\beta_{16,1}$	0.328619284
	$\gamma_{12,11}$	-0.01134747		$\gamma_{16,11}$	-0.09190366
	$\gamma_{12,12}$	-0.01134747		$\gamma_{16,12}$	-0.09190366
	$b_{12,1}$	0.591487491		$b_{16,1}$	-2.75840204
	$a_{0,12,1}$	-9.02571003		$a_{0,16,1}$	-16.1603856
	$a_{1,12,1}$	-0.50253006		$a_{1,16,1}$	7.425567243

Lampiran 4. *Output* Estimasi Parameter (Lanjutan)

Provinsi	Parameter	Estimasi Parameter	Provinsi	Parameter	Estimasi Parameter
Bali ($y_{17,j}$)	$\beta_{0,17}$	-2.15E-14	Kalimantan Tengah ($y_{21,j}$)	$\beta_{0,21}$	-7.5481E-17
	$\beta_{17,1}$	-0.02462487		$\beta_{21,1}$	-0.00470695
	$\gamma_{17,11}$	0.032197385		$\gamma_{21,11}$	0.004189352
	$\gamma_{17,12}$	0.032197385		$\gamma_{21,12}$	0.004189352
	$b_{17,1}$	-1.46561869		$b_{21,1}$	0.36737283
	$a_{0,17,1}$	16.73443255		$a_{0,21,1}$	-2.69107437
	$a_{1,17,1}$	2.368655317		$a_{1,21,1}$	-0.55024405
Nusa Tenggara Barat ($y_{18,j}$)	$\beta_{0,18}$	8.12E-14	Kalimantan Selatan ($y_{22,j}$)	$\beta_{0,22}$	-7.8998E-15
	$\beta_{18,1}$	-0.13221727		$\beta_{22,1}$	-0.22561696
	$\gamma_{18,11}$	-0.21301593		$\gamma_{22,11}$	0.12464169
	$\gamma_{18,12}$	-0.21301593		$\gamma_{22,12}$	0.12464169
	$b_{18,1}$	-1.11240474		$b_{22,1}$	-17.078884
	$a_{0,18,1}$	99.01812251		$a_{0,22,1}$	188.3173899
	$a_{1,18,1}$	-6.06213544		$a_{1,22,1}$	21.90341902
Nusa Tenggara Timur ($y_{19,j}$)	$\beta_{0,19}$	3.13E-14	Kalimantan Timur ($y_{23,j}$)	$\beta_{0,23}$	-2.2803E-15
	$\beta_{19,1}$	-0.08306523		$\beta_{23,1}$	0.045493486
	$\gamma_{19,11}$	-0.03339505		$\gamma_{23,11}$	-0.01665646
	$\gamma_{19,12}$	-0.03339505		$\gamma_{23,12}$	-0.01665646
	$b_{19,1}$	-0.0695197		$b_{23,1}$	-0.63839165
	$a_{0,19,1}$	18.03606241		$a_{0,23,1}$	5.541537661
	$a_{1,19,1}$	-1.38861667		$a_{1,23,1}$	0.309465802
Kalimantan Barat ($y_{20,j}$)	$\beta_{0,20}$	-4.04E-16	Sulawesi Utara ($y_{24,j}$)	$\beta_{0,24}$	-2.4201E-15
	$\beta_{20,1}$	0.081577397		$\beta_{24,1}$	-0.08961834
	$\gamma_{20,11}$	-0.01956662		$\gamma_{24,11}$	-0.03882697
	$\gamma_{20,12}$	-0.01956662		$\gamma_{24,12}$	-0.03882697
	$b_{20,1}$	-0.20024065		$b_{24,1}$	-6.03492401
	$a_{0,20,1}$	-7.19420697		$a_{0,24,1}$	116.0134386
	$a_{1,20,1}$	-0.61511212		$a_{1,24,1}$	-9.54486537

Lampiran 4. *Output* Estimasi Parameter (Lanjutan)

Provinsi	Parameter	Estimasi Parameter	Provinsi	Parameter	Estimasi Parameter
Sulawesi Tengah ($y_{25,j}$)	$\beta_{0,25}$	1.15126E-14	Sulawesi Barat ($y_{29,j}$)	$\beta_{0,29}$	3.69884E-14
	$\beta_{25,1}$	-0.35020028		$\beta_{29,1}$	-0.15091239
	$\gamma_{25,11}$	0.116509648		$\gamma_{29,11}$	0.074501404
	$\gamma_{25,12}$	0.116509648		$\gamma_{29,12}$	0.074501404
	$b_{25,1}$	0.387044735		$b_{29,1}$	2.682120591
	$a_{0,25,1}$	38.15638156		$a_{0,29,1}$	-42.6459591
	$a_{1,25,1}$	-1.39792697		$a_{1,29,1}$	-3.39164105
Sulawesi Selatan ($y_{26,j}$)	$\beta_{0,26}$	3.1293E-15	Maluku ($y_{30,j}$)	$\beta_{0,30}$	4.41409E-16
	$\beta_{26,1}$	-0.02512405		$\beta_{30,1}$	-0.02481809
	$\gamma_{26,11}$	-0.00042617		$\gamma_{30,11}$	-0.02995873
	$\gamma_{26,12}$	-0.00042617		$\gamma_{30,12}$	-0.02995873
	$b_{26,1}$	-0.01590801		$b_{30,1}$	0.08732293
	$a_{0,26,1}$	6.96333482		$a_{0,30,1}$	4.751130273
	$a_{1,26,1}$	0.130616597		$a_{1,30,1}$	0.42167946
Sulawesi Tenggara ($y_{27,j}$)	$\beta_{0,27}$	7.83843E-16	Maluku Utara ($y_{31,j}$)	$\beta_{0,31}$	-1.16E-15
	$\beta_{27,1}$	0.009341699		$\beta_{31,1}$	0.051516605
	$\gamma_{27,11}$	0.010122691		$\gamma_{31,11}$	0.013527428
	$\gamma_{27,12}$	0.010122691		$\gamma_{31,12}$	0.013527428
	$b_{27,1}$	-0.04697451		$b_{31,1}$	-0.05822782
	$a_{0,27,1}$	0.76198007		$a_{0,31,1}$	-6.99902121
	$a_{1,27,1}$	-0.08265978		$a_{1,31,1}$	-0.37144379
Gorontalo ($y_{28,j}$)	$\beta_{0,28}$	-6.3309E-14			
	$\beta_{28,1}$	-1.34687726			
	$\gamma_{28,11}$	-0.9747487			
	$\gamma_{28,12}$	-0.9747487			
	$b_{28,1}$	-0.08860915			
	$a_{0,28,1}$	351.6222761			
	$a_{1,28,1}$	14.36371189			

Lampiran 5. *Output* Estimasi Variabel Respon (y)

Provinsi	Tahun	y	\hat{y}	Provinsi	Tahun	y	\hat{y}
Aceh	2010	0.41	0.436179	Sumatera Selatan	2010	2.77	3.228855
	2011	0.45	0.475926		2011	1.37	2.324962
	2012	0.39	0.416773		2012	2.85	2.106845
	2013	0.51	0.423048		2013	5.31	3.510749
	2014	0.3	0.296614		2014	2.28	3.579437
	2015	0.27	0.283605		2015	4.17	2.612685
	2016	0.33	0.348003		2016	1.68	1.887005
	2017	0.38	0.359851		2017	1.54	2.719463
Sumatera Utara	2010	3.47	2.368438	Bengkulu	2010	0.17	3.255619
	2011	1.59	2.415025		2011	14.46	10.07348
	2012	1.32	1.668185		2012	0.89	5.526157
	2013	1.12	1.387681		2013	0.81	0.686378
	2014	1.75	1.009089		2014	0.76	0.866542
	2015	1.13	1.36299		2015	0.68	-0.457
	2016	0.52	0.48976		2016	0.29	-2.57006
	2017	0.35	0.548832		2017	0.27	0.948879
Sumatera Barat	2010	2.1	2.326936	Lampung	2010	1.23	1.125137
	2011	1.5	1.354762		2011	0.16	0.411944
	2012	1.81	2.149275		2012	0.83	0.844906
	2013	2.35	2.205056		2013	1.2	1.059953
	2014	2.72	2.661453		2014	1.21	0.819619
	2015	2.81	2.491283		2015	0.82	1.11555
	2016	2.54	2.342964		2016	0.97	1.161357
	2017	1.95	2.248272		2017	0.81	0.691535
Riau	2010	1.32	2.461919	Kep. Bangka Belitung	2010	3.05	3.039923
	2011	1.2	2.076329		2011	3.09	3.24533
	2012	1.44	2.094624		2012	3.73	4.013418
	2013	5.77	3.569609		2013	5.26	5.295431
	2014	0.98	-0.20452		2014	5.71	4.616945
	2015	1.35	2.06344		2015	6.03	6.143057
	2016	1.55	1.690211		2016	5.5	5.186689
	2017	0.84	0.69839		2017	4.21	5.039207
Jambi	2010	1.35	1.073703	Kep. Riau	2010	0.16	0.65169
	2011	0.35	1.171851		2011	2.26	1.172504
	2012	1.91	1.962322		2012	0.85	1.31647
	2013	1.85	2.23095		2013	0.94	0.965983
	2014	1.79	1.55288		2014	1.02	1.12747
	2015	2.36	1.77125		2015	0.36	0.214549
	2016	1.72	1.728395		2016	0.55	0.367611
	2017	1.54	1.378649		2017	0.58	0.903723

Lampiran 5. *Output* Estimasi Variabel Respon (y) (Lanjutan)

Provinsi	Tahun	y	\hat{y}	Provinsi	Tahun	y	\hat{y}
DKI Jakarta	2010	1.96	3.120089	Banten	2010	0.97	1.566372
	2011	4.67	3.940042		2011	8.86	6.437249
	2012	3.23	2.67059		2012	2.15	1.568516
	2013	2.94	4.789124		2013	3.72	5.69061
	2014	4.77	4.47328		2014	3.29	4.298598
	2015	8.13	5.197592		2015	2.96	1.81911
	2016	4.81	5.283452		2016	2.5	3.290435
	2017	4.72	5.755831		2017	2.49	2.269111
Jawa Barat	2010	5.41	5.117861	Bali	2010	1.46	1.710713
	2011	4.69	5.009302		2011	1.52	1.406526
	2012	4.72	4.427231		2012	2.38	2.384344
	2013	3.85	4.504894		2013	1.58	1.81098
	2014	4.59	4.018901		2014	1.91	1.693643
	2015	4.11	4.156408		2015	2.51	2.400745
	2016	3.99	3.869453		2016	2.06	1.788654
	2017	2.91	3.16595		2017	1.63	1.854396
Jawa Tengah	2010	1.19	1.568646	Nusa Tenggara Barat	2010	6.46	7.122907
	2011	1.67	1.649529		2011	0.12	0.83131
	2012	2.94	2.684801		2012	5.62	5.013513
	2013	2.2	1.8627		2013	5.52	4.268124
	2014	3.79	3.526326		2014	4.96	4.474014
	2015	1.44	1.70601		2015	4.42	5.000205
	2016	2.2	1.646609		2016	5.15	4.352659
	2017	1.94	2.725379		2017	3.71	4.897267
DI Yogyakarta	2010	0.72	5.097451	Nusa Tenggara Timur	2010	0.6	0.781677
	2011	38.76	31.58641		2011	1.59	1.547466
	2012	2.01	-10.032		2012	1.02	1.062033
	2013	1.53	6.200519		2013	0.35	0.430195
	2014	1.44	1.523198		2014	0.82	0.718875
	2015	1.37	4.822618		2015	1.88	1.722194
	2016	1.15	1.981194		2016	0.67	0.955853
	2017	0.42	6.22066		2017	0.58	0.291707
Jawa Timur	2010	1.74	0.806786	Kalimantan Barat	2010	0.77	0.805741
	2011	0.09	1.571652		2011	1.31	1.292113
	2012	2.19	2.064021		2012	0.81	0.814709
	2013	2.65	2.69806		2013	0.7	0.715294
	2014	3.5	2.914774		2014	0.59	0.656334
	2015	3.28	3.004025		2015	0.97	0.92325
	2016	3.22	3.014242		2016	0.51	0.386163
	2017	2.27	2.86644		2017	0.34	0.406395

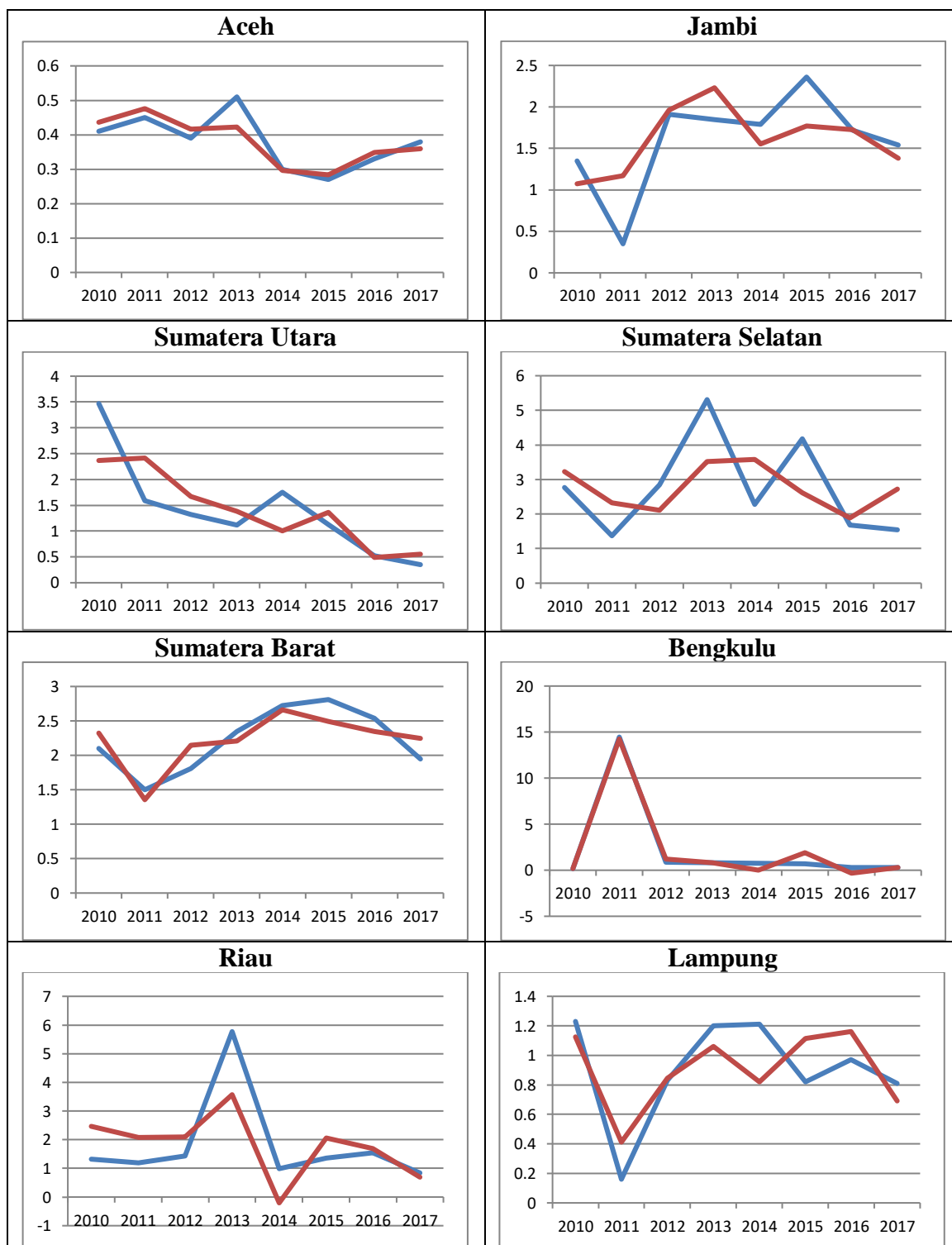
Lampiran 5. *Output* Estimasi Variabel Respon (y) (Lanjutan)

Provinsi	Tahun	y	\hat{y}	Provinsi	Tahun	y	\hat{y}
Kalimantan Tengah	2010	0.47	0.406843	Sulawesi Selatan	2010	0.94	1.010789
	2011	0.24	0.325985		2011	0.71	0.688557
	2012	0.21	0.218549		2012	1.14	1.113343
	2013	0.13	0.148474		2013	1.03	0.994298
	2014	0.21	0.207574		2014	0.56	0.871329
	2015	0.22	0.17917		2015	1.35	1.023609
	2016	0.18	0.211481		2016	0.66	0.855235
	2017	0.19	0.151923		2017	0.74	0.572841
Kalimantan Selatan	2010	4.42	3.033743	Sulawesi Tenggara	2010	0.4	0.855865
	2011	0.49	1.60864		2011	1.52	1.069967
	2012	3.99	4.022078		2012	1.43	1.505308
	2013	3.9	3.48636		2013	2.09	1.235639
	2014	5.13	5.456219		2014	1.83	1.147578
	2015	3.87	4.487586		2015	0.44	1.145522
	2016	3.26	2.925863		2016	1.05	1.305176
	2017	3.64	3.679511		2017	0.87	1.364945
Kalimantan Timur	2010	0.72	0.741155	Gorontalo	2010	0.69	2.7832
	2011	1.9	1.749076		2011	37.16	32.98442
	2012	1.98	1.990178		2012	2.69	3.606083
	2013	2.03	1.280536		2013	4.36	2.301406
	2014	0.2	1.475358		2014	1.52	6.418164
	2015	2.16	1.990921		2015	1.73	6.007088
	2016	2.05	1.692454		2016	3.68	6.678227
	2017	1.94	2.060322		2017	2.67	-6.27859
Sulawesi Utara	2010	2.7	2.162034	Sulawesi Barat	2010	3.41	3.67384
	2011	1.4	1.46776		2011	4.62	4.426663
	2012	0.5	1.068141		2012	1.3	1.249353
	2013	0.34	1.482831		2013	1.15	0.984579
	2014	6.1	1.637214		2014	0.2	0.42571
	2015	0.45	2.382743		2015	1.17	0.906733
	2016	0.27	1.31235		2016	1.74	1.675903
	2017	0.14	0.386927		2017	1.22	1.467217
Sulawesi Tengah	2010	2.44	1.963776	Maluku	2010	1.22	1.073631
	2011	2.87	2.877328		2011	0.22	0.423103
	2012	2.76	2.393125		2012	0.53	0.495041
	2013	2.52	3.519044		2013	0.35	0.148277
	2014	1.9	2.2788		2014	0.37	0.247727
	2015	6.1	5.633133		2015	0.12	0.442723
	2016	3.53	3.721962		2016	0.26	0.20968
	2017	2.78	2.512833		2017	0.52	0.549817

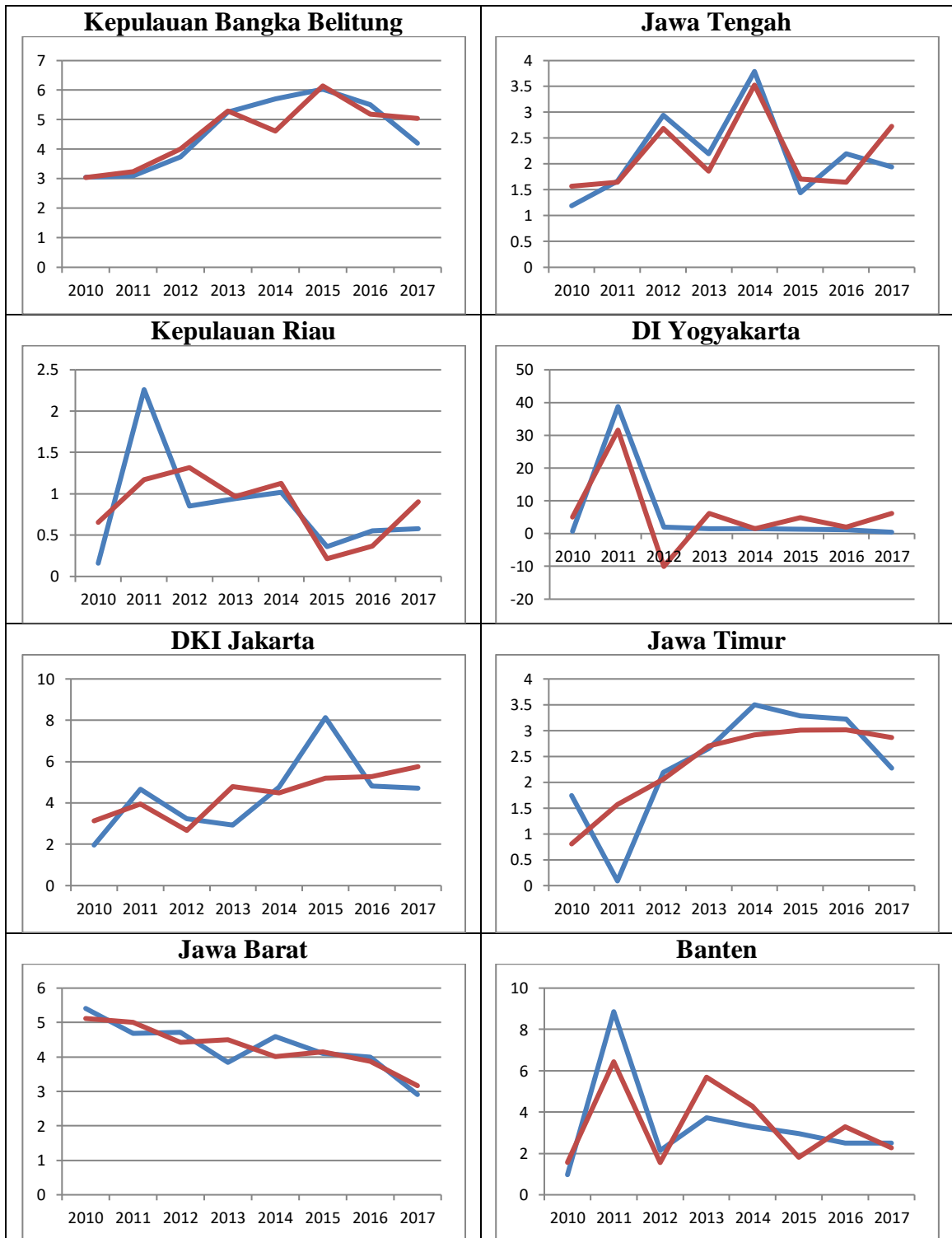
Lampiran 5. *Output* Estimasi Variabel Respon (y) (Lanjutan)

Provinsi	Tahun	y	\hat{y}
Maluku Utara	2010	0.96	1.169558
	2011	1.82	1.421631
	2012	0.97	0.988334
	2013	0.32	0.788961
	2014	0.71	0.643696
	2015	0.62	0.43466
	2016	0.44	0.69421
	2017	0.7	0.398951

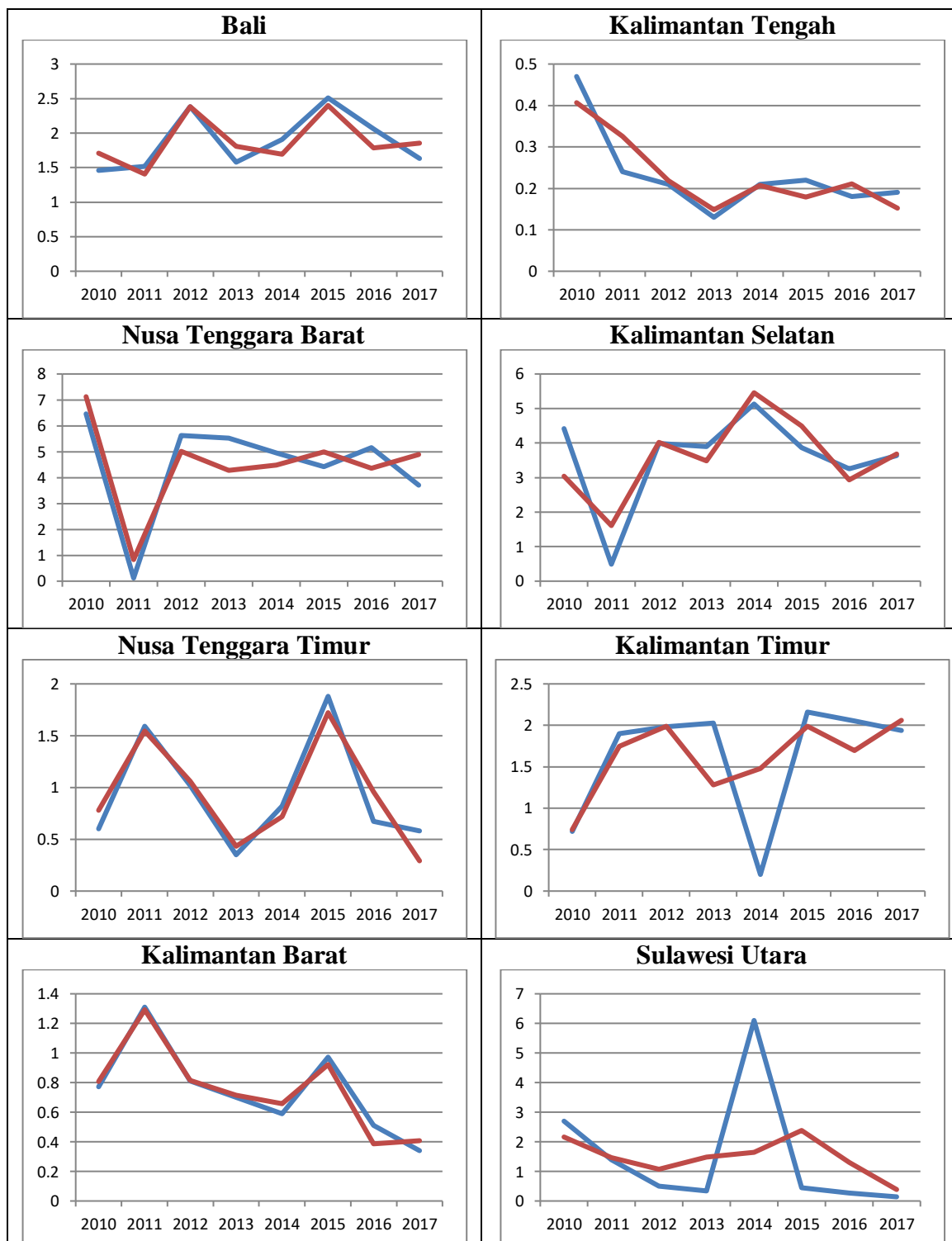
Lampiran 6. Plot antara y dan \hat{y} di Masing-Masing Provinsi di Indonesia



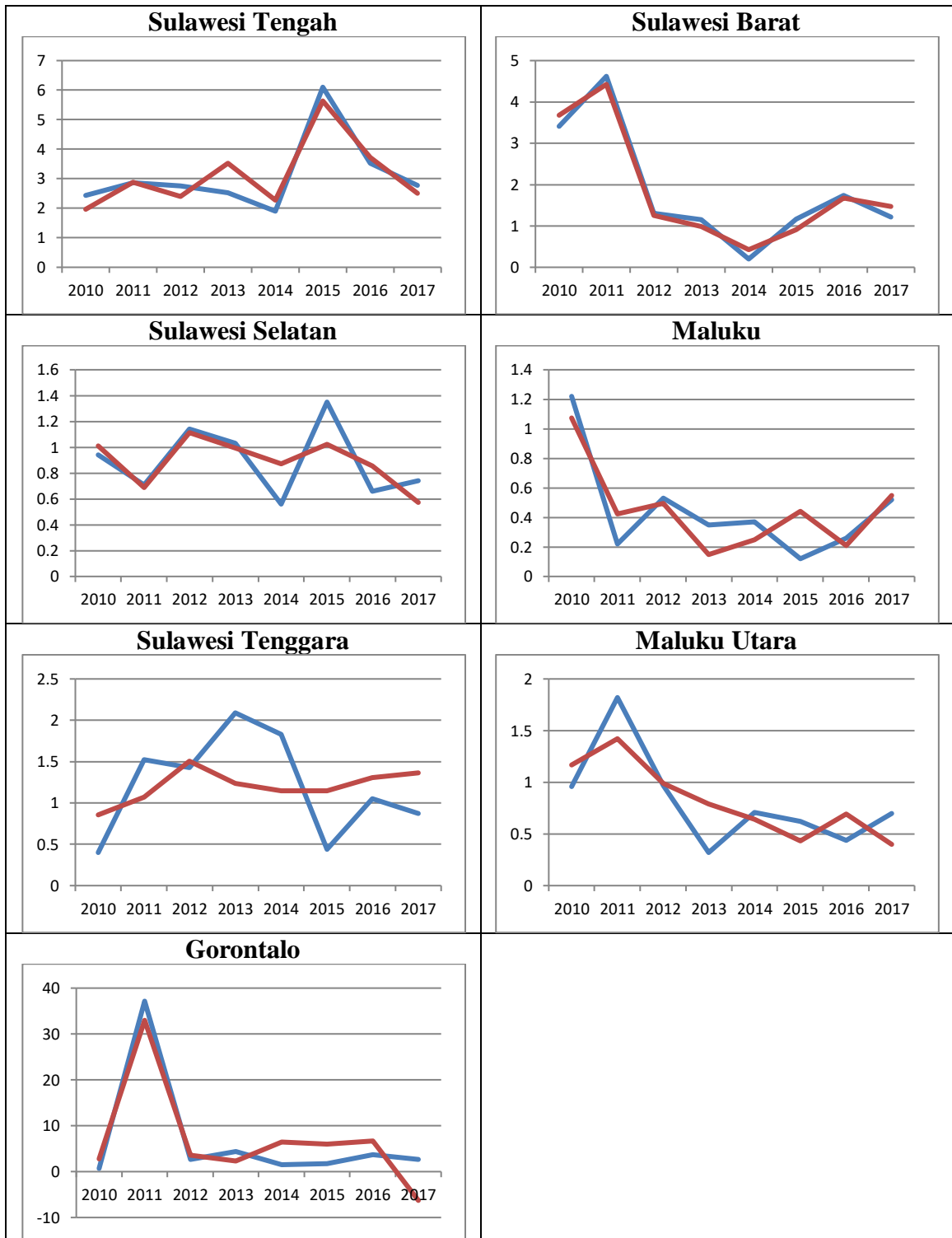
Lampiran 6. Plot antara y dan \hat{y} di Masing-Masing Provinsi di Indonesia (Lanjutan)



Lampiran 6. Plot antara y dan \hat{y} di Masing-Masing Provinsi di Indonesia (Lanjutan)



Lampiran 6. Plot antara y dan \hat{y} di Masing-Masing Provinsi di Indonesia (Lanjutan)



Lampiran 7. Estimasi Model Regresi Semiparametrik dengan Estimator Campuran Spline Truncated dan Deret Fourier Berdasarkan Data Series

1.	$\hat{y}_{i,2010} = 3,13 \times 10^{-16} + 0,0769x_{i,2010} - 0,03t_{i,2010} - 0,01246(t_{i,2010} - 59,5671)_+ -$ $0,03192z_{i,2010} - \frac{1}{2} \times 6,08434 - 0,19713 \cos z_{i,2010}$
2.	$\hat{y}_{i,2011} = 1,42 \times 10^{-15} - 0,1667x_{i,2011} + 0,24719t_{i,2011} + 0,07741(t_{i,2011} - 60,3971)_+ +$ $0,74471z_{i,2011} - \frac{1}{2} \times 5,38573 + 2,21647 \cos z_{i,2011}$
3.	$\hat{y}_{i,2012} = -1,88 \times 10^{-16} + 0,03179x_{i,2012} + 0,02294t_{i,2012} + 0,00088(t_{i,2012} - 60,4143)_+ +$ $0,06364z_{i,2012} - \frac{1}{2} \times 5,04841 - 0,77141 \cos z_{i,2012}$
4.	$\hat{y}_{i,2013} = -1,17 \times 10^{-15} + 0,07002x_{i,2013} - 0,02193t_{i,2013} - 0,01245(t_{i,2013} - 60,4143)_+ -$ $0,06324z_{i,2013} - \frac{1}{2} \times 3,7396 - 0,26609 \cos z_{i,2013}$
5.	$\hat{y}_{i,2014} = 2,48 \times 10^{-16} + 0,01514x_{i,2014} + 0,02592t_{i,2014} + 0,00971(t_{i,2014} - 57,1143)_+ -$ $0,05967z_{i,2014} + \frac{1}{2} \times 0,09443 - 0,42797 \cos z_{i,2014}$
6.	$\hat{y}_{i,2015} = 5,91 \times 10^{-16} + 0,05286x_{i,2015} + 0,01476t_{i,2015} + 0,0087(t_{i,2015} - 57,9257)_+ -$ $0,07461z_{i,2015} - \frac{1}{2} \times 4,53414 - 0,75501 \cos z_{i,2015}$
7.	$\hat{y}_{i,2016} = -1,52 \times 10^{-15} + 0,05231x_{i,2016} + 0,01181t_{i,2016} + 0,0084(t_{i,2016} - 58,98)_+ -$ $0,00811z_{i,2016} - \frac{1}{2} \times 6,82598 - 0,27679 \cos z_{i,2016}$
8.	$\hat{y}_{i,2017} = -2,31 \times 10^{-16} + 0,042358x_{i,2017} + 0,00051t_{i,2017} + 0,000411(t_{i,2017} - 58,983)_+ -$ $0,04118z_{i,2017} - \frac{1}{2} \times 3,40824 - 0,09032 \cos z_{i,2017}$

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini, mahasiswa Departemen Statistika FSAD ITS,

Nama : Anjar Wicitra Wening
NRP : 06211750010029
Program : Magister

Menyatakan bahwa data yang digunakan dalam Tesis ini merupakan data sekunder yang diambil dari publikasi Kementerian Kesehatan Republik Indonesia (Kemenkes RI) yaitu Profil Kesehatan Indonesia dari tahun 2010 sampai tahun 2017.

Surat ini dibuat dengan sebenar-benarnya. Apabila ada pemalsuan data maka saya siap menerima sanksi dengan aturan yang berlaku.

Surabaya, 31 Januari 2020

Mahasiswa

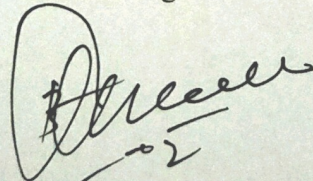


Anjar Wicitra Wening

NRP. 06211750010029

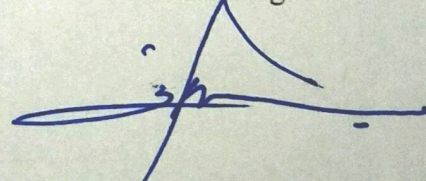
Mengetahui

Pembimbing I



Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
NIP. 19650603 198903 1 003

Pembimbing II



Dr. Ismaini Zain, M.Si.
NIP. 19600525 198803 2 001

BIODATA PENULIS



Anjar Wicitra Wening lahir di Sumenep pada tanggal 24 September 1995. Penulis merupakan anak pertama dari pasangan Bapak Mustofa dan Ibu Kusniah. Penulis memulai jenjang pendidikan formal di Sekolah Dasar Negeri 1 Batuan, Kabupaten Sumenep pada tahun 2001-2007. Kemudian penulis melanjutkan pendidikan ke Sekolah Menengah Pertama Negeri 4 Sumenep tahun 2007-2010. Penulis menyelesaikan pendidikan yang selanjutnya di Sekolah Menengah Atas Negeri 3 Pamekasan pada tahun 2010- 2013. Penulis melanjutkan pendidikan Sarjana (S1) di Universitas Airlangga Surabaya pada program studi Statistika sampai selesai pada tahun 2017. Pada tahun yang sama melanjutkan studi Pascasarjana (S2) di Institut Teknologi Sepuluh Nopember (ITS) Suraabaya, Fakultas Sains dan Analitika Data (FSAD), Departemen Statistika. Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini bisa menghubungi penulis melalui email: anjarwicitra@gmail.com.