

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam statistika, data longitudinal mengacu pada kumpulan data yang berasal dari pengukuran berulang terhadap sejumlah individu (*unit cross-sectional*) dalam rentang waktu tertentu secara berturut-turut (Hermawan, 2021). Data jenis ini kerap digunakan dalam berbagai disiplin ilmu, seperti di bidang kesehatan untuk memantau kondisi pasien secara berkelanjutan, di bidang ekonomi untuk menganalisis dinamika pendapatan atau pola konsumsi rumah tangga, di bidang lingkungan untuk mengamati perubahan kualitas udara atau iklim, serta di ranah sosial untuk menelusuri evolusi opini publik dari waktu ke waktu. Penelitian longitudinal memberikan keunggulan metodologis karena mampu merekam dinamika perubahan seiring waktu serta mengidentifikasi efek keterlambatan (*lag effect*), yang umumnya tidak dapat terdeteksi melalui pendekatan data *cross-sectional* (Dani, Ni'matuzzahroh, et al., 2021).

Dalam konteks data yang melibatkan dimensi waktu, data longitudinal perlu dibedakan dari data *time series* dan data panel agar tidak menimbulkan kekeliruan konseptual. Data *time series* merupakan data yang diamati secara berurutan menurut waktu pada satu unit pengamatan, sehingga secara umum hanya dinyatakan dengan satu indeks waktu yaitu y_t , misalnya inflasi bulanan Indonesia atau harga saham harian suatu perusahaan (Esling et al., 2026). Berbeda dengan itu, data panel merupakan data yang memiliki dua indeks, yaitu indeks unit (i) dan indeks waktu (t), sehingga observasinya dapat dinyatakan dalam bentuk y_{it} . Dalam pengertian ini, data panel menekankan struktur data yang menggabungkan dimensi

cross-sectional dan dimensi waktu.. Sementara itu, data longitudinal juga memiliki bentuk y_{it} , tetapi penekanannya bukan semata-mata pada struktur data, melainkan pada analisis perubahan variabel respons pada unit yang sama secara berulang dari waktu ke waktu (Taddei, 2025). Dalam struktur ini terdapat identitas unit (i), waktu pengamatan (t), dan variabel respons (y_{it}) yang diamati pada unit ke- i pada waktu ke- t . Dengan demikian, fokus utama data longitudinal bukan sekadar adanya dua dimensi data, melainkan pada analisis perubahan nilai respons dalam unit yang sama dari waktu ke waktu atau perubahan intraunit. Dengan kata lain, setiap data longitudinal dapat dipandang sebagai data panel secara struktur, tetapi tidak setiap data panel secara substantif berorientasi pada analisis perubahan intraunit sebagaimana dalam studi longitudinal.

Karakteristik dasar dari data longitudinal terletak pada adanya pengamatan berulang terhadap unit yang sama, sehingga memungkinkan munculnya heterogenitas antarindividu dan ketergantungan antarobservasi dalam unit yang sama sepanjang waktu. Dalam praktiknya, sebagian data longitudinal dapat memperlihatkan pola nonstasioner, misalnya ketika rata-rata atau varians berubah seiring waktu, namun kondisi tersebut bukan merupakan syarat yang harus selalu melekat pada seluruh data longitudinal. Selain itu, data ini juga sering memperlihatkan perilaku *burst like*, yakni munculnya lonjakan tiba-tiba atau variabilitas ekstrem yang terjadi dalam kurun waktu tertentu (Hyun et al., 2022; Metwally et al., 2022).

Karakteristik data longitudinal yang melibatkan pengamatan berulang terhadap unit yang sama dalam rentang waktu tertentu secara alami menyebabkan munculnya ketergantungan temporal antar observasi. Hal ini terjadi karena status

atau nilai variabel yang diamati pada suatu waktu umumnya masih dipengaruhi oleh kondisi sebelumnya, sehingga nilai-nilai observasi tidak lagi bersifat independen. Karena hal tersebut, autokorelasi menjadi masalah utama yang sering muncul dalam analisis data longitudinal, yaitu korelasi antara nilai suatu variabel pada satu waktu dengan nilai variabel yang sama pada waktu sebelumnya (Puspitasari et al., 2023). Autokorelasi muncul karena adanya ketergantungan waktu alami dalam fenomena yang diamati, dan ini tidak dapat dihindari dalam struktur data longitudinal. Pengabaian terhadap autokorelasi dalam analisis data longitudinal dapat mengakibatkan bias pada hasil estimasi serta menurunkan efisiensinya. Selain itu, autokorelasi dalam data semacam ini tidak selalu bersifat konstan, melainkan dapat bervariasi antar fase waktu atau periode pengamatan, sehingga menambah tingkat kompleksitas yang harus ditangani oleh metode analisis yang diterapkan.

Salah satu metode analisis dalam statistika yang digunakan untuk mengeksplorasi pola hubungan antara satu variabel dengan satu atau lebih variabel, yaitu analisis regresi. Dalam proses estimasi kurva regresi, dikenal tiga pendekatan utama, yakni regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Dalam penerapannya, pendekatan parametrik seperti *Linear Mixed Models* dan *Generalized Estimating Equations* sering digunakan untuk mengakomodasi autokorelasi dalam data longitudinal. Namun, pendekatan ini memiliki sejumlah keterbatasan, khususnya ketika data memperlihatkan pola non-stasioner atau hubungan non-linier yang kompleks (Garcia & Marder, 2017). Metode parametrik umumnya mengasumsikan bentuk hubungan yang spesifik antar variabel, serta struktur varians dan korelasi yang konstan sepanjang waktu, yang kurang tepat untuk menangani data dengan dinamika *burst* atau perubahan antar

fase waktu. Pendekatan regresi semiparametrik digunakan ketika sebagian struktur fungsi regresi telah diketahui secara eksplisit, sementara sisanya tetap tidak ditentukan secara pasti. Sebaliknya, regresi nonparametrik diterapkan dalam kondisi di mana bentuk kurva regresi dianggap sepenuhnya tidak diketahui dan perlu diestimasi secara fleksibel dari data. (Dani, Ni'matuzzahroh, et al., 2021). Dengan demikian, pendekatan regresi nonparametrik menjadi alternatif yang lebih fleksibel karena tidak memerlukan asumsi eksplisit mengenai bentuk hubungan antara variabel bebas dan respon.

Dalam pemodelan regresi nonparametrik, serangkaian estimator telah diinovasi oleh para peneliti, diantaranya yaitu *Spline*, Kernel, dan Deret Fourier (Sriliana et al., 2024). Dalam penelitian ini, perhatian difokuskan pada pendekatan *Spline* dan *Kernel Regression* karena keduanya memiliki fleksibilitas yang tinggi dalam menangkap pola hubungan yang kompleks, termasuk ketika perubahan respons terhadap waktu tidak mengikuti bentuk parametrik tertentu. (Metwally et al., 2022). Pada konteks data longitudinal, fleksibilitas tersebut menjadi relevan karena pola perkembangan respons pada unit yang diamati berulang dapat bersifat beragam dan tidak selalu mudah direpresentasikan dengan model parametrik sederhana. Terdapat beberapa jenis *spline* yang umum digunakan, seperti *Cubic Spline*, *B-spline*, *Penalized Spline*, *Spline Truncated*, dan masih banyak lagi. Masing-masing jenis memiliki karakteristik tersendiri dalam menangani kompleksitas struktural pada data. *Cubic Spline* menghasilkan kurva halus dengan kekontinuan hingga turunan kedua dan cocok digunakan pada data dengan tren gradual (Sofiyani & Permanasari, 2023). *B-spline* memiliki basis lokal yang efisien secara komputasi dan mampu meredam efek global dari perubahan lokal data (ARIESTA et al., 2021).

Penalized Spline (P-spline) merupakan pengembangan dari *B-spline* yang dikombinasikan dengan penalti terhadap kekasaran fungsi untuk menghindari *overfitting*. *P-spline* banyak digunakan dalam konteks regresi karena mampu menghasilkan kurva yang lebih halus dengan mengontrol kompleksitas model secara langsung melalui parameter penalti (Kurniasari et al., 2019). Namun demikian, *Cubic Spline*, *B-spline*, dan *P-spline* memiliki keterbatasan ketika diterapkan pada data longitudinal yang menunjukkan perubahan lokal ekstrem dan pola autokorelasi yang tidak konstan. Kurva yang dihasilkan cenderung terlalu halus secara global atau terlalu dikendalikan oleh penalti, sehingga kurang adaptif dalam menangkap dinamika temporal yang tajam.

Dalam konteks tersebut, *Spline Truncated* lebih sesuai untuk dianalisis karena dibangun dari fungsi basis polinomial yang dipotong (*truncated*), yang hanya aktif setelah titik *knot* tertentu. Struktur ini memberikan fleksibilitas tinggi untuk memodelkan perubahan lokal yang tajam dan menangkap dinamika temporal kompleks dalam data longitudinal (Puspitasari et al., 2023). Keunggulan metode ini terletak pada kemampuannya menangani hubungan non-linear serta autokorelasi antar pengamatan secara fleksibel. Namun, metode ini memiliki kekurangan berupa kompleksitas komputasi tinggi dan sensitivitas terhadap pemilihan titik *knot*. Sementara itu, regresi *Kernel Regression* merupakan pendekatan nonparametrik lain yang populer. Metode regresi *Kernel Regression* di definisikan sebagai pendekatan nonparametrik yang memanfaatkan fungsi Kernel untuk mengestimasi hubungan antara variabel tanpa menetapkan bentuk kurva regresi secara eksplisit (Fadilah & Kudus, 2023). Kelebihan dalam metode ini ialah fleksibel dalam menghadapi data yang fluktuatif dan mampu menghasilkan estimasi yang stabil

meskipun pada data tersensor. Namun, kekurangannya terletak pada sensitivitas terhadap pemilihan parameter *bandwidth* (nilai h), yang dapat memengaruhi akurasi hasil imputasi jika tidak dipilih dengan tepat.

Berdasarkan penelitian yang dilakukan oleh Dani et al (2021) mengungkap bahwa metode *Spline Truncated* efektif dalam memodelkan hubungan non-linier dan tren jangka panjang pada data kesehatan lansia. Pendekatan ini memungkinkan identifikasi perubahan gradual yang tidak terdeteksi oleh model parametrik konvensional. Di sisi lain, studi yang dilakukan oleh Anuradha et al (2024) menunjukkan bahwa *Kernel Regression* memiliki keunggulan dalam mendeteksi pola lokal dan fluktuasi temporal tajam (*burst-like*), yang sangat berguna dalam analisis spasial-temporal data kecelakaan lalu lintas. Hasil-hasil tersebut memperkuat argumen bahwa kedua metode memiliki karakteristik dan kekuatan yang berbeda. Namun demikian, literatur yang secara eksplisit membandingkan kedua metode tersebut dalam konteks data longitudinal yang mengandung variasi tingkat autokorelasi masih sangat terbatas. Sebagian besar penelitian hanya mengevaluasi salah satu metode dalam konteks data spesifik, tanpa mempertimbangkan keunggulan relatif metode alternatif dalam kondisi data yang berbeda. Selain itu, fokus analisis yang terlalu terpusat pada aspek akurasi prediktif telah menyebabkan kurangnya perhatian terhadap kemampuan masing-masing metode dalam mengidentifikasi struktur pola temporal yang mendasari data, seperti dinamika autokorelasi.

Keterbatasan kajian komparatif yang sistematis dan berbasis pada keragaman karakteristik data menimbulkan adanya kesenjangan ilmiah dalam memahami secara utuh performa kedua metode tersebut pada data longitudinal yang memiliki

tingkat autokorelasi berbeda. Oleh karena itu, penelitian ini dilakukan untuk membandingkan kedua metode tersebut melalui studi simulasi pada beberapa skenario autokorelasi, sehingga dapat diperoleh gambaran yang lebih sistematis mengenai keunggulan dan keterbatasan masing-masing metode dalam menghadapi karakteristik data yang beragam. Selain melalui simulasi, penelitian ini juga menerapkan kedua metode pada data riil longitudinal sebagai pengujian empiris tambahan, agar hasil yang diperoleh tidak hanya berlaku pada kondisi yang terkontrol, tetapi juga memiliki relevansi pada situasi nyata. Berangkat dari fenomena tersebut, penulis termotivasi untuk melakukan penelitian mengenai Analisis Data Longitudinal dengan Berbagai Tingkat Autokorelasi dengan Metode Nonparametrik *Spline Truncated* dan *Kernel Regression*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka rumusan masalah penelitian ini adalah sebagai berikut.

1. Bagaimana performa metode nonparametrik *Spline Truncated* dalam menganalisis data longitudinal dengan berbagai tingkat autokorelasi?
2. Bagaimana performa metode nonparametrik *Kernel Regression* dalam menganalisis data longitudinal dengan berbagai tingkat autokorelasi?
3. Metode manakah yang memberikan kinerja paling baik berdasarkan kriteria evaluasi yang digunakan pada kondisi tersebut?
4. Sejauh mana hasil evaluasi komparatif pada data simulasi bersifat konsisten ketika kedua metode diterapkan pada data longitudinal empiris (data riil), sehingga mendukung validitas eksternal temuan simulasi?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengevaluasi performa metode nonparametrik *Spline Truncated* dalam menganalisis data longitudinal dengan variasi tingkat autokorelasi.
2. Mengevaluasi performa metode nonparametrik *Kernel Regression* dalam menganalisis data longitudinal dengan variasi tingkat autokorelasi.
3. Membandingkan kinerja kedua metode dan menentukan metode yang paling optimal pada setiap kondisi autokorelasi sesuai ukuran evaluasi.
4. Menerapkan kedua metode pada data longitudinal empiris dan mengevaluasi apakah pola/keunggulan metode yang diperoleh dari simulasi tetap konsisten pada data nyata.

1.4 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode statistik secara praktis maupun teoritis.

1. Manfaat Teoritis

Diharapkan penelitian ini berkontribusi dalam perkembangan ilmu statistik, khususnya dalam hal analisis data longitudinal menggunakan metode regresi nonparametris. Selain itu, hasil penelitian ini diharapkan menjadi referensi bagi akademisi dan peneliti dalam memilih teknik yang paling sesuai berdasarkan karakteristik data, serta mendorong pengembangan metode yang lebih optimal dalam menangani struktur data yang kompleks.

2. Manfaat Praktis

Penelitian ini diharapkan dapat memberikan manfaat praktis, yakni:

a. Bagi Peneliti dan Akademisi

Menyediakan panduan metodologis bagi peneliti dan akademisi dalam memilih metode *Spline Truncated* atau *Kernel Regression* yang paling sesuai untuk analisis data longitudinal dengan berbagai tingkat autokorelasi.

b. Bagi Praktisi di Berbagai Bidang

Memberikan wawasan bagi ekonom, analis kesehatan, dan ilmuwan sosial dalam memilih teknik analisis yang tepat untuk mengolah data longitudinal serta mempermudah pengambilan keputusan berbasis data dalam ekonomi, kesehatan, dan transportasi, terutama yang melibatkan pola perubahan seiring waktu.

c. Bagi Pengambil Kebijakan di Bidang Terkait

Dengan pemilihan metode yang tepat, hasil analisis data longitudinal menjadi lebih akurat dan andal, membantu pengambil kebijakan dalam membuat keputusan berbasis bukti serta mendukung pengembangan sistem prediksi di berbagai bidang, seperti tren ekonomi, perencanaan kesehatan masyarakat, dan analisis keselamatan transportasi.

1.5 Pembatasan Masalah

Adapun yang menjadi batas permasalahan dalam penelitian ini adalah sebagai berikut.

1. Jumlah subjek (unit pengamatan) yang digunakan dalam data simulasi dibatasi sebanyak 30 subjek, dengan masing-masing subjek memiliki 5 titik waktu pengamatan.

2. Data yang dibangkitkan hanya mencakup satu variabel respons dan satu variabel prediktor.
3. Struktur autokorelasi yang digunakan dalam pembangkitan data terbatas pada autokorelasi linier.
4. Pada setiap tingkat autokorelasi yang ditetapkan, proses pembangkitan data simulasi direplikasi sebanyak 50 kali guna memperoleh estimasi kinerja metode yang lebih stabil serta meminimalkan pengaruh variasi acak.

