



Evaluasi Kinerja Regresi Lokal pada Pemodelan Hubungan Nonlinier (Studi Simulasi Monte Carlo)

Linda Rassiyanti^{1*}, Rohimatul Anwar², Sella Nofriska³

¹Program Studi Sains Data, Institut Teknologi Sumatera, Indonesia

²Program Studi Kimia, Universitas Lampung, Indonesia

³Program Studi Kimia, Institut Agama Islam Negeri Sorong, Indonesia

*Penulis Korespondensi: linda.rassiyanti@sd.itera.ac.id

Abstract. LOESS (*Locally Estimated Scatterplot Smoothing*) is a nonparametric approach for modeling nonlinear relationships without explicitly specifying a functional form. Its performance depends on the span, sample size, and error variability. This study evaluates local regression against classical linear regression and cubic polynomial regression across three types of nonlinear relationships (sinusoidal, local spike, and Runge function) through a Monte Carlo simulation with 300 replications per scenario. The factors varied include sample size ($n = 30, 50, 100, 200$), sigma (0.15 and 0.40), and LOESS span (0.3, 0.5, and 0.7). Model performance is evaluated using the Mean Squared Error (MSE), along with its decomposition into squared bias and variance. The results show that a narrow span (0.3) consistently yields the lowest MSE for functions with sharp curvature and local spikes, with an average MSE reduction of up to 91.7% compared with linear regression for the Runge function. For relatively smooth functions, cubic polynomial regression slightly outperforms LOESS. Larger sample sizes reduce the MSE of nonparametric methods by decreasing variance, whereas excessively wide spans increase bias due to oversmoothing. These findings highlight the importance of adaptive span selection and demonstrate the advantages of local regression in modeling complex and heterogeneous nonlinear relationships.

Keywords: Bias–Variance; Local Regression; LOESS; Monte Carlo Simulation; Nonlinear Relationships.

Abstrak. LOESS (*Locally Estimated Scatterplot Smoothing*) adalah pendekatan nonparametrik untuk memodelkan hubungan nonlinier tanpa menetapkan bentuk fungsional secara eksplisit. Kinerjanya bergantung pada span, ukuran sampel, dan keragaman galat. Penelitian ini mengevaluasi dan membandingkan regresi lokal, regresi linear klasik, dan polinomial derajat tiga pada tiga bentuk hubungan nonlinier (sinusoidal, lonjakan lokal, fungsi Runge) melalui simulasi Monte Carlo dengan 300 replikasi per skenario. Faktor yang divariasikan yaitu, ukuran sampel ($n = 30, 50, 100, 200$), sigma (0,15 dan 0,40), dan span LOESS (0,3; 0,5; 0,7). Kinerja dinilai dengan *Mean Squared Error* (MSE) beserta dekomposisi bias kuadrat dan varians. Hasilnya, span sempit (0,3) konsisten memberikan MSE terendah pada fungsi berkelengkungan tajam dan berlonjakan lokal, dengan penurunan MSE rata-rata hingga 91,7 persen dibandingkan regresi linear pada fungsi Runge. Pada fungsi yang relatif mulus, polinomial derajat tiga sedikit mengungguli LOESS. Ukuran sampel besar menurunkan MSE metode nonparametrik melalui reduksi varians, sedangkan span terlalu lebar meningkatkan bias akibat *oversmoothing*. Temuan ini menegaskan pentingnya pemilihan span adaptif dan keunggulan regresi lokal untuk hubungan nonlinier kompleks dan heterogen.

Kata kunci: Bias-Varians; Hubungan Nonlinier; LOESS; Regresi Lokal; Simulasi Monte Carlo.

1. LATAR BELAKANG

Analisis regresi merupakan alat statistik mendasar untuk memodelkan hubungan antara peubah respons dan prediktor. Regresi linear klasik banyak digunakan karena mudah diinterpretasikan dan dihitung, tetapi banyak fenomena di bidang sains, teknik, ekonomi, maupun ilmu sosial menunjukkan pola yang tidak dapat diestimasi oleh model linear, misalnya pola musiman, kurva pertumbuhan yang melandai, atau perubahan rezim pada titik tertentu.

Ketika bentuk fungsional tidak diketahui, pendekatan parametrik berisiko mengalami kesalahan spesifikasi. Regresi polinomial rentan terhadap osilasi pada derajat tinggi (fenomena Runge) dan kurang fleksibel untuk perubahan kelengkungan lokal. Regresi lokal menawarkan

alternatif dengan mengestimasi fungsi di sekitar titik evaluasi menggunakan subset data yang berdekatan, tanpa menetapkan bentuk tertentu.

Salah satu teknik regresi lokal yang paling banyak diterapkan adalah LOESS/LOWESS (*Locally Weighted Scatterplot Smoothing*), yaitu regresi tertimbang lokal dengan bobot yang menurun seiring jarak dari titik evaluasi (James *et al.*, 2021). Fleksibilitasnya bergantung pada parameter pemulusan (*span*) yang mengatur proporsi data pada setiap estimasi lokal. Pemilihan *span* yang terlalu kecil menghasilkan varians tinggi (*overfitting*), sedangkan *span* yang terlalu besar menghasilkan bias tinggi (*oversmoothing*). *Trade-off* bias-variens inilah inti permasalahan penerapan regresi lokal (James *et al.*, 2021; Barreiro-Ures *et al.*, 2026).

Kajian teoritis mengenai sifat asimtotik regresi lokal telah berkembang pesat (Cattaneo *et al.*, 2024; Kharoua *et al.*, 2024), tetapi kajian empiris pada sampel berhingga (*finite sample*) yang membandingkannya dengan metode parametrik pada berbagai bentuk hubungan nonlinier masih terbatas. Simulasi Monte Carlo memungkinkan evaluasi yang terkontrol karena fungsi regresi sebenarnya diketahui, sehingga kinerja estimator dapat diukur langsung melalui *mean squared error* (MSE), bias, dan varians (Boulesteix *et al.*, 2020; Morris *et al.*, 2019).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengevaluasi kinerja regresi lokal (LOESS) dibandingkan dengan regresi linear dan regresi polinomial derajat tiga dalam memodelkan berbagai bentuk hubungan nonlinier, menganalisis pengaruh ukuran sampel dan tingkat keragaman galat terhadap akurasi estimasi masing-masing metode, dan mengkaji *trade-off bias-variens* pada regresi lokal akibat variasi parameter *span*.

2. KAJIAN TEORITIS

Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) merupakan metode estimasi parameter regresi yang bertujuan meminimumkan jumlah kuadrat residual antara nilai observasi dan nilai prediksi model. Menurut Montgomery *et al.* (2021), metode kuadrat terkecil merupakan pendekatan mendasar dalam analisis regresi untuk mengestimasi hubungan antara variabel respons dan satu atau lebih variabel prediktor. Dalam regresi linear sederhana, hubungan tersebut dinyatakan sebagai

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

dengan Y_i merupakan variabel respons, X_i merupakan variabel prediktor, β_0 adalah intersep, β_1 adalah koefisien regresi, dan ε_i adalah galat acak. Estimasi parameter diperoleh dengan meminimumkan fungsi objektif

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

Montgomery *et al.* (2021) menguraikan estimasi, pengujian hipotesis, dan evaluasi kecocokan model regresi linear, sedangkan Fox dan Weisberg (2019) menekankan pentingnya analisis diagnostik untuk memeriksa asumsi model. OLS mudah diinterpretasikan, tetapi ketepatannya bergantung pada kesesuaian bentuk fungsional yang diasumsikan; pada hubungan nonlinier, model linear tidak mampu menangkap kelengkungan data. Meskipun demikian, OLS tetap relevan sebagai tolok ukur (*baseline*) bagi metode yang lebih fleksibel (James *et al.*, 2021).

Regresi Nonparametrik

Regresi nonparametrik merupakan pendekatan statistik yang digunakan untuk memodelkan hubungan antara variabel respons dan prediktor tanpa menetapkan bentuk fungsional tertentu secara ketat sejak awal. Secara umum, model regresi nonparametrik dapat dituliskan sebagai

$$Y_i = f(X_i) + \varepsilon_i$$

dengan $f(X_i)$ merupakan fungsi regresi yang tidak diketahui dan diestimasi berdasarkan data pengamatan, sedangkan ε_i merupakan galat acak. Berbeda dengan regresi parametrik yang menggunakan bentuk fungsi tertentu dengan sejumlah parameter terbatas, regresi nonparametrik memberikan fleksibilitas yang lebih besar dalam menangkap pola hubungan yang kompleks (Efron & Hastie, 2016; James *et al.*, 2021).

Regresi kernel, regresi lokal, dan *smoothing splines* merupakan pendekatan nonparametrik yang umum digunakan (Eilers & Marx, 2021; James *et al.*, 2021; Wood, 2017). Pada seluruh pendekatan tersebut, parameter penghalusan menentukan karakter kurva: penghalusan yang terlalu kuat menghilangkan pola penting, sedangkan yang terlalu lemah membuat kurva mengikuti variasi acak (Araveeporn, 2019; James *et al.*, 2021). Oleh karena itu, parameter penghalusan umumnya dipilih secara *data-driven*, misalnya melalui validasi silang (Araveeporn, 2019; Barreiro-Ures *et al.*, 2026).

Regresi Lokal

Regresi lokal merupakan salah satu pendekatan regresi nonparametrik yang mengestimasi fungsi regresi berdasarkan pengamatan di sekitar titik prediksi. Berbeda dari regresi global yang menggunakan satu fungsi untuk merepresentasikan seluruh data, regresi lokal membentuk estimasi pada setiap titik berdasarkan pengamatan di sekitarnya. Pendekatan ini memungkinkan model mengikuti perubahan pola hubungan yang terjadi pada rentang

prediktor yang berbeda (James *et al.*, 2021). Sifat asimtotik estimator regresi linear lokal, termasuk konvergensi hampir lengkap dan konsistensi seragam, telah dikaji secara formal (Kharoua *et al.*, 2024), dan kerangka regresi lokal juga telah diperluas untuk mengestimasi fungsi distribusi dan kerapatan yang adaptif terhadap titik batas (Cattaneo *et al.*, 2024). Regresi polinomial lokal juga banyak dimanfaatkan dalam ekonometrika, misalnya pada *regression discontinuity designs* (Cattaneo & Titiunik, 2022), dan inferensi yang memperhitungkan efek estimasi bias telah dikembangkan untuk metode ini (Calonico *et al.*, 2018).

Pada *locally weighted regression*, pengamatan yang lebih dekat dengan titik estimasi diberi bobot lebih besar sehingga kurva dapat menyesuaikan karakteristik data pada berbagai rentang prediktor tanpa satu bentuk fungsi global (James *et al.*, 2021). Secara matematis, regresi lokal dapat dirumuskan melalui minimisasi fungsi kuadrat terbobot sebagai berikut:

$$\min_{\beta(x)} \sum_{i=1}^n \omega_i(x) \left[Y_i - \sum_{j=0}^p \beta_j(x) (X_i - x)^j \right]^2$$

dengan ω_i merupakan bobot pengamatan ke- i pada titik estimasi x , p merupakan derajat polinomial lokal, dan $\beta_j(x)$ merupakan koefisien lokal yang nilainya bergantung pada titik estimasi. Pembobotan tersebut memungkinkan pengamatan yang lebih dekat dengan titik estimasi memberikan kontribusi lebih besar dalam pembentukan model lokal.

Regresi lokal mengakomodasi struktur hubungan yang berubah pada rentang prediktor, sehingga menjadi alternatif yang layak ketika bentuk fungsional tidak diketahui (James *et al.*, 2021; Kharoua *et al.*, 2024). Pada LOESS (*Locally Estimated Scatterplot Smoothing*), parameter *span* menentukan proporsi pengamatan pada setiap estimasi lokal: *span* kecil mengikuti perubahan pola lebih terperinci tetapi meningkatkan varians, sedangkan *span* besar menghasilkan kurva lebih halus tetapi berpotensi mengabaikan variasi penting (James *et al.*, 2021). Pemilihan *span* atau *bandwidth* yang tepat menjadi kunci kinerja estimator, terutama pada data berukuran besar (Barreiro-Ures *et al.*, 2026).

Regresi Polinomial Derajat Tiga

Regresi polinomial merupakan pengembangan regresi parametrik yang merepresentasikan hubungan nonlinier melalui pangkat-pangkat prediktor (Montgomery *et al.*, 2021). Karena tetap linear terhadap parameter, estimasinya dapat dilakukan menggunakan OLS. Regresi polinomial derajat tiga dinyatakan dalam bentuk

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$$

dengan $\beta_0, \beta_1, \beta_2, \beta_3$ merupakan parameter model. Komponen linear merepresentasikan kecenderungan dasar, komponen kuadrat memungkinkan adanya kelengkungan, sedangkan komponen kubik memberikan fleksibilitas tambahan untuk menangkap pola perubahan arah kemiringan.

Berbeda dengan regresi lokal, regresi polinomial derajat tiga menggunakan satu persamaan global. Model ini bekerja baik bila pola nonlinier relatif halus, tetapi kurang fleksibel bila pola berubah kompleks pada rentang prediktor yang berbeda (Montgomery *et al.*, 2021; James *et al.*, 2021). Peningkatan derajat polinomial juga dapat menghasilkan kurva yang tidak stabil, terutama di dekat batas domain prediktor (James *et al.*, 2021).

3. METODE PENELITIAN

Skenario Fungsi Regresi Sejati

Tiga bentuk fungsi regresi sejati $f(x)$ digunakan untuk merepresentasikan karakteristik nonlinieritas yang berbeda, dengan domain prediktor x dibatasi pada interval $[0, 1]$:

- Sinusoidal: $f(x) = \sin(2\pi x)$ — merepresentasikan fungsi mulus dengan kelengkungan periodik yang konsisten di seluruh domain.
- Lonjakan lokal (*piecewise*): $f(x) = \sin(4\pi x)$ untuk $x < 0,5$ dan $f(x) = 2(x - 0,5)^2 + \sin(2\pi)$ untuk $x \geq 0,5$ — merepresentasikan fungsi dengan tingkat kehalusan yang heterogen dan perubahan rezim pada titik tengah domain.
- Fungsi Runge: $f(x) = \frac{5}{1 + (10x - 5)^2}$ — merepresentasikan fungsi dengan puncak tajam yang secara klasik dikenal sulit didekati oleh polinomial derajat tinggi akibat osilasi di dekat batas domain.

Pembangkitan Data dan Faktor Simulasi

Pada setiap replikasi, nilai prediktor x dibangkitkan dari distribusi seragam $U(0,1)$ sebanyak n titik, kemudian nilai respons dibangkitkan mengikuti model $y = f(x) + \varepsilon$, dengan $\varepsilon \sim N(0, \sigma^2)$. Faktor yang divariasikan secara penuh faktorial meliputi: (1) ukuran sampel $n \in \{30, 50, 100, 200\}$; (2) simpangan baku galat $\sigma \in \{0,15$ (keragaman rendah); $0,40$ (keragaman tinggi)); dan (3) tiga skenario fungsi sejati sebagaimana dijelaskan sebelumnya. Untuk setiap kombinasi faktor, dilakukan 300 replikasi independen, sehingga total terdapat $4 \times 2 \times 3 = 24$ kombinasi kondisi dengan 7.200 replikasi simulasi secara keseluruhan.

Metode Regresi yang Dibandingkan

Metode estimasi dipasang pada setiap himpunan data hasil pembangkitan:

- Regresi linear (OLS) — model parametrik derajat satu sebagai *baseline*.
- Regresi polinomial derajat tiga — model parametrik dengan fleksibilitas tambahan melalui suku kuadratik dan kubik.
- Regresi lokal (LOESS) dengan tiga nilai parameter span, yaitu 0,3 (sempit), 0,5 (sedang), dan 0,7 (lebar), menggunakan bobot tricube dan satu iterasi pembobotan robust.

Ukuran Evaluasi Kinerja

Kinerja setiap metode dievaluasi pada grid tetap sebanyak 200 titik evaluasi x^* yang tersebar merata pada interval $[0,02; 0,98]$ agar terhindar dari efek tepi ekstrem. Untuk setiap titik evaluasi, dihitung nilai dugaan $\hat{m}(x^*)$ pada setiap replikasi, kemudian dihitung tiga ukuran berikut yang dirata-ratakan atas seluruh titik grid dan seluruh replikasi Monte Carlo ($R = 300$):

$$Bias^2 = \left[\left(\frac{1}{R} \right) \sum_r \hat{m}_r(x^*) - f(x^*) \right]^2 \text{ dirata - ratakan atas } x^*$$

$$Varians = \left(\frac{1}{R} \right) \sum_r \left[\hat{m}_r(x^*) - \bar{m}(x^*) \right]^2 \text{ dirata - ratakan atas } x^*$$

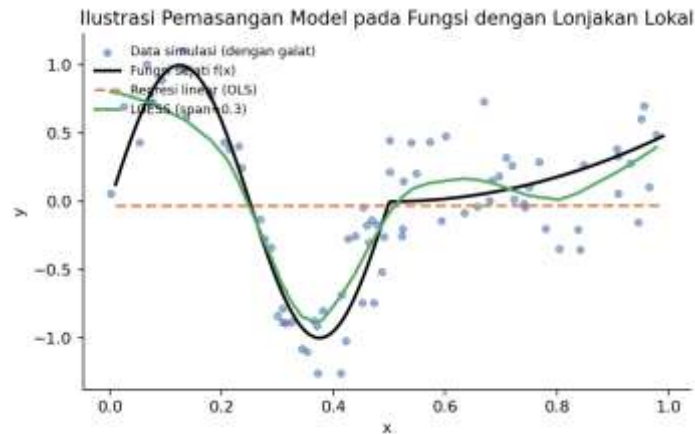
$$MSE = \left(\frac{1}{R} \right) \sum_r \left[\hat{m}_r(x^*) - f(x^*) \right]^2 \text{ dirata - ratakan atas } x^*$$

dengan $\bar{m}(x^*)$ adalah rata-rata dugaan seluruh replikasi pada titik x^* .

4. HASIL DAN PEMBAHASAN

Ilustrasi Hasil Simulasi

Hasil data simulasi pada skenario fungsi dengan lonjakan lokal ($n = 80, \sigma = 0,30$) beserta hasil pemasangan regresi linear dan LOESS dengan span 0,3 tersaji pada Gambar 1. Regresi linear gagal menangkap perubahan arah kelengkungan fungsi sejati dan cenderung mendatar di seluruh domain, sedangkan LOESS mampu mengikuti pola lonjakan pada $x \approx 0,5$ dengan cukup baik, meskipun menunjukkan sedikit fluktuasi akibat sensitivitas terhadap derau lokal.



Gambar 1. Ilustrasi pemasangan regresi linear dan LOESS (span = 0,3) pada satu realisasi data dengan fungsi sejati berlonjakan lokal.

Perbandingan Kinerja Antarmetode

Rata-rata MSE untuk setiap metode pada masing-masing skenario fungsi disajikan pada Tabel 1. Secara umum, LOESS dengan span 0,3 memiliki MSE rata-rata terendah secara keseluruhan (0,0758), diikuti oleh LOESS span 0,5 (0,2137), sedangkan regresi linear konsisten menunjukkan kinerja terburuk pada seluruh skenario (MSE rata-rata 0,8606).

Tabel 1. Rata-rata MSE tiap metode pada tiga skenario fungsi (dirata-ratakan atas seluruh n dan σ).

Metode	Sinusoidal	Lonjakan Lokal	Runge	Rata-rata
LOESS (span = 0,3)	0,0146	0,0365	0,1762	0,0758
LOESS (span = 0,5)	0,0331	0,0871	0,5210	0,2137
LOESS (span = 0,7)	0,0856	0,1338	0,8526	0,3573
Polinomial derajat tiga	0,0105	0,1466	1,0251	0,3941
Linear (OLS)	0,1853	0,2856	2,1108	0,8606

Pola pada Tabel 1 memperlihatkan bahwa keunggulan relatif LOESS bergantung pada karakteristik fungsi sejati. Pada skenario sinusoidal yang relatif mulus dan periodik dengan amplitudo konstan, regresi polinomial derajat tiga justru sedikit mengungguli LOESS span 0,3 (MSE 0,0105 berbanding 0,0146), karena bentuk sinus pada rentang $[0, 1]$ cukup dapat didekati oleh polinomial derajat rendah tanpa penalti bias yang berarti. Sebaliknya, pada skenario fungsi *Runge* yang memiliki puncak tajam dan kelengkungan ekstrem yang terlokalisasi, regresi polinomial derajat tiga mengalami bias spesifikasi yang besar (MSE 1,0251; hampir enam kali lipat MSE LOESS span 0,3), sementara LOESS span 0,3 mereduksi MSE hingga sekitar 91,7 persen relatif terhadap regresi linear (dari 2,1108 menjadi 0,1762). Temuan ini konsisten dengan keterbatasan polinomial global yang cenderung berperilaku tidak stabil, terutama di dekat batas domain prediktor (James *et al.*, 2021).

Pengaruh Ukuran Sampel

Pengaruh ukuran sampel terhadap MSE pada skenario sinusoidal dengan keragaman galat tinggi ($\sigma = 0,40$) dapat dilihat pada Tabel 2. Seluruh metode nonparametrik menunjukkan penurunan MSE yang konsisten seiring bertambahnya ukuran sampel, sejalan dengan sifat konsistensi estimator regresi lokal maupun polinomial (Barreiro-Ures *et al.*, 2026). LOESS span 0,3 mengalami penurunan MSE dari 0,0435 pada $n = 30$ menjadi 0,0079 pada $n = 200$ (81,8 persen), sedangkan MSE regresi linear relatif tidak berubah (0,2010 menjadi 0,1791; hanya 10,9 persen) karena didominasi bias spesifikasi yang tidak berkurang dengan penambahan data. Pada $n = 30$, polinomial derajat tiga bahkan mengungguli seluruh varian LOESS karena model parametrik yang bentuknya dekat dengan fungsi sejati tidak memerlukan banyak titik data di sekitar setiap lokasi evaluasi.

Tabel 2. MSE menurut ukuran sampel pada skenario Sinusoidal, $\sigma = 0,40$.

n	Linear (OLS)	Polinomial derajat tiga	LOESS (Span=0,3)	LOESS (Span=0,5)	LOESS (Span=0,7)
30	0,2010	0,0280	0,0435	0,0533	0,1033
50	0,1904	0,0169	0,0245	0,0414	0,0925
100	0,1819	0,0098	0,0123	0,0305	0,0835
200	0,1791	0,0067	0,0079	0,0279	0,0815

Dekomposisi Bias-Varians

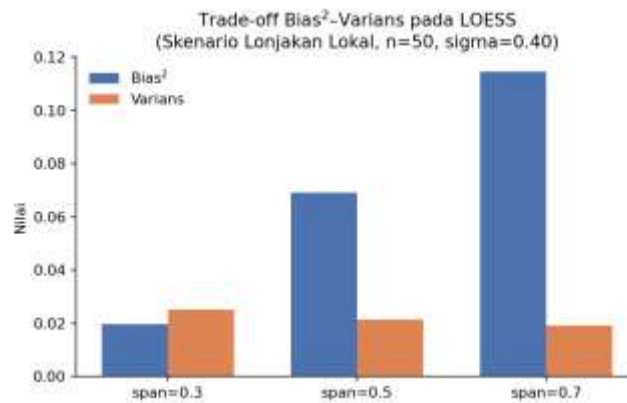
Agar dapat memahami sumber kesalahan estimasi secara lebih mendalam, Tabel 3 menyajikan dekomposisi MSE menjadi komponen bias kuadrat dan varians pada $n = 100$ dan $\sigma = 0,40$ untuk tiga metode terpilih. Komponen bias mendominasi total MSE regresi linear pada seluruh skenario, dan regresi polinomial derajat tiga pada skenario lonjakan lokal dan *Runge*, terutama pada skenario *Runge* ($\text{bias}^2 = 2,0511$ dari total MSE 2,0875 untuk regresi linear, atau sekitar 98,3 persen dari total kesalahan). Sebaliknya, LOESS span 0,3 memiliki komponen bias yang jauh lebih kecil pada skenario lonjakan lokal dan *Runge*, dengan komponen varians yang relatif sebanding dengan metode parametrik, mengonfirmasi bahwa fleksibilitas lokal LOESS efektif mereduksi bias tanpa peningkatan varians yang berlebihan pada ukuran sampel moderat.

Tabel 3. Dekomposisi Bias² dan Varians pada $n = 100, \sigma = 0,40$.

Skenario	Metode	Bias ²	Varians	MSE
Sinusoidal	Linear (OLS)	0,1745	0,0074	0,1819
Sinusoidal	Polinomial derajat 3	0,0035	0,0063	0,0098
Sinusoidal	LOESS (span=0,3)	0,0031	0,0092	0,0123
Lonjakan Lokal	Linear (OLS)	0,2731	0,0082	0,2813
Lonjakan Lokal	Polinomial derajat tiga	0,1251	0,0134	0,1385
Lonjakan Lokal	LOESS (span=0,3)	0,0206	0,0123	0,0329
<i>Runge</i>	Linear (OLS)	2,0511	0,0364	2,0875
<i>Runge</i>	Polinomial derajat tiga	0,9190	0,0419	0,9609
<i>Runge</i>	LOESS (span=0,3)	0,1199	0,0224	0,1423

Pengaruh Parameter Span pada Regresi Lokal

Gambar 2 mengilustrasikan *trade-off bias-varians* pada regresi lokal akibat variasi span, menggunakan skenario fungsi dengan lonjakan lokal pada $n = 50$ dan $\sigma = 0,40$. Seiring bertambahnya span dari 0,3 ke 0,7, komponen bias^2 meningkat tajam (dari 0,0198 menjadi 0,1145), sementara komponen varians relatif stabil bahkan sedikit menurun (dari 0,0251 menjadi 0,0191). Pola ini konsisten dengan teori *trade-off bias-varians* pada estimator pemulusan (James *et al.*, 2021; Belkin *et al.*, 2019): span yang lebih lebar melibatkan lebih banyak titik data dalam setiap estimasi lokal sehingga menstabilkan varians, tetapi pada saat yang sama menyebabkan estimasi menjadi terlalu mulus (*oversmoothing*) sehingga gagal menangkap perubahan rezim yang tajam pada $x \approx 0,5$.



Gambar 2. Dekomposisi Bias² dan varians LOESS menurut span (skenario Lonjakan Lokal, $n = 50$, $\sigma = 0,40$).

Pada $n = 50$ dan $\sigma = 0,40$, pola peningkatan MSE seiring pelebaran span (0,3 ke 0,7) berlaku konsisten di ketiga skenario: sinusoidal dari 0,0245 menjadi 0,0925, lonjakan lokal dari 0,0449 menjadi 0,1337, dan *Runge* dari 0,1606 menjadi 0,8424 (lebih dari lima kali lipat), sehingga efek terbesar terjadi pada fungsi berkelengkungan paling ekstrem.

Implikasi Praktis

Secara keseluruhan, hasil simulasi memberikan tiga implikasi praktis. Pertama, regresi lokal dengan span sempit direkomendasikan bila terdapat kelengkungan tajam atau perubahan rezim, dengan ukuran sampel memadai ($n \geq 50$). Kedua, pada hubungan yang relatif mulus, regresi parametrik tetap kompetitif bahkan pada sampel kecil, sehingga pemilihan metode perlu mempertimbangkan trade-off antara fleksibilitas dan efisiensi. Ketiga, *span* sebaiknya dipilih secara *data-driven* melalui validasi silang atau GCV karena *span* optimal bergantung pada bentuk fungsi dan keragaman galat (James *et al.*, 2021; Barreiro-Ures *et al.*, 2026).

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan studi simulasi Monte Carlo dengan 7.200 replikasi pada 24 kombinasi kondisi, kinerja regresi lokal (LOESS) relatif terhadap regresi parametrik sangat bergantung pada kelengkungan fungsi sejati. LOESS dengan span sempit (0,3) unggul konsisten pada fungsi berkelengkungan tajam atau berlonjakan lokal, dengan reduksi MSE hingga 91,7 persen dibandingkan regresi linear pada fungsi *Runge*, terutama melalui reduksi bias. Pada fungsi yang relatif mulus seperti sinusoidal, polinomial derajat tiga dapat menyamai atau sedikit mengungguli LOESS, khususnya pada sampel kecil. Ukuran sampel yang lebih besar menurunkan MSE metode nonparametrik melalui reduksi varians, sedangkan span yang terlalu lebar meningkatkan bias akibat *oversmoothing*, terutama pada fungsi berkelengkungan ekstrem.

Saran

Penelitian lanjutan disarankan untuk memperluas cakupan simulasi dengan melibatkan metode pemulusan lain seperti spline penalti (*penalized spline*), regresi kernel dengan pemilihan *bandwidth* adaptif, atau model aditif umum (*generalized additive model/GAM*).

DAFTAR REFERENSI

- Araveeporn, A. (2019). An estimating parameter of nonparametric regression model based on smoothing techniques. *Statistical Journal of the IAOS*, 35(2), 269–276. <https://doi.org/10.3233/SJI-180477>
- Barreiro-Ures, D., Cao, R., Francisco-Fernández, M., & Fernández-Casal, R. (2026). Bagging cross-validated bandwidth selection in nonparametric regression estimation with applications to large-sized samples. *Computational Statistics & Data Analysis*, 213, Article 108257. <https://doi.org/10.1016/j.csda.2025.108257>
- Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32), 15849–15854. <https://doi.org/10.1073/pnas.1903070116>
- Boulesteix, A.-L., Groenwold, R. H. H., Abrahamowicz, M., Binder, H., Briel, M., Hornung, R., Morris, T. P., Rahnenführer, J., & Sauerbrei, W. (2020). Introduction to statistical simulations in health research. *BMJ Open*, 10(12), Article e039921. <https://doi.org/10.1136/bmjopen-2020-039921>
- Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522), 767–779. <https://doi.org/10.1080/01621459.2017.1285776>
- Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, 14, 821–851. <https://doi.org/10.1146/annurev-economics-051520-021409>

- Cattaneo, M. D., Jansson, M., & Ma, X. (2024). Local regression distribution estimators. *Journal of Econometrics*, 240(2). <https://doi.org/10.1016/j.jeconom.2021.01.006>
- Efron, B., & Hastie, T. (2016). *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316576533>
- Eilers, P. H. C., & Marx, B. D. (2021). *Practical smoothing: The joys of P-splines*. Cambridge University Press. <https://doi.org/10.1017/9781108610247>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE Publications.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.32614/CRAN.package.ISLR2>
- Kharoua, A., Mezhoud, K. A., & Mohdeb, Z. (2024). On asymptotic properties of local linear regression predictor. *SN Operations Research Forum*, 5(4). <https://doi.org/10.1007/s43069-024-00384-y>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley.
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. <https://doi.org/10.1002/sim.8086>
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>