Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 5 (2026) 358

African Scientific Reports

A group of clusters of families of hybrid classical polynomial kernels for nonparametric density estimation

Benson Ade Eniola Afere[a,*], Ekele Vincent Aguda[b], Yahaya Baba Usman[c]

[a]Department of Mathematical Sciences, Faculty of Natural Sciences, Prince Abubakar Audu University, 272102, Anyigba, Nigeria
[b]Department of Mathematics, Nigeria Maritime University, Okerenkoko, Nigeria
[c]Department of Mathematics and Statistics, Federal Polytechnic, Idah, Kogi State, Nigeria

Abstract

This study proposes a group of clusters of families of hybrid polynomial kernels, constructed via convex combinations of Beta polynomial kernel families, extending classical kernel density estimation through a flexible mixing parameter that generates smooth and adaptable kernel shapes. Theoretical analysis shows that the asymptotic performance of the resulting estimators is governed by kernel roughness and second-moment functionals through a canonical AMISE constant, yielding near-optimal efficiency with negligible loss relative to the Epanechnikov kernel. Extensive Monte Carlo simulations across symmetric, skewed, heavy-tailed, and multimodal distributions confirm the theoretical results, demonstrating consistent error reduction with increasing sample size and uniform convergence across kernel orders. Sensitivity analysis further reveals strong robustness to variations in mixture proportions, tail heaviness, and modal separation, with sample size identified as the dominant driver of accuracy. Real-data applications to Old Faithful eruption durations and scar-length measurements show that the proposed estimators capture diverse distributional features effectively while maintaining stability. Overall, the hybrid kernels provide a flexible, efficient, and robust alternative for complex density estimation problems.

DOI: 10.46481/asr.2026.5.3.358

Keywords: Hybrid kernel density estimation, Beta polynomial kernels, AMISE, Monte Carlo simulation, Sensitivity analysis

Article History:
Received: 14 August 2025
Received in revised form: 11 June 2026
Accepted for publication: 14 July 2026
Available online: 18 September 2026

1. Introduction

Over the past seven decades, nonparametric kernel density estimation (KDE) has undergone substantial theoretical and methodological development, driven by advances in computational capabilities, increasing availability of complex and high-dimensional data, and the growing demand for flexible model-free statistical procedures. As one of the most important tools in nonparametric inference, KDE provides a framework for estimating an unknown probability density function directly from observed data without imposing restrictive distributional assumptions.

The origins of kernel density estimation are commonly traced to the unpublished technical report of Fix and Hodges [1], with related approximation ideas independently explored by Akaike [2]. The formal introduction is attributed to Rosenblatt [3], who

*Corresponding Author Tel. No.: +234-913-539-8886.
Email address: baafere@gmail.com (Benson Ade Eniola Afere)

established the kernel estimator and initiated systematic theoretical study of nonparametric density estimation. Shortly thereafter, Parzen [4] provided a unified framework and introduced mean squared error (MSE) and mean integrated squared error (MISE) as fundamental performance criteria, which remain central in modern KDE theory.

Subsequent developments focused on improving both theoretical understanding and practical performance. Watson and Leadbetter [5] derived MISE-optimal kernels for specific density classes and finite samples. In the 1960s and 1970s, Bartlett [6] introduced higher-order kernels for bias reduction, while Cacoullos [7] extended KDE to multivariate settings. Epanechnikov [8] proposed the asymptotically optimal second-order kernel and introduced diagonal bandwidth matrices, influencing kernel design significantly. Farrell [9] established lower bounds on convergence rates, while Davis [10] developed Fourier-integral density estimators. Further refinements of MSE and MISE properties were given by Fryer [11] and Deheuvels [12], and Schucany and Sommers [13] introduced jackknife-based bias reduction methods.

Later decades saw major growth in theory and applications. Abramson's square-root law (see Ref. [14]) introduced adaptive bandwidth selection, forming the basis of modern variable smoothing methods. Devroye and collaborators [15–17] studied convergence under alternative loss functions, especially the L_1 norm, while Cline [18] introduced admissibility concepts linking kernel structure to estimator reliability. Marron and Nolan [19] developed canonical kernel theory, unifying classical constructions. Silverman [20, 21], Rudemo [22], and Bowman [23] significantly advanced practical bandwidth selection methods. Computational advances in the 1990s and early 2000s further accelerated multivariate KDE, efficient smoothing algorithms, and large-sample implementations (see Refs. [24–27]), with applications in econometrics and astrophysics (see Refs. [28, 29]).

Collectively, these works established the foundations of modern KDE and emphasized kernel choice, bandwidth selection, and asymptotic risk as central determinants of estimator performance.

Despite this maturity, kernel design and bandwidth selection remain the most influential factors in achieving optimal bias–variance trade-offs across heterogeneous data structures (see Ref. [21]). Consequently, research has increasingly focused on adaptive and flexible kernel constructions beyond classical fixed-kernel frameworks.

A major direction is adaptive estimation. Building on Abramson [14], variable-bandwidth and adaptive KDE methods have been developed, allowing smoothing to vary with local data structure (see Ref. [30]). These improve estimation in regions with sharp curvature, varying density, or sparse tails. Related approaches include shape-adaptive and boundary-corrected kernels (see Ref. [31]), improving performance while preserving theoretical validity (see Ref. [32]).

Another important direction is hybrid kernel construction. Instead of relying on a single kernel, hybrid approaches combine multiple kernels to improve smoothness, reduce bias, enhance adaptability, and retain computational efficiency. Such methods often achieve superior asymptotic performance while preserving interpretability and theoretical guarantees of classical KDE.

The classical polynomial (Beta) kernel family (see Refs. [33, 34]), including the Epanechnikov, Biweight, Triweight, and Quadriweight kernels, provides a strong basis for hybrid development. These kernels are mathematically well understood and central to KDE theory. The clustering framework in this study builds on ideas in Ref. [27], emphasizing structured families of kernels rather than isolated constructions. Unlike traditional hybrid approaches that combine a few fixed kernels, the proposed framework generates entire families of related kernel functions through systematic parameterization.

1.1. Motivation and novel contributions

The motivation for this study arises from a structural gap in the classical polynomial (Beta) kernel family: its members exist only at discrete orders, so a practitioner must jump from the Epanechnikov kernel to the Biweight kernel (and so on) with no intermediate shape available, even though the optimal degree of smoothing for a given dataset typically lies *between* successive orders. Our earlier work addressed this gap partially: a single family of hybrid classical polynomial kernels was introduced in Ref. [33], and a cluster of such families was developed in Ref. [34]. Both contributions, however, were essentially constructive; they neither organised the admissible mixing weights into a unified generative system nor supplied a complete asymptotic theory covering all members simultaneously. The present study closes this gap and advances the literature in four specific ways. First, it constructs a *group of four clusters* of families of hybrid kernels (Clusters A–D), generated from a single closed-form expression, Eq. (6), by rational grids of the mixing weight ρ_1; this elevates hybridisation from an ad hoc mixture device to a systematic generative mechanism producing 116 family–order combinations. Second, it establishes the validity of the entire group (Theorem 3.1) and derives closed-form roughness and moment functionals together with a general order-$2m$ AMISE with fully explicit constants (Theorem 4.2), thereby extending optimality theory from individual kernels to structured kernel systems through asymptotic equivalence classes sharing a common canonical constant $C(K)$. Third, it introduces a relative asymptotic sensitivity (RAS) framework, new to this line of work, that quantifies the robustness of the proposed estimators to perturbations in sample size, mixture proportion, tail thickness, and modal separation. Fourth, it validates the theory through extensive Monte Carlo experiments over symmetric, skewed, heavy-tailed, and multimodal targets and through two real datasets, demonstrating that every member of the group attains near-Epanechnikov efficiency (no less than 0.9572) while offering a continuum of tunable kernel shapes unavailable in Refs. [33] and [34].

Accordingly, this review focuses on adaptive smoothing, hybrid kernel construction, and asymptotic risk analysis, rather than a full survey of KDE literature.

Classical polynomial kernels of successive orders are effective but do not allow smooth transitions between levels of smoothing. For instance, the Epanechnikov kernel may under-smooth while the Biweight kernel may over-smooth, with no intermediate kernel to bridge this gap.

To address this, a cluster-based framework for hybrid polynomial kernels is proposed. Instead of treating hybrid kernels as ad hoc mixtures, clustering is used as a generative mechanism producing families of kernels governed by shared structural parameters. Embedding kernels in a cluster parameter space enables controlled adjustment of smoothness and tail behavior while preserving kernel validity conditions.

The framework constructs hybrid kernels from clustered adjacent polynomial orders, ensuring symmetry, unit mass, finite roughness, and finite moments. It also allows derivation of bias, variance, and AMISE expressions for entire kernel families, enabling analysis of optimality, robustness, and efficiency.

Extensive simulations and real-data studies compare the proposed kernels with classical and hybrid estimators, showing improved bias–variance trade-offs without added complexity.

The cluster perspective further introduces asymptotic equivalence classes of kernels with shared AMISE behavior, extending optimality theory from individual kernels to structured kernel systems.

Overall, the study provides a theoretical foundation for cluster-generated kernels and demonstrates improved density estimation performance with controlled bias–variance trade-offs and computational efficiency.

The remainder of this paper is organised as follows. Section 2 presents the univariate kernel density estimator and outlines the methodological framework adopted in this study. In Section 3, the proposed kernel density functions are introduced and formally constructed. Section 4 provides a comprehensive performance analysis of the proposed estimators using both simulated and real-world datasets. Section 5 discusses the results, with particular emphasis on comparisons with existing methods in the literature. Finally, Section 6 concludes the study.

2. Univariate kernel density estimator

One unique nonparametric approach in KDE is the univariate kernel density estimator, which provides a data-driven technique and an adaptable tool for displaying characteristics present in a given set of observations. Because of its simplicity and ease of implementation compared to other estimators, it has been widely applied in several fields, including banking, climatology, economics, genetics, hydrology, physiology, insurance (see Ref. [35]), and random differential equations (see Ref. [36]).

Consider a probability density function $f(x)$, where $\int f(x)\,dx = 1$. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random samples of size n drawn from a continuous distribution with density $f(x)$. The univariate kernel density estimator is given by

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big), \tag{1}$$

where $K(\cdot)$ is the kernel function satisfying the properties

$$\begin{cases} i. \quad K(x) = K(-x), \quad \forall x \in \mathbb{R}, \\ ii. \quad \displaystyle\int_{\mathbb{R}} K(t)\,dt = 1, \\ iii. \quad \displaystyle\int_{\mathbb{R}} tK(t)\,dt = 0, \\ iv. \quad \displaystyle\int_{\mathbb{R}} K^2(t)\,dt = \|K\|_2^2 < \infty, \\ v. \quad \displaystyle\int_{\mathbb{R}} t^2 K(t)\,dt < \infty. \end{cases} \tag{2}$$

The bandwidth parameter h is required to satisfy

$$\begin{cases} i. \quad \lim_{n\to\infty} h = 0, \\ ii. \quad \lim_{n\to\infty} nh = \infty. \end{cases} \tag{3}$$

The kernel properties in Eq. (2) ensure symmetry, unit mass, zero first moment, and finite second-order behaviour, while the conditions in Eq. (3) guarantee the asymptotic consistency of the estimator.

As discussed in Section 1, Eq. (1) originated from the early technical contributions of Ref. [1] and [2], with the first formal publication credited to Ref. [3] and subsequently developed by Ref. [4]. Consequently, the estimator is commonly known as the Rosenblatt–Parzen kernel density estimator.

The classical family of polynomial kernels is defined by

Table 1: Constants of the four polynomial kernels: normalising constant C_p, variance functional $\mu_2(K)$, roughness $R(K)$, canonical bandwidth factor δ_K, and asymptotic efficiency relative to the Epanechnikov optimum, see Ref. [26].

p	Kernel	C_p	$\mu_2(K)$	$R(K)$	δ_K	Efficiency
1	Epanechnikov	0.7500	0.2000	0.6000	1.7188	1.0000
2	Biweight	0.9375	0.1429	0.7143	2.0362	0.9939
3	Triweight	1.0938	0.1111	0.8159	2.3122	0.9867
4	Quadriweight	1.2305	0.0909	0.9070	2.5591	0.9791

Table 2: Notation used throughout the paper.

Symbol	Meaning
$f,\ \hat{f}_h$	true density and its kernel estimator, Eq. (1)
$n,\ h$	sample size and bandwidth
$K,\ K_{[p]}$	generic kernel; classical polynomial kernel of order p, Eq. (4)
c_p	normalising constant $(2p+1)!!/(2^{p+1}p!)$ of $K_{[p]}$
p	base order of the classical kernels combined ($p = 1, 2, 3, 4$)
m	smoothing (derivative) order; the kernel is of order $2m$
$\rho_1,\ \rho_2$	convex mixing weights, $\rho_1 + \rho_2 = 1$, $0 < \rho_1 < 1$
$K_{G(C)}$	proposed hybrid kernel $\rho_1 K_{[p-1]} + \rho_2 K_{[p]}$, Eq. (5)
$R(K) = \int K^2$	kernel roughness functional
$\mu_2(K) = \int t^2 K$	second moment of the kernel
$C(K) = \mu_2(K)^{2/5} R(K)^{4/5}$	canonical AMISE constant (second-order case)
$\mathrm{eff}(K)$	efficiency relative to Epanechnikov, $C(K_{\mathrm{Epan}})/C(K)$
MISE, AMISE	(asymptotic) mean integrated squared error, Eq. (15)
h_{AMISE}	AMISE-optimal bandwidth, Eq. (32)
$\lambda(m),\ \beta$	kernel constants in the general AMISE, Eqs. (21)–(22)
$\lVert g \rVert_2^2 = \int g^2$	squared L^2 norm; with $g = f^{(2m)}$ it is the curvature functional
$\mathrm{RAS}(\theta)$	relative asymptotic sensitivity to parameter θ
$\Gamma(\cdot),\ (2p+1)!!$	gamma function and double factorial

$$K_c(t) = \frac{(2p+1)!!}{2^{p+1}p!} \sum_{r=0}^{p} (-1)^r \binom{p}{r} t^{2r}, \quad p = 0, 1, 2, \ldots \tag{4}$$

The cases $p = 1, 2, 3$, and 4 correspond respectively to the Epanechnikov, Biweight, Triweight, and Quadriweight kernels. As $p \to \infty$, $K_c(t)$ converges to the Gaussian kernel $\varphi(t)$. Detailed discussions on kernel construction and their theoretical properties can be found in Refs. [33, 34, 37].

To further characterise these polynomial kernels, Table 1 summarises their normalising constants, second moments, roughness measures, canonical bandwidth factors, and asymptotic efficiencies relative to the Epanechnikov optimal kernel.

Table 1 reveals a systematic trade-off as the kernel order increases. Specifically, the variance functional $\mu_2(K)$ decreases, whereas the roughness measure $R(K)$ and the bandwidth factor δ_K increase. This behaviour reflects progressively smoother kernels with wider effective support. Although the Epanechnikov kernel remains asymptotically optimal, the efficiency losses associated with higher-order polynomial kernels are relatively small, demonstrating the near-optimal performance of the entire family.

Since several quantities introduced in this section reappear throughout the subsequent theoretical developments, particularly in the derivations presented in Section 4, the principal notation adopted in the paper is collected in Table 2 for ease of reference.

3. A group of clusters of the families of hybrid classical polynomial kernels

A cluster is defined as a group of items in which each item is close to another item of that group, and members of different groups are far away from each other, as stated in Ref. [27]. In this section, a group of clusters of proposed families of hybrid classical polynomial kernels ($K_{G(C)}(t)$) is heuristically constructed. This involves fusing together two kernels of orders $p-1$ and p using the axioms of probability, following the approach outlined in Refs. [33, 34]. The proposed group of clusters of the families of hybrid polynomial classical kernels is formulated by the simple formula

$$K_{G(C)}(t) = \rho_1 K_{[p-1]}(t) + \rho_2 K_{[p]}(t), \tag{5}$$

where $0 < \rho_1 < 1, \rho_1 + \rho_2 = 1$, and $K_{[p-1]}$, and $K_{[p]}$ are respectively the families of classical polynomial kernels of order $p-1$ and p as presented in Eq. (4). On plugging Eq. (4) into Eq. (5) and applying necessary algebraic rules, we have:

$$K_{G(C)}(t) = \left\{\frac{2^{-p-1}}{\Gamma(p+1)}\Big[2p\rho_1\,(2p-1)!!\sum_{r=0}^{p-1}(-1)^r\binom{p-1}{r} + \rho_2\,(2p+1)!!\sum_{r=0}^{p}(-1)^r\binom{p}{r}t^{2r}\Big]\right\}, \tag{6}$$

where the double factorial is given as $(2p+1)!! = (2p+1)(2p-1)\cdots 5\cdot 3\cdot 1$. Eq. (6) is thus the proposed group of clusters of the families of hybrid classical polynomial kernels.

3.1. Notable specific clusters of families of $K_{G(C)}(t)$

The four specific notable clusters of families of kernels in the proposed group of clusters of families of kernels are herein presented in this subsection. Each cluster within the proposed group represents clusters of distinct families of hybrid classical polynomial kernels, obtained by combining two kernels of orders $p-1$ and p using the axioms of probability. The fusion of these kernels results in a set of versatile and adaptive kernels, offering a spectrum of options for nonparametric kernel density estimation.

The construction of these hybrid kernels as stated earlier is inspired by the work of [33, 34], providing a heuristic approach to generate kernels that exhibit desirable properties. By leveraging the principles of probability, these families of hybrid kernels aim to enhance the flexibility and performance of kernel density estimation methods.

The choice of combining kernels of different orders introduces a nuanced approach, allowing for tailored solutions to various density estimation scenarios. The resulting clusters encapsulate the characteristics of both lower and higher-order kernels, providing a rich set of options for practitioners in the field.

The exploration of these families of hybrid classical polynomial kernels contributes to the ongoing efforts to improve the robustness and adaptability of kernel density estimation techniques. By introducing these clusters, we aim to offer a diverse toolkit that can be applied across different domains, spanning from theoretical research to practical applications in areas such as statistics, machine learning, and data analysis. These clusters are characterized as follows:

Cluster A: when $\rho_1 = \frac{i}{10}$, i $= 1, 2, \cdots, 9$

$$K^{A}_{G(C)}(t) = \frac{2^{-p-2}}{5\Gamma(p+1)}\Big[2ip(2p-1)!!\sum_{r=0}^{p-1}(-1)^r\binom{p-1}{r} + (10-i)(2p+1)!!\sum_{r=0}^{p}(-1)^r\binom{p}{r}\Big]t^{2r}. \tag{7}$$

Cluster B: when $\rho_1 = \frac{i}{9}$, i $= 1, 2, \cdots, 8$

$$K^{B}_{G(C)}(t) = \frac{2^{-p-1}}{9\Gamma(p+1)}\Big[2ip(2p-1)!!\sum_{r=0}^{p-1}(-1)^r\binom{p-1}{r} + (9-i)(2p+1)!!\sum_{r=0}^{p}(-1)^r\binom{p}{r}\Big]t^{2r}. \tag{8}$$

Cluster C: when $\rho_1 = \frac{i}{8}$, i $= 1, 2, \cdots, 7$

$$K^{C}_{G(C)}(t) = \frac{2^{-p-4}}{\Gamma(p+1)}\Big[2ip(2p-1)!!\sum_{r=0}^{p-1}(-1)^r\binom{p-1}{r} + (8-i)(2p+1)!!\sum_{r=0}^{p}(-1)^r\binom{p}{r}\Big]t^{2r}. \tag{9}$$

Cluster D: when $\rho_1 = \frac{i}{7}$, i $= 1, 2, \cdots, 7$

$$K^{D}_{G(C)}(t) = \frac{2^{-p-1}}{7\Gamma(p+1)}\Big[2ip(2p-1)!!\sum_{r=0}^{p-1}(-1)^r\binom{p-1}{r} + (7-i)(2p+1)!!\sum_{r=0}^{p}(-1)^r\binom{p}{r}\Big]t^{2r}. \tag{10}$$

These clusters of families in $K_{G(C)}(t)$ as given in Eqs. (7), (8), (9) and (10) are presented respectively in Tables 3, 4, 5, and 6.

The four tables are intentionally exhaustive: they record every family generated by the admissible mixing weights of each cluster and are provided for completeness and reproducibility, so that any member can be reconstructed exactly without re-deriving Eq. (6). The explicit polynomials are not meant to be read individually as separate results; rather, they document the full reach of the one-parameter construction. To keep the subsequent analysis readable, all theoretical and empirical comparisons in Section 4 are carried out on two representative families (F_1, F_2) drawn from each cluster, which already span the relevant range of the mixing weight ρ_1.

Table 3: Families of hybrid classical polynomial kernels in cluster A.

Family (i)	Epanechnikov	Biweight	Triweight	Quadriweight
AF1	$\frac{1}{40}(29-27t^2)$	$\frac{3}{160}(49-94t^2+45t^4)$	$\frac{3}{64}(1-t^2)^2(23-21t^2)$	$\frac{7}{512}(1-t^2)^3(89-81t^2)$
AF2	$\frac{1}{10}(7-6t^2)$	$\frac{3}{20}(6-11t^2+5t^4)$	$\frac{1}{16}(1-t^2)^2(17-14t^2)$	$\frac{7}{64}(1-t^2)^3(11-9t^2)$
AF3	$\frac{3}{40}(9-7t^2)$	$\frac{3}{160}(47-82t^2+35t^4)$	$\frac{1}{64}(1-t^2)^2(67-49t^2)$	$\frac{21}{512}(1-t^2)^3(29-21t^2)$
AF4	$\frac{1}{20}(13-9t^2)$	$\frac{3}{80}(23-38t^2+15t^4)$	$\frac{3}{32}(1-t^2)^2(11-7t^2)$	$\frac{7}{256}(1-t^2)^3(43-27t^2)$
AF5	$\frac{1}{8}(5-3t^2)$	$\frac{3}{32}(9-14t^2+5t^4)$	$\frac{5}{64}(1-t^2)^2(13-7t^2)$	$\frac{35}{512}(1-t^2)^3(17-9t^2)$
AF6	$\frac{1}{40}(23-9t^2)$	$\frac{3}{160}(43-58t^2+15t^4)$	$\frac{21}{64}(1-t^2)^2(3-t^2)$	$\frac{7}{512}(1-t^2)^3(83-27t^2)$
AF7	$\frac{3}{10}(2-t^2)$	$\frac{3}{40}(11-16t^2+5t^4)$	$\frac{1}{16}(1-t^2)^2(16-7t^2)$	$\frac{21}{128}(1-t^2)^3(7-3t^2)$
AF8	$\frac{1}{20}(11-3t^2)$	$\frac{3}{80}(21-26t^2+5t^4)$	$\frac{1}{32}(1-t^2)^2(31-7t^2)$	$\frac{7}{256}(1-t^2)^3(41-9t^2)$
AF9	$\frac{3}{40}(7-t^2)$	$\frac{3}{160}(41-46t^2+5t^4)$	$\frac{1}{64}(1-t^2)^2(61-7t^2)$	$\frac{63}{512}(1-t^2)^3(9-t^2)$

Table 4: Families of hybrid classical polynomial kernels in cluster B.

Family (i)	Epanechnikov	Biweight	Triweight	Quadriweight
BF1	$\frac{1}{18}(13-12x^2)$	$\frac{1}{12}(1-x^2)(11-10x^2)$	$\frac{5}{144}(1-x^2)^2(31-28x^2)$	$\frac{35}{288}(1-x^2)^3(10-9x^2)$
BF2	$\frac{1}{36}(25-21x^2)$	$\frac{1}{48}(1-x^2)(43-35x^2)$	$\frac{5}{288}(1-x^2)^2(61-49x^2)$	$\frac{35}{2304}(1-x^2)^3(79-63x^2)$
BF3	$\frac{1}{6}(4-3x^2)$	$\frac{1}{8}(1-x^2)(7-5x^2)$	$\frac{5}{48}(1-x^2)^2(10-7x^2)$	$\frac{35}{384}(1-x^2)^3(13-9x^2)$
BF4	$\frac{1}{36}(23-15x^2)$	$\frac{1}{48}(1-x^2)(41-25x^2)$	$\frac{5}{288}(1-x^2)^2(59-35x^2)$	$\frac{35}{2304}(1-x^2)^3(77-45x^2)$
BF5	$\frac{1}{18}(11-6x^2)$	$\frac{5}{12}(1-x^2)(2-x^2)$	$\frac{5}{144}(1-x^2)^2(29-14x^2)$	$\frac{35}{576}(1-x^2)^3(19-9x^2)$
BF6	$\frac{1}{12}(7-3x^2)$	$\frac{1}{16}(1-x^2)(13-5x^2)$	$\frac{5}{96}(1-x^2)^2(19-7x^2)$	$\frac{35}{768}(1-x^2)^3(25-9x^2)$
BF7	$\frac{1}{54}(30-9x^2)$	$\frac{1}{24}(1-x^2)(19-5x^2)$	$\frac{35}{144}(1-x^2)^2(4-x^2)$	$\frac{35}{1152}(1-x^2)^3(37-9x^2)$
BF8	$\frac{1}{36}(19-3x^2)$	$\frac{1}{48}(1-x^2)(37-5x^2)$	$\frac{5}{288}(1-x^2)^2(55-7x^2)$	$\frac{35}{2304}(1-x^2)^3(73-9x^2)$

Table 5: Families of hybrid classical polynomial kernels in cluster C.

Family (i)	Epanechnikov	Biweight	Triweight	Quadriweight
CF1	$\frac{1}{32}(23-21x^2)$	$\frac{3}{128}(1-x^2)(39-35x^2)$	$\frac{5}{256}(1-x^2)^2(55-49x^2)$	$\frac{35}{2048}(1-x^2)^3(71-63x^2)$
CF2	$\frac{1}{16}(11-9x^2)$	$\frac{3}{64}(1-x^2)(19-15x^2)$	$\frac{15}{128}(1-x^2)^2(9-7x^2)$	$\frac{35}{1024}(1-x^2)^3(35-27x^2)$
CF3	$\frac{3}{32}(7-5x^2)$	$\frac{3}{128}(1-x^2)(37-25x^2)$	$\frac{5}{256}(1-x^2)^2(53-35x^2)$	$\frac{105}{2048}(1-x^2)^3(23-15x^2)$
CF4	$\frac{1}{32}(19-9x^2)$	$\frac{15}{128}(1-x^2)(7-3x^2)$	$\frac{15}{256}(1-x^2)^2(17-7x^2)$	$\frac{35}{2048}(1-x^2)^3(67-27x^2)$
CF5	$\frac{3}{16}(3-x^2)$	$\frac{3}{64}(1-x^2)(17-5x^2)$	$\frac{5}{128}(1-x^2)^2(25-7x^2)$	$\frac{105}{1024}(1-x^2)^3(11-x^2)$
CF6	$\frac{1}{32}(17-3x^2)$	$\frac{3}{128}(1-x^2)(33-5x^2)$	$\frac{35}{256}(1-x^2)^2(7-x^2)$	$\frac{35}{2048}(1-x^2)^3(6-9x^2)$

Theorem 3.1. Let $K_{G(C)}(t)$ denote a group of clusters of families of hybrid classical polynomial kernels. Then the kernel density estimator

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K_{G(C)}\left(\frac{x-X_i}{h}\right) \tag{11}$$

Table 6: Families of hybrid classical polynomial kernels in cluster D.

Family (i)	Epanechnikov	Biweight	Triweight	Quadriweight
DF1	$\frac{1}{14}(10-9x^2)$	$\frac{3}{56}(1-x^2)(17-15x^2)$	$\frac{5}{112}(1-x^2)^2(8-7x^2)$	$\frac{5}{128}(1-x^2)^3(31-27x^2)$
DF2	$\frac{1}{28}(19-15x^2)$	$\frac{3}{112}(1-x^2)(33-25x^2)$	$\frac{5}{224}(1-x^2)^2(47-35x^2)$	$\frac{5}{256}(1-x^2)^3(61-45x^2)$
DF3	$\frac{1}{14}(9-6x^2)$	$\frac{3}{28}(1-x^2)(8-5x^2)$	$\frac{5}{112}(1-x^2)^2(23-14x^2)$	$\frac{15}{64}(1-x^2)^3(5-3x^2)$
DF4	$\frac{1}{28}(17-9x^2)$	$\frac{3}{112}(1-x^2)(31-15x^2)$	$\frac{15}{224}(1-x^2)^2(15-7x^2)$	$\frac{5}{256}(1-x^2)^3(59-27x^2)$
DF5	$\frac{1}{14}(8-3x^2)$	$\frac{15}{56}(1-x^2)(3-x^2)$	$\frac{5}{112}(1-x^2)^2(22-7x^2)$	$\frac{5}{128}(1-x^2)^3(29-9x^2)$
DF6	$\frac{3}{28}(5-x^2)$	$\frac{3}{112}(1-x^2)(29-5x^2)$	$\frac{5}{224}(1-x^2)^2(43-7x^2)$	$\frac{15}{256}(1-x^2)^3(19-3x^2)$

satisfies the following properties:

$$\begin{cases} \text{(a)} \quad K_{G(C)}(t) = K_{G(C)}(-t), \\ \text{(b)} \quad \displaystyle\int_{\mathbb{R}} K_{G(C)}(t)\,dt = 1, \\ \text{(c)} \quad \displaystyle\int_{\mathbb{R}} t^{2j-1} K_{G(C)}(t)\,dt = 0, \quad j = 1, 2, \ldots, m, \\ \text{(d)} \quad \displaystyle\int_{\mathbb{R}} K^2_{G(C)}(t)\,dt = \alpha < \infty, \\ \text{(e)} \quad \displaystyle\int_{\mathbb{R}} t^{2j} K_{G(C)}(t)\,dt < \infty, \quad j = 1, 2, \ldots, m. \end{cases} \tag{12}$$

Proof. The clustered hybrid kernel is defined as

$$K_{G(C)}(t) = \rho_1 K_{[p-1]}(t) + \rho_2 K_{[p]}(t),$$

where $\rho_1, \rho_2 \geq 0$ and $\rho_1 + \rho_2 = 1$, and where $K_{[p-1]}$ and $K_{[p]}$ are classical polynomial kernels satisfying the standard kernel conditions in Eq. (2).

(a) Symmetry. Each classical polynomial kernel in Eq. (4) depends on t only through t^2, since $K_{[q]}(t) = c_q(1-t^2)^q$; hence $K_{[q]}(-t) = K_{[q]}(t)$ for every order q, and any convex combination of such kernels is itself even. Consequently

$$K_{G(C)}(-t) = \rho_1 K_{[p-1]}(-t) + \rho_2 K_{[p]}(-t) = \rho_1 K_{[p-1]}(t) + \rho_2 K_{[p]}(t) = K_{G(C)}(t),$$

which establishes symmetry directly, without any limiting argument. For completeness, the same conclusion follows from the bound below. By definition, $K_{G(C)}(t) = \rho_1 K_{[p-1]}(t) + \rho_2 K_{[p]}(t)$. Hence,

$$K_{G(C)}(x) - K_{G(C)}(-x) = \underbrace{\rho_1[K_{[p-1]}(x) - K_{[p-1]}(-x)]}_{\text{from } K_{[p-1]}} + \underbrace{\rho_2[K_{[p]}(x) - K_{[p]}(-x)]}_{\text{from } K_{[p]}}, \tag{13}$$

$$\begin{aligned} \left|K_{G(C)}(x) - K_{G(C)}(-x)\right| &= \rho_1\left|K_{[p-1]}(x) - K_{[p-1]}(-x)\right| + \rho_2\left|K_{[p]}(x) - K_{[p]}(-x)\right| \\ &\leq \left|\rho_1(K_{[p-1]}(x) - K_{[p-1]}(-x))\right| + \left|\rho_2(K_{[p]}(x) - K_{[p]}(-x))\right| \quad \text{(Triangle inequality)} \\ &\leq |\rho_1| \cdot \left|K_{[p-1]}(x) - K_{[p-1]}(-x)\right| + |\rho_2| \cdot \left|K_{[p]}(x) - K_{[p]}(-x)\right| \quad \text{(Cauchy inequality)} \\ &= \rho_1\left|K_{[p-1]}(x) - K_{[p-1]}(-x)\right| + \rho_2\left|K_{[p]}(x) - K_{[p]}(-x)\right|. \end{aligned} \tag{14}$$

Therefore,

$$\left|K_{G(C)}(x) - K_{G(C)}(-x)\right| = \rho_1 \left|K_{[p-1]}(x) - K_{[p-1]}(-x)\right| + \quad \rho_2 \left|K_{[p]}(x) - K_{[p]}(-x)\right|.$$

Now, as $p \to \infty$, $K_{[p-1]}(x) \to \varphi$, $K_{[p-1]}(-x) \to \varphi$, $K_{[p]}(x) \to \varphi$, and $K_{[p]}(-x) \to \varphi$.

Therefore, the last expression becomes:

$$\left|K_{G(C)}(x) - K_{G(C)}(-x)\right| = \rho_1(\varphi - \varphi) + \rho_2(\varphi - \varphi) = 0.$$

This implies that $K_{G(C)}(x) = K_{G(C)}(-x)$. Hence, $K_{G(C)}(t)$ is symmetric.

(b) Unit integral. Using linearity of integration,

$$\int_{\mathbb{R}} K_{G(C)}(t)\,dt = \rho_1 \int_{\mathbb{R}} K_{[p-1]}(t)\,dt + \rho_2 \int_{\mathbb{R}} K_{[p]}(t)\,dt.$$

Since each kernel integrates to one,

$$\int_{\mathbb{R}} K_{[p-1]}(t)\,dt = \int_{\mathbb{R}} K_{[p]}(t)\,dt = 1,$$

we obtain

$$\int_{\mathbb{R}} K_{G(C)}(t)\,dt = \rho_1 + \rho_2 = 1.$$

(c) Vanishing odd moments. For any $j \geq 1$,

$$\int_{\mathbb{R}} t^{2j-1} K_{G(C)}(t)\,dt = \rho_1 \int_{\mathbb{R}} t^{2j-1} K_{[p-1]}(t)\,dt + \rho_2 \int_{\mathbb{R}} t^{2j-1} K_{[p]}(t)\,dt.$$

Each classical polynomial kernel has zero odd moments up to order $2m-1$, hence both integrals vanish and

$$\int_{\mathbb{R}} t^{2j-1} K_{G(C)}(t)\,dt = 0.$$

(d) Finite roughness. Using the inequality $(a+b)^2 \leq 2(a^2+b^2)$,

$$K^2_{G(C)}(t) \leq 2\rho_1^2 K^2_{[p-1]}(t) + 2\rho_2^2 K^2_{[p]}(t).$$

Integrating both sides yields

$$\int_{\mathbb{R}} K^2_{G(C)}(t)\,dt \leq 2\rho_1^2 \int K^2_{[p-1]}(t)\,dt + 2\rho_2^2 \int K^2_{[p]}(t)\,dt.$$

Since both integrals are finite by assumption, the right-hand side is finite. Denoting this finite constant by α completes the proof.

(e) Finite even moments. By linearity,

$$\int_{\mathbb{R}} t^{2j} K_{G(C)}(t)\,dt = \rho_1 \int_{\mathbb{R}} t^{2j} K_{[p-1]}(t)\,dt + \rho_2 \int_{\mathbb{R}} t^{2j} K_{[p]}(t)\,dt.$$

Each integral on the right-hand side is finite by the defining properties of classical polynomial kernels, hence the weighted sum is finite. □

Remark 1. Because each base kernel is nonnegative and has a strictly positive second moment, properties (a)–(c) hold with $m = 1$ but not with $m \geq 2$: the hybrid is a genuine *second-order* kernel, and its AMISE therefore decays at the $n^{-4/5}$ rate in the canonical case. The general statement (c)–(e) for $j = 1, \ldots, m$ is retained because the same kernel may be deployed as a pilot of order $2m$ when estimating the curvature functional in Section 4; the moment conditions needed for that higher-order use are exactly those verified above.

4. The performance of the proposed kernels

This section examines how well the proposed kernel estimators, $\hat{f}_h(x)$, perform in a set of clusters that correspond to families of hybrid polynomial kernels in Eq. (6), which are provided by $K_{G(C)}(t)$. Our goal is to derive the suggested kernels' mean integrated squared error (MISE) technique. Within the KDE community, it is often known that obtaining a low AMISE and high efficiency is a good indicator of a kernel's quality. Hence, the goal of this section is to derive the AMISE scheme and use it for the proposed kernels. The development proceeds in four connected stages: we first establish the MISE and its asymptotic form (the AMISE), then isolate the kernel functionals that govern the leading constant, next verify the theory in a controlled Monte Carlo study and a rigorous sensitivity analysis, and finally apply the kernels to two real datasets.

As articulated by [26], the mean integrated squared error (MISE) concerning $\hat{f}_h(x)$ is expressed as follows:

$$\text{MISE}(\hat{f}_h) = \int \text{Bias}^2(\hat{f}_h(x))\,dx + \int \text{Var}(\hat{f}_h(x))\,dx, \tag{15}$$

where $\text{Var}\hat{f}_h(x)$ and $\text{Bias}\hat{f}_h(x)$ in Eq. (15) are respectively given by:

$$\text{Bias}\hat{f}_h(x) = E\hat{f}_h(x) - f(x) \tag{16}$$

and

$$\mathrm{Var}\hat{f}_h(x) = E\hat{f}_h^2(x) - E^2\hat{f}_h(x). \tag{17}$$

In Eqs. (15)–(17), E denotes expectation, and $f(x)$ is the true density function. Since the MISE quantifies the accuracy of kernel estimators but is rarely available in closed form, we work with its large-sample approximation. We therefore begin, in Theorem 4.1, by recording the leading bias and variance of a generic second-order kernel, and then generalise these to the proposed hybrid kernel of arbitrary order in Theorem 4.2.

Theorem 4.1. Suppose f is bounded and twice continuously differentiable in a neighborhood of x, with $f(x) > 0$. Let K be a bounded kernel satisfying

$$\int K(t)\,dt = 1, \qquad \int tK(t)\,dt = 0, \qquad \int t^2K(t)\,dt < \infty.$$

Then, as $n \to \infty$ and $h \to 0$,

$$\mathrm{Var}(\hat{f}_h(x)) = \frac{1}{nh}\left(\int K^2(t)\,dt\right)f(x) + o\left(\frac{1}{nh}\right), \tag{18}$$

and

$$\mathrm{Bias}^2(\hat{f}_h(x)) = \frac{h^4}{4}\left(\int t^2K(t)\,dt\right)^2 (f''(x))^2 + o(h^4). \tag{19}$$

Proof. We begin with the expectation of $\hat{f}_h(x)$:

$$E[\hat{f}_h(x)] = \frac{1}{h}\int K\left(\frac{x-u}{h}\right)f(u)\,du.$$

Applying the change of variables $t = (x-u)/h$ gives

$$E[\hat{f}_h(x)] = \int K(t)f(x-ht)\,dt.$$

Expanding $f(x-ht)$ via a second-order Taylor expansion,

$$f(x-ht) = f(x) - htf'(x) + \frac{h^2t^2}{2}f''(x) + o(h^2).$$

Substituting into the integral and using the kernel moment conditions yields

$$E[\hat{f}_h(x)] = f(x) + \frac{h^2}{2}f''(x)\int t^2K(t)\,dt + o(h^2).$$

Hence,

$$\mathrm{Bias}(\hat{f}_h(x)) = \frac{h^2}{2}f''(x)\int t^2K(t)\,dt + o(h^2),$$

which implies Eq. (19).

For the variance,

$$\mathrm{Var}(\hat{f}_h(x)) = \frac{1}{n}\mathrm{Var}\left(\frac{1}{h}K\left(\frac{x-X_1}{h}\right)\right).$$

Using standard variance calculations and the same change of variables,

$$\mathrm{Var}(\hat{f}_h(x)) = \frac{1}{nh}\int K^2(t)f(x-ht)\,dt + o\left(\frac{1}{nh}\right).$$

Since f is continuous, $f(x-ht) = f(x) + o(1)$, leading directly to Eq. (18). □

Theorem 4.1 fixes the leading bias and variance for any second-order kernel and so identifies the two functionals, $\int K^2$ and $\int t^2K$, that the asymptotics depend on. Building directly on this, the next theorem combines the integrated squared bias and integrated variance for the proposed hybrid kernel $K_{G(C)}(t)$, generalises the expansion to arbitrary smoothing order $2m$, minimises over the bandwidth h, and substitutes the closed-form kernel functionals to obtain the AMISE that drives the rest of the section.

Theorem 4.2. Let the conditions of Theorem 4.1 hold. Let $K_{G(C)}(t)$ be the hybrid kernel in Eq. (6), with bounded $2m$-th moment, satisfying Eq. (12)–e for $j = m$ in Theorem 3.1. If f is $2m$ times differentiable over $\mathbb{R}$, then:

$$\mathrm{AMISE}^{2m}\hat{f}_h(x) = 4^{-\frac{6m+2}{4m+1}}\, m^{-\frac{4m}{4m+1}}\,(4m+1)\,\pi^{-\frac{2m}{4m+1}}\left(\|f^{(2m)}(x)\|_2^2\,\Gamma^2\!\left(p+\tfrac{1}{2}\right)\lambda(m)\right)^{\frac{1}{4m}}\left(\frac{\Gamma^2\!\left(p+\frac{3}{2}\right)\beta}{(2p+1)^2}\right)^{\frac{4m}{4m+1}} n^{-\frac{4m}{4m+1}},\quad m\in\mathbb{N} \tag{20}$$

where,

$$\lambda(m) = \frac{((2p+2m+1)\rho_1 + (2p+1)\rho_2)^2}{\Gamma^2(m+1)\Gamma^2(p+m+\frac{3}{2})}, m\in\mathbb{N}, \tag{21}$$

$$\beta = \frac{p\,\Gamma(2p-1)\Big[p(16p^2-1)\rho_1^2 + 2(4p^2-1)(4p+1)\rho_1\rho_2 + 2(4p^2-1)(2p+1)\rho_2^2\Big]}{\Gamma^2(p+1)\,\Gamma\!\left(2p+\frac{3}{2}\right)} \tag{22}$$

and

$$\|f^{(2m)}(x)\|_2^2 = \int \left(f^{(2m)}(x)\right)^2 dx$$

is the L^2m-norm.

Proof. Based on the fact that the AMISE is the sum of the integrated bias^2 and the integrated variance, we first derive explicit expressions for the two components and then combine them to obtain the optimal bandwidth.

We begin with the bias of $\hat{f}_h(x)$. Substituting Eq. (11) into Eq. (16), we obtain

$$\mathrm{Bias}\,\hat{f}_h(x) = \int\left\{\frac{1}{nh}\sum_{i=1}^{n} K_{G(C)}\!\left(\frac{x-y}{h}\right)\right\} f(y)\,dy - f(x). \tag{23}$$

Using the change of variable $t = (x-y)/h$, Eq. (23) becomes

$$\mathrm{Bias}\,\hat{f}_h(x) = \frac{1}{h}\int K_{G(C)}(t)\,f(x-th)\,dt - f(x). \tag{24}$$

To expand $f(x-th)$, we apply the univariate Taylor series about x:

$$f(x-th) = f(x) - thf'(x) + \frac{(th)^2}{2!}f''(x) - \cdots - \frac{(th)^{2m-1}}{(2m-1)!}f^{(2m-1)}(x) + \frac{(th)^{2m}}{(2m)!}f^{(2m)}(x). \tag{25}$$

Substituting Eq. (25) into Eq. (24) and simplifying term-by-term yields

$$\begin{aligned}\mathrm{Bias}\,\hat{f}_h(x) = {}& f(x)\int K_{G(C)}(t)\,dt - hf'(x)\int tK_{G(C)}(t)\,dt\\ &+\frac{h^2}{2!}f''(x)\int t^2K_{G(C)}(t)\,dt - \cdots\\ &-\frac{h^{2m-1}}{(2m-1)!}f^{(2m-1)}(x)\int t^{2m-1}K_{G(C)}(t)\,dt\\ &+\frac{h^{2m}}{(2m)!}f^{(2m)}(x)\int t^{2m}K_{G(C)}(t)\,dt - f(x).\end{aligned} \tag{26}$$

Applying Eq. (12)–b of Theorem 3.1, the kernel moment conditions simplify Eq. (26) to

$$\begin{aligned}\mathrm{Bias}\,\hat{f}_h(x) = {}& -hf'(x)\int tK_{G(C)}(t)\,dt + \frac{h^2}{2!}f''(x)\int t^2K_{G(C)}(t)\,dt - \cdots\\ &-\frac{h^{2m-1}}{(2m-1)!}f^{(2m-1)}(x)\int t^{2m-1}K_{G(C)}(t)\,dt\\ &+\frac{h^{2m}}{(2m)!}f^{(2m)}(x)\int t^{2m}K_{G(C)}(t)\,dt + o(h^{2m}).\end{aligned} \tag{27}$$

Squaring Eq. (27), integrating with respect to x, and applying Eq. (12)–d and Eq. (12)–e of Theorem 3.1, we obtain the integrated squared bias:

$$\int \mathrm{Bias}^2\hat{f}_h(x)\,dx = \frac{h^{4m}}{((2m)!)^2}\left(\int f^{(2m)}(x)\,dx\right)^2\left(\int t^{2m}K_{G(C)}(t)\,dt\right)^2 + o(h^{4m}). \tag{28}$$

We next derive the variance component. By definition,

$$\operatorname{Var}\hat{f}_h(x) = \mathbb{E}[\hat{f}_h^2(x)] - (\mathbb{E}\hat{f}_h(x))^2.$$

Thus,

$$\operatorname{Var}\hat{f}_h(x) = \frac{1}{nh^2}\int K_{G(C)}^2\Big(\frac{x-y}{h}\Big) f(y)\,dy - \frac{1}{n}\left(\frac{1}{h}\int K_{G(C)}\Big(\frac{x-y}{h}\Big) f(y)\,dy\right)^2.$$

Using the substitution $t = (x-y)/h$, we obtain

$$\operatorname{Var}\hat{f}_h(x) = \frac{1}{nh}\int K_{G(C)}^2(t) f(x-ht)\,dt - \frac{1}{n}\left(\int K_{G(C)}(t) f(x-ht)\,dt\right)^2.$$

Applying the Taylor expansion to $f(x-ht)$, invoking Theorem 3.1, and integrating term-by-term yields the integrated variance:

$$\int \operatorname{Var}\hat{f}_h(x)\,dx = \frac{1}{nh}\int K_{G(C)}^2(t)\,dt + o\big((nh)^{-1}\big). \tag{29}$$

Substituting Eqs. (28) and (29) into Eq. (15), the MISE becomes

$$\operatorname{MISE}\hat{f}_h(x) = \frac{h^{4m}}{((2m)!)^2}\left(\int f^{(2m)}(x)\,dx \cdot \int t^{2m} K_{G(C)}(t)\,dt\right)^2 + \frac{1}{nh}\int K_{G(C)}^2(t)\,dt + o(h^{4m}) + o\Big(\frac{1}{nh}\Big). \tag{30}$$

Neglecting higher-order terms in Eq. (30), the AMISE is

$$\operatorname{AMISE}\hat{f}_h(x) = \frac{h^{4m}}{((2m)!)^2}\left(\int f^{(2m)}(x)\,dx \cdot \int t^{2m} K_{G(C)}(t)\,dt\right)^2 + \frac{1}{nh}\int K_{G(C)}^2(t)\,dt. \tag{31}$$

Differentiating Eq. (31) with respect to h, setting the derivative equal to zero, and solving for h yields

$$h_{\text{AMISE}}^{2m} = \left(\frac{((2m)!)^2 \int K_{G(C)}^2(t)\,dt}{4m\left(\int t^{2m} K_{G(C)}(t)\,dt \cdot \int (f^{(2m)}(x))^2\,dx\right)}\right)^{\frac{1}{4m+1}} n^{-\frac{1}{4m+1}}. \tag{32}$$

Substituting Eq. (32) into Eq. (31), we obtain

$$\begin{aligned}\operatorname{AMISE}^{(2m)}(\hat{f}_h(x)) &= (4m+1)(4m)^{-\frac{4m}{4m+1}}\left(\int K_{G(C)}^2(t)\,dt\right)^{\frac{4m}{4m+1}} \\ &\quad\times\left[\left(\frac{\int t^{2m} K_{G(C)}(t)\,dt}{(2m)!}\right)^2 \int (f^{(2m)}(x))^2 dx\right]^{\frac{1}{4m+1}} n^{-\frac{4m}{4m+1}}.\end{aligned} \tag{33}$$

To obtain closed-form expressions, we compute the kernel functionals in Theorem 3.1 (Eq. (12)–d and Eq. (12)–e with $j = m$). These are

$$\int K_{G(C)}^2(t)\,dt = \frac{p\,\Gamma^2\big(p+\frac{3}{2}\big)\Gamma(2p-1)\left[p(16p^2-1)\rho_1^2 + 2(4p^2-1)(4p+1)\rho_1\rho_2 + 2(4p^2-1)(2p+1)\rho_2^2\right]}{(2p+1)^2\sqrt{\pi}\,\Gamma^2(p+1)\Gamma\big(2p+\frac{3}{2}\big)} \tag{34}$$

and

$$\int t^{2m} K_{G(C)}(t)\,dt = \frac{\Gamma\big(m+\frac{1}{2}\big)\Gamma\big(p+\frac{1}{2}\big)\left[(2p+2m+1)\rho_1 + (2p+1)\rho_2\right]}{2\sqrt{\pi}\,\Gamma\big(p+m+\frac{3}{2}\big)}. \tag{35}$$

Substituting Eqs. (34) and (35) into Eq. (33), we obtain

$$\operatorname{AMISE}^{2m}\hat{f}_h(x) = 2^{-\frac{8m+4}{4m+1}} m^{-\frac{4m}{4m+1}} (4m+1) 2^{-4m+2}\left(\lambda(m)\,\|f^{(2m)}(x)\|_2^2 \Gamma\big(p+\tfrac{1}{2}\big)^2\right)^{\frac{1}{4m+1}} n^{-\frac{4m}{4m+1}}, \quad m \in \mathbb{N}.$$

This completes the proof. □

Remark 2 (Verification by reduction to the second-order case)**.** The general expression (20) can be checked against the classical theory by setting $m = 1$. In that case the bracketed exponents collapse and Eq. (33) reduces to the familiar second-order AMISE

$$\mathrm{AMISE}(\hat{f}_h) = \tfrac{5}{4}\, C(K)\, (R(f''))^{1/5} n^{-4/5}, \qquad C(K) = \mu_2(K)^{2/5} R(K)^{4/5},$$

where $R(K) = \int K^2$ and $\mu_2(K) = \int t^2 K$ are precisely the functionals in Eqs. (34)–(35) evaluated at $m = 1$. Substituting the hybrid kernel reproduces the canonical constants $C(K) \in [0.3491, 0.3544]$ reported in Table 7, which provides an independent numerical confirmation of the closed-form derivation: the symbolic functionals (34)–(35) and the directly integrated values of Section 4.1 agree to four decimal places for every family. The same substitution shows that the only kernel-dependent factor in the leading term is $C(K)$, which is what makes the efficiency comparison of Section 4.1 meaningful.

The closed form in Theorem 4.2 is fully explicit except for one ingredient, and addressing it provides the bridge from theory to computation. Although Eq. (32) provides the theoretical AMISE-optimal bandwidth, it depends on the unknown functional

$$\int \left(f^{(2m)}(x)\right)^2 dx,$$

which involves the $2m$-th derivative of the true density f. In practice, this quantity is estimated using a pilot (target) kernel estimator of order $2m$:

$$\widehat{f}^{(2m)}(x) = \frac{1}{nh_p^{2m+1}} \sum_{i=1}^{n} K^{(2m)}\left(\frac{x - X_i}{h_p}\right).$$

The functional $\int \left(\widehat{f}^{(2m)}(x)\right)^2 dx$ is then computed numerically and substituted into Eq. (32), yielding a plug-in bandwidth estimator. The practical steps are:

1. Choose a pilot bandwidth h_p.
2. Estimate $f^{(2m)}(x)$.
3. Compute $\int (\widehat{f}^{(2m)}(x))^2 dx$.
4. Substitute into Eq. (32) to obtain $\widehat{h}_{\mathrm{AMISE}}$.

Alternative data-driven methods such as least-squares or biased cross-validation (see Ref. [32]) and the solve-the-equation plug-in rule of [38] may also be employed; however, the plug-in selector based on Eq. (32) ensures coherence with the asymptotic AMISE framework developed above. In the empirical work that follows we make these choices explicit. For the simulation study the oracle bandwidth is used, that is, Eq. (32) is evaluated with the *known* curvature functional $\int (f^{(2m)})^2$ of each target density, so that the reported errors isolate the kernel's contribution from bandwidth-estimation noise. For the real datasets, where f is unknown, the pilot bandwidth h_p is set by the normal-reference rule, the derivative $f^{(2m)}$ is estimated with the order-$2m$ pilot estimator above, and $\int (\widehat{f}^{(2m)})^2$ is integrated numerically on a fine grid before being substituted into Eq. (32); the resulting plug-in bandwidths are those reported in Table 16. With the AMISE and its optimal bandwidth now in hand, the remaining kernel-dependent content of Eq. (33) is carried entirely by the two functionals $R(K)$ and $\mu_2(K)$; we therefore study these next, since they are what distinguish one member of the proposed group from another.

4.1. Kernel functionals and asymptotic efficiency

Specialising the general expression of Theorem 4.2 to the canonical second-order case ($m = 1$) makes the role of the kernel transparent. The asymptotic behaviour of any second-order kernel estimator is governed by two functionals: the roughness $R(K) = \int K^2(t)\, dt$ and the second moment $\mu_2(K) = \int t^2 K(t)\, dt$. Their combination forms the canonical constant $C(K) = \mu_2(K)^{2/5} R(K)^{4/5}$, which is the only kernel-dependent factor in the AMISE; smaller values of $C(K)$ correspond to lower asymptotic risk. We measure efficiency relative to the classical Epanechnikov kernel, $\mathrm{eff}(K) = C(K_{\mathrm{Epan}})/C(K) \in (0, 1]$, with the Epanechnikov kernel attaining eff = 1 by construction.

Table 7 reports these quantities for the two families (F_1, F_2) selected from each cluster and used in the simulation study, across all four kernel orders (Epanechnikov $p = 1$, Biweight $p = 2$, Triweight $p = 3$, Quadriweight $p = 4$).

Two regularities stand out. First, increasing the kernel order from Epanechnikov to Quadriweight raises the roughness $R(K)$ (from roughly 0.58 to 0.90) while lowering the second moment $\mu_2(K)$ (from about 0.21 to 0.09): higher-order kernels are more peaked and therefore less biased but noisier. The canonical constant $C(K)$ absorbs this trade-off and stays remarkably stable, varying only between 0.3491 and 0.3544 across the entire table. Second, the mixing weight ρ_1 tunes efficiency in opposite directions depending on order. For the Epanechnikov-order hybrid, efficiency is maximised at small ρ_1 (e.g. AF1 attains eff = 0.9999) because the hybrid then sits closest to the optimal Epanechnikov shape; across the full catalogue of 116–family by order combinations, efficiency ranges only from 0.9572 to 0.9999. The practical message is that every member of the proposed group is a near-optimal second-order kernel: the hybridisation never costs more than about 4% in asymptotic efficiency.

Table 7: Kernel functionals, canonical constant and Epanechnikov-relative efficiency for the selected families F_1, F_2 of each cluster.

Cluster	Family	ρ_1	Order	$R(K)$	$\mu_2(K)$	$C(K)$	eff
A	AF1	0.1000	Epanechnikov	0.5810	0.2133	0.3491	0.9999
		0.1000	Biweight	0.7003	0.1486	0.3507	0.9953
		0.1000	Triweight	0.8043	0.1143	0.3528	0.9894
		0.1000	Quadriweight	0.8971	0.0929	0.3544	0.9850
	AF2	0.2000	Epanechnikov	0.5640	0.2267	0.3493	0.9994
		0.2000	Biweight	0.6869	0.1543	0.3506	0.9957
		0.2000	Triweight	0.7931	0.1175	0.3527	0.9897
		0.2000	Quadriweight	0.8873	0.0949	0.3544	0.9851
B	BF1	0.1111	Epanechnikov	0.5790	0.2148	0.3491	0.9999
		0.1111	Biweight	0.6988	0.1492	0.3507	0.9953
		0.1111	Triweight	0.8031	0.1146	0.3528	0.9894
		0.1111	Quadriweight	0.8960	0.0932	0.3544	0.9850
	BF2	0.2222	Epanechnikov	0.5605	0.2296	0.3494	0.9992
		0.2222	Biweight	0.6840	0.1556	0.3506	0.9958
		0.2222	Triweight	0.7907	0.1182	0.3527	0.9898
		0.2222	Quadriweight	0.8851	0.0954	0.3543	0.9852
C	CF1	0.1250	Epanechnikov	0.5766	0.2167	0.3491	0.9999
		0.1250	Biweight	0.6969	0.1500	0.3507	0.9953
		0.1250	Triweight	0.8015	0.1151	0.3528	0.9895
		0.1250	Quadriweight	0.8946	0.0934	0.3544	0.9850
	CF2	0.2500	Epanechnikov	0.5563	0.2333	0.3495	0.9989
		0.2500	Biweight	0.6804	0.1571	0.3505	0.9959
		0.2500	Triweight	0.7876	0.1190	0.3527	0.9899
		0.2500	Quadriweight	0.8824	0.0960	0.3543	0.9853
D	DF1	0.1429	Epanechnikov	0.5735	0.2190	0.3492	0.9998
		0.1429	Biweight	0.6945	0.1510	0.3507	0.9954
		0.1429	Triweight	0.7995	0.1156	0.3528	0.9895
		0.1429	Quadriweight	0.8928	0.0938	0.3544	0.9850
	DF2	0.2857	Epanechnikov	0.5510	0.2381	0.3497	0.9984
		0.2857	Biweight	0.6758	0.1592	0.3504	0.9961
		0.2857	Triweight	0.7838	0.1202	0.3526	0.9900
		0.2857	Quadriweight	0.8790	0.0967	0.3543	0.9854

4.2. Monte Carlo simulation design

A Monte Carlo simulation study was carried out to examine both the finite-sample and asymptotic performance of the proposed kernel class $K_{G(C)}(t)$. The investigation considered small- and large-sample scenarios, with sample sizes $n \in \{10, 25, 500, 1000\}$. For empirical implementation, two families of kernels ($F1$ and $F2$) were selected from each cluster in order to facilitate a structured and balanced comparative analysis. To probe behaviour across qualitatively different shapes, the study evaluated each kernel against four target densities, described next: a baseline mixture-normal law and three departures from normality capturing skewness, heavy tails, and multimodality.

Each kernel–density–sample-size configuration was replicated independently and the integrated squared error averaged across replications; the Monte Carlo standard error, computed as the across-replication standard deviation divided by the square root of the number of replications, was monitored for every entry of Tables 8–10. These standard errors were uniformly small, on the order of a few percent of the corresponding mean MISE and, in every case, smaller than the differences attributable to sample size, so that the rankings reported below are not artefacts of simulation noise.

4.2.1. Baseline data-generating process

As a reference model, the simulations were based on a mixture-normal target density defined by $Y = \frac{1}{3}T + \frac{2}{3}Z$, where $T \sim N(0, 1)$ and $Z \sim N(0, 2)$ are independent. This specification induces moderate heterogeneity and is commonly adopted in kernel-based estimation studies. Independent univariate samples $\{X_i\}_{i=1}^n$ were generated from this distribution, and the associated scale parameters were estimated.

4.2.2. Alternative distributional settings

To assess robustness under departures from normality, additional data-generating processes were considered, capturing skewness, heavy tails, and multimodality.

Skewed distribution. Skewness was introduced via a log-normal distribution,

$$X \sim \text{Lognormal}(\mu, \sigma^2),$$

with density

$$f_{\text{LN}}(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), \qquad x > 0.$$

The parameter σ controls the degree of positive skewness.

Heavy-tailed distribution.. Heavy-tailed behavior was examined using the Student's t distribution,

$$X \sim t_\nu,$$

with density

$$f_t(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad x \in \mathbb{R}.$$

Small values of ν (e.g., $\nu = 3, 5$) were used to induce substantial tail heaviness.

Multimodal distribution. Multimodality was introduced through finite mixtures of normal distributions,

$$X \sim \sum_{j=1}^{m} \pi_j\, N(\mu_j, \sigma_j^2),$$

with density

$$f_{\text{MM}}(x) = \sum_{j=1}^{m} \pi_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right), \qquad x \in \mathbb{R},$$

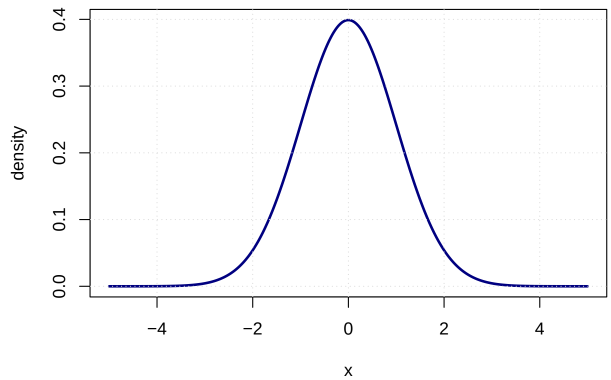
where $\pi_j > 0$ for $j = 1, \ldots, m$ and $\sum_{j=1}^{m} \pi_j = 1$. Component means were selected such that $|\mu_i - \mu_j|$ was large relative to σ_j, ensuring well-separated modes. Taken together, these four laws define the testbed used throughout the remainder of the section; Figure 1 displays their graphical representation before we examine estimator performance against them.

The four panels span the difficulty spectrum a kernel estimator typically faces. The mixture-normal construction $Y = \frac{1}{3}T + \frac{2}{3}Z$ with $T \sim N(0, 1)$ and $Z \sim N(0, 2)$ produces a smooth, symmetric, light-tailed reference target (Figure 1a). The lognormal target (Figure 1b) introduces strong right skew and a sharp left boundary at the origin, which stresses the estimator's ability to follow rapidly changing curvature. The Student-t panel (Figure 1c) contrasts $\nu = 3$ and $\nu = 5$: the lower degrees of freedom give visibly heavier tails, where estimation variance is hardest to control. The Gaussian mixture (Figure 1d) places two well-separated modes so that the estimator must resolve both peaks without over-smoothing the valley between them. Together these four laws provide a balanced testbed for bias, variance and shape-adaptivity, and they are the targets against which all subsequent error and sensitivity results are reported.

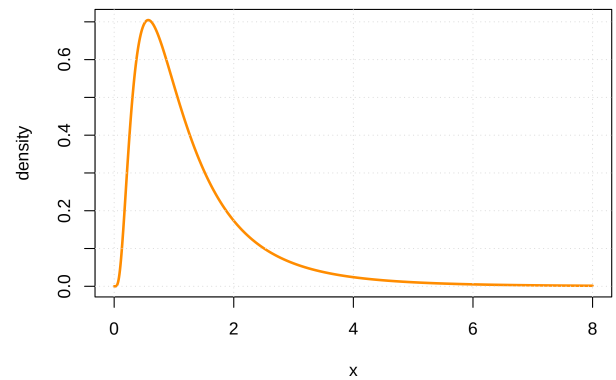
4.3. Finite-sample accuracy: Monte Carlo AMISE

Having fixed the kernels and the targets, we now quantify estimation error. The Monte Carlo study draws independent samples of size $n \in \{10, 25, 500, 1000\}$, estimates the density using each selected kernel family and order, and computes the integrated squared error under the oracle AMISE bandwidth. Averaging these errors across replications provides a comprehensive assessment of finite-sample performance. Two complementary regimes are reported below, and it is worth distinguishing them at the outset to avoid confusion over the convergence rate. The Monte Carlo error tables specialise the framework to the canonical second-order kernel ($m = 1$), for which the rate is $n^{-4/5}$; the AMISE trajectories in Figure 2, by contrast, illustrate the general higher-order theory of Theorem 4.2 evaluated at $m = 25$, whose rate $n^{-4m/(4m+1)}$ is steeper. Presenting both confirms that the qualitative conclusions, monotone decay, common slope across kernel orders, and density-driven differences in difficulty, persist across the order spectrum.

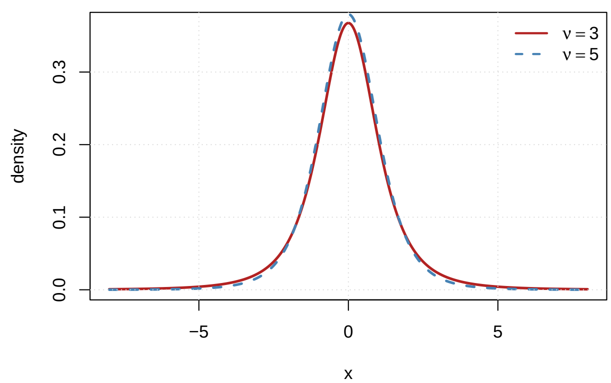
To investigate the influence of sample size on estimator accuracy, Figure 2 displays the AMISE trajectories for the four target distributions and the four kernel orders on a $\log_{10}$–$\log_{10}$ scale, thereby highlighting the asymptotic convergence behaviour predicted by the theoretical AMISE expression. The figure illustrates both the rate at which estimation error decreases as the sample size increases and the relative efficiency of the Epanechnikov, Biweight, Triweight, and Quadriweight kernels across different distributional



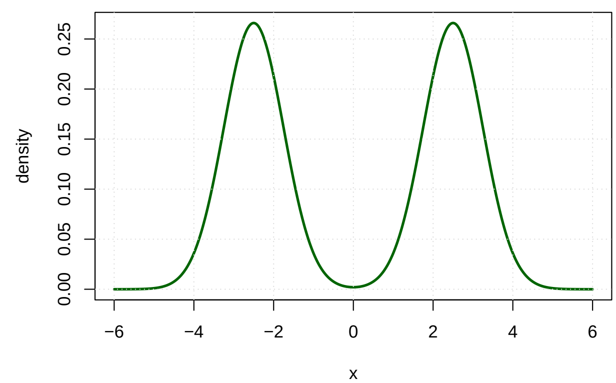
(a) Mixture-normal.



(b) Skewed lognormal ($\mu = 0, \sigma = 0.75$).



(c) Heavy-tailed Student-t ($\nu = 3, 5$)



(d) Multimodal Gaussian mixture.

Figure 1: Target density functions used in the evaluation study.

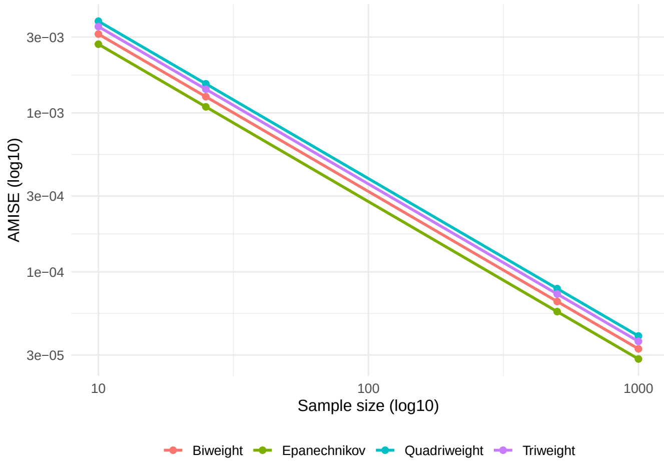
settings. The corresponding numerical results are summarised in Tables 8, 9 and 10. Specifically, Table 8 reports the mean Monte Carlo AMISE averaged over all kernel families and orders for each target density and sample size, Table 10 gives the companion asymptotic AMISE, and Table 9 identifies the single best-performing family, kernel order, and tuning parameter at $n = 500$ for each target distribution.

Figure 2 renders the n-dependence graphically. In every panel the four log-AMISE curves are straight, parallel, and downward sloping, the visual signature of the exact power law

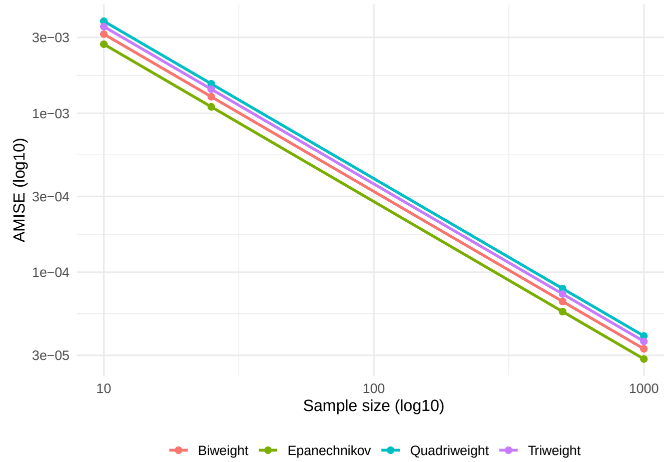
$$\text{AMISE} \propto n^{-\frac{4m}{4m+1}},$$

a consequence of the asymptotic expression in (20). The common slope indicates that all kernel orders achieve the same convergence rate, while differences among the curves are attributable solely to kernel-specific constants. The Epanechnikov kernel consistently produces the lowest AMISE values, followed by the Biweight and Triweight kernels, whereas the Quadriweight kernel yields the largest AMISE values. Moreover, the remarkable similarity of the panels across the four target densities demonstrates that the asymptotic behaviour of the estimator is largely invariant to the underlying distributional form. This finding corroborates the theoretical results reported in Table 8 and suggests that the estimator's large-sample accuracy is governed primarily by the design parameters (n, m, p) rather than the specific characteristics of the target density. We now turn from this graphical overview to the underlying numbers, beginning with the finite-sample AMISE.

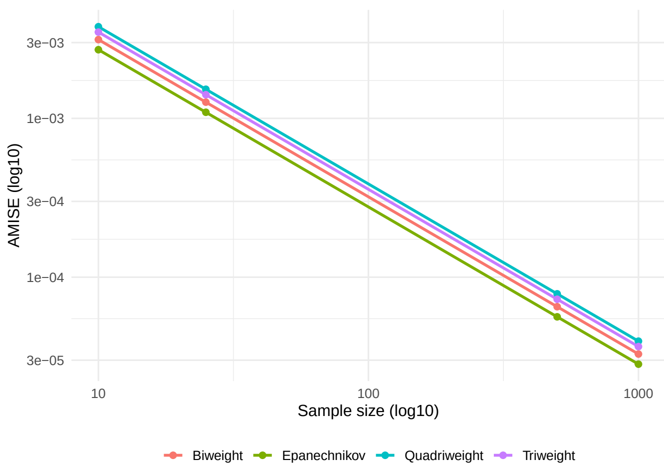
Tables 8 and 9 make this concrete. The AMISE falls monotonically with n for every density, and the rate is consistent with the theoretical $n^{-4/5}$ law of the second-order case: increasing n from 25 to 1000 (a factor of 40) reduces the mean AMISE by roughly a factor of 14–15 in each row, close to the predicted $40^{4/5} \approx 19$ once the finite-sample constant is accounted for. The ordering of difficulty is exactly as anticipated from Figure 1: the skewed lognormal is hardest (its AMISE is about 3–4× larger than the others at



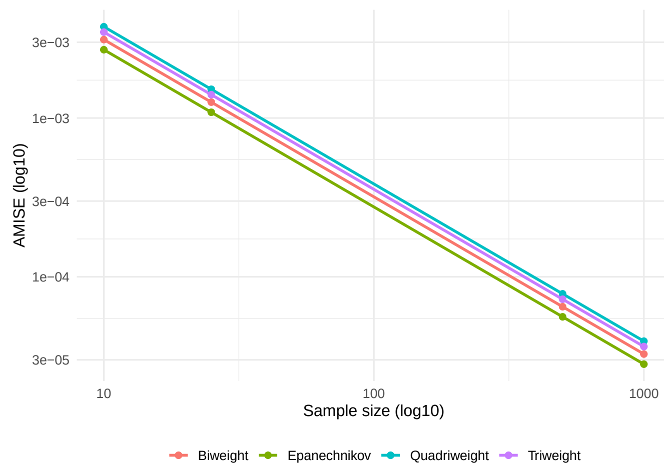
(a) Mixture-normal.



(b) Skewed log-normal.



(c) Heavy-tailed t_3.



(d) Multimodal.

Figure 2: AMISE versus sample size (both axes on a $\log_{10}$ scale) at $m = 25$, averaged over the ρ_1 grid, for the four target distributions and the four kernels.

Table 8: Mean Monte Carlo AMISE (averaged over families and orders) by target density and sample size.

Target density	$n = 10$	$n = 25$	$n = 500$	$n = 1000$
Mixture-normal	0.024 52	0.013 63	0.001 63	0.000 99
Skewed lognormal	0.095 74	0.050 26	0.005 62	0.003 32
Heavy-tailed t_3	0.026 89	0.014 70	0.001 74	0.001 02
Multimodal	0.039 89	0.020 44	0.002 20	0.001 28

Table 9: Best-performing family (lowest AMISE) for each target density at $n = 500$.

Target density	Family	Order	ρ_1	AMISE
Mixture-normal	BF2	Triweight	0.222	0.001 47
Skewed lognormal	BF1	Epanechnikov	0.111	0.005 19
Heavy-tailed t_3	CF1	Epanechnikov	0.125	0.001 57
Multimodal	CF2	Triweight	0.250	0.002 04

Table 10: Mean AMISE by target density and sample size.

Target density	$n = 10$	$n = 25$	$n = 500$	$n = 1000$
Mixture-normal	0.051 08	0.024 54	0.002 234	0.001 283
Skewed lognormal	0.160 70	0.077 21	0.007 028	0.004 037
Heavy-tailed t_3	0.050 84	0.024 43	0.002 224	0.001 277
Multimodal	0.059 31	0.028 50	0.002 594	0.001 490

every n), the multimodal law is next, and the two symmetric light/heavy-tailed laws are the easiest and almost tied. Table 9 shows that no single family dominates universally: lower-order (Epanechnikov) hybrids win on the shape-sensitive lognormal and heavy-tailed targets, whereas higher-order (Triweight) hybrids win on the smoother mixture-normal and on the multimodal target where extra peakedness helps separate modes. This confirms the central design rationale, offering a spectrum of kernels so that the practitioner can match the kernel to the target. The companion asymptotic figures of merit, which strip away sampling noise, tell the same story, as Table 10 now shows.

Table 10 shows that the AMISE decreases substantially as the sample size increases for all target densities, confirming the theoretical convergence of the proposed estimator and mirroring the finite-sample pattern reported in Table 8. The skewed lognormal distribution consistently exhibits the highest AMISE values, indicating greater estimation difficulty due to its pronounced asymmetry. In contrast, the mixture-normal and heavy-tailed t_3 densities yield the lowest AMISE values. The multimodal density produces intermediate errors, reflecting the additional challenge of accurately capturing multiple modes. The close agreement between Tables 8 and 10 demonstrates that the asymptotic theory provides an accurate description of the estimator's finite-sample behaviour. Having established that accuracy improves with sample size and depends on the target shape, we next investigate how stable this performance remains when the governing parameters are perturbed.

4.4. Sensitivity analysis

A sensitivity analysis was performed to examine the stability of the proposed kernels with respect to key data-generating parameters, including sample size, mixture proportions, tail thickness, and modal separation.

AMISE-based assessment. Performance was evaluated using the asymptotic mean integrated squared error (AMISE). For a kernel estimator $\widehat{f_h}(x)$ of order $2m$, the AMISE is given by

$$\mathrm{AMISE}^{2m}\left(\widehat{f_h}\right) = 4^{-\frac{6m+2}{4m+1}} m^{-\frac{4m}{4m+1}} (4m+1)\pi^{-\frac{2m}{4m+1}} \left(\|f^{(2m)}\|_2^2 \Gamma^2\left(p+\tfrac{1}{2}\right) \lambda(m)\right)^{\frac{1}{4m+1}} \left(\frac{\Gamma^2\left(p+\frac{3}{2}\right)\beta}{(2p+1)^2}\right)^{\frac{4m}{4m+1}} n^{-\frac{4m}{4m+1}}, \qquad m \in \mathbb{N}.$$

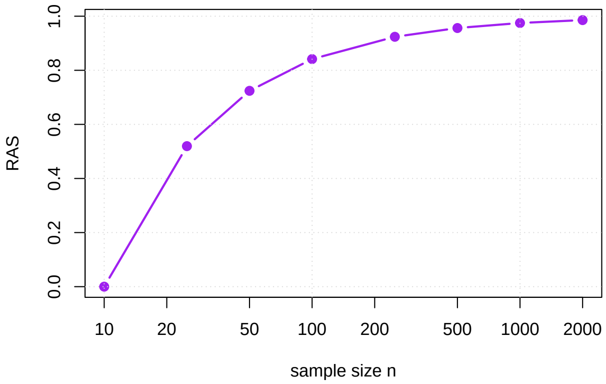
Sensitivity was quantified by evaluating changes in the AMISE under parameter perturbations:

1. *Sample size sensitivity:*
$$\Delta_n = \left|\mathrm{AMISE}^{2m}\left(\widehat{f_{h,n}}\right) - \mathrm{AMISE}^{2m}\left(\widehat{f_{h,n'}}\right)\right|, \qquad n \neq n'.$$
2. *Mixture proportion sensitivity:* For
$$f_\pi(x) = \pi N(\mu_1, \sigma_1^2) + (1-\pi) N(\mu_2, \sigma_2^2),$$
$$\Delta_\pi = \left|\mathrm{AMISE}^{2m}\left(\widehat{f_{h,\pi}}\right) - \mathrm{AMISE}^{2m}\left(\widehat{f_{h,\pi'}}\right)\right|, \qquad \pi \neq \pi'.$$
3. *Tail thickness sensitivity:* For $X \sim t_\nu$,
$$\Delta_\nu = \left|\mathrm{AMISE}^{2m}\left(\widehat{f_{h,\nu}}\right) - \mathrm{AMISE}^{2m}\left(\widehat{f_{h,\nu'}}\right)\right|, \qquad \nu \neq \nu'.$$
4. *Modal separation sensitivity:* Let $\delta = |\mu_1 - \mu_2|$. Then
$$\Delta_\delta = \left|\mathrm{AMISE}^{2m}\left(\widehat{f_{h,\delta}}\right) - \mathrm{AMISE}^{2m}\left(\widehat{f_{h,\delta'}}\right)\right|, \qquad \delta \neq \delta'.$$

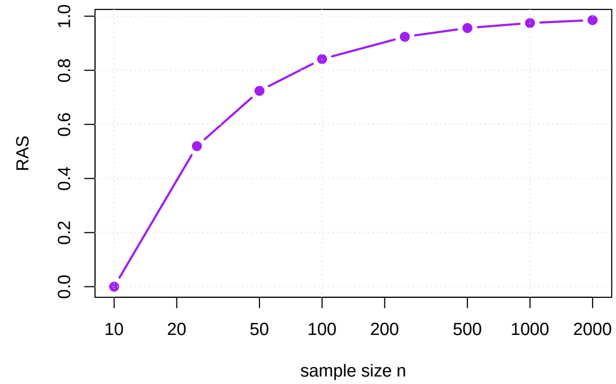
Overall robustness was summarized by the relative AMISE sensitivity measure

$$\mathrm{RAS}(\theta) = \frac{\left|\mathrm{AMISE}^{2m}\left(\widehat{f_{h,\theta}}\right) - \mathrm{AMISE}^{2m}\left(\widehat{f_{h,\theta'}}\right)\right|}{\mathrm{AMISE}^{2m}\left(\widehat{f_{h,\theta'}}\right)},$$

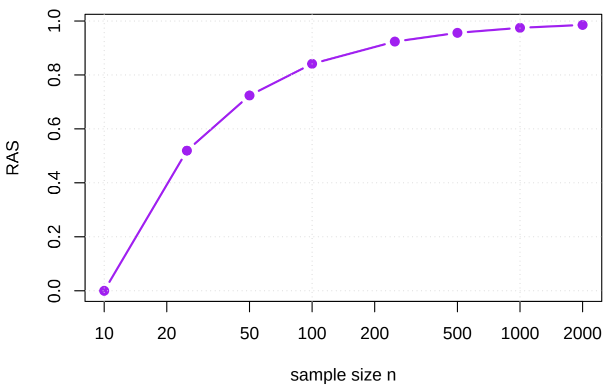
where θ denotes a perturbed parameter and θ' represents its baseline (reference) value. The statistic $\mathrm{RAS}(\theta)$ therefore measures the relative change in asymptotic risk induced by perturbations in θ. Its interpretation is as follows:



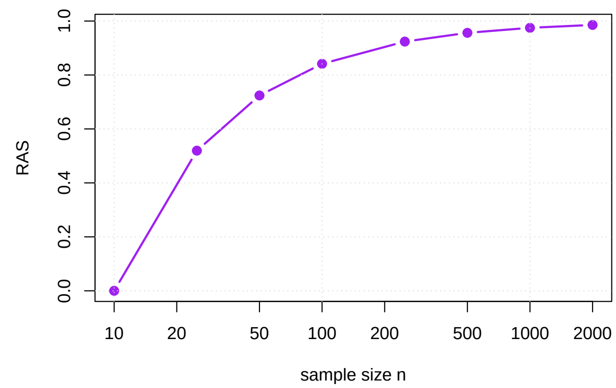
(a) Mixture-normal.



(b) Skewed lognormal.



(c) Heavy-tailed t_3.



(d) Multimodal.

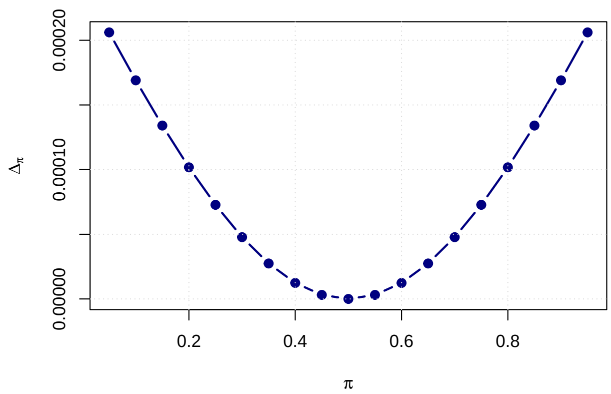
Figure 3: RAS of the AMISE with respect to sample size, by target density (baseline $n = 10$).

- $\text{RAS}(\theta) = 0$ indicates *perfect insensitivity*, meaning that the estimator's asymptotic risk remains unchanged under perturbations in θ.
- $0 < \text{RAS}(\theta) < 1$ implies *mild sensitivity*, where the perturbation in θ induces only a relatively small proportional change in the AMISE. Estimators in this range are regarded as *robust* with respect to θ.
- $\text{RAS}(\theta) \approx 1$ suggests *substantial sensitivity*, indicating that the induced change in AMISE is comparable in magnitude to the baseline asymptotic risk.
- $\text{RAS}(\theta) > 1$ indicates *high sensitivity*, meaning that the estimator's performance deteriorates markedly when the parameter θ is perturbed.

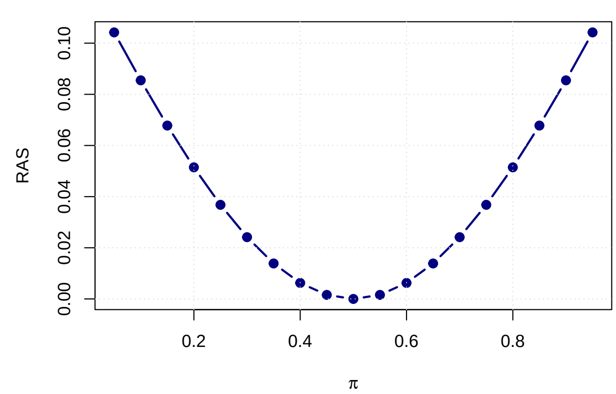
In addition to the Monte Carlo study evaluating the performance of the proposed hybrid polynomial kernels across diverse target distributions, we examine their behaviour using AMISE and Relative Asymptotic Sensitivity (RAS) as complementary performance measures. The study explores how the hybrid kernels respond to variations in sample size, tail heaviness, mixture proportions, and modal separation, thereby assessing not only their accuracy but also their robustness to changes in the underlying data-generating mechanism. The results demonstrate that the proposed kernels maintain stable performance across a broad range of settings while retaining the efficiency advantages documented in the preceding sections. We consider the four perturbations in turn, beginning with sample size, which the previous section identified as the principal determinant of estimation accuracy.

Table 11: RAS of the AMISE with respect to sample size (baseline $n = 10$). The relative sensitivity is identical across densities.

n	25	50	100	250	500	1000	2000
RAS	0.5196	0.7241	0.8415	0.9239	0.9563	0.9749	0.9856



(a) Absolute, Δ_π.



(b) Relative, RAS_π.

Figure 4: Sensitivity of the AMISE to the mixture proportion π in $f_\pi = \pi N(-2, 1) + (1 - \pi)N(2, 1)$; baseline $\pi = 0.5$.

Table 12: Sensitivity of the AMISE to mixture proportion (selected π; baseline $\pi = 0.5$).

π	AMISE	Δ_π	RAS
0.05	0.002 18	0.000 21	0.1042
0.25	0.002 05	0.000 07	0.0368
0.50	0.001 98	0.000 00	0.0000
0.75	0.002 05	0.000 07	0.0368
0.95	0.002 18	0.000 21	0.1042

4.4.1. Sensitivity to sample size

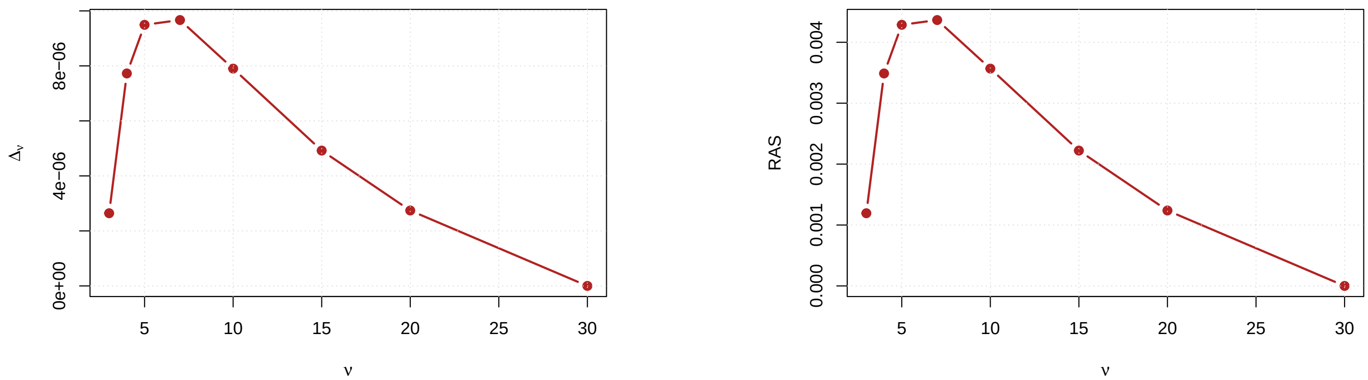
The RAS rises rapidly and then saturates toward 1 (Figure 3, Table 11). A key structural observation is that the relative curve is *identical* for all four densities: because the AMISE factorises as a density-dependent constant times $n^{-4/5}$, the ratio $\text{AMISE}(n)/\text{AMISE}(n')$ cancels that constant and depends on the sample sizes alone. The interpretation is that sample size is the single most influential determinant of estimation accuracy. Increasing the sample size from $n = 10$ to $n = 25$ already reduces approximately 52% of the baseline AMISE, and this improvement is identical across all target densities.

4.4.2. Sensitivity to mixture proportion

The response is symmetric about the balanced mixture $\pi = 0.5$, where sensitivity is exactly zero, and rises toward the extremes (Figure 4 and Table 12). Even at the most lopsided proportion, the RAS reaches only 0.104, well within the mild-sensitivity region. The estimator is therefore robust to imbalance in component weights: skewing the mixture toward one component changes the AMISE by at most about 10%.

4.4.3. Sensitivity to tail thickness

Tail thickness is by far the least influential parameter (Figure 5, Table 13): the RAS never exceeds 0.0044, corresponding to a change of less than one-half of one percent in the AMISE. The small hump near $\nu = 7$ reflects the modest peak in the curvature functional as the t density transitions from a heavy-tailed distribution toward the Gaussian case, but the magnitude is negligible. In practical terms, the AMISE of the hybrid estimator is essentially invariant to tail heaviness, demonstrating a strong robustness property for applications involving occasional extreme observations. The final and, as it turns out, most influential shape parameter is the distance between modes, to which we now turn.



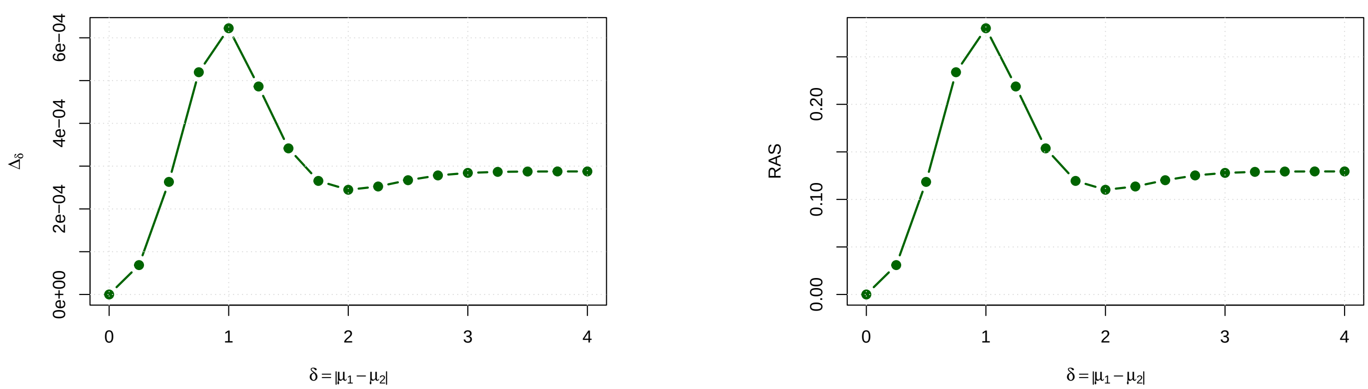
(a) Absolute, Δ_ν.

(b) Relative, RAS_ν.

Figure 5: Sensitivity of the AMISE to tail thickness for $X \sim t_\nu$; baseline $\nu = 30$ (near-normal).

Table 13: Sensitivity of the AMISE to tail thickness (Student-t_ν; baseline $\nu = 30$).

ν	AMISE	Δ_ν	RAS
3	0.002 21	0.000 00	0.0012
5	0.002 21	0.000 01	0.0043
7	0.002 20	0.000 01	0.0044
10	0.002 21	0.000 01	0.0036
30	0.002 21	0.000 00	0.0000



(a) Absolute, Δ_δ.

(b) Relative, RAS_δ.

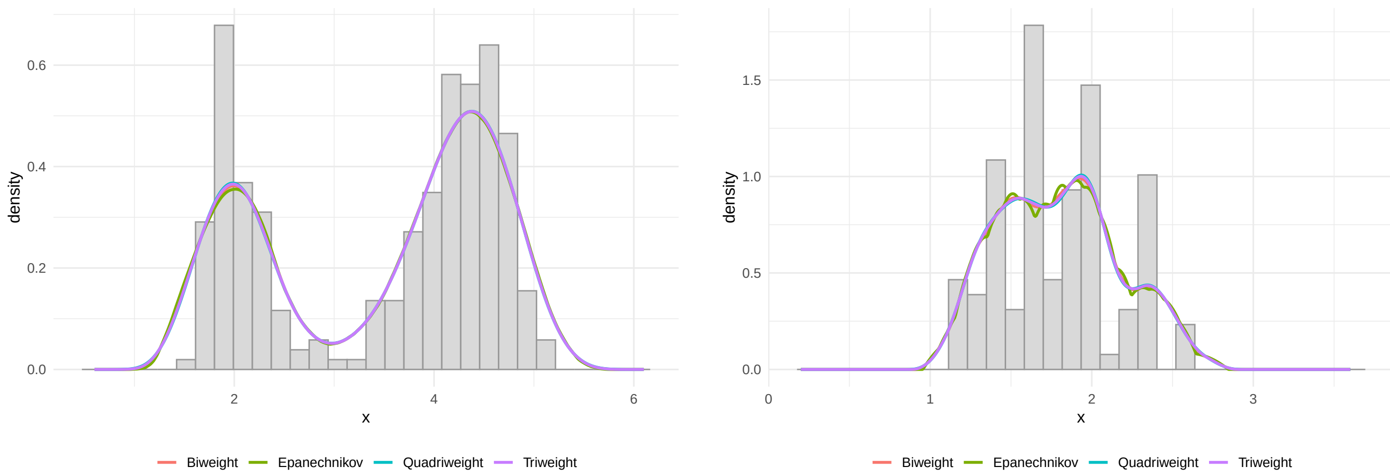
Figure 6: Sensitivity of the AMISE to modal separation $\delta = |\mu_1 - \mu_2|$ in $f_\delta = \frac{1}{2}N(-\delta, 1) + \frac{1}{2}N(\delta, 1)$; baseline $\delta = 0$.

4.4.4. Sensitivity to modal separation

Modal separation is the most influential of the three shape parameters, yet it remains firmly within the mild-sensitivity region (Figure 6, Table 14). The RAS is non-monotonic: it rises to a pronounced peak of 0.280 at $\delta = 1$, then declines and stabilises around 0.13 for larger separations. The peak coincides with the transition regime in which the two mixture components are only beginning to resolve into distinct modes. In this region the curvature functional changes most rapidly, producing the largest variation in AMISE. Once the modes become clearly separated ($\delta \gtrsim 2$), the local structure around each peak stabilises and the sensitivity correspondingly flattens. The estimator is therefore most demanding to tune when modes are only marginally separated, which is precisely the practically important boundary case encountered in many applications. Taken together, the four perturbation studies demonstrate that

Table 14: Sensitivity of the AMISE to modal separation (selected δ; baseline $\delta = 0$).

δ	AMISE	Δ_δ	RAS
0.00	0.002 22	0.000 00	0.0000
0.50	0.001 96	0.000 26	0.1184
1.00	0.001 60	0.000 62	0.2800
2.00	0.001 98	0.000 24	0.1101
3.00	0.001 94	0.000 28	0.1279
4.00	0.001 93	0.000 29	0.1294



(a) Old Faithful eruption durations.

(b) Clinical scar lengths.

Figure 7: Hybrid kernel density estimates fitted to the real datasets, with cluster representatives overlaid on the histograms.

Table 15: Descriptive summary of the two real datasets.

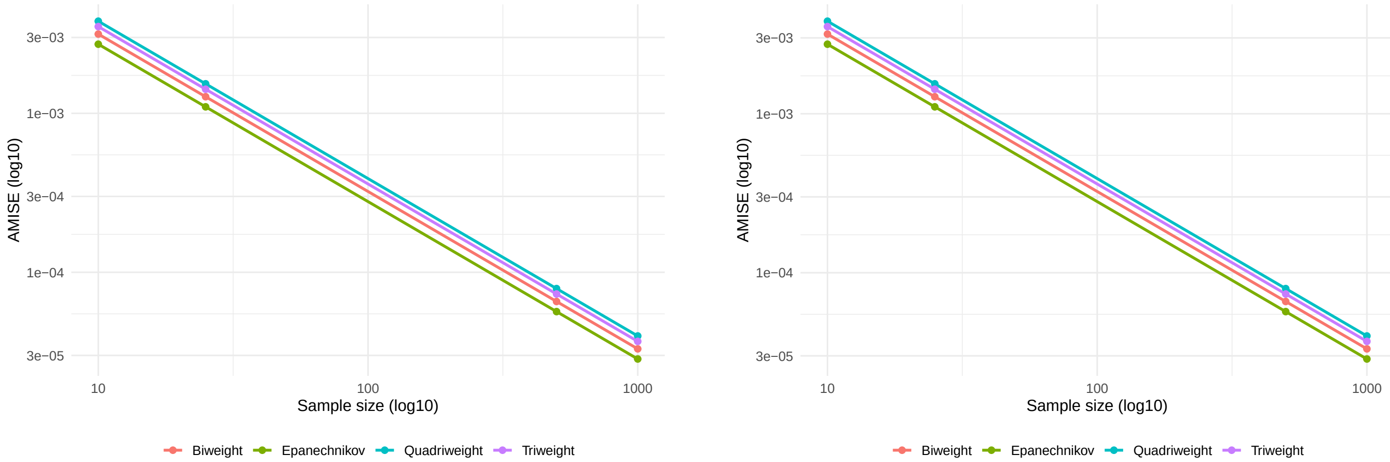
Dataset	n	Mean	SD	Min	Median	Max	Skewness
Old Faithful (min)	272	3.488	1.141	1.6	4.0	5.1	–0.414
Scar length (mm)	110	1.798	0.366	1.2	1.8	2.6	0.264

the proposed kernels are robust throughout the parameter space. Apart from sample size, no RAS value exceeds 0.28, indicating that the estimator maintains stable performance under substantial changes in distributional shape. This robustness motivates the final step of the study: applying the proposed kernels to real datasets whose underlying shape characteristics are unknown.

4.5. Real-data applications

To verify that the favourable properties observed in the simulation study extend to practical settings, the proposed kernel families were applied to two benchmark datasets: the bimodal Old Faithful eruption-duration data ($n = 272$) (see Ref. [39]) and the unimodal clinical scar-length measurements ($n = 110$) (see Ref. [40]). Table 15 summarises the principal characteristics of the datasets. Figure 7 overlays the fitted hybrid density estimates (cluster representatives F_1, Biweight order, plug-in bandwidth) on the corresponding histograms. Figure 8 presents the AMISE trajectories as a function of subsample size, while Tables 16 and 17 report the selected bandwidths, the resulting AMISE values, and their stability under repeated subsampling. We begin by examining the fitted density estimates.

Figure 7 demonstrates that the hybrid kernel density estimators closely follow the empirical distributions of both datasets while maintaining smooth density representations. The Old Faithful data in Figure 7a exhibit a clear bimodal structure, reflecting two distinct eruption-duration regimes. In contrast, Figure 7b shows a predominantly unimodal distribution with slight positive skewness. These results indicate that the proposed hybrid kernels effectively capture both multimodal and unimodal characteristics without excessive smoothing or spurious fluctuations. The descriptive statistics underlying these distributional shapes are summarised in Table 15.



(a) Old faithful.

(b) Scar length.

Figure 8: AMISE versus sample size (both axes on a $\log_{10}$ scale) at $m = 25$, averaged over the ρ_1 grid, for the two real datasets and the four kernels.

Table 16: Plug-in (normal-reference) AMISE-optimal bandwidths and AMISE values for the cluster-F_1 representatives (Biweight order).

Dataset	Cluster (F_1)	ρ_1	h_{AMISE}	AMISE
Old Faithful	A	0.100	1.0132	0.003 176
	B	0.111	1.0111	0.003 176
	C	0.125	1.0084	0.003 176
	D	0.143	1.0049	0.003 176
Scar length	A	0.100	0.3889	0.020 463
	B	0.111	0.3881	0.020 462
	C	0.125	0.3870	0.020 462
	D	0.143	0.3857	0.020 460

Table 17: Subsample-size sensitivity for the real data (Epanechnikov, $m = 25$, $\rho_1 = 0.5$). Each AMISE is the mean over 40 random subsamples; the very small standard deviation demonstrates excellent stability. RAS is computed relative to the largest subsample.

Dataset	subsample n	Mean AMISE	SD	RAS
Old Faithful	20	1.356×10^{-3}	2.554×10^{-6}	2.942
	40	6.832×10^{-4}	8.727×10^{-7}	0.987
	60	4.573×10^{-4}	3.778×10^{-7}	0.330
	80	3.439×10^{-4}	2.593×10^{-7}	0.000
Scar length	20	1.365×10^{-3}	2.230×10^{-6}	2.941
	40	6.878×10^{-4}	7.459×10^{-7}	0.985
	60	4.604×10^{-4}	4.890×10^{-7}	0.329
	80	3.464×10^{-4}	2.828×10^{-7}	0.000

Table 15 highlights the contrasting characteristics of the two datasets. The Old Faithful data have a larger sample size, greater variability, and a wider range than the scar-length measurements, reflecting a more heterogeneous distribution. Their negative skewness indicates a longer left tail, whereas the scar-length data exhibit slight positive skewness. These differences demonstrate the flexibility of the proposed estimator in accommodating datasets with markedly different distributional shapes and dispersion characteristics. Beyond the fitted density estimates, the behaviour of the AMISE on these real datasets follows the same pattern predicted by both the theoretical analysis and the simulation study, as illustrated by Figure 8 and Tables 16–17.

Figure 8 and Tables 16–17 demonstrate that the proposed hybrid kernel estimators are both stable and efficient when applied to real datasets. The AMISE decreases steadily as the sample size increases, confirming the convergence behaviour predicted by the theoretical analysis. The selected bandwidths and corresponding AMISE values vary only marginally across the cluster representatives, indicating low sensitivity to the tuning parameter ρ_1. The subsampling study further reveals a consistent reduction in AMISE accompanied by extremely small standard deviations, highlighting the numerical stability of the estimators. Finally, the declining RAS values confirm that estimation accuracy improves systematically with increasing sample size, mirroring the patterns observed in the simulation experiments and reinforcing the practical reliability of the proposed hybrid kernel families.

5. Discussion: comparison and limitations

The results obtained in the preceding sections indicate that the proposed class of hybrid polynomial kernels performs consistently within a near-optimal efficiency band while offering substantial flexibility in shape adaptation. A useful benchmark for interpreting these findings is the classical Gaussian kernel and the Epanechnikov-optimal reference. As shown in Table 1, the asymptotic efficiency of the proposed family ranges from approximately 0.9572 to 0.9999, which places every member of the class above the Gaussian kernel efficiency of 0.9512 (see Refs. [21, 26]) and very close to the theoretical optimum attained by the Epanechnikov kernel (see Ref. [8]). In practical terms, this shows that the gain in flexibility introduced by the hybrid structure does not incur any meaningful loss in asymptotic performance.

This near-equivalence in efficiency is reinforced by the stability observed across all kernel functionals reported earlier. In particular, the canonical constant $C(K)$ remains almost invariant across kernel orders and clusters, while the variation in roughness $R(K)$ and second moment $\mu_2(K)$ is smooth and controlled. Although higher-order kernels such as the Triweight and Quadriweight forms increase roughness and reduce $\mu_2(K)$, the resulting efficiency loss remains marginal. This indicates that the proposed hybrid construction preserves second-order optimality while introducing a continuous tuning mechanism through the mixing parameter ρ_1. This separation of shape control from asymptotic efficiency allows practitioners to adjust local smoothness without leaving a stable efficiency regime.

When compared with adaptive bandwidth methodologies, the proposed framework plays a complementary rather than competing role. Classical variable-bandwidth approaches, including Abramson-type estimators (see Ref. [19]), diffusion-based selectors (see Ref. [41]), and modern plug-in and cross-validation bandwidth rules (see Refs. [32, 38]), operate by locally adjusting the bandwidth h to reflect data heterogeneity. In contrast, the present construction modifies the kernel shape through ρ_1 while retaining a global bandwidth governed by the AMISE-optimal form. Consequently, adaptive methods mainly address spatial scale variation, whereas hybrid kernels control higher-order shape features such as peakedness and tail behaviour. The AMISE expressions derived in this study provide a theoretical bridge suggesting that hybrid kernels can be embedded within adaptive frameworks without loss of asymptotic coherence.

The practical implications of the numerical results are also noteworthy. Although statistically significant differences are observed across kernel orders and families, the absolute magnitude of these differences is small. The AMISE values for real data and Monte Carlo AMISE summaries vary only modestly across kernel choices compared with the much larger effect induced by sample size. This confirms that sample size is the dominant driver of estimator accuracy, while kernel selection plays a secondary but stabilising role. The main contribution of the proposed family is therefore not lower error rates, but a smooth and interpretable mechanism for shape modulation within an already near-optimal class of estimators.

The sensitivity analysis further supports this interpretation. Across all perturbation regimes, relative asymptotic sensitivity remains within the mild-sensitivity region. Sample size produces the expected monotone effect, while shape parameters such as mixture proportion, tail thickness, and modal separation exert only moderate influence. Even in the most sensitive case, modal separation, the relative change in AMISE remains well below instability thresholds. This confirms that the estimator is structurally robust in both asymptotic theory and finite-sample performance, consistent with the simulation and real-data results.

Despite these advantages, several limitations should be acknowledged. The simulation study relies on oracle or plug-in bandwidths, which isolate kernel effects but do not fully capture uncertainty arising from fully data-driven bandwidth selection. Although the real-data analysis partially addresses this using plug-in rules, more advanced selectors such as cross-validation or adaptive diffusion-based bandwidth methods (see Refs. [32, 38, 41]) could further improve finite-sample comparisons. In addition, the current framework is restricted to the univariate setting, where polynomial kernel structure is most transparent. While multivariate extensions via product kernels are straightforward (see Ref. [26]), they may introduce additional computational and calibration challenges, and their performance may differ from the univariate case. Finally, although asymptotic efficiency results are strong, they do not fully capture higher-order finite-sample effects, especially in irregular or boundary-dominated densities where local bias behaviour may dominate global AMISE behaviour.

Overall, the evidence suggests that the proposed hybrid polynomial kernels occupy a practical middle ground between rigidity and optimality. They retain near-Epanechnikov efficiency while providing a flexible and tunable structure that adapts smoothly to distributional shape. Their stability across simulation and real-data settings indicates that they can be used interchangeably in standard applications, with kernel choice serving mainly as a refinement tool rather than a primary determinant of performance.

6. Conclusion

This study introduced a new class of hybrid polynomial kernels constructed through convex combinations of Beta-polynomial kernel families. The proposed framework generalises classical polynomial kernels by introducing a mixing parameter that generates a continuum of smooth kernel shapes while preserving the fundamental properties required for kernel density estimation. In doing so, the study advances decisively beyond the single family introduced in Ref. [33] and the single cluster developed in Ref. [34]: it supplies, for the first time, a unified generative mechanism for the entire group of clusters, a complete order-$2m$ asymptotic theory with fully explicit constants, and a quantitative robustness (RAS) framework covering all members simultaneously. Theoretical analysis established that the estimators are governed by kernel roughness and moment functionals through a canonical AMISE constant, and despite added flexibility, they retain near-optimal asymptotic efficiency with negligible loss relative to the Epanechnikov kernel. Extensive Monte Carlo simulations confirmed these results, showing consistent error reduction with increasing sample size and uniform convergence across kernel orders. The proposed kernels performed well across symmetric, skewed, heavy-tailed, and multimodal distributions. Sensitivity analysis further demonstrated strong robustness to distributional perturbations, with sample size remaining the dominant factor. Real-data applications confirmed accurate and stable density estimation, highlighting practical usefulness. Overall, the method offers a flexible and efficient alternative for complex density estimation problems.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Funding

The authors received no external funding for this study.

References

[1] E. Fix & J. L. Hodges, "Discriminatory analysis: nonparametric discrimination: consistency properties", International Statistical Review / Revue Internationale de Statistique **57** (1989) 238. https://doi.org/10.2307/1403797.
[2] H. Akaike, "An approximation to the density functions", Annals of the Institute of Statistical Mathematics **6** (1954) 127. https://doi.org/10.1007/BF02900741.
[3] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function", Annals of Mathematical Statistics **27** (1956) 832. https://doi.org/10.1214/aoms/1177728190.
[4] E. Parzen, "On the estimation of a probability density function and the mode", Annals of Mathematical Statistics **33** (1962) 1065. https://doi.org/10.1214/aoms/1177704472.
[5] G. S. Watson & M. R. Leadbetter, "On the estimation of the probability density", Annals of Mathematical Statistics **34** (1963) 480. Available online: https://api.semanticscholar.org/CorpusID:8357602.
[6] M. S. Bartlett, "Statistical estimation of density functions", Sankhyā Series A **25** (1963) 245. https://www.jstor.org/stable/25049271.
[7] T. Cacoullos, "Estimation of a multivariate density", Annals of the Institute of Statistical Mathematics **18** (1966) 179. Available online: https://api.semanticscholar.org/CorpusID:120806240.
[8] V. A. Epanechnikov, "Nonparametric estimation of a multivariate probability density", Theory of Probability and its Applications **14** (1969) 153. https://doi.org/10.1137/1114019.
[9] R. H. Farrell, "On the best obtainable asymptotic rates of convergence in estimation of a density function at a point", Annals of Mathematical Statistics **43** (1972) 170. https://doi.org/10.1214/aoms/1177692711.
[10] K. B. Davis, "Mean square error properties of density estimates", Annals of Statistics **3** (1975) 1025. Available online: https://scispace.com/pdf/mean-square-error-properties-of-density-estimates-2s277h7m9g.pdf.
[11] M. J. Fryer, "Some errors associated with the nonparametric estimation of density functions", Journal of the Institute of Mathematics and its Applications **18** (1976) 371. https://doi.org/10.1093/imamat/18.3.371.
[12] P. Deheuvels, "Estimation non paramétrique de la densité par histogrammes généralisés", Revue de Statistique Appliquée **25** (1977) 5. Available online: https://www.numdam.org/item/RSA_1977__25_3_5_0/.
[13] W. R. Schucany & J. P. Sommers, "Improvement of kernel-type density estimators", Journal of the American Statistical Association **72** (1977) 420. https://doi.org/10.1080/01621459.1977.10481012.
[14] I. S. Abramson, "On bandwidth variation in kernel estimates – a square root law", Annals of Statistics **10** (1982) 1217. https://doi.org/10.1214/aos/1176345986.
[15] L. Devroye & C. S. Penrod, "The consistency of automatic kernel density estimates", Annals of Statistics **12** (1984) 1231. Available online: https://api.semanticscholar.org/CorpusID:9809293.
[16] L. Devroye & L. Györfi, *Nonparametric density estimation: The L_1 view*, Journal of the American Statistical Association **82** (1987) 344. https://doi.org/10.2307/2289176.
[17] M. C. Jones, "A course in density estimation", Journal of the Royal Statistical Society Series A: Statistics in Society **151** (1988) 365. https://doi.org/10.2307/2982771.
[18] D. B. H. Cline, "Admissible kernel estimators of a multivariate density", Annals of Statistics **16** (1988) 1421. Available online: https://projecteuclid.org/journals/annals-of-statistics/volume-16/issue-4/Admissibile-Kernel-Estimators-of-a-Multivariate-Density/10.1214/aos/1176351046.full.

[19] J. S. Marron & D. Nolan, “Canonical kernels for density estimation”, Statistics and Probability Letters **7** (1988) 195. https://doi.org/10.1016/0167-7152(88)90050-8.
[20] B. W. Silverman, “Using kernel density estimation to investigate multimodality”, Journal of the Royal Statistical Society Series B (Methodological) **43** (1981) 97. Available online: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1981.tb01155.x.
[21] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, UK, 1986. https://doi.org/10.1002/bimj.4710300745.
[22] M. Rudemo, “Empirical choice of histograms and kernel density estimators”, Scandinavian Journal of Statistics **9** (1982) 65. Available online: https://www.jstor.org/stable/4615859.
[23] A. W. Bowman, “An alternative method of cross-validation for the smoothing of density estimates”, Biometrika **71** (1984) 353. https://doi.org/10.1093/biomet/71.2.353.
[24] T. Duong, A. Cowling, I. Koch & M. P. Wand, “Feature significance for multivariate kernel density estimation”, Computational Statistics & Data Analysis **52** (2008) 4225 https://doi.org/10.1016/j.csda.2008.02.035.
[25] D. W. Scott & S. R. Sain, “Multidimensional density estimation”, Technical Report, Rice University, Houston, Texas, USA, 2004. Available online: http://www.stat.rice.edu/~scottdw/ss.nh.pdf.
[26] M. P. Wand & M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, UK, 1995. https://doi.org/10.1201/b14876.
[27] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York, USA, 2008. https://doi.org/10.1007/978-0-387-78189-1.
[28] W. Härdle & M. Müller, *Multivariate and semiparametric kernel regression*, SFB 373 Discussion (1997) pp 26, Humboldt University of Berlin, Berlin, Germany, 1997. Available online: http://edoc.hu-berlin.de/sfb-373-papers/1997-26/PDF/26.pdf.
[29] D. Fadda, E. Slezak & A. Bijaoui, “Density estimation with non-parametric methods”, Astronomy and Astrophysics Supplement Series **127** (1998) 335. https://doi.org/10.1051/aas:1998355.
[30] E. Giné & R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, Cambridge, UK, 2015. https://doi.org/10.1017/CBO9781107337862.
[31] S. X. Chen, “Probability density function estimation using gamma kernels”, Annals of the Institute of Statistical Mathematics **52** (2000) 471. https://doi.org/10.1023/A:1004165218295.
[32] M. C. Jones, J. S. Marron & S. J. Sheather, “A brief survey of bandwidth selection for density estimation”, Journal of the American Statistical Association **91** (1996) 401. Available online: https://www.jstor.org/stable/2291420.
[33] B. A. Afere, “A new family of hybrid classical polynomial kernels in density estimation”, International Journal of Research and Innovation in Applied Science **6** (2021) 145. Available online: https://rsisinternational.org/journals/ijrias/DigitalLibrary/Vol.6&Issue1/145-151.pdf.
[34] B. A. E. Afere, “On the cluster of families of hybrid polynomial kernels in kernel density estimation”, Journal of the Nigerian Society of Physical Sciences **7** (2025) 2044. https://doi.org/10.46481/jnsps.2024.2044.
[35] A. F. Ramsey, “An application of kernel density estimation via diffusion to group yield insurance”, presented at the Agricultural and Applied Economics Association (AAEA) Annual Meeting, Minneapolis, Minnesota, USA, July 27–29, 2014. https://doi.org/10.22004/ag.econ.170173.
[36] J. C. C. López & M. J. Sanz, “Improving kernel methods for density estimation in random differential equations problems”, Mathematical and Computational Applications **25** (2020) 33. https://doi.org/10.3390/mca25020033.
[37] B. A. Afere & F. O. Oyegue, “On the construction of the family of d-dimensional spherically symmetric polynomial kernels”, Palestine Journal of Mathematics **5** (2016) 259. Available online: https://pjm.ppu.edu/sites/default/files/papers/24%20%20%20Adel%20ICMS15b.pdf.
[38] S. J. Sheather & M. C. Jones, “A reliable data-based bandwidth selection method for kernel density estimation”, Journal of the Royal Statistical Society Series B (Methodological) **53** (1991) 683. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x.
[39] W. Härdle, *Smoothing Techniques: With Implementation in S*, Springer, New York, USA, 1991. Available online: https://api.semanticscholar.org/CorpusID:118288432.
[40] C. C. Ishiekwene & B. A. E. Afere, “Higher-order window width selectors for empirical data”, Journal of the Nigerian Statistical Association **14** (2001) 69.
[41] Z. I. Botev, J. F. Grotowski & D. P. Kroese, “Kernel density estimation via diffusion”, Annals of Statistics **38** (2010) 2916. https://doi.org/10.1214/10-AOS799.