



KAJIAN REGRESI PANEL SEMIPARAMETRIK PADA POLA NONLINEAR DAN PENERAPANNYA PADA DATA KEMISKINAN

CAHYANI ULFAH



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Kajian Regresi Panel Semiparametrik pada Pola Nonlinear dan Penerapannya pada Data Kemiskinan” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2024

Cahyani Ulfah
NIM G1501201003

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

CAHYANI ULFAH. Kajian Regresi Panel Semiparametrik pada Pola Nonlinear dan Penerapannya pada Data Kemiskinan. Dibimbing oleh ANIK DJURAI DAH dan HARI WIJAYANTO.

Data panel merupakan gabungan data antara data lintas individu dan waktu. Regresi panel biasanya menggunakan pendekatan parametrik dengan asumsi hubungan antar peubah bersifat linear. Namun, peubah respon dan peubah penjelas tidak selalu memiliki hubungan linear, melainkan juga bisa memiliki hubungan nonlinear. Data dengan pola nonlinear dapat ditangani menggunakan metode regresi nonparametrik. Salah satu metode regresi nonparametrik yang dapat digunakan adalah regresi spline terpenalti (P-spline). Metode P-spline dapat diformulasikan dan diduga ke dalam model campuran linear.

Komponen parametrik dan nonparametrik secara bersamaan dimungkinkan terdapat dalam model regresi panel, sehingga pendekatan semiparametrik bisa digunakan sebagai penyesuaian yang lebih baik terhadap karakteristik data. Model regresi panel semiparametrik menggunakan metode *within group* P-spline dengan pendekatan model campuran linear. Metode tersebut menggunakan P-spline ke dalam model campuran linear untuk menangkap pola nonlinear, sementara *within group* digunakan untuk mengendalikan heterogenitas antar unit yang tidak teramati. Oleh sebab itu, penelitian ini bertujuan untuk mengkaji kinerja regresi panel dengan pendekatan semiparametrik. Kajian dilakukan terhadap data simulasi yang memuat pola nonlinear dengan ukuran ragam yang berbeda. Hasil kajian tersebut kemudian diterapkan pada data kemiskinan di Pulau Jawa selama tahun 2018 sampai 2021. Lebih lanjut, penelitian ini juga mengidentifikasi peubah yang berpengaruh terhadap tingkat kemiskinan di Pulau Jawa.

Kajian simulasi dibatasi pada regresi panel pengaruh tetap dengan pendekatan parametrik, nonparametrik, dan semiparametrik. Simulasi dilakukan dengan membangkitkan tiga pola peubah penjelas yaitu pola linear, pola rata-rata tanpa tren, dan pola rata-rata dengan tren untuk masing-masing ukuran ragam (σ^2) yaitu 1 dan 5. Ketiga pola tersebut dikombinasikan dengan ukuran ragam yang diberikan sehingga membentuk 6 skenario model. Setiap skenario model diulang sebanyak 1000 kali perulangan.

Hasil kajian simulasi menunjukkan bahwa model regresi panel semiparametrik lebih baik digunakan pada pola rata-rata dengan tren ketika ragam $\sigma^2 = 1$. Model tersebut memiliki nilai *mean square error* (MSE) dan korelasi yang lebih baik dibandingkan regresi panel dengan pendekatan parametrik dan nonparametrik. Artinya, model regresi panel semiparametrik mampu menghasilkan prediksi yang cukup akurat dalam mengikuti pola nonlinear dan dapat menjelaskan sebagian besar keragaman data. Hasil dari kajian terapan menunjukkan bahwa regresi panel semiparametrik menghasilkan nilai MSE, R^2 , dan adjusted- R^2 yang lebih baik dibandingkan regresi panel.

Kata kunci: p-spline, regresi panel, semiparametrik, *within group*



SUMMARY

CAHYANI ULFAH. A Study of Semiparametric Panel Regression on Nonlinear Patterns and Its Application to Poverty Data. Supervised by ANIK DJURAIDAH, and HARI WIJAYANTO.

Panel data is a combination of data between individuals and time. Panel data regression usually uses a parametric approach with the assumption that the relationship between variables is linear. However, response variables and explanatory variables do not always have a linear relationship, but can also have a nonlinear relationship. Data with nonlinear patterns can be resolved using nonparametric regression methods. One of the nonparametric regression method that can be used is penalized spline regression (P-spline). The P-spline method can be formulated and estimated into a linear mixed model.

Parametric and nonparametric components are simultaneously possible in the panel regression, so that a semiparametric approach can be used as a better adjustment to the characteristics of the data. Semiparametric panel regression model uses the within-group P-spline method with a linear mixed model approach. The method uses P-spline into linear mixed model to capture nonlinear patterns, while within-group is used to control for unobserved heterogeneity between units. Therefore, this research aims to examine the performance of panel regression using a semiparametric approach. The study was carried out on simulated data that contained nonlinear patterns with different variance sizes. The results of this study were then applied to poverty data on the island of Java from 2018 to 2021. Furthermore, this research also identified variables that influence the level of poverty on the island of Java.

The simulation study is limited to fixed effect panel regression with parametric, nonparametric and semiparametric approaches. The simulation is carried out by generating three patterns of explanatory variables, namely linear pattern, average patterns without trend, and average patterns with trend for each variance (σ^2), which are 1 and 5. These three patterns will be combined with the given variance to form 6 model scenarios. Each model scenario was repeated 1000 times.

The results of the simulation study show that the semiparametric panel regression model is better used for average patterns with trends when the variance $\sigma^2 = 1$. This model has better mean square error (MSE) and correlation values compared to panel regression with parametric and nonparametric approaches. This means that the semiparametric panel regression model is able to produce quite accurate predictions following nonlinear patterns and can explain most of the variability in the data. The results of the applied study show that semiparametric panel regression produces better MSE, R^2 , and adjusted- R^2 values than panel regression.

Keywords: panel regression, p-spline, semiparametric, within-group



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

KAJIAN REGRESI PANEL SEMIPARAMETRIK PADA POLA NONLINEAR DAN PENERAPANNYA PADA DATA KEMISKINAN

CAHYANI ULFAH

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

IPB University

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Ir. Indahwati, M.Si



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Tesis : Kajian Regresi Panel Semiparametrik pada Pola Nonlinear dan Penerapannya pada Data Kemiskinan

Nama : Cahyani Ulfah

NIM : G1501201003

Disetujui oleh

Pembimbing 1:

Prof. Dr. Ir. Anik Djuraidah, M.S.



Pembimbing 2:

Prof. Dr. Ir. Hari Wijayanto, M.Si.



Diketahui oleh

Ketua Program Studi:

Dr. Agus M. Soleh, S.Si., M.T.

NIP 197503151999031004



Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam:

Dr. Berry Juliandi, S.Si., M.Si.

NIP 197807232007011001



Tanggal Ujian: 3 Juni 2024

Tanggal Lulus:



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga tesis ini berhasil diselesaikan. Topik yang dipilih dalam penelitian ini adalah pemodelan pada data berpola nonlinear, dengan judul “Kajian Regresi Panel Semiparametrik pada Pola Nonlinear dan Penerapannya pada Data Kemiskinan”.

Penyelesaian tesis ini tidak terlepas dari doa, motivasi, bimbingan, dan kerja sama berbagai pihak. Penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Ir. Anik Djuraidah, M.S. selaku ketua komisi pembimbing dan Prof. Dr. Ir. Hari Wijayanto, M.Si. selaku anggota komisi pembimbing yang telah membimbing dengan penuh kesabaran dan banyak memberikan arahan serta saran yang membangun.
2. Dr. Ir. Indahwati, M.Si. selaku penguji luar komisi pembimbing dan Dr. Farit Mochamad Afendi, S.Si., M.Si. selaku pemimpin ujian tesis yang telah memberikan masukan dan saran yang sangat membangun untuk perbaikan tesis ini.
3. Ayahanda H. Syahidin Rambe dan Ibunda tercinta Hj. Ir. Rosmawati, M.Si. atas dukungan, doa, motivasi dan kasih sayangnya.
4. Segenap dosen dan staf Departemen Statistika dan Sains Data IPB atas ilmu, bantuan, dan kerjasamanya selama perkuliahan.
5. Teman-teman Program Studi Statistika dan Sains Data IPB angkatan 2020 yang telah memberikan semangat dan bersedia berbagi pengalaman dan ilmunya yang sangat luar biasa.
6. Terakhir penulis ucapkan terima kasih kepada diri sendiri yang telah bertahan dan berhasil melalui proses panjang perkuliahan, penelitian, dan menyelesaikan program magister ini dengan sebaik-baiknya.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2024

Cahyani Ulfah



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Model Umum Data Panel	3
2.2 Regresi Panel Pengaruh Tetap	4
2.3 Regresi Spline Terpenalti (P-spline)	4
2.4 Kesetaraan P-spline dengan Model Campuran Linear	6
2.5 Regresi Panel Pengaruh Tetap dengan Pendekatan Semiparametrik	7
III METODE	9
3.1 Data Simulasi	9
3.2 Data Empiris	10
3.3 Tahapan Analisis Data Simulasi	10
3.4 Tahapan Analisis Data Empiris	12
IV HASIL DAN PEMBAHASAN	13
4.1 Hasil Kajian Simulasi	13
4.1.1 Perbandingan Nilai MSE	13
4.1.2 Perbandingan Korelasi	14
4.2 Hasil Penerapan Data Empiris	15
4.2.1 Eksplorasi Data	15
4.2.2 Menentukan Komponen Parametrik dan Nonparametrik	16
4.2.3 Spesifikasi Model Panel	17
4.2.4 Pemilihan Jumlah Simpul Spline	18
4.2.5 Perbandingan Model	18
V SIMPULAN	20
5.1 Simpulan	20
DAFTAR PUSTAKA	21
LAMPIRAN	24
RIWAYAT HIDUP	28



DAFTAR TABEL

1	Struktur data panel	3
2	Skenario simulasi	9
3	Peubah yang digunakan	10
4	Korelasi antara peubah respon untuk setiap peubah penjelas	17
5	Hasil uji Chow dan uji Hausman	17
6	Nilai VIF peubah penjelas	17
7	Hasil optimasi jumlah simpul untuk semua derajat spline	18
8	Perbandingan model	18

DAFTAR GAMBAR

1	Pola rataaan tanpa tren	10
2	Pola rataaan dengan tren	10
3	Diagram alir tahapan analisis untuk data simulasi	11
4	Diagram alir tahapan analisis untuk data empiris	12
5	MSE seluruh skenario model pada derajat spline 1, 2 dan 3	14
6	Korelasi seluruh skenario model pada derajat spline 1, 2 dan 3	15
7	Peta tematik tingkat kemiskinan pulau Jawa 2018-2021	16
8	Diagram pencar antara peubah respon dengan peubah penjelas	16

LAMPIRAN

1	Plot pemulus spline pada derajat spline 1, derajat spline 2, dan derajat spline 3	25
2	Penduga parameter efek tetap p-spline campuran linear pada regresi panel nonparametrik derajat spline 2	26
3	Penduga parameter efek acak p-spline campuran linear pada regresi panel nonparametrik derajat spline 2	26