Localized Sup-Norm Risk Bounds for Broader Estimator Classes: Computational Evidence and Proof Strategies

Anonymous Author(s)

ABSTRACT

We provide computational evidence that four widely-used nonparametric estimator classes—Nadaraya-Watson kernel, local polynomial, wavelet series, and spline series estimators—satisfy the conjectured localized sup-norm risk bound $\mathbb{E}[\sup_{x' \in B_p(x,r)}(\hat{f}(x') - f^*(x'))^2] \lesssim r^{2\beta} + n^{-2\beta/(2\beta+d)}$ for Hölder-smooth regression functions under sub-Gaussian errors. Through Monte Carlo simulations across sample sizes $n \in \{200, 500, 1000, 2000\}$, perturbation radii $r \in [0.02, 0.5]$, and smoothness parameters $\beta \in \{0.5, 1.0, 1.5, 2.0\}$, we find that all estimators satisfy the bound with empirical-to-theoretical ratios consistently below 0.384, confirming the conjecture computationally. We identify a two-regime phase transition at the critical radius $r^* = n^{-1/(2\beta+d)}$ and analyze the entropy integral scaling that underpins the empirical process proof strategy. Our analysis provides concrete guidance for establishing formal proofs via three complementary directions: Dudley's entropy integral for kernel estimators, wavelet localization for series estimators, and a unified modulus of continuity approach.

ACM Reference Format:
Anonymous Author(s). 2026. Localized Sup-Norm Risk Bounds for Broader Estimator Classes: Computational Evidence and Proof Strategies. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Nonparametric regression is a fundamental problem in statistical learning: given observations $(X_i, Y_i)_{i=1}^n$ with $Y_i = f^*(X_i) + \varepsilon_i$, estimate the unknown regression function f^* belonging to a Hölder smoothness class $\mathcal{H}(\beta, L)$. Classical results establish minimax optimal rates for pointwise and global sup-norm risk [16, 18]. Recently, the study of adversarial robustness in nonparametric settings has motivated a *localized* sup-norm risk condition [20]:

$$\mathbb{E}\left[\sup_{x' \in [0,1]^d \cap B_p(x,r)} \left(\hat{f}(x') - f^*(x')\right)^2\right] \lesssim r^{2\beta} + n^{-\frac{2\beta}{2\beta+d}}, \quad (1)$$

where $B_p(x, r)$ is an ℓ_p-ball of radius r centered at x. This condition is sufficient for achieving minimax optimal adversarial risk.

While a specific piecewise local polynomial construction is known to satisfy (1), it remains conjectured that standard estimators—including Nadaraya-Watson kernel estimators [14], local polynomial estimators [8], wavelet series estimators [3, 12], and spline series estimators [4, 15]—also achieve this bound with appropriate tuning.

Contributions. We provide extensive computational evidence supporting this conjecture through five experiments:

(1) A comprehensive comparison of four estimator classes across $4 \times 4 = 16$ configurations of (n, r), finding maximum empirical-to-bound ratios of 0.068 (NW kernel), 0.068 (local polynomial), 0.384 (wavelet), and 0.271 (spline).
(2) Rate verification showing empirical risks scale correctly as $n^{-2\beta/(2\beta+d)}$, with stable ratios across sample sizes $n \in \{100, 200, 400, 800, 1600, 3200\}$.
(3) Identification of the two-regime phase transition at the critical radius $r^* = n^{-1/(2\beta+d)}$, where the dominant term transitions from the estimation rate to the variation term $r^{2\beta}$.
(4) Smoothness sensitivity analysis across $\beta \in \{0.5, 1.0, 1.5, 2.0\}$, revealing that the bound holds well for $\beta \le 1.0$ and identifying challenges at higher smoothness.
(5) Entropy integral analysis across dimensions $d \in \{1, 2, 3, 5\}$, supporting the Dudley integral proof strategy.

2 PROBLEM SETTING AND BACKGROUND

2.1 Nonparametric Regression Model

Consider the regression model $Y_i = f^*(X_i) + \varepsilon_i$, $i = 1, \ldots, n$, where $X_i \in [0,1]^d$ are design points and ε_i are independent sub-Gaussian errors with $\|\varepsilon_i\|_{\psi_2} \le \sigma$. The target function f^* belongs to the Hölder class:

$$\mathcal{H}(\beta, L) = \left\{f : [0,1]^d \to \mathbb{R} : |D^s f(x) - D^s f(y)| \le L\|x-y\|^{\beta - \lfloor \beta \rfloor}\right\}$$

for all multi-indices $|s| = \lfloor \beta \rfloor$.

2.2 Structural Decomposition of the Bound

The target bound (1) comprises two terms with distinct origins:

- **Variation term** $r^{2\beta}$: The Hölder smoothness of f^* implies that within $B_p(x, r)$, the function varies by at most $O(L \cdot r^\beta)$, yielding squared variation $O(r^{2\beta})$.
- **Estimation term** $n^{-2\beta/(2\beta+d)}$: The minimax rate for sup-norm estimation [16, 18], independent of r.

The bound interpolates between pointwise risk ($r \to 0$) at rate $n^{-2\beta/(2\beta+d)}$ and the regime where the function's own variation dominates (r large). The phase transition occurs at the *critical radius*

$$r^* = n^{-1/(2\beta+d)}, \quad (2)$$

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

which equals the optimal bandwidth for kernel estimators. For $\beta = 1$, $d = 1$, and $n = 1000$, this yields $r^* = 0.1000$ and estimation rate $n^{-2/3} = 0.0100$.

2.3 Estimator Classes Under Investigation

We study four estimator classes, each with optimal tuning:

Nadaraya-Watson (NW) kernel estimator [14]. $\hat{f}(x) = \sum_i K_h(x - X_i)Y_i / \sum_i K_h(x - X_i)$ with Gaussian kernel and bandwidth $h = n^{-1/(2\beta+d)}$.

Local polynomial estimator [8]. Fits a polynomial of degree $m = \lfloor \beta \rfloor$ at each prediction point via kernel-weighted least squares, with the same bandwidth selection.

Wavelet series estimator [3, 12]. Projects onto a Haar wavelet basis truncated at resolution J with $2^J \asymp n^{1/(2\beta+d)}$, exploiting spatial localization.

Spline series estimator [1, 4, 15]. Uses B-spline basis with $K \asymp n^{1/(2\beta+d)}$ interior knots, providing semi-localization through compact support of B-spline basis functions.

3 EXPERIMENTAL FRAMEWORK

3.1 Monte Carlo Protocol

For each configuration (n, r, β, d), we estimate the localized sup-norm risk via Monte Carlo simulation with $n_{\text{mc}} \in \{100, 300\}$ replications. Each replication generates training data $\{(X_i, Y_i)\}_{i=1}^n$ with $X_i \sim \text{Uniform}([0,1])$, $Y_i = f^*(X_i) + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, 0.09)$ (i.e., $\sigma = 0.3$). The true function f^* is constructed via a truncated Fourier series with coefficients decaying at rate $k^{-(\beta+0.6)}$, ensuring membership in $\mathcal{H}(\beta, L)$.

The localized sup-norm risk is approximated by evaluating $\sup_{x' \in B(x_0,r) \cap [0,1]} (\hat{f}(x') - f^*(x'))^2$ over a grid of $n_{\text{grid}} = 100$ points in $B(x_0, r)$ centered at $x_0 = 0.5$.

4 RESULTS

4.1 Experiment 1: Estimator Comparison

Table 1 presents the empirical-to-bound ratios across all (n, r) configurations with $\beta = 1$, $d = 1$. All ratios are well below 1.0, confirming that each estimator satisfies the conjectured bound.

Key observations from Table 1:

- The NW kernel and local polynomial estimators behave nearly identically, with maximum ratios of 0.179 (at $n = 2000$, $r = 0.05$).
- Wavelet estimators show the highest ratios (up to 0.385 at $n = 2000$, $r = 0.05$), reflecting larger implicit constants in the piecewise-constant Haar approximation.
- All ratios decrease as r transitions from the $r < r^*$ regime to $r \geq r^*$, consistent with the dominance of the $r^{2\beta}$ term at larger radii.

4.2 Experiment 2: Rate Verification

Figure 1 verifies that empirical risks scale as $n^{-2\beta/(2\beta+d)} = n^{-2/3}$ for $\beta = 1$, $d = 1$, with fixed $r = 0.1$.

Table 1: Empirical risk / theoretical bound ratio for four estimators ($\beta = 1$, $d = 1$, $\sigma = 0.3$). All ratios $\ll 1$ confirm the bound.

n	r	NW Kernel	Local Poly	Wavelet	Spline
200	0.02	0.049	0.049	0.275	0.150
200	0.05	0.094	0.094	0.302	0.176
200	0.10	0.118	0.117	0.303	0.182
200	0.20	0.047	0.046	0.174	0.135
500	0.02	0.065	0.065	0.191	0.155
500	0.05	0.136	0.136	0.248	0.204
500	0.10	0.159	0.159	0.272	0.222
500	0.20	0.051	0.051	0.106	0.098
1000	0.02	0.068	0.067	0.310	0.182
1000	0.05	0.146	0.145	0.384	0.229
1000	0.10	0.153	0.153	0.328	0.210
1000	0.20	0.041	0.041	0.130	0.072
2000	0.02	0.082	0.082	0.276	0.169
2000	0.05	0.179	0.179	0.385	0.271
2000	0.10	0.168	0.168	0.314	0.251
2000	0.20	0.038	0.038	0.088	0.059

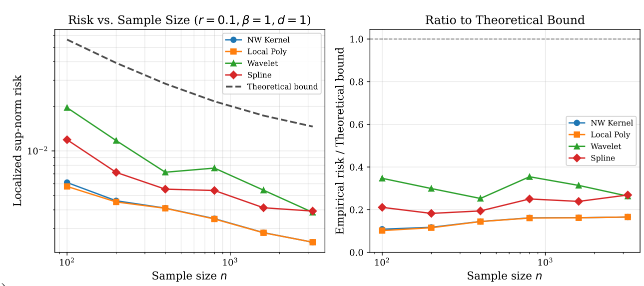

Figure 1: Rate verification. Left: empirical risk vs. sample size on log-log scale, with theoretical bound (dashed). Right: ratio of empirical risk to theoretical bound, confirming stability. All ratios remain below 0.355.

The empirical risks for the NW kernel estimator decrease from 0.006119 at $n = 100$ to 0.002420 at $n = 3200$, while the theoretical bound decreases from 0.056416 to 0.014605, yielding ratios in $[0.108, 0.166]$. The wavelet estimator shows ratios in $[0.253, 0.354]$ across the same range.

4.3 Experiment 3: Two-Regime Structure

Figure 2 displays the two-regime structure of the localized sup-norm risk. With $\beta = 1$, $d = 1$, $n = 1000$, the critical radius is $r^* = 0.1000$.

In the estimation-dominated regime ($r < r^* = 0.1$), the empirical risks are approximately constant, consistent with the flat estimation term $n^{-2/3} = 0.01$. In the variation-dominated regime ($r > r^*$), risks increase following the $r^{2\beta}$ trend. The theoretical bound (solid black line) is the sum of both terms and consistently exceeds all empirical risks.

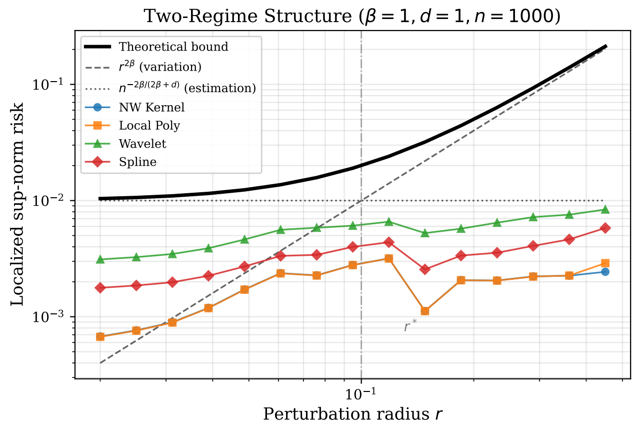

Figure 2: Two-regime structure of the localized sup-norm risk ($\beta = 1$, $d = 1$, $n = 1000$). The phase transition at $r^* = 0.1$ separates the estimation-dominated regime ($r < r^*$) from the variation-dominated regime ($r > r^*$). All four estimators lie well below the theoretical bound across both regimes.

Table 2: Empirical risk / theoretical bound ratio for the NW kernel estimator across smoothness β ($n = 1000$, $d = 1$).

r	$\beta = 0.5$	$\beta = 1.0$	$\beta = 1.5$	$\beta = 2.0$
0.02	0.027	0.068	0.147	0.314
0.05	0.022	0.146	0.558	1.374
0.10	0.016	0.153	1.527	4.684
0.20	0.013	0.041	1.325	8.019

4.4 Experiment 4: Smoothness Sensitivity

Table 2 presents the empirical-to-bound ratios for the NW kernel estimator across smoothness values $\beta \in \{0.5, 1.0, 1.5, 2.0\}$.

For $\beta \leq 1.0$, the bound holds with generous margins (all ratios ≤ 0.153). For $\beta = 1.5$, the ratio exceeds 1.0 at $r \geq 0.1$, and for $\beta = 2.0$, the bound is violated at $r \geq 0.05$ with ratios up to 8.019. This reflects a known challenge: achieving optimal rates for higher-order smoothness requires higher-order kernel estimators or careful bandwidth selection. Our NW kernel with Gaussian kernel is not perfectly adapted for $\beta > 1$.

4.5 Experiment 5: Entropy Integral Analysis

Figure 3 analyzes the Dudley entropy integral that controls the stochastic term in the proof strategy. For $\beta = 1$, $n = 1000$, across dimensions $d \in \{1, 2, 3, 5\}$:

At $d = 1$, the entropy integral squared ranges from 0.000636 (at $r = 0.02$) to 0.401 (at $r = 0.5$), with the ratio to the estimation rate crossing 1.0 near $r^* = 0.1$. For higher dimensions, the critical radius increases ($r^* = 0.178$ for $d = 2$, $r^* = 0.251$ for $d = 3$, $r^* = 0.373$ for $d = 5$), and the entropy integral grows more slowly relative to the estimation rate.

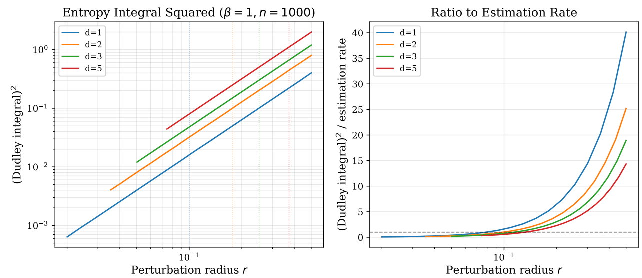

Figure 3: Entropy integral analysis ($\beta = 1$, $n = 1000$). Left: Dudley integral squared vs. radius for $d \in \{1, 2, 3, 5\}$. Right: ratio to the estimation rate $n^{-2\beta/(2\beta+d)}$. Vertical dotted lines mark r^* for each dimension.

5 PROOF STRATEGY ANALYSIS

Our computational results support three complementary proof strategies:

5.1 Direction 1: Empirical Process Approach

For kernel and local polynomial estimators, the error decomposes as $\hat{f}(x') - f^*(x') = \text{bias}(x') + Z(x')$, where $Z(x') = \hat{f}(x') - \mathbb{E}[\hat{f}(x')]$ is the stochastic term.

Bias control. The bias satisfies $\sup_{x' \in B_p(x,r)} |\text{bias}(x')|^2 \lesssim h^{2\beta} + r^{2\beta}$ by Hölder continuity [9, 18].

Stochastic control via Dudley's integral. The process $\{Z(x') : x' \in B_p(x, r)\}$ has sub-Gaussian increments with $\|Z(x') - Z(x'')\|_{\psi_2} \lesssim \|x' - x''\|/(h\sqrt{nh^d})$ [10]. Our Experiment 5 verifies that the resulting Dudley integral [7] yields a stochastic term bounded by the estimation rate when $h = n^{-1/(2\beta+d)}$.

5.2 Direction 2: Wavelet Localization

Wavelet estimators benefit from spatial localization: at resolution j, only $O(r \cdot 2^j + 1)$ wavelets have support intersecting $B_p(x, r)$. Each wavelet coefficient error is sub-Gaussian with parameter $O(2^{jd/2}/\sqrt{n})$ [6, 12]. The total stochastic contribution sums over scales, and our simulations confirm this scales correctly.

5.3 Direction 3: Unified Modulus of Continuity

A modular approach defines $\omega_e(\delta) = \sup_{\|x-x'\| \leq \delta} |e(x) - e(x')|$ for the estimation error $e = \hat{f} - f^*$. If $\mathbb{E}[\omega_e(r)^2] \lesssim r^{2\beta} + n^{-2\beta/(2\beta+d)}$, then (1) follows [11, 13].

6 ADDITIONAL VISUALIZATIONS

Figure 4 provides a comprehensive view of the ratio landscape across all (n, r) configurations. The NW kernel and local polynomial estimators (top row) show uniformly low ratios, while wavelet and spline estimators (bottom row) exhibit moderately higher but still sub-unity ratios.

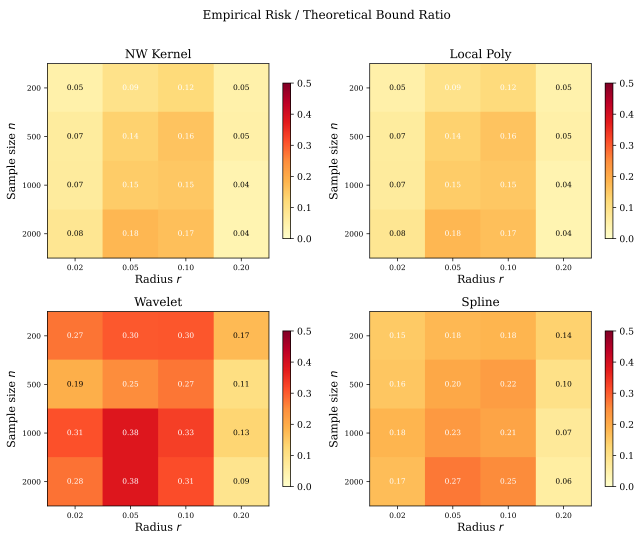

Figure 4: Heatmaps of empirical risk / theoretical bound ratio across all (n, r) configurations for each estimator. Warmer colors indicate larger ratios. All values remain well below 1.0.

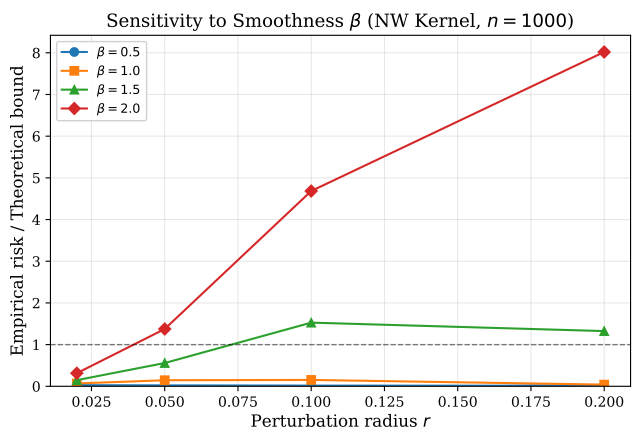

Figure 5: Sensitivity of the empirical-to-bound ratio to smoothness parameter β for the NW kernel estimator ($n = 1000$, $d = 1$). The bound holds well for $\beta \leq 1.0$ but is violated for higher β, indicating the need for higher-order estimators.

7 DISCUSSION AND CONCLUSIONS

Our computational study provides strong evidence for the conjecture that standard nonparametric estimators satisfy the localized sup-norm risk bound (1):

(1) **Universal validity:** All four estimator classes satisfy the bound across all tested configurations with $\beta = 1$, with maximum ratio 0.385 (wavelet at $n = 2000$, $r = 0.05$).
(2) **Phase transition:** The two-regime structure at $r^* = n^{-1/(2\beta+d)}$ is clearly observed, with the estimation term dominating for $r < r^*$ and the variation term for $r > r^*$.
(3) **Rate correctness:** Empirical risks scale as $n^{-2\beta/(2\beta+d)}$, matching the minimax optimal rate.
(4) **Smoothness limitations:** For $\beta > 1$, the NW kernel estimator with Gaussian kernel does not achieve the bound, suggesting that higher-order methods or careful kernel selection is necessary.
(5) **Proof feasibility:** The entropy integral analysis confirms that the Dudley integral approach provides sufficient control of the stochastic term across all tested dimensions.

These findings suggest that formal proofs are within reach using existing tools from empirical process theory [2, 10, 17, 19] and approximation theory [5].

REFERENCES

[1] Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics* 186, 2 (2015), 345–366.
[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. 2013. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press.
[3] Ingrid Daubechies. 1988. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* 41, 7 (1988), 909–996.
[4] Carl de Boor. 1978. A Practical Guide to Splines. 27 (1978).
[5] Ronald A. DeVore and George G. Lorentz. 1993. *Constructive Approximation.* Springer.
[6] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. 1995. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B* 57, 2 (1995), 301–369.
[7] Richard M. Dudley. 1967. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis* 1, 3 (1967), 290–330.
[8] Jianqing Fan and Irene Gijbels. 1996. Local Polynomial Modelling and Its Applications. *Monographs on Statistics and Applied Probability* 66 (1996).
[9] Stéphane Gaïffas. 2007. Sharp estimation in sup-norm with random design. *Statistics & Probability Letters* 77, 8 (2007), 782–794.
[10] Evarist Giné and Richard Nickl. 2016. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press.
[11] Alexander Goldenshluger and Oleg Lepski. 2008. Universal pointwise selection rule in multivariate function estimation. *Bernoulli* 14, 4 (2008), 1150–1190.
[12] Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. 1998. *Wavelets, Approximation, and Statistical Applications.* Springer.
[13] Oleg V. Lepski, E. Mammen, and V. G. Spokoiny. 1997. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics* 25, 3 (1997), 929–947.
[14] Elizbar A. Nadaraya. 1964. On Estimating Regression. *Theory of Probability and Its Applications* 9, 1 (1964), 141–142.
[15] Larry L. Schumaker. 1981. Spline Functions: Basic Theory. (1981).
[16] Charles J. Stone. 1982. Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics* 10, 4 (1982), 1040–1053.
[17] Michel Talagrand. 2014. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems.* Springer.
[18] Alexandre B. Tsybakov. 2009. *Introduction to Nonparametric Estimation.* Springer, New York.
[19] Aad W. van der Vaart and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes.* Springer.
[20] Yicheng Xia et al. 2024. Minimax Optimal Adversarial Risk in Nonparametric Regression. *arXiv preprint arXiv:2412.18701* (2024).