Direct Estimation of Schrödinger Bridge Time-Series Drifts: Finite-Sample, Asymptotic, and Adaptive Guarantees

Othmane Mazhar
Institut Louis Bachelier
Fondation du Risque
Paris, France
othmane.mazhar@institutlouisbachelier.org

Huyên Pham
CMAP, École Polytechnique
Palaiseau, France
huyen.pham@polytechnique.edu

Abstract

We study nonparametric estimation of Schrödinger bridge (SB) drifts from i.i.d. samples of endpoint pairs observed on a single time interval. Starting from the conditional-ratio representation of the Schrödinger bridge time-series (SBTS) drift, we analyze a direct Nadaraya–Watson plug-in estimator based on kernelized numerator and denominator terms. Unlike approaches based on entropic-OT potentials, Sinkhorn iterations, or iterative bridge solvers, our method targets the drift directly and isolates *statistical error* from optimization, approximation, and discretization errors.

Under Hölder regularity, a marginal-density floor, and bounded support, we establish a uniform finite-sample bound for admissible bandwidth pairs. Under genuine undersmoothing, we prove a pointwise central limit theorem and consistency of a plug-in covariance estimator, yielding asymptotically valid pointwise Wald confidence intervals. We also establish an adaptive bandwidth-selection oracle inequality and a pivot-local minimax lower bound. Through an explicit uniform pivot, we derive a global minimax lower bound under compatibility conditions, showing that the adaptive estimator attains the minimax rate up to logarithmic factors. Synthetic experiments provide theorem-targeted diagnostics for finite-sample scaling, Gaussian approximation, and adaptive behavior.

1 Introduction

Schrödinger bridges (SBs) interpolate between probability distributions by minimizing relative entropy on path space and connect stochastic control, entropic optimal transport, and diffusion-based generative modeling [5, 11, 17, 20, 30]. They have recently been used for generative modeling, sampling, and time-series generation [1, 8, 16, 31]. Yet the *statistical theory* of data-driven SB estimators remains limited, especially for *direct nonparametric drift estimation*.

Most existing analyses do not treat the drift as the primary statistical object. They instead proceed through entropic-OT potentials, Sinkhorn iterations, or iterative bridge solvers, so the resulting guarantees are intertwined with optimization, approximation, or discretization error. We take a different route: starting from the closed-form conditional-ratio representation of the SB drift, we analyze a direct kernel plug-in estimator and study the *statistical error of the drift itself*. This drift-level viewpoint is also essential for sharpness: our lower bound perturbs the drift directly while preserving normalization, denominator separation, and the global smoothness class.

Problem and scope. Fix one interval $[s, u]$ with $0 < s < u$. In the first-order SBTS notation of [16], the drift on $[s, u)$ depends on the observed past through $\xi = X_s$. Given M i.i.d. samples

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

$\{(X_s^{(m)}, X_u^{(m)})\}_{m=1}^M \sim \nu$, our goal is to estimate $a^*(t, x; \xi)$ and quantify its accuracy. This first-order notation is only for readability: fixed-memory extensions replace $\xi \in \mathbb{R}^d$ by a lag block in $(\mathbb{R}^d)^L$, with d replaced by Ld, and finite-grid bounds follow intervalwise by a union bound.

Why this is nontrivial. The drift is a *ratio* of conditional expectations, so numerator and denominator errors must be transferred stably and empirical denominators must inherit deterministic population floors. Moreover, the kernelized terms are weighted by an SB exponential with both *contracting* and *expanding* factors, and the drift is singular as $t \uparrow u$. Bounded support is therefore central: it controls the expanding factor and makes the weighted empirical class analyzable.

Contributions. Our contributions are fourfold.

1. **Direct drift-level analysis.** We analyze a kernel plug-in estimator derived from the SBTS conditional-ratio drift formula, decoupling statistical error from Sinkhorn stability, iterative solver error, and SDE discretization.
2. **Uniform finite-sample guarantees.** Under Hölder smoothness, a marginal-density floor, and bounded support, we prove a uniform non-asymptotic bound over space and time away from the terminal singularity, with rate $(\log M/M)^{\beta/(2\beta+d)}$ up to constants, the terminal factor, and logarithmic terms.
3. **Pointwise inference.** We prove $\sqrt{Mh^d}\,(\hat{a}(t, x; \xi) - a^*(t, x; \xi)) \Rightarrow \mathcal{N}(0, \Sigma(t, x; \xi))$ under genuine undersmoothing and establish consistency of a plug-in covariance estimator, yielding asymptotically valid pointwise Wald intervals.
4. **Adaptive minimax optimality.** We construct a data-driven bandwidth selector satisfying an oracle inequality. A pivot-local Le Cam lower bound shows that the rate cannot be improved around interior laws, and an explicit uniform pivot yields a global minimax lower bound under transparent compatibility conditions. Thus the adaptive selector is globally minimax-rate optimal up to logarithmic factors.

The empirical section is theorem-aligned rather than benchmark-driven: synthetic experiments with known drifts assess finite-sample scaling, pointwise Gaussian approximation, confidence-interval coverage, and oracle-competitiveness of the adaptive selector.

2 Related Work

Recent SB theory and algorithms typically proceed through entropic optimal transport, Sinkhorn-type iterations, or iterative bridge solvers; see [5, 11, 17, 20, 30] for background and [7, 8, 23, 31] for modern computational, generative, sampling, and bridge-matching perspectives. In that line of work, the drift is usually obtained only after estimating potentials or solving an auxiliary optimization problem, so the final guarantees are coupled with computational stability, iteration error, or discretization error. Our paper takes a different route: starting from the closed-form SBTS drift representation of [16], we study the resulting conditional-ratio drift formula and estimate the drift directly.

Recent statistical and nonparametric analyses in the SB/entropic-OT literature include sample-complexity and CLT results for entropic OT and Sinkhorn divergences [12, 25], plug-in estimation of Schrödinger bridges via entropic-OT potentials [29], and forward–reverse kernel methods for Schrödinger-type quantities [2]. In the time-series setting, [16] provides the SBTS path-space formalism, optimal-drift characterization, and conditional-ratio representation used here, while [1] provides extensive empirical benchmarks for SBTS generation. Our repeated-sample formulation matches this SBTS data model, is standard in sample-based OT/SB studies [23, 29], and parallels copies-based drift estimation [24, 28]. These works are complementary to ours: our target is the drift estimator itself, our main setting is the SBTS single-interval formulation, and our guarantees are stated directly at the drift level.

Technical distinction from standard kernel theory. Our analysis uses classical kernel and pointwise-asymptotic ideas [15, 32], together with adaptive bandwidth-selection ideas [14, 21], but these tools do not apply off the shelf. The estimand is a drift-level ratio of weighted conditional moments, and the SB weight contains both a contracting Gaussian term and an expanding exponential

term. The proof therefore needs a deterministic ratio-transfer step, empirical denominator localization, and a weighted-class entropy argument showing that the SB-weighted kernel class retains the entropy order of translated kernels. The lower bound also differs from standard regression lower bounds: the perturbation must change the drift while preserving normalization, the Hölder ball, the marginal-density floor, and the automatic SB denominator floor. These are the SB-specific ingredients that allow the finite-sample, CLT, adaptive, and minimax results to hold directly at the drift level.

3 Problem Setup and Estimator

Recall the SBTS construction of [16]. For an observation grid $\mathcal{T} = \{t_1, \dots, t_N\}$ and a target time-series law μ on $(\mathbb{R}^d)^N$, the bridge is the path-space entropy projection

$$P^* \in \arg\min_P \{H(P \mid W) : (X_{t_1}, \dots, X_{t_N})_\# P = \mu\},$$

where W is the reference Wiener law. The constraint is therefore imposed on the full joint law of the observed time series. Writing $\mu_{\mathcal{T}}^W$ for the grid law under W and $\varrho := d\mu/d\mu_{\mathcal{T}}^W$, the optimizer is the tilt $dP^*/dW = \varrho(X_{t_1}, \dots, X_{t_N})$. The stochastic-control formulation gives an adapted log-gradient drift, and [16] derive from it the conditional-ratio formula used below. We study this formula on a fixed interval $[s, u]$, with $0 < s < u$. Write $\Delta := u - s$ and $\Delta(t) := u - t$ for $t \in [s, u)$. We observe M i.i.d. copies $\{(X_s^{(m)}, X_u^{(m)})\}_{m=1}^M$ of an $\mathbb{R}^d \times \mathbb{R}^d$-valued random pair (X_s, X_u) with joint law ν. Throughout, $B \subset \mathbb{R}^d$ denotes a bounded set containing the support of both X_s and X_u.

For readability, we use the first-order notation of the SBTS drift formula: on the observation grid, the conditional law of the next state depends on the past only through the current endpoint. Thus, with $s = t_i$, $u = t_{i+1}$, and $\xi = X_s$, the drift on $[s, u)$ is

$$a^*(t, x; \xi) = \frac{1}{\Delta(t)} \left(\frac{\mathbb{E}[X_u \, F(t, \xi, x, X_u) \mid X_s = \xi]}{\mathbb{E}[F(t, \xi, x, X_u) \mid X_s = \xi]} - x \right), \; F(t, \xi, x, y) := e^{-\frac{|y-x|^2}{2\Delta(t)} + \frac{|y-\xi|^2}{2\Delta}}. \tag{1}$$

This conditional-ratio formula is the starting point of the paper: we estimate the drift directly from (1), rather than through potentials, Sinkhorn iterations, or an auxiliary bridge solver.

Let p denote the joint density of (X_s, X_u) on $B \times B$, let $f(\xi) := \int_B p(\xi, y)\, dy$ be the marginal density of X_s, and define

$$g_1(t, x; \xi) := \int_B F(t, \xi, x, y) p(\xi, y)\, dy = f(\xi) \mathbb{E}[F(t, \xi, x, X_u) \mid X_s = \xi],$$
$$g_2(t, x; \xi) := \int_B y \, F(t, \xi, x, y) p(\xi, y)\, dy = f(\xi) \mathbb{E}[X_u F(t, \xi, x, X_u) \mid X_s = \xi].$$

Set $D^*(t, x; \xi) := g_1(t, x; \xi)/f(\xi)$, $N^*(t, x; \xi) := g_2(t, x; \xi)/f(\xi)$, and $Q^*(t, x; \xi) := N^*(t, x; \xi)/D^*(t, x; \xi)$. Then

$$a^*(t, x; \xi) = \frac{1}{\Delta(t)} (Q^*(t, x; \xi) - x) = \frac{1}{\Delta(t)} \left(\frac{N^*(t, x; \xi)}{D^*(t, x; \xi)} - x \right).$$

Kernel plug-in estimator. Fix a possibly signed kernel $K : \mathbb{R}^d \to \mathbb{R}$ and bandwidths $h_1, h_2 > 0$, with $K_h(z) := h^{-d} K(z/h)$. For $(t, x, \xi) \in [s, u) \times \mathbb{R}^d \times B$, define

$$\hat{f}_j(\xi) := \frac{1}{M} \sum_{m=1}^M K_{h_j}(X_s^{(m)} - \xi), \qquad j \in \{1, 2\},$$
$$\hat{g}_1(t, x; \xi) := \frac{1}{M} \sum_{m=1}^M F(t, \xi, x, X_u^{(m)}) \, K_{h_1}(X_s^{(m)} - \xi),$$
$$\hat{g}_2(t, x; \xi) := \frac{1}{M} \sum_{m=1}^M X_u^{(m)} \, F(t, \xi, x, X_u^{(m)}) \, K_{h_2}(X_s^{(m)} - \xi).$$

Let $\hat{D}(t,x;\xi) := \hat{g}_1(t,x;\xi)/\hat{f}_1(\xi)$ and $\hat{N}(t,x;\xi) := \hat{g}_2(t,x;\xi)/\hat{f}_2(\xi)$. The direct Nadaraya–Watson plug-in estimator [26, 34] is

$$\hat{a}(t,x;\xi) := \frac{1}{\Delta(t)} \left(\frac{\hat{N}(t,x;\xi)}{\hat{D}(t,x;\xi)} - x \right). \tag{2}$$

For risk statements we use the denominator-floor convention stated in Appendix E; it is inactive on the high-probability events used below. We allow $h_1 \neq h_2$ because the two weighted functionals may have different effective envelopes and variance levels. In the main-text CLT and adaptive results, however, we specialize to the diagonal choice $h_1 = h_2 = h$ to keep the bandwidth parameter one-dimensional. The unequal-bandwidth CLT, and adaptive versions are given in Appendix D to F.

3.1 Assumptions

We state assumptions for the *SBTS single-interval model* above.

Assumption 3.1 (Kernel)**.** Let $\ell_\beta := \max\{m \in \mathbb{N}_0 : m < \beta\}$ and $\theta_\beta := \beta - \ell_\beta \in (0,1]$. The kernel $K : \mathbb{R}^d \to \mathbb{R}$ is bounded, compactly supported, Lipschitz with constant L_K, integrates to one, and is a multivariate kernel of order ℓ_β: $\int_{\mathbb{R}^d} z^\alpha K(z)\,dz = 0$, $1 \leq |\alpha| \leq \ell_\beta$. Moreover, $\mu_\beta(K) := \int_{\mathbb{R}^d} |z|^\beta |K(z)|\,dz < \infty$.

Assumption 3.2 (Pair-law regularity)**.** The pair (X_s, X_u) admits a joint density p supported on $B \times B$, where $B \subset \mathbb{R}^d$ is bounded. There exist constants $\beta > 0$, $L_p > 0$, and $f_{\min} > 0$ such that:

1. for every $y \in B$, the map $\xi \mapsto p(\xi, y)$ on B is the restriction of a function $\bar{p}(\cdot, y) \in \Sigma(\beta, L_p)$, with $\Sigma(\beta, L_p)$ defined in Appendix A, on a neighborhood of B, uniformly in $y \in B$;
2. the marginal density $f(\xi) := \int_B p(\xi, y)\,dy$ satisfies $\inf_{\xi \in B} f(\xi) \geq f_{\min} > 0$.

The lower bound on f is a population observability condition ensuring adequate local design mass; it is distinct from the numerical safeguards used in the implementation. Consequently, denominator separation holds automatically on compact query sets. Indeed, for every $R > 0$ and $\eta \in (0, \Delta)$, writing $B_R := \{x \in \mathbb{R}^d : |x| \leq R\}$, the explicit form of F and the boundedness of B imply that there exists a deterministic constant $D_{\min} = D_{\min}(B, R, \eta, \Delta) > 0$ such that $\inf_{\xi \in B,\, x \in B_R,\, t \in [s, u-\eta]} D^*(t,x;\xi) = \inf_{\xi \in B,\, x \in B_R,\, t \in [s, u-\eta]} \mathbb{E}[F(t,\xi,x,X_u) \mid X_s = \xi] \geq 2D_{\min}$. Thus $D_{\min}$ denotes this deterministic constant and is not an additional model parameter.

Remark 3.3 (Role of bounded support). Bounded support is essential for the present unmodified estimator and uniform proof, because it provides deterministic envelopes for the SB-weighted empirical classes. Its role is time-dependent. Expanding the exponent gives

$$-\frac{|y-x|^2}{2(u-t)} + \frac{|y-\xi|^2}{2(u-s)} = -\frac{1}{2}\left(\frac{1}{u-t} - \frac{1}{u-s}\right)|y|^2 + \left(\frac{x}{u-t} - \frac{\xi}{u-s}\right)^\top y + c.$$

Hence, on $t \in [s+\kappa, u-\eta]$, the strictly negative quadratic term provides Gaussian damping. At $t = s$, the quadratic terms cancel and conditional sub-Gaussianity gives fixed moments but only lognormal-type tails, not a deterministic envelope. Thus the original ratio estimator may remain analyzable in the damped interior, whereas a result uniform down to $t = s$ requires truncation or robustification. Appendix G.8 gives the detailed calculation and relates it to the wide-support stress test. In particular, $C_F := \sup_{\xi, y \in B} e^{\frac{|y-\xi|^2}{2\Delta}} < \infty$, $C_Y := \sup_{y \in B} |y| < \infty$.

Remark 3.4 (Fixed finite-memory extensions). The first-order display is used only to keep notation readable. For any fixed memory length L, replace the conditioning variable $\xi \in \mathbb{R}^d$ by the lag block

$$h_i^{(L)} = (X_{t_{i-L+1}}, \ldots, X_{t_i}) \in (\mathbb{R}^d)^L$$

and replace the pair law of $(X_{t_i}, X_{t_{i+1}})$ by the joint law of $(h_i^{(L)}, X_{t_{i+1}})$. The ratio algebra is unchanged, and the kernel bias, entropy, concentration, variance, CLT, and adaptive arguments carry over with the conditioning dimension d replaced by Ld. For a fixed interval i, the same applies to the full finite history, with d replaced by id. What is not covered here is estimation of the SB drift from a single discretely observed trajectory; that problem has dependent design points and requires different empirical-process and time-series arguments.

Bandwidth scale. The oracle bandwidth scale is $h_1 \asymp h_2 \asymp (\log M/M)^{1/(2\beta+d)}$, up to model-dependent constants and the precise logarithmic term in the finite-sample theorem. For the pointwise CLT, however, the bandwidth must be genuinely undersmoothed and therefore must decay strictly faster than this rate-optimal scale.

4 Main Results

We now present finite-sample, asymptotic, and adaptive guarantees for the estimator (2) in the SBTS single-interval setting of Section 3. Fix $R > 0$, $\eta \in (0, \Delta)$, and $\rho > 0$, and write $B_R := \{x \in \mathbb{R}^d : |x| \le R\}$. If $\mathrm{supp}(K) \subset B(0, R_K)$, set $B_\rho := \{\xi \in B : \mathrm{dist}(\xi, \partial B) \ge \rho\}$, $Q_{R,\eta,\rho} := [s, u-\eta] \times B_R \times B_\rho$. All uniform suprema below are taken over $Q_{R,\eta,\rho}$, and bandwidths satisfy $h_j \le \rho/R_K$. Constants may also depend on ρ, but not on (M, h_1, h_2).

4.1 Deterministic ratio transfer

The following proposition isolates the deterministic step that turns estimation error for the kernelized building blocks into estimation error for the drift.

Proposition 4.1 (Deterministic ratio-stability transfer)**.** *Assume Assumption 3.2. Let*

$$\mathcal{E} := \left\{ \inf_{\xi \in B_\rho} \hat{f}_1(\xi) \ge \frac{f_{\min}}{2}, \quad \inf_{\xi \in B_\rho} \hat{f}_2(\xi) \ge \frac{f_{\min}}{2}, \quad \inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} \hat{D}(t,x;\xi) \ge \frac{D_{\min}}{2} \right\}.$$

Then on $\mathcal{E}$, for every $(t, x, \xi) \in Q_{R,\eta,\rho}$,

$$|\hat{a}(t,x;\xi) - a^*(t,x;\xi)| \le \frac{4}{\Delta(t) D_{\min}} \left[\frac{f(\xi)|\hat{g}_2(t,x;\xi) - g_2(t,x;\xi)| + |g_2(t,x;\xi)|\,|\hat{f}_2(\xi) - f(\xi)|}{f_{\min}^2} \right.$$
$$\left. + |Q^*(t,x;\xi)| \frac{f(\xi)|\hat{g}_1(t,x;\xi) - g_1(t,x;\xi)| + |g_1(t,x;\xi)|\,|\hat{f}_1(\xi) - f(\xi)|}{f_{\min}^2} \right].$$

Under Assumption 3.2, the quantities f, g_1, g_2, and Q^* are uniformly bounded on $Q_{R,\eta,\rho}$, so Proposition 4.1 yields the simpler upper bound for $|\hat{a}(t,x;\xi) - a^*(t,x;\xi)|$ on $\mathcal{E}$,

$$\frac{C}{\Delta(t)} \Big(|\hat{g}_2(t,x;\xi) - g_2(t,x;\xi)| + |\hat{f}_2(\xi) - f(\xi)| + |\hat{g}_1(t,x;\xi) - g_1(t,x;\xi)| + |\hat{f}_1(\xi) - f(\xi)| \Big).$$

4.2 Finite-sample uniform bound

Let $C_K > 0$ be the entropy constant from Appendix A and define

$$\Lambda(h,\delta) := d \log\Big(\frac{C_K}{h}\Big) + \log\Big(\frac{2}{\delta}\Big),\ \widetilde{\Lambda}(h,\delta) := d \log\Big(\frac{C_K}{h}\Big) + \log\Big(\frac{8}{\delta}\Big),\ r_{e,j}(\delta) := h_j^\beta + \sqrt{\frac{\widetilde{\Lambda}(h_j,\delta)}{M h_j^d}},$$

for $j \in \{1, 2\}$. We call (h_1, h_2) δ-admissible if $r_{e,1}(\delta) + r_{e,2}(\delta) \le c_0$.

Theorem 4.2 (Uniform finite-sample bound)**.** *Under Assumptions 3.1 and 3.2, there exist constants $c_0, C > 0$ such that, for every $\delta \in (0,1)$ and every δ-admissible bandwidth pair (h_1, h_2), with probability at least $1 - \delta$,*

$$\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |\hat{a}(t,x;\xi) - a^*(t,x;\xi)| \le \frac{C}{\eta} \left[h_1^\beta + h_2^\beta + \sqrt{\frac{\Lambda(h_1,\delta)}{M h_1^d}} + \sqrt{\frac{\Lambda(h_2,\delta)}{M h_2^d}} \right].$$

The key difficulty is that the SB weight is not a standard bounded regression weight: it contains both a contracting Gaussian and an expanding exponential. Our proof handles this by factorizing $F(t,\xi,x,y) = \chi(t,x,y) G(\xi,y)$ with $\chi(t,x,y) := e^{-|y-x|^2/(2\Delta(t))} \in [0,1]$ and $G(\xi,y) := e^{|y-\xi|^2/(2\Delta)} \le C_F$ on $B \times B$. Bounded support controls the expanding term, while the contraction term yields both Lipschitz control in the parameters and decay in the unbounded state variable x. As a result, the weighted empirical-process class has the same entropy order as the classical translated-kernel class $\{z \mapsto K_h(z - \xi) : \xi \in B_\rho\}$.

Corollary 4.3 (Oracle bandwidth)**.** *If* $h_1 = h_2 = h^\star := \big((d\log(C_K M) + \log(2/\delta))/M\big)^{1/(2\beta+d)}$, *then for each fixed* $\delta \in (0,1)$ *there exists* $M_0(\delta) \geq 1$, *depending on* δ *and the model constants, such that for all* $M \geq M_0(\delta)$ *the pair* $(h^\star, h^\star)$ *is* δ*-admissible and, with probability at least* $1-\delta$,

$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |\hat{a}(t,x;\xi) - a^*(t,x;\xi)| \leq \frac{C}{\eta}\left(\frac{d\log(C_K M) + \log(2/\delta)}{M}\right)^{\beta/(2\beta+d)}.$$

In particular, for $\delta = M^{-\gamma}$ *with fixed* $\gamma > 0$,

$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |\hat{a} - a^*| = O_{\mathbb{P}}\left(\left(\tfrac{\log M}{M}\right)^{\beta/(2\beta+d)}\right).$$

Remark 4.4 (Finite observation grids). For a grid $0 < t_1 < \cdots < t_N = T$ with first-order Markov observations, construct the interval-specific estimator $\hat{a}_i$ from the pairs $(X_{t_i}, X_{t_{i+1}})$. If the interval laws satisfy Assumption 3.2 with constants and support set uniform in i, Theorem 4.2 applies on each interval with the same constants. Applying it with confidence level $\delta/(N-1)$ and taking a union bound gives the simultaneous intervalwise bound with

$$\Lambda_N(h,\delta) = d\log\left(\frac{C_K}{h}\right) + \log\left(\frac{2(N-1)}{\delta}\right).$$

Thus finite-grid extensions cost only the logarithmic factor $\log(N-1)$.

Remark 4.5 (Time-edge singularity). The factor $1/\eta$ reflects the intrinsic singularity of the SB drift as $t \uparrow u$. The statistical rate in M remains the classical kernel rate up to logarithmic factors.

Remark 4.6 (One versus two bandwidths). Both blocks have optimal order $(\log M/M)^{1/(2\beta+d)}$, so unequal bandwidths do not improve the stated exponent, although block-specific constants may favor $h_1 \neq h_2$ in finite samples [6]. We retain the unequal-bandwidth appendices but use $h_1 = h_2 = h$ in the main text and experiments.

4.3 Pointwise inference

In the main text we specialize to $h_1 = h_2 = h$; the unequal-bandwidth statement is given in Appendix D. For fixed $(t,x,\xi) \in [s, u-\eta] \times B_R \times \mathrm{int}(B)$, let $R(K) := \int_{\mathbb{R}^d} K(z)^2\, dz$ and define

$$\psi_{t,x,\xi}(y) := \big(y - x - \Delta(t)a^*(t,x;\xi)\big)F(t,\xi,x,y),$$
$$\Sigma(t,x;\xi) := \frac{R(K)\mathrm{Var}(\psi_{t,x,\xi}(X_u) \mid X_s = \xi)}{f(\xi)\Delta(t)^2 D^*(t,x;\xi)^2}.$$

Let $\hat{\Sigma}_h(t,x;\xi)$ be the plug-in covariance estimator defined in (11).

Theorem 4.7 (Pointwise inference)**.** *Assume Assumptions 3.1 and 3.2, and let* $h_1 = h_2 = h \to 0$ *satisfy* $Mh^d \to \infty$ *and* $h^\beta\sqrt{Mh^d} \to 0$. *Then, for every fixed* $(t,x,\xi) \in [s, u-\eta] \times B_R \times \mathrm{int}(B)$,

$$\sqrt{Mh^d}\big(\hat{a}(t,x;\xi) - a^*(t,x;\xi)\big) \Rightarrow \mathcal{N}\big(0, \Sigma(t,x;\xi)\big),$$
$$|\hat{\Sigma}_h(t,x;\xi) - \Sigma(t,x;\xi)|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0.$$

Consequently, for every fixed $v \in \mathbb{R}^d$ *such that* $v^\top\Sigma(t,x;\xi)v > 0$, *we have* $v^\top\hat{\Sigma}_h(t,x;\xi)v > 0$ *with probability tending to one and*

$$\frac{\sqrt{Mh^d}\, v^\top(\hat{a}(t,x;\xi) - a^*(t,x;\xi))}{\sqrt{v^\top\hat{\Sigma}_h(t,x;\xi)v}} \Rightarrow \mathcal{N}(0,1).$$

Thus, for every $\alpha \in (0,1)$, $v^\top\hat{a}(t,x;\xi) \pm z_{1-\alpha/2}\sqrt{v^\top\hat{\Sigma}_h(t,x;\xi)v/(Mh^d)}$ *is an asymptotically valid pointwise* $(1-\alpha)$ *Wald interval for* $v^\top a^*(t,x;\xi)$. *This result does not establish simultaneous confidence bands.*

Remark 4.8 (Undersmoothing is genuinely required). The condition $h^\beta\sqrt{Mh^d} \to 0$ is the scaled-bias requirement for the CLT. Taking only a smaller constant multiple of the rate-optimal bandwidth does not suffice as it preserves the same exponent and therefore does not make the scaled bias vanish.

4.4 Adaptive oracle inequality and minimax optimality

Again following Remark 4.6, we specialize in the main text to $h_1 = h_2 = h$. Let $\mathcal{H} = \{h_0 2^{-k}\}_{k=0}^{K_M} \subset (0, 1]$ be a finite geometric bandwidth grid with largest element $h_0 \asymp 1$ and smallest element of order $M^{-1/d}$. For $h \in \mathcal{H}$, let $\hat{a}_h$ denote (2) with $h_1 = h_2 = h$, and define

$$\mathcal{R}_\nu(\hat{a}_h) := \sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} \mathbb{E}_\nu\big[|\hat{a}_h(t, x; \xi) - a^*(t, x; \xi)|\big].$$

Let $\mathfrak{P}(\beta, L_p, f_{\min}, B)$ denote the class of pair laws satisfying Assumption 3.2 with the displayed constants.

Theorem 4.9 (Oracle inequality)**.** *Under Assumptions 3.1 and 3.2, there exist constants $M_0 \geq 1$ and $C_1, C_2 > 0$ such that, for all $M \geq M_0$, the diagonal GL/Lepski selector $\widehat{h} \in \mathcal{H}$ described in Appendix E satisfies*

$$\sup_{\nu\in\mathfrak{P}(\beta,L_p,f_{\min},B)} \mathcal{R}_\nu(\hat{a}_{\widehat{h}}) \leq C_1 \inf_{h\in\mathcal{H}} \left(h^\beta + \sqrt{\frac{\log M}{Mh^d}} + \frac{\log M}{Mh^d} \right) + C_2 M^{-1}.$$

Corollary 4.10 (Adaptive minimax optimality)**.** *If $\mathcal{H} = \{h_0 2^{-k} : k = 0, 1, \ldots, K_M\}$, with $h_0 \asymp 1$ and smallest element of order $M^{-1/d}$, then for all sufficiently large M,*

$$\sup_{\nu\in\mathfrak{P}(\beta,L_p,f_{\min},B)} \mathcal{R}_\nu(\hat{a}_{\widehat{h}}) \lesssim \left(\frac{\log M}{M}\right)^{\beta/(2\beta+d)}.$$

Moreover, for every interior pivot law $\nu_0 \in \mathfrak{P}(\beta, L_p, f_{\min}, B)$ satisfying the slack conditions of Proposition F.1, there exist constants $Q_0 > 0$, $c_{\mathrm{loc}} > 0$, and $c > 0$ such that, for all large M,

$$\inf_{\widetilde{a}} \sup_{\nu\in\mathfrak{P}^{\cap}_{\mathrm{loc},M}(\nu_0;Q_0,c_{\mathrm{loc}}M^{-\beta/(2\beta+d)})} \mathcal{R}_\nu(\widetilde{a}) \geq cM^{-\beta/(2\beta+d)}.$$

Thus the adaptive rate is locally sharp around every such interior pivot.

If, in addition, $f_{\min} < |B|^{-1}$ and $L_p > |B|^{-2}$, then the same local construction, instantiated at the explicit uniform pivot, yields the global minimax lower bound

$$\inf_{\widetilde{a}} \sup_{\nu\in\mathfrak{P}(\beta,L_p,f_{\min},B)} \mathcal{R}_\nu(\widetilde{a}) \gtrsim M^{-\beta/(2\beta+d)}.$$

Thus, $\hat{a}_{\widehat{h}}$ is globally minimax-rate optimal over $\mathfrak{P}(\beta, L_p, f_{\min}, B)$, up to logarithmic factors.

Remark 4.11 (Oracle interpretation and minimax interpretation). The adaptive theorem controls the estimator up to multiplicative constants and lower-order terms. Thus it supports rate-level competitiveness with oracle tuning, but not finite-sample equality with the oracle bandwidth.

The lower-bound argument has two layers. Proposition F.1 is a pivot-local Le Cam construction: around any interior law satisfying the slack conditions, no estimator can improve on the rate $M^{-\beta/(2\beta+d)}$ for the uniform risk. Corollary F.2 then instantiates this construction at an explicit uniform pivot contained in the global class, under transparent compatibility conditions. Since that local neighborhood is a subset of the global model, the same construction yields a global minimax lower bound over $\mathfrak{P}(\beta, L_p, f_{\min}, B)$.

Thus the adaptive selector is not only locally sharp around interior pivots; under the explicit compatibility conditions of Corollary F.2, it is globally minimax-rate optimal up to logarithmic factors.

Remark 4.12 (Experiment-facing proxy). The adaptive theorem controls uniform risk over (t, x, ξ). Exact evaluation of that risk is not available in experiments, so we use the fixed-grid sup error

$$E_{\infty,\mathcal{G}}(t, \xi; h) := \max_{x\in\mathcal{G}_x} |\hat{a}_h(t, x; \xi) - a^*(t, x; \xi)| \qquad \text{for fixed prescribed interior pairs } (t, \xi). \quad (3)$$

It tests the same one-sided GL behavior on a theorem-facing proxy, but should not be read as a direct finite-sample verification of the full uniform oracle inequality.

Table 1: Estimated oracle slopes of $\log \overline{E}_{\infty,\mathcal{G}}(M)$ versus $\log M$.

Testbed	d	Theory	Oracle-h_{or}
GG	1	-0.349379	-0.348899
GG	2	-0.291150	-0.243435
MM	1	-0.349379	-0.390229
MM	2	-0.291150	-0.385828

5 Experiments

We use controlled synthetic experiments in the SBTS single-interval setting to test the theorems' predictions: finite-sample scaling, pointwise Gaussian approximation under undersmoothing, adaptive bandwidth behavior, and terminal-edge deterioration. The goal is not to re-benchmark SBTS as an end-to-end generator. Alouadi et al. [1] compare the same finite-history SBTS construction with TSGM-VP/subVP [22], ImagenTime [27], COT-GAN [35], and GAN/VAE baselines including TimeGAN [36] and TimeVAE [9]. These path-level predictive and discriminative benchmarks complement our known-drift experiments, which isolate statistical drift-estimation error from optimization, discretization, simulation, and potential-to-drift recovery error. We therefore use synthetic bridge families with cached ground-truth drifts.

Setups and metrics. We consider two synthetic families with numerically cached ground-truth drifts: a Gaussian-to-Gaussian bridge (GG) and a Mixture-to-Mixture bridge (MM). We report results for $d \in \{1, 2\}$ on fixed bounded evaluation boxes $\mathcal{X}_{\mathrm{eval}} \subset \mathbb{R}^d$, using the product Epanechnikov kernel and sample sizes $M \in \{10^3, 2 \times 10^3, 4 \times 10^3, 8 \times 10^3\}$; exact model parameters, grids, repetition counts, and interior query points are given in Appendix G.

For rate and adaptivity, the primary loss is the sup-grid error $E_{\infty,\mathcal{G}}(t_0, \xi_0; h)$ defined in (3), which is structurally closer to the theorem-side uniform risk than the integrated squared error. For rate verification we fit the slope of $\log \overline{E}_{\infty,\mathcal{G}}(M)$ versus $\log M$ and compare it to the finite-range slope induced by $(\log M/M)^{\beta/(2\beta+d)}$; since the implementation uses the product Epanechnikov kernel, the benchmark is the $\beta = 2$ absolute-error rate.

For the CLT experiment we use $Z_M := \sqrt{Mh_M^d}\,(\hat{a}_{h_M}(t_0, x_0; \xi_0) - a^*(t_0, x_0; \xi_0))/\sqrt{\hat{\Sigma}(t_0, x_0; \xi_0)}$ as the standardized statistic, with $\hat{\Sigma}$ the plug-in variance estimator from Appendix G. The bandwidth sequence h_M is chosen from a genuinely undersmoothed regime satisfying $Mh_M^d \to \infty$ and $h_M^\beta \sqrt{Mh_M^d} \to 0$. In the final reported one-dimensional runs we fix $h_M = M^{-\alpha}$ with $\alpha = 0.22$ for GG1 and $\alpha = 0.28$ for MM1 after a short pilot screen on the larger sample sizes. The final CLT summaries use $N_{\mathrm{rep}} = 300$ Monte Carlo repetitions.

For adaptivity we compare the empirical oracle bandwidth h_{or}, obtained by minimizing the empirical error against the cached truth on a fixed geometric grid, with a single canonical theorem-aligned selector $\hat{h}$ on the same grid. We use the raw-max, one-sided GL-type rule with $\kappa_{\mathrm{pair}} = \kappa_{\mathrm{final}} = 2$ and restrict the search to the stable regime $Mh^d \geq 81$; selector calibration and nearby variants are summarized in Appendix G.4. The main diagnostic is the oracle ratio $\Gamma_{\nu,M}^{(r)} := E_{\infty,\mathcal{G}}^{(r)}(t_0, \xi_0; \hat{h})/E_{\infty,\mathcal{G}}^{(r)}(t_0, \xi_0; h_{\mathrm{or}}) \geq 1$, together with its repetition average $\bar{\Gamma}_\nu(M)$ and the four-point summaries $\widehat{C}_\nu^{\max} := \max_{M \in \mathcal{M}} \bar{\Gamma}_\nu(M)$ and $\widehat{C}_\nu^{\mathrm{avg}} := |\mathcal{M}|^{-1} \sum_{M \in \mathcal{M}} \bar{\Gamma}_\nu(M)$ for $\mathcal{M} = \{10^3, 2 \times 10^3, 4 \times 10^3, 8 \times 10^3\}$.

Rate verification. Figure 1 and Table 1 show finite-range decay compatible with Corollary 4.3. The replicate-level 95% bootstrap intervals from 5,000 resamples are $[-0.443, -0.248]$, $[-0.342, -0.148]$, $[-0.457, -0.324]$, and $[-0.483, -0.288]$ for GG1, GG2, MM1, and MM2; each contains the corresponding finite-range benchmark. Because the slopes use only four sample sizes, we do not interpret them as identification of the exact asymptotic exponent. Adaptivity is assessed separately through oracle-ratio diagnostics.

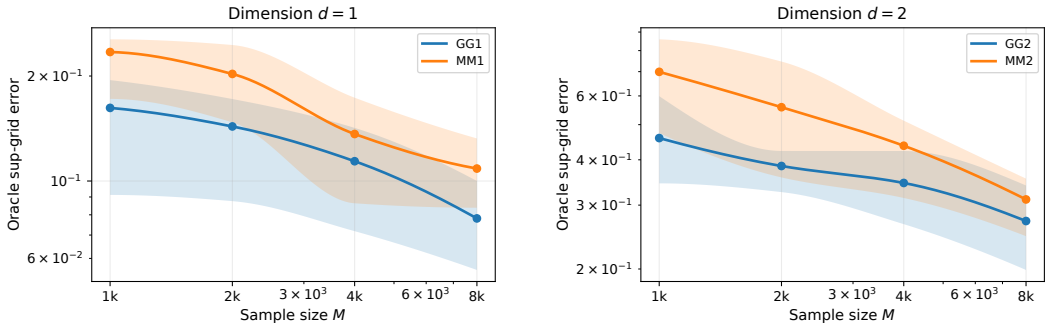


Figure 1: Oracle sup-grid error by conditioning dimension. Left: GG1 and MM1 in $d = 1$. Right: GG2 and MM2 in $d = 2$. Markers denote the evaluated sample sizes, curves are shape-preserving spline interpolants, and shaded bands show repetition-level interquartile ranges. The fitted oracle slopes are compared with the finite-range theoretical slope induced by $(\log M/M)^{\beta/(2\beta+d)}$; adaptive performance is assessed separately through oracle-ratio diagnostics.

Table 2: Empirical coverage (%) of nominal 95% confidence intervals in the final $N_{\mathrm{rep}} = 300$ CLT runs.

Testbed	10^3	2×10^3	4×10^3	8×10^3
GG	92.67	96.33	96.67	96.67
MM	92.33	96.67	96.33	96.00

Pointwise Gaussian approximation and adaptivity. Figure 2 and Table 2 examine the CLT and the practical selector. The final one-dimensional CLT runs show near-nominal coverage from $M = 2 \times 10^3$ onward, approximately linear QQ-plots at the two largest sample sizes, and formal Shapiro–Wilk / Anderson–Darling diagnostics that do not reject normality at the largest sample size in either testbed; empirical means and variances are reported in Appendix G. We therefore interpret the CLT experiment as showing a clear finite-sample Gaussian footprint under genuinely undersmoothed bandwidth sequences, while emphasizing that the evidence is pointwise and design-specific.

For adaptivity, the selector remains in a low-single-digit oracle-ratio regime across all four testbeds, with average ratios between 1.314 and 1.960, maximum ratios between 1.431 and 2.175, and essentially zero boundary-hit rates. This is the correct oracle-inequality benchmark: the issue is not oracle-tightness but whether the selector stays moderate and stable across the tested range [14, 18]. Under that criterion, the canonical raw-max one-sided selector is practically oracle-competitive once the bandwidth grid is restricted to $Mh^d \geq 81$. Across all 6,000 candidate-bandwidth evaluations, neither 10^{-12} numerical safeguard was activated; the global minima were $\min \widehat{f}_h = 0.1225$ and $\min \widehat{D}_h = 8.20 \times 10^{-4}$. Per-testbed diagnostics are given in Appendix G.4, and the complete candidate-level audit is included in the supplement.

Terminal singularity and reproducibility. To visualize the factor $\Delta(t)^{-1}$, we also evaluate the estimator at times approaching u. The raw error $E_{\infty,\mathcal{G}}(t, \xi_0; \hat{h})$ increases as $t \uparrow u$, whereas the rescaled quantity $\Delta(t) E_{\infty,\mathcal{G}}(t, \xi_0; \hat{h})$ remains substantially flatter; see Figure 4. This matches Remark 4.5 and indicates that the deterioration near u is driven mainly by the intrinsic SB singularity. Appendix G.8 reports a bounded-support stress test: in the easier GG1 testbed the wider-support perturbation has only a mild effect, but in a stronger MM1 wide-support variant it sharply increases the variability of the weighted empirical terms, produces nontrivial boundary-hit events, and substantially worsens the selected-estimator error. The code and configuration files reproducing the experiments, figures, and tables are publicly available at `https://github.com/OthmaneAnalytics/sb-drift-experiments`. The repository includes synthetic-model definitions, ground-truth computation, bandwidth configurations, random seeds, stress-test variants, evaluation grids, and saved experimental outputs.

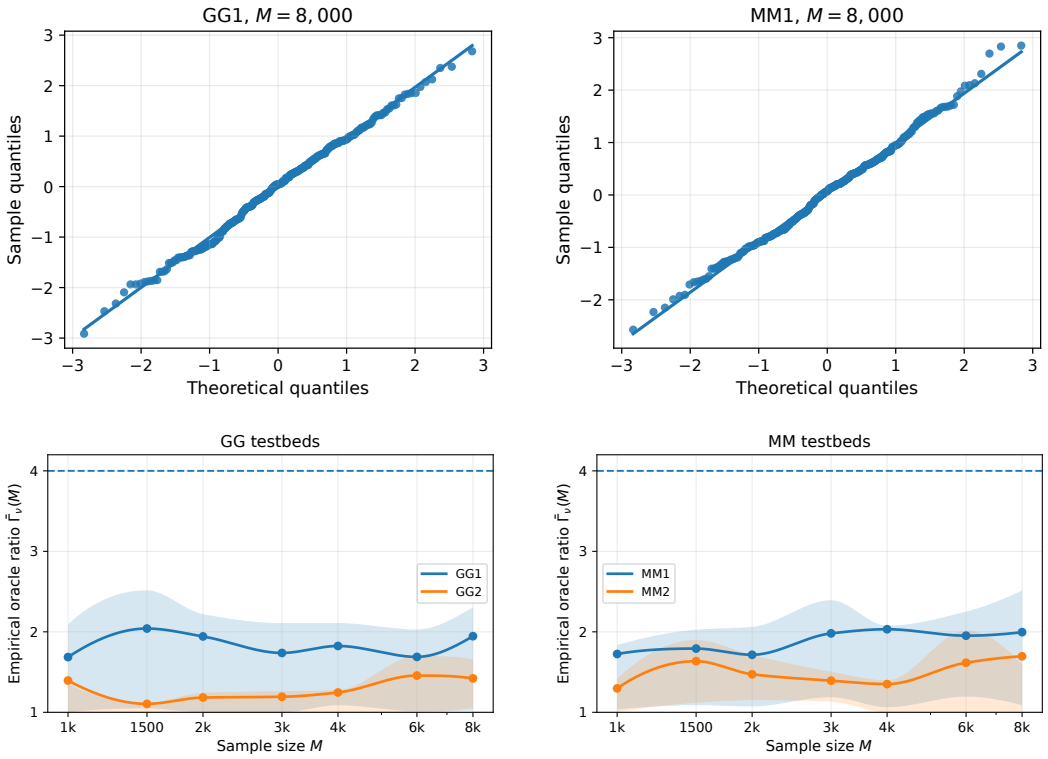


Figure 2: Pointwise Gaussian approximation and adaptive behavior. Top: QQ-plots for the final one-dimensional CLT runs at $M = 8 \times 10^3$ in GG1 and MM1. Bottom: empirical oracle ratios for the canonical selector on the denser sample-size grid, separated into GG and MM testbeds. Markers denote the evaluated sample sizes, curves are shape-preserving spline interpolants, shaded bands show interquartile ranges across repetitions, and the dashed line marks the GL reference level 4.

6 Discussion and Limitations

The paper gives a sharp theory for *direct drift estimation in the SBTS single-interval* setting. Bounded support is central to the current proof because it controls the expanding exponential factor in the SB weight and makes the weighted empirical-process analysis tractable. As a nonparametric method, the estimator also inherits the usual curse of dimensionality, and adaptive selection adds the computational cost of evaluating a bandwidth grid.

The theory is uniform on the query set $Q_{R,\eta,\rho} = [s, u-\eta] \times B_R \times B_\rho$ and necessarily deteriorates as $t \uparrow u$. The main results are proved for i.i.d. trajectory samples in the single-interval model, while the finite-sample ratio analysis works on a natural high-probability event where empirical denominators are separated from zero. Fixed finite-memory extensions are standard, replacing the conditioning variable by a lag block and the ambient conditioning dimension d by Ld. Estimating the SB drift from one discretely observed trajectory, where the effective samples are dependent and the past may grow with the grid, requires new tools and is left for future work.

The lower-bound result is local in its construction but global in its implication under the compatibility conditions of Corollary F.2. Proposition F.1 proves a sharp two-point Le Cam lower bound inside a shrinking neighborhood of any interior pivot law. Corollary F.2 then exhibits an explicit uniform pivot satisfying these interiority conditions whenever the model constants are compatible with the bounded support class. Since the resulting local neighborhood is contained in the original global model, the same construction yields a global minimax lower bound over $\mathfrak{P}(\beta, L_p, f_{\min}, B)$. Thus the adaptive selector is globally minimax-rate optimal up to logarithmic factors, while the proof also gives the stronger local interpretation around interior laws.

7 Conclusion

We studied direct nonparametric estimation of Schrödinger bridge time-series drifts. Unlike approaches based on entropic-OT potentials, Sinkhorn iterations, or iterative bridge solvers, our analysis works directly at the drift level and isolates the *statistical error* of the estimator. Under Hölder regularity, a marginal-density floor, and bounded support, we proved an interior uniform finite-sample bound at the classical kernel rate up to logarithmic factors, a pointwise CLT under genuine undersmoothing, and an adaptive selector satisfying an oracle inequality. We also proved a pivot-local minimax lower bound and, by exhibiting an explicit uniform pivot in the global class, derived a global minimax lower bound under transparent compatibility conditions. This shows that the adaptive selector is minimax-rate optimal up to logarithmic factors. Together, these results provide a self-contained statistical theory for direct kernel estimation of SBTS drifts.

Acknowledgements and Disclosures

Funding and support.

Othmane Mazhar acknowledges support from the BNP Paribas Chair "Futures of Quantitative Finance" and from Jump Trading through the Chair in Artificial Intelligence and Quantitative Methods for Finance at Institut Louis Bachelier, hosted within Fondation du Risque.

Huyên Pham acknowledges support from the Chair "Financial Risks", FiME (Laboratory of Finance and Energy Markets), and the EDF–CACIB Chair "Finance and Sustainable Development".

Competing interests.

The authors declare no competing interests beyond the funding and financial support disclosed above.

References

[1] A. Alouadi, B. Barreau, L. Carlier, and H. Pham. Robust time series generation via Schrödinger Bridge: a comprehensive evaluation. In *Proceedings of the 6th ACM International Conference on AI in Finance*, ICAIF '25, pages 906–914, 2025. URL `https://doi.org/10.1145/3768292.3770391`.

[2] D. Belomestny and J. Schoenmakers. Forward reverse kernel regression for the schrödinger bridge problem. *arXiv preprint arXiv:2507.00640*, 2025.

[3] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[4] O. Bousquet. A bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus Mathematique*, 334(6):495–500, 2002. doi: 10.1016/S1631-073X(02)02292-6.

[5] Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrödinger bridge. *SIAM Review*, 63(2):249–313, 2021. doi: 10.1137/20M1339982.

[6] F. Comte and N. Marie. On a Nadaraya–Watson estimator with two bandwidths. *Electronic Journal of Statistics*, 15(1):2566–2607, 2021. doi: 10.1214/21-EJS1849.

[7] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013. URL `https://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport`.

[8] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, 2021.

[9] A. Desai, C. Freeman, Z. Wang, and I. Beaver. TimeVAE: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*, 2021. doi: 10.48550/arXiv.2111.08095.

[10] U. Einmahl and D. M. Mason. Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3):1380–1403, 2005.

[11] H. Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Mathematics*, pages 101–203. Springer, 1988. doi: 10.1007/BFb0086180.

[12] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. Sample complexity of sinkhorn divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 2019. URL https://proceedings.mlr.press/v89/genevay19a.html.

[13] E. Giné and A. Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 38(6):907–921, 2002.

[14] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011. doi: 10.1214/11-AOS883.

[15] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002. doi: 10.1007/978-1-4613-0003-1.

[16] M. Hamdouche, P. Henry-Labordère, and H. Pham. Nonparametric generative modeling for time series via Schrödinger bridge. *Journal of Machine Learning Research*, 27(58):1–23, 2026. URL https://www.jmlr.org/papers/v27/23-1162.html.

[17] B. Jamison. The markov processes of schrödinger. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32:323–331, 1975. doi: 10.1007/BF00535844.

[18] C. Lacour and P. Massart. Minimal penalty for goldenshluger–lepski method. *Stochastic Processes and their Applications*, 126(12):3774–3789, 2016. doi: 10.1016/j.spa.2016.04.015.

[19] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer, New York, 1986.

[20] C. Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems - A*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533.

[21] O. V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, 25(6):2512–2546, 1997. doi: 10.1214/aos/1030741083.

[22] H. Lim, J. Lee, S. Park, M. Kim, and N. Park. TSGM: Regular and irregular time-series generation using score-based generative models. In *2025 IEEE International Conference on Big Data (BigData)*, pages 54–63. IEEE, 2025. doi: 10.1109/BIGDATA66926.2025.11401567.

[23] G.-H. Liu, Y. Lipman, M. Nickel, B. Karrer, E. A. Theodorou, and R. T. Q. Chen. Generalized schrödinger bridge matching. In *International Conference on Learning Representations*, 2024.

[24] N. Marie and A. Rosier. Nadaraya–watson estimator for i.i.d. paths of diffusion processes. *Scandinavian Journal of Statistics*, 50(2):589–637, 2023. doi: 10.1111/sjos.12593.

[25] G. Mena and J. Weed. Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5acdc9ca5d99ae66afdfe1eea0e3b26b-Abstract.html.

[26] E. A. Nadaraya. On estimating regression. *Theory of Probability and Its Applications*, 9(1):141–142, 1964. doi: 10.1137/1109020.

[27] I. Naiman, N. Berman, I. Pemper, I. Arbiv, G. Fadlon, and O. Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024. doi: 10.52202/079017-3868.

[28] F. Panloup, S. Tindel, and M. Varvenne. A general drift estimation procedure for stochastic differential equations with additive fractional noise. *Electronic Journal of Statistics*, 14(1): 1075–1136, 2020. doi: 10.1214/20-EJS1685.

[29] A. Pooladian and J. Niles-Weed. Plug-in estimation of schrödinger bridges. *arXiv preprint arXiv:2408.11686*, 2024.

[30] E. Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, pages 144–153, 1931.

[31] A. J. Stromme. Sampling from a schrödinger bridge. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 4058–4067. PMLR, 2023. URL `https://proceedings.mlr.press/v206/stromme23a.html`.

[32] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009. doi: 10.1007/b13794.

[33] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996. doi: 10.1007/978-1-4757-2545-2.

[34] G. S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

[35] T. Xu, L. K. Wenliang, M. Munn, and B. Acciaio. COT-GAN: Generating sequential data via causal optimal transport. In *Advances in Neural Information Processing Systems*, volume 33, pages 8798–8809. Curran Associates, Inc., 2020.

[36] J. Yoon, D. Jarrett, and M. van der Schaar. Time-series generative adversarial networks. In *Advances in Neural Information Processing Systems*, volume 32, pages 5509–5519. Curran Associates, Inc., 2019.

[37] B. Yu. Assouad, fano, and le cam. In D. Pollard, E. Torgersen, and G. L. Yang, editors, *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997. doi: 10.1007/978-1-4612-1880-7_29.

A Proof roadmap and auxiliary notation

Throughout the Appendices, we keep the notation of Sections 3–4. In particular, $g_1, g_2, D^*, N^*, Q^*, a^*, \widehat{f}_j, \widehat{g}_j, \widehat{D}, \widehat{N}, \widehat{a}$ retain their meanings from the main text. We also write $C_f := |f|_\infty$, $C_B := \lambda(B)$, and $|K|_2 := (\int_{\mathbb{R}^d} K(z)^2\,dz)^{1/2}$ and $\mu_\infty(K) := |K|_\infty$. For fixed (t, x, ξ), we write $\Delta_t := \Delta(t)$ and $F_{t,x,\xi}(y) := F(t, \xi, x, y)$ when convenient.

For an open neighborhood U of B, write $\Sigma_U(\beta, L)$ for the Hölder ball of functions $r : U \to \mathbb{R}$ satisfying

$$\sum_{|\alpha|\le \ell_\beta} |D^\alpha r|_\infty + \sum_{|\alpha|=\ell_\beta} \sup_{z\neq z'\in U} \frac{|D^\alpha r(z) - D^\alpha r(z')|}{|z - z'|^{\beta-\ell_\beta}} \le L.$$

When the neighborhood U is clear, we write $\Sigma(\beta, L)$.

A.1 Uniform KDE bound

Lemma A.1 (Uniform KDE bound for $\widehat{f}_j$)**.** *For each $j \in \{1, 2\}$, there exist constants $C_1, C_2, C_K > 0$, depending only on $K, d, \mathrm{diam}(B)$, such that for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\sup_{\xi\in B_\rho} \big|\widehat{f}_j(\xi) - f(\xi)\big| \le \frac{C_B L_p}{\ell_\beta!}\mu_\beta(K)h_j^\beta + C_1\sqrt{C_f}\,|K|_2\sqrt{\frac{d\log\big(\frac{C_K}{h_j}\big) + \log\big(\frac{2}{\delta}\big)}{Mh_j^d}} + C_2\frac{\mu_\infty(K)}{h_j^d}\frac{d\log\big(\frac{C_K}{h_j}\big) + \log\big(\frac{2}{\delta}\big)}{M}.$$

Proof. This is a standard uniform kernel-density estimate for the translated-kernel class given as $\{K_{h_j}(\cdot - \xi) : \xi \in B_\rho\}$ under Hölder regularity of f. For $h_j \le \rho/R_K$, the kernel window stays inside B for $\xi \in B_\rho$. The bias is the standard Tsybakov-order expansion: all monomials $1 \le |\alpha| \le \ell_\beta$ vanish, and since $f(\xi) = \int_B p(\xi, y)\,dy$ and $\sup_{y\in B} |\bar{p}(\cdot, y)|_{C^\beta} \le L_p$, the marginal satisfies $|f|_{C^\beta} \le C_B L_p$. Hence the Hölder remainder is bounded by $(C_B L_p/\ell_\beta!)h_j^\beta\mu_\beta(K)$. The stochastic term is the usual uniform empirical-process bound for translated kernels. Results of this type may be found, for example, in Giné and Guillou [13] and Einmahl and Mason [10]. Applied to the present fixed-bandwidth setting, they yield a bound of the displayed form, so we omit the routine details. □

A.2 Uniform weighted bound for $\widehat{g}_1$

Lemma A.2 (Uniform weighted bound for $\widehat{g}_1$)**.** *For any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\sup_{t\in[s,u-\eta],\, x\in\mathbb{R}^d,\, \xi\in B_\rho} \big|\widehat{g}_1(t, x; \xi) - g_1(t, x; \xi)\big| \le \frac{C_B C_F L_p}{\ell_\beta!}\mu_\beta(K)h_1^\beta + C_1\,C_F\,\sqrt{C_f}\,|K|_2\sqrt{\frac{d\log\big(\frac{C_K}{h_1}\big) + \log\big(\frac{2}{\delta}\big)}{Mh_1^d}} + C_2\,\frac{C_F\,\mu_\infty(K)}{h_1^d}\frac{d\log\big(\frac{C_K}{h_1}\big) + \log\big(\frac{2}{\delta}\big)}{M}.$$

Proof. Let v_d denote the volume of the Euclidean unit ball in $\mathbb{R}^d$, and choose $R_K > 0$ such that $\mathrm{supp}(K) \subseteq B(0, R_K)$. Fix (t, x, ξ) and write, with the change of variables $u = (z - \xi)/h_1$,

$$\begin{aligned}\mathbb{E}\hat{g}_1(t, x; \xi) &= \iint F(t, \xi, x, y)\,K_{h_1}(z - \xi)\,p(z, y)\,dz\,dy \\ &= \int_{\mathbb{R}^d} K(u)\Big(\int_B F(t, \xi, x, y)\,\bar{p}(\xi + h_1 u, y)\,dy\Big)\,du,\end{aligned}$$

where $\bar{p}(\cdot, y)$ is the extension from Assumption 3.2(i). Define $H_{t,x,\xi}(z) := \int_B F(t, \xi, x, y)\,\bar{p}(z, y)\,dy$. By bounded support and the definition of C_F in Remark 3.3, we have $F(t, \xi, x, y) \le C_F$ for

all $(t,x,\xi,y) \in [s,u-\eta] \times \mathbb{R}^d \times B \times B$. By Assumption 3.2 (i), the map $z \mapsto \bar{p}(z,y)$ is Hölder–β uniformly in $y \in B$. Hence $H_{t,x,\xi} \in \Sigma(\beta, C_B C_F L_p)$ in its z-argument, uniformly over $(t,x,\xi) \in [s,u-\eta] \times \mathbb{R}^d \times B_\rho$. More explicitly, for every multi-index α with $|\alpha| \le \ell_\beta$,

$$|D^\alpha H_{t,x,\xi}(z)| \le \int_B F(t,\xi,x,y)\, |D^\alpha \bar{p}(z,y)|\, dy \le C_B C_F L_p,$$

and for $|\alpha| = \ell_\beta$, $|D^\alpha H_{t,x,\xi}(z) - D^\alpha H_{t,x,\xi}(z')| \le C_B C_F L_p\, |z-z'|^{\beta-\ell_\beta}$.

With $\ell_\beta = \max\{m \in \mathbb{N}_0 : m < \beta\}$, Taylor expansion of $H_{t,x,\xi}$ at ξ to order ℓ_β, together with the moment conditions $\int_{\mathbb{R}^d} u^\alpha K(u)\, du = 0,\ 1 \le |\alpha| \le \ell_\beta$, yields

$$\begin{aligned}\big|\mathbb{E}\,\hat{g}_1(t,x;\xi) - H_{t,x,\xi}(\xi)\big| &= \Big| \int K(u)\big(H_{t,x,\xi}(\xi + h_1 u) - H_{t,x,\xi}(\xi)\big)\, du \Big| \\ &\le \frac{C_B C_F L_p}{\ell_\beta!}\, h_1^\beta \int |u|^\beta |K(u)|\, du.\end{aligned}$$

As $\xi \in B$ and $\bar{p}(\xi,y) = p(\xi,y)$ on B: $H_{t,x,\xi}(\xi) = \int_B F(t,\xi,x,y)\, p(\xi,y)\, dy = g_1(t,x;\xi)$. We conclude

$$\sup_{t\in[s,u-\eta],\, x\in\mathbb{R}^d,\, \xi\in B_\rho} |\mathbb{E}\hat{g}_1(t,x;\xi) - g_1(t,x;\xi)| \le \frac{C_B C_F L_p}{\ell_\beta!}\mu_\beta(K) h_1^\beta.$$

Now define the weighted class

$$\mathcal{H}_{h_1} := \big\{ h_{t,x,\xi}(z,y) := F(t,\xi,x,y) K_{h_1}(z-\xi) : (t,x,\xi) \in [s,u-\eta] \times \mathbb{R}^d \times B_\rho \big\}$$

and the centered process $\mathbb{G}_M h := \frac{1}{M}\sum_{m=1}^M \big(h(X_s^{(m)}, X_u^{(m)}) - \mathbb{E}h(X_s,X_u)\big)$. Then

$$\sup_{t\in[s,u-\eta],\, x\in\mathbb{R}^d,\, \xi\in B_\rho} |\hat{g}_1(t,x;\xi) - \mathbb{E}\hat{g}_1(t,x;\xi)| = \sup_{h\in\mathcal{H}_{h_1}} |\mathbb{G}_M h|.$$

The class $\mathcal{H}_{h_1}$ has envelope $U_{h_1} := \sup_{h\in\mathcal{H}_{h_1}} |h|_\infty \le C_F\, \mu_\infty(K)/h_1^d$. Moreover, if P denotes the law of (X_s, X_u), then $\sigma_{h_1} := \sup_{h\in\mathcal{H}_{h_1}} |h|_{L_2(P)} \le C_F \sqrt{C_f}\, |K|_2 / h_1^{d/2}$, because

$$\begin{aligned}|h_{t,x,\xi}|_{L_2(P)}^2 &= \iint F(t,\xi,x,y)^2\, K_{h_1}(z-\xi)^2\, p(z,y)\, dz\, dy \\ &\le C_F^2 \int K_{h_1}(z-\xi)^2\, f(z)\, dz \le C_F^2 C_f\, |K_{h_1}|_{L_2(\lambda)}^2.\end{aligned}$$

Write $F(t,\xi,x,y) = \chi(t,x,y)\, G(\xi,y)$, where $\chi(t,x,y) := e^{-\frac{|y-x|^2}{2\Delta(t)}}$ and $G(\xi,y) := e^{\frac{|y-\xi|^2}{2\Delta}}$. On $[s,u-\eta] \times \mathbb{R}^d \times B$, the function χ is globally Lipschitz in (t,x): there exists $L_\chi > 0$, depending only on η, Δ, such that

$$|\chi(t,x,y) - \chi(t',x',y)| \le L_\chi\big(|t-t'| + |x-x'|\big) \qquad \text{for all } t,t' \in [s,u-\eta],\ x,x' \in \mathbb{R}^d,\ y \in B.$$

Indeed, $\sup_{r\ge 0,\ \tau\in[\eta,\Delta]} \frac{r}{\tau} e^{-r^2/(2\tau)} < \infty$, $\sup_{r\ge 0,\ \tau\in[\eta,\Delta]} \frac{r^2}{2\tau^2} e^{-r^2/(2\tau)} < \infty$. Also, G is Lipschitz in ξ on $B \times B$: there exists $L_G > 0$, depending only on Δ and $\mathrm{diam}(B)$, such that

$$|G(\xi,y) - G(\xi',y)| \le L_G |\xi - \xi'| \qquad \text{for all } \xi,\xi',y \in B.$$

Now fix (t,x,ξ) and (t',x',ξ'). Then

$$\begin{aligned}|h_{t,x,\xi}(z,y) - h_{t',x',\xi'}(z,y)| \le\ & |\chi(t,x,y) - \chi(t',x',y)|\, G(\xi,y)\, |K_{h_1}(z-\xi)| \\ &+ \chi(t',x',y)\, |G(\xi,y) - G(\xi',y)|\, |K_{h_1}(z-\xi)| \\ &+ \chi(t',x',y)\, G(\xi',y)\, |K_{h_1}(z-\xi) - K_{h_1}(z-\xi')|.\end{aligned}$$

Using $\chi \le 1$, $G \le C_F$, the two Lipschitz bounds above, and $|K_{h_1}(\cdot-\xi)|_{L_2(P)} \le \sqrt{C_f}\, |K|_2 / h_1^{d/2}$, together with $|K_{h_1}(\cdot-\xi) - K_{h_1}(\cdot-\xi')|_{L_2(P)} \le \sqrt{C_f}\, \frac{L_K \sqrt{2^d v_d}\, R_K^{d/2}}{h_1^{d/2+1}}\, |\xi-\xi'|$, we obtain

$$|h_{t,x,\xi} - h_{t',x',\xi'}|_{L_2(P)} \le \sigma_{h_1}\Big(A_1|t-t'| + A_2|x-x'| + A_3\frac{|\xi-\xi'|}{h_1}\Big),$$

for constants $A_1, A_2, A_3 > 0$ depending only on K, d, η, Δ, and $\mathrm{diam}(B)$.

For the unbounded x-parameter by exponential decay, let $\mathrm{dist}(x, B) := \inf_{y\in B} |x-y|$. Since $\Delta(t) \le \Delta$ and $y \in B$, we have $\chi(t, x, y) \le e^{-\mathrm{dist}(x,B)^2/(2\Delta)}$, thus $|h_{t,x,\xi}|_{L_2(P)} \le \sigma_{h_1}\, e^{-\mathrm{dist}(x,B)^2/(2\Delta)}$ uniformly in (t, ξ). Given $\varepsilon \in (0, 1]$, define $r_\varepsilon := \sqrt{2\Delta \log(4/\varepsilon)}$. If $\mathrm{dist}(x, B) \ge r_\varepsilon$, then $|h_{t,x,\xi}|_{L_2(P)} \le \frac{\varepsilon}{4}\sigma_{h_1}$. Therefore, up to an additional ball of radius $\varepsilon\sigma_{h_1}/4$ around the zero function, it suffices to cover the parameter set with $x \in B^+_{r_\varepsilon} := \{x \in \mathbb{R}^d : \mathrm{dist}(x, B) \le r_\varepsilon\}$.

Choose a grid in $t \in [s, u-\eta]$ with mesh $\varepsilon/(4A_1)$, a grid in $x \in B^+_{r_\varepsilon}$ with mesh $\varepsilon/(4A_2)$, and a grid in $\xi \in B_\rho$ with mesh $\varepsilon h_1/(4A_3)$.

By the Lipschitz estimate, this induces an $L_2(P)\, \varepsilon\sigma_{h_1}$-net of $\mathcal{H}_{h_1}$. Equivalently, the covering number is bounded by the product of coverings in the t-, x-, and ξ-coordinates: $N(\varepsilon\sigma_{h_1}, \mathcal{H}_{h_1}, L_2(P)) \le N_t(\varepsilon)\, N_x(\varepsilon)\, N_\xi(\varepsilon)$, with $\log N_t(\varepsilon) \lesssim \log(C/\varepsilon)$, $\log N_x(\varepsilon) \lesssim d\log(C/\varepsilon)$, $\log N_\xi(\varepsilon) \lesssim d\log(C/(\varepsilon h_1))$. Since $r_\varepsilon \lesssim 1 + \sqrt{\log(1/\varepsilon)}$, there exists $C_K > 0$, depending only on K, d, η, Δ, and $\mathrm{diam}(B)$, such that $\log N(\varepsilon\sigma_{h_1}, \mathcal{H}_{h_1}, L_2(P)) \le d\log\Big(\frac{C_K}{\varepsilon h_1}\Big) + C\log\Big(\frac{C_K}{\varepsilon}\Big)$, for a universal constant $C > 0$. In particular, after enlarging C_K if necessary, $\log N(\varepsilon\sigma_{h_1}, \mathcal{H}_{h_1}, L_2(P)) \le C\, d\log\Big(\frac{C_K}{\varepsilon h_1}\Big)$.

We now apply standard symmetrization and Dudley's entropy integral [33], together with Bousquet's inequality [4], using envelope U_{h_1} and variance proxy $\sigma^2_{h_1}$; see also [3]. By symmetrization and Dudley's entropy integral, there is a universal $c_0 > 0$ such that

$$\mathbb{E} \sup_{h\in\mathcal{H}_{h_1}} |\mathbb{G}_M h| \le \frac{2c_0}{\sqrt{M}} \int_0^{\sigma_{h_1}} \sqrt{\log N(\tau, \mathcal{H}_{h_1}, L_2(P))}\, d\tau.$$

Using the entropy bound above and the change of variables $\tau = \varepsilon\sigma_{h_1}$,

$$\mathbb{E} \sup_{h\in\mathcal{H}_{h_1}} |\mathbb{G}_M h| \le C\, C_F\, \sqrt{C_f}\, |K|_2 \sqrt{\frac{d\big(1 + \log(C_K/h_1)\big)}{M h_1^d}},$$

for a universal constant $C > 0$.

Finally, Bousquet's inequality for the class $\mathcal{H}_{h_1}$ with envelope U_{h_1} and variance proxy $\sigma^2_{h_1}$ yields that, with probability at least $1 - \delta$,

$$\sup_{h\in\mathcal{H}_{h_1}} |\mathbb{G}_M h| \le C_1\, C_F\, \sqrt{C_f}\, |K|_2 \sqrt{\frac{d\log\big(\frac{C_K}{h_1}\big) + \log\big(\frac{2}{\delta}\big)}{M h_1^d}} + C_2\, \frac{C_F\, \mu_\infty(K)}{h_1^d}\, \frac{d\log\big(\frac{C_K}{h_1}\big) + \log\big(\frac{2}{\delta}\big)}{M}.$$

Combining this stochastic bound with the bias bound proved above proves the lemma. □

A.3 Uniform weighted bound for $\widehat{g}_2$

Lemma A.3 (Uniform weighted bound for $\widehat{g}_2$)**.** *Recall C_F and C_Y from Remark 3.3. Then, with ℓ_β as in Assumption 3.1, there exist constants $C_1, C_2, C_K > 0$, where C_K depends only on K, d, η, Δ, and $\mathrm{diam}(B)$, such that for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\begin{aligned}\sup_{t\in[s,u-\eta],\, x\in\mathbb{R}^d,\, \xi\in B_\rho} \big|\widehat{g}_2(t, x; \xi) - g_2(t, x; \xi)\big| \le{}& \frac{C_B C_Y C_F L_p}{\ell_\beta!}\, \mu_\beta(K)\, h_2^\beta \\ &+ C_1\, C_Y C_F\, \sqrt{C_f}\, |K|_2 \sqrt{\frac{d\log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2}{\delta}\big)}{M h_2^d}} \\ &+ C_2\, \frac{C_Y C_F\, \mu_\infty(K)}{h_2^d}\, \frac{d\log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2}{\delta}\big)}{M}.\end{aligned}$$

Proof. For $r \in 1 : d$, define $\widehat{g}_{2,r}(t, x; \xi) := \frac{1}{M}\sum_{m=1}^M X^{(m)}_{u,r}\; F(t, \xi, x, X^{(m)}_u)\; K_{h_2}(X^{(m)}_s - \xi)$, $g_{2,r}(t, x; \xi) := \int_B y_r\, F(t, \xi, x, y)\, p(\xi, y)\, dy$. Then $\widehat{g}_2 = (\widehat{g}_{2,1}, \ldots, \widehat{g}_{2,d})$ and $g_2 = (g_{2,1}, \ldots, g_{2,d})$.

Fix $r \in \{1, \dots, d\}$, $t \in [s, u-\eta]$, $x \in \mathbb{R}^d$, and $\xi \in B_\rho$. With the change of variables $u = (z-\xi)/h_2$,

$$\mathbb{E}\hat{g}_{2,r}(t,x;\xi) = \int_{\mathbb{R}^d} K(u) \Big(\int_B y_r F(t,\xi,x,y)\, \bar{p}(\xi + h_2 u, y)\, dy \Big)\, du,$$

where $\bar{p}(\cdot, y)$ is the extension from Assumption 3.2(i). Define $H_{t,x,\xi,r}(z) := \int_B y_r F(t,\xi,x,y)\, \bar{p}(z,y)\, dy$. Since $|y_r| \le C_Y$ on B and $\bar{p}(\cdot, y) \in \Sigma(\beta, L_p)$ uniformly in $y \in B$, the map $z \mapsto H_{t,x,\xi,r}(z)$ belongs to $\Sigma(\beta, C_B C_Y C_F L_p)$ uniformly over $(t,x,\xi) \in [s, u-\eta] \times \mathbb{R}^d \times B_\rho$. With $\ell_\beta = \max\{m \in \mathbb{N}_0 : m < \beta\}$, Taylor expansion to order ℓ_β, together with $\int_{\mathbb{R}^d} u^\alpha K(u)\, du = 0$, $1 \le |\alpha| \le \ell_\beta$, yields

$$|\mathbb{E}\hat{g}_{2,r}(t,x;\xi) - H_{t,x,\xi,r}(\xi)| \le \frac{C_B C_Y C_F L_p}{\ell_\beta!} \mu_\beta(K) h_2^\beta.$$

Since $\xi \in B_\rho \subset B$ and $\bar{p}(\xi, y) = p(\xi, y)$ on B, we have $H_{t,x,\xi,r}(\xi) = g_{2,r}(t,x;\xi)$. Thus

$$\sup_{t \in [s,u-\eta],\, x \in \mathbb{R}^d,\, \xi \in B_\rho} |\mathbb{E}\hat{g}_{2,r}(t,x;\xi) - g_{2,r}(t,x;\xi)| \le \frac{C_B C_Y C_F L_p}{\ell_\beta!} \mu_\beta(K) h_2^\beta.$$

Now consider the scalar weighted class

$$\mathcal{H}_{h_2}^{(r)} := \Big\{ h_{t,x,\xi}^{(r)}(z,y) := y_r F(t,\xi,x,y) K_{h_2}(z-\xi) : (t,x,\xi) \in [s, u-\eta] \times \mathbb{R}^d \times B_\rho \Big\}.$$

The proof of Lemma A.2 applies verbatim to $\mathcal{H}_{h_2}^{(r)}$, the only change being the extra scalar factor y_r, which is uniformly bounded by C_Y on B. Therefore, we have $U_{h_2}^{(r)} \le C_Y C_F\, \mu_\infty(K)/h_2^d$, $\sigma_{h_2}^{(r)} \le C_Y C_F \sqrt{C_f}\, |K|_2 / h_2^{d/2}$, and the entropy bound has the same form as in Lemma A.2, namely

$$\log N\Big(\varepsilon\, \sigma_{h_2}^{(r)}, \mathcal{H}_{h_2}^{(r)}, L_2(P) \Big) \le C\, d \log\Big(\frac{C_K}{\varepsilon h_2} \Big).$$

Hence, by the same symmetrization, Dudley-integral, and Bousquet-inequality argument as in Lemma A.2, for any $\delta \in (0,1)$, with probability at least $1 - \delta/d$,

$$\sup_{t \in [s,u-\eta],\, x \in \mathbb{R}^d,\, \xi \in B_\rho} \big| \widehat{g}_{2,r}(t,x;\xi) - \mathbb{E}\widehat{g}_{2,r}(t,x;\xi) \big| \le C_1'\, C_Y C_F\, \sqrt{C_f}\, |K|_2 \sqrt{\frac{d \log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2d}{\delta}\big)}{M h_2^d}} + C_2'\, \frac{C_Y C_F\, \mu_\infty(K)}{h_2^d}\, \frac{d \log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2d}{\delta}\big)}{M},$$

where $C_1', C_2' > 0$ depend only on universal numerical constants.

Combining the last two bounds, we get that, with probability at least $1 - \delta/d$,

$$\sup_{t \in [s,u-\eta],\, x \in \mathbb{R}^d,\, \xi \in B_\rho} \big| \widehat{g}_{2,r}(t,x;\xi) - g_{2,r}(t,x;\xi) \big| \le \frac{C_B C_Y C_F L_p}{\ell_\beta!}\, \mu_\beta(K)\, h_2^\beta + C_1'\, C_Y C_F\, \sqrt{C_f}\, |K|_2 \sqrt{\frac{d \log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2d}{\delta}\big)}{M h_2^d}} + C_2'\, \frac{C_Y C_F\, \mu_\infty(K)}{h_2^d}\, \frac{d \log\big(\frac{C_K}{h_2}\big) + \log\big(\frac{2d}{\delta}\big)}{M}. \tag{4}$$

By a union bound over $r = 1, \dots, d$, with probability at least $1 - \delta$, the bound (4) holds simultaneously for all coordinates. On that event,

$$\sup_{t,x,\xi} \big| \widehat{g}_2(t,x;\xi) - g_2(t,x;\xi) \big| \le \sqrt{d}\, \max_{1 \le r \le d} \sup_{t,x,\xi} \big| \widehat{g}_{2,r}(t,x;\xi) - g_{2,r}(t,x;\xi) \big|.$$

Since the paper allows constants to depend on d, the factors $\sqrt{d}$ and $\log d$ can be absorbed into C_1, C_2 and C_K by enlarging them if necessary. Thus, with probability at least $1 - \delta$, the displayed bound in the statement of the lemma holds, which proves the result. $\square$

B Deterministic Ratio Stability

Lemma B.1 (Deterministic denominator stability)**.** *Under Assumption 3.2(ii), fix* $(t,x,\xi) \in Q_{R,\eta,\rho}$*. On the event* $\Big\{\hat{f}_1(\xi) \geq \frac{1}{2} f_{\min}\Big\}$*, we have*

$$\big|\hat{D}(t,x;\xi) - D^*(t,x;\xi)\big| \leq \frac{2}{f_{\min}^2}\Big(f(\xi)\,\big|\hat{g}_1(t,x;\xi) - g_1(t,x;\xi)\big| + |g_1(t,x;\xi)|\,\big|\hat{f}_1(\xi) - f(\xi)\big|\Big).$$

Consequently, with $D_{\min}$ *defined in Assumption 3.2, on the event* $\Big\{\inf_{\xi\in B_\rho} \hat{f}_1(\xi) \geq \frac{1}{2} f_{\min}\Big\}$*, if*

$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} \frac{2}{f_{\min}^2}\Big(f(\xi)\,\big|\hat{g}_1(t,x;\xi) - g_1(t,x;\xi)\big| + |g_1(t,x;\xi)|\,\big|\hat{f}_1(\xi) - f(\xi)\big|\Big) \leq \frac{D_{\min}}{2},$$

then

$$\inf_{(t,x,\xi)\in Q_{R,\eta,\rho}} \hat{D}(t,x;\xi) \geq \frac{D_{\min}}{2}.$$

Proof. Fix $(t,x,\xi) \in Q_{R,\eta,\rho}$, we have

$$\hat{D} - D^* = \frac{\hat{g}_1}{\hat{f}_1} - \frac{g_1}{f(\xi)} = \frac{(\hat{g}_1 - g_1)\,f(\xi) - g_1\,(\hat{f}_1 - f(\xi))}{\hat{f}_1\,f(\xi)}.$$

On the event $\{\hat{f}_1(\xi) \geq f_{\min}/2\}$, Assumption 3.2 gives $\hat{f}_1(\xi)\,f(\xi) \geq \frac{f_{\min}^2}{2}$, and therefore

$$\big|\hat{D} - D^*\big| \leq \frac{2}{f_{\min}^2}\Big(f(\xi)\,\big|\hat{g}_1 - g_1\big| + |g_1|\,\big|\hat{f}_1 - f(\xi)\big|\Big).$$

This proves the first claim. Now, assume $\inf_{\xi\in B_\rho} \hat{f}_1(\xi) \geq \frac{f_{\min}}{2}$ and that the displayed supremum is at most $D_{\min}/2$. Then, for every $(t,x,\xi) \in Q_{R,\eta,\rho}$, the first part gives $\big|\hat{D}(t,x;\xi) - D^*(t,x;\xi)\big| \leq \frac{D_{\min}}{2}$. Using Assumption 3.2, $D^*(t,x;\xi) \geq D_{\min}$, hence

$$\hat{D}(t,x;\xi) \geq D^*(t,x;\xi) - \big|\hat{D}(t,x;\xi) - D^*(t,x;\xi)\big| \geq D_{\min} - \frac{D_{\min}}{2} = \frac{D_{\min}}{2}.$$

Taking the infimum over $(t,x,\xi) \in Q_{R,\eta,\rho}$ proves the claim. □

Proposition B.2 (Deterministic ratio-stability transfer)**.** *Fix* $(t,x,\xi) \in Q_{R,\eta,\rho}$*. On the event*

$$\mathcal{E}(t,x,\xi) := \Big\{\hat{D}(t,x;\xi) \geq \tfrac{1}{2} D_{\min},\ \hat{f}_1(\xi) \geq \tfrac{1}{2} f_{\min},\ \hat{f}_2(\xi) \geq \tfrac{1}{2} f_{\min}\Big\},$$

we have

$$|\hat{a} - a^*| \leq \frac{4}{\Delta(t)\,D_{\min}}\left(\frac{f(\xi)\,|\hat{g}_2 - g_2| + |g_2|\,|\hat{f}_2 - f(\xi)|}{f_{\min}^2} + |Q^*|\,\frac{f(\xi)\,|\hat{g}_1 - g_1| + |g_1|\,|\hat{f}_1 - f(\xi)|}{f_{\min}^2}\right).$$

Proof. Fix $(t,x,\xi) \in Q_{R,\eta,\rho}$ and write $\Delta(t) := u - t$. Recall $|\hat{a} - a^*| = \frac{1}{\Delta(t)}\,|\hat{Q} - Q^*|$. Using

$$\hat{Q} - Q^* = \frac{\hat{N}}{\hat{D}} - \frac{N^*}{D^*} = \frac{\hat{N} - N^*}{\hat{D}} + N^*\left(\frac{1}{\hat{D}} - \frac{1}{D^*}\right) = \frac{\hat{N} - N^*}{\hat{D}} - \frac{Q^*}{\hat{D}}(\hat{D} - D^*),$$

we obtain on the event $\{\hat{D} \geq D_{\min}/2\}$,

$$|\hat{Q} - Q^*| \leq \frac{2}{D_{\min}}\Big(|\hat{N} - N^*| + |Q^*|\,|\hat{D} - D^*|\Big).$$

Taking absolute values and using the event $\mathcal{E}(t,x,\xi)$ (so $\hat{D}(t,x;\xi) \geq D_{\min}/2$),

$$|\hat{Q} - Q^*| \ \leq\ \frac{2}{D_{\min}}\Big(|\hat{N} - N^*| + |Q^*|\,|\hat{D} - D^*|\Big). \tag{5}$$

For $N^*(t,x;\xi) = g_2(t,x;\xi)/f(\xi)$, $\hat{N} - N^* = \frac{\hat{g}_2}{\hat{f}_2} - \frac{g_2}{f(\xi)} = \frac{(\hat{g}_2 - g_2)\, f(\xi) - g_2\, (\hat{f}_2 - f(\xi))}{\hat{f}_2\, f(\xi)}$. On $\mathcal{E}(t,x,\xi)$, $\hat{f}_2(\xi) \geq f_{\min}/2$ and Assumption 3.2 gives $f(\xi) \geq f_{\min}$, so $\hat{f}_2(\xi)\, f(\xi) \geq \frac{f_{\min}^2}{2}$, and therefore

$$|\hat{N} - N^*| \;\leq\; \frac{2}{f_{\min}^2}\Big(f(\xi)\,|\hat{g}_2 - g_2| + |g_2|\,|\hat{f}_2 - f(\xi)|\Big). \tag{6}$$

Similarly, for $D^* = \frac{g_1}{f(\xi)}$, $\hat{D} - D^* = \frac{\hat{g}_1}{\hat{f}_1} - \frac{g_1}{f(\xi)} = \frac{(\hat{g}_1 - g_1)\, f(\xi) - g_1\, (\hat{f}_1 - f(\xi))}{\hat{f}_1\, f(\xi)}$, and on $\mathcal{E}(t,x,\xi)$, $\hat{f}_1(\xi)\, f(\xi) \geq \frac{f_{\min}^2}{2}$, hence

$$|\hat{D} - D^*| \;\leq\; \frac{2}{f_{\min}^2}\Big(f(\xi)\,|\hat{g}_1 - g_1| + |g_1|\,|\hat{f}_1 - f(\xi)|\Big). \tag{7}$$

Plug (6)–(7) into (5) to get

$$|\hat{Q} - Q^*| \;\leq\; \frac{4}{D_{\min}}\left(\frac{f(\xi)\,|\hat{g}_2 - g_2| + |g_2|\,|\hat{f}_2 - f(\xi)|}{f_{\min}^2} \;+\; |Q^*|\,\frac{f(\xi)\,|\hat{g}_1 - g_1| + |g_1|\,|\hat{f}_1 - f(\xi)|}{f_{\min}^2}\right).$$

Finally, multiply by $1/\Delta(t)$ to pass from Q to the drift:

$$|\hat{a} - a^*| \;\leq\; \frac{4}{\Delta(t)\, D_{\min}}\left(\frac{f(\xi)\,|\hat{g}_2 - g_2| + |g_2|\,|\hat{f}_2 - f(\xi)|}{f_{\min}^2} \;+\; |Q^*|\,\frac{f(\xi)\,|\hat{g}_1 - g_1| + |g_1|\,|\hat{f}_1 - f(\xi)|}{f_{\min}^2}\right).$$

This is the claimed bound. □

C Uniform Kernel Deviations

Let $C_K > 0$ denote the entropy constant appearing in Lemmas A.1 to A.3, and define

$$\widetilde{r}_j(\delta) := h_j^{\beta} + \sqrt{\frac{d\log(\frac{C_K}{h_j}) + \log(\frac{8}{\delta})}{M h_j^d}}, \qquad j \in \{1,2\}.$$

We call a bandwidth pair (h_1, h_2) δ-*admissible* if

$$\widetilde{r}_1(\delta) + \widetilde{r}_2(\delta) \leq c_0.$$

Theorem C.1 (Uniform finite-sample bound)**.** *Assume Assumptions 3.1 and 3.2. Then there exist constants*

$$c_0 > 0, \qquad C = C(\beta, L_K, \mu_\beta(K), \mu_\infty(K), |K|_2, C_f, d, L_p, f_{\min}, D_{\min}, R, \eta, \rho, \Delta, \operatorname{diam}(B))$$

such that, for every δ-admissible bandwidth pair (h_1, h_2), with probability at least $1 - \delta$,

$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |\hat{a} - a^*| \leq \frac{C}{\eta}\left(h_1^{\beta} + h_2^{\beta} + \sqrt{\frac{d\log(\frac{C_K}{h_1}) + \log(\frac{2}{\delta})}{M h_1^d}} + \sqrt{\frac{d\log(\frac{C_K}{h_2}) + \log(\frac{2}{\delta})}{M h_2^d}}\right).$$

Remark C.2. Since $\Delta(t) \geq \eta$ on $Q_{R,\eta,\rho}$, Theorem C.1 implies Theorem 4.2 for every δ-admissible bandwidth pair (h_1, h_2).

Proof. By bounded support and continuity, the quantities $f(\xi),\ g_1(t,x;\xi),\ g_2(t,x;\xi),\ Q^*(t,x;\xi)$ are uniformly bounded on $Q_{R,\eta,\rho}$. Hence there exist deterministic constants $B_f,\ B_{g_1},\ B_{g_2},\ B_Q > 0$ such that

$$\sup_{\xi\in B_\rho} f(\xi) \leq B_f, \qquad \sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |g_1(t,x;\xi)| \leq B_{g_1},$$
$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |g_2(t,x;\xi)| \leq B_{g_2}, \qquad \sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |Q^*(t,x;\xi)| \leq B_Q.$$

By Lemmas A.1 to A.3, there exists a constant $A_1,\ A_2,\ A_3,\ A_4 > 0$ and events of probability at least $1 - \delta/4$ for each such that the following inequalities hold,

$$\sup_{t\in[s,u-\eta],\ x\in\mathbb{R}^d,\ \xi\in B_\rho} |\hat{g}_1 - g_1| \leq A_1\,\widetilde{r}_1(\delta), \qquad \sup_{t\in[s,u-\eta],\ x\in\mathbb{R}^d,\ \xi\in B_\rho} |\hat{g}_2 - g_2| \leq A_2\,\widetilde{r}_2(\delta), \tag{8}$$
$$\sup_{\xi\in B_\rho} |\hat{f}_1 - f| \leq A_3\,\widetilde{r}_1(\delta), \qquad \sup_{\xi\in B_\rho} |\hat{f}_2 - f| \leq A_4\,\widetilde{r}_2(\delta).$$

Since (h_1, h_2) is δ-admissible and c_0 is chosen small, the quantities $\widetilde{\Lambda}(h_j, \delta)/(Mh_j^d),\ j \in \{1, 2\}$, are bounded by 1. Therefore

$$\frac{\widetilde{\Lambda}(h_j, \delta)}{Mh_j^d} \le \sqrt{\frac{\widetilde{\Lambda}(h_j, \delta)}{Mh_j^d}}, \qquad j \in \{1, 2\},$$

so the lower-order linear concentration remainders in Lemmas A.1 to A.3 are absorbed into the constants multiplying $\widetilde{r}_j(\delta)$.

Let Ω denote the intersection of the four events in (8). By a union bound, $\Pr(\Omega) \ge 1 - \delta$.

We now verify the denominator-floor event on Ω. First, on Ω,

$$\inf_{\xi \in B_\rho} \hat{f}_1(\xi) \ge f_{\min} - A_3\, \widetilde{r}_1(\delta), \qquad \inf_{\xi \in B_\rho} \hat{f}_2(\xi) \ge f_{\min} - A_4\, \widetilde{r}_2(\delta).$$

Next, by Lemma B.1, on the event $\Big\{ \inf_{\xi \in B_\rho} \hat{f}_1(\xi) \ge \frac{1}{2} f_{\min} \Big\}$, we have

$$\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |\hat{D} - D^*| \le \frac{2}{f_{\min}^{\,2}} \Big(B_f \sup_{t,x,\xi} |\hat{g}_1 - g_1| + B_{g_1} \sup_{\xi} |\hat{f}_1 - f| \Big),$$

and therefore, on Ω, $\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |\hat{D} - D^*| \le \frac{2}{f_{\min}^{\,2}} \Big(B_f A_1 + B_{g_1} A_3 \Big)\, \widetilde{r}_1(\delta)$.

Let $c_0 := \min\Big\{ \frac{f_{\min}}{2A_3},\ \frac{f_{\min}}{2A_4},\ \frac{D_{\min} f_{\min}^{\,2}}{4(B_f A_1 + B_{g_1} A_3)} \Big\}$. If (h_1, h_2) is δ-admissible, then $\widetilde{r}_1(\delta) + \widetilde{r}_2(\delta) \le c_0$, hence, in particular, $A_3 \widetilde{r}_1(\delta) \le \frac{f_{\min}}{2},\ A_4 \widetilde{r}_2(\delta) \le \frac{f_{\min}}{2}$, and $\frac{2}{f_{\min}^{\,2}} \Big(B_f A_1 + B_{g_1} A_3 \Big)\, \widetilde{r}_1(\delta) \le \frac{D_{\min}}{2}$. Therefore, on Ω, $\inf_{\xi \in B_\rho} \hat{f}_1(\xi) \ge \frac{f_{\min}}{2},\ \inf_{\xi \in B_\rho} \hat{f}_2(\xi) \ge \frac{f_{\min}}{2},\ \inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} \hat{D}(t, x; \xi) \ge \frac{D_{\min}}{2}$. In other words, on Ω the event

$$\mathcal{E} := \Big\{ \inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} \hat{D}(t, x; \xi) \ge \tfrac{1}{2} D_{\min},\ \inf_{\xi \in B_\rho} \hat{f}_1(\xi) \ge \tfrac{1}{2} f_{\min},\ \inf_{\xi \in B_\rho} \hat{f}_2(\xi) \ge \tfrac{1}{2} f_{\min} \Big\}$$

holds.

Now fix $(t, x, \xi) \in Q_{R,\eta,\rho}$. On $\mathcal{E}$, Proposition B.2 gives

$$|\hat{a} - a^*| \le \frac{4}{\Delta(t)\, D_{\min}} \left(\frac{f(\xi)\, |\hat{g}_2 - g_2| + |g_2|\, |\hat{f}_2 - f|}{f_{\min}^{\,2}} + |Q^*|\, \frac{f(\xi)\, |\hat{g}_1 - g_1| + |g_1|\, |\hat{f}_1 - f|}{f_{\min}^{\,2}} \right). \tag{9}$$

Since $\Omega \subset \mathcal{E}$, we may insert the uniform bounds (8) into (9). Using the deterministic bounds $f(\xi) \le B_f,\ |g_1| \le B_{g_1},\ |g_2| \le B_{g_2},\ |Q^*| \le B_Q$, we obtain on Ω

$$|\hat{a} - a^*| \le \frac{C}{\eta} \big(\widetilde{r}_1(\delta) + \widetilde{r}_2(\delta) \big),$$

for a constant $C > 0$ depending only on

$$(\beta, L_K, \mu_\beta(K), \mu_\infty(K), |K|_2, C_f, d, L_p, f_{\min}, D_{\min}, R, \eta, \Delta, \operatorname{diam}(B)).$$

Expanding the definition of $\widetilde{r}_1(\delta)$ and $\widetilde{r}_2(\delta)$, and absorbing the harmless difference between $\log(8/\delta)$ and $\log(2/\delta)$ into the constant C, gives the claimed bound. □

Corollary C.3 (Oracle bandwidth choice)**.** *Set $h_1 = h_2 = h$ and let*

$$L := d \log(C_K M) + \log(2/\delta), \qquad h^\star := \left(\frac{L}{M} \right)^{1/(2\beta + d)}.$$

Then for each fixed $\delta \in (0, 1)$ there exists $M_0(\delta) \ge 1$, depending on δ and the model constants, such that for all $M \ge M_0(\delta)$, the pair $(h^\star, h^\star)$ is δ-admissible and, with probability at least $1 - \delta$,

$$\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |\hat{a}(t, x; \xi) - a^*(t, x; \xi)| \le \frac{C}{\eta} \left(\frac{L}{M} \right)^{\beta/(2\beta + d)}.$$

Moreover, for $\delta = M^{-\gamma}$ with fixed $\gamma > 0$,

$$\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |\hat{a}(t, x; \xi) - a^*(t, x; \xi)| = O_{\mathbb{P}}\left(\left(\frac{\log M}{M} \right)^{\beta/(2\beta + d)} \right).$$

Proof. Since $(h^\star)^\beta = (L/M)^{\beta/(2\beta+d)}$, we have

$$\log\Big(\frac{C_K}{h^\star}\Big) = \log C_K + \frac{1}{2\beta+d}\log\Big(\frac{M}{L}\Big) \le \log(C_K M).$$

Moreover,

$$d\log\Big(\frac{C_K}{h^\star}\Big) + \log\Big(\frac{8}{\delta}\Big) \le L + \log 4.$$

For all sufficiently large M, $L \ge \log 4$, and hence the last display is at most $2L$. Therefore,

$$\sqrt{\frac{d\log(C_K/h^\star) + \log(8/\delta)}{M(h^\star)^d}} \le \sqrt{2}\left(\frac{L}{M}\right)^{\beta/(2\beta+d)}.$$

Consequently, for $j \in \{1,2\}$,

$$\widetilde{r}_j(\delta) \le (1+\sqrt{2})\left(\frac{L}{M}\right)^{\beta/(2\beta+d)}.$$

Choose $M_0(\delta)$ sufficiently large that, for every $M \ge M_0(\delta)$,

$$2(1+\sqrt{2})\left(\frac{L}{M}\right)^{\beta/(2\beta+d)} \le c_0, \qquad \left(\frac{L}{M}\right)^{1/(2\beta+d)} \le \frac{\rho}{R_K}, \qquad L \ge \log 4.$$

Such a finite $M_0(\delta)$ exists because $L/M \to 0$. The first condition makes $(h^\star, h^\star)$ δ-admissible, while the second gives $h^\star \le \rho/R_K$. Hence Theorem C.1 applies and, with probability at least $1-\delta$,

$$\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |\hat{a}(t,x;\xi) - a^*(t,x;\xi)| \le \frac{C}{\eta}\left(\frac{L}{M}\right)^{\beta/(2\beta+d)}.$$

If $\delta = M^{-\gamma}$, then $L = d\log(C_K M) + \gamma\log M \le (d+\gamma)\log M + O(1)$, and therefore

$$\left(\frac{L}{M}\right)^{\beta/(2\beta+d)} = O\left(\left(\frac{\log M}{M}\right)^{\beta/(2\beta+d)}\right).$$

This proves the final claim. □

D Unequal-bandwidth pointwise inference

In this Appendix we prove the pointwise CLT, consistency of a plug-in covariance estimator, the studentized limit, and pointwise Wald validity for possibly different bandwidths h_1 and h_2. The main-text Theorem 4.7 is recovered by the diagonal specialization $h_1 = h_2 = h$; see Remark D.2.

Fix $(t,x,\xi) \in [s, u-\eta] \times B_R \times \mathrm{int}(B)$. Write $U(y) := y\,F_{t,x,\xi}(y) \in \mathbb{R}^d$, $V(y) := F_{t,x,\xi}(y) \in \mathbb{R}$. Then

$$N^*(t,x;\xi) = \mathbb{E}\big[U(X_u) \mid X_s = \xi\big], \qquad D^*(t,x;\xi) = \mathbb{E}\big[V(X_u) \mid X_s = \xi\big],$$
$$Q^*(t,x;\xi) = \frac{N^*(t,x;\xi)}{D^*(t,x;\xi)}, \qquad a^*(t,x;\xi) = \frac{Q^*(t,x;\xi) - x}{\Delta_t}.$$

Also let $R(K) := \int_{\mathbb{R}^d} K(u)^2\,du$, $R_\lambda(K) := \int_{\mathbb{R}^d} K(u)K(\lambda u)\,du$, $\lambda > 0$. Writing $U := U(X_u)$ and $V := V(X_u)$ for brevity, define

$$n_U := \mathbb{E}[U \mid X_s = \xi], \qquad n_V := \mathbb{E}[V \mid X_s = \xi], \qquad n_{UV} := \mathbb{E}[UV \mid X_s = \xi],$$
$$m_{UU} := \mathbb{E}[UU^\top \mid X_s = \xi], \qquad m_{V^2} := \mathbb{E}[V^2 \mid X_s = \xi].$$

Then the block covariance matrix is

$$\Omega_\lambda(t,x;\xi) := f(\xi)\begin{bmatrix} \lambda^{-d}R(K)\,m_{UU} & R_\lambda(K)\,n_{UV} & R_\lambda(K)\,n_U & \lambda^{-d}R(K)\,n_U \\ R_\lambda(K)\,n_{UV}^\top & R(K)\,m_{V^2} & R(K)\,n_V & R_\lambda(K)\,n_V \\ R_\lambda(K)\,n_U^\top & R(K)\,n_V & R(K) & R_\lambda(K) \\ \lambda^{-d}R(K)\,n_U^\top & R_\lambda(K)\,n_V & R_\lambda(K) & \lambda^{-d}R(K) \end{bmatrix}.$$

Let $\Gamma_t(a,b,c,e) := \frac{1}{\Delta_t}\left(\frac{ac}{be} - x\right)$, $(a,b,c,e) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, where ac denotes scalar multiplication of the vector a by the scalar c. Denote by $J_\Gamma(t,x;\xi) := D\Gamma_t\big(g_2(t,x;\xi),\ g_1(t,x;\xi),\ f(\xi),\ f(\xi)\big)$ the Jacobian of Γ_t at the population point. Explicitly, since $g_1(t,x;\xi) = f(\xi)D^*(t,x;\xi)$, set

$$A := \big(f(\xi)D^*(t,x;\xi)\big)^{-1} = g_1(t,x;\xi)^{-1}, \qquad B := f(\xi)^{-1}.$$

Then

$$J_\Gamma(t,x;\xi) = \frac{1}{\Delta_t}\Big[AI_d,\ -AQ^*(t,x;\xi),\ BQ^*(t,x;\xi),\ -BQ^*(t,x;\xi)\Big].$$

This is viewed as a $d \times (d+3)$ block row. Finally, define $\Sigma_\lambda(t,x;\xi) := J_\Gamma(t,x;\xi)\ \Omega_\lambda(t,x;\xi)\ J_\Gamma(t,x;\xi)^\top$.

For any bounded scalar-, vector-, or matrix-valued function ϕ, set

$$\widehat{\mathbb{E}}_{h_1}[\phi \mid \xi] := \frac{\sum_{m=1}^M \phi(X_u^{(m)}) K_{h_1}(X_s^{(m)} - \xi)}{\sum_{m=1}^M K_{h_1}(X_s^{(m)} - \xi)}.$$

Write $\lambda_M := h_2/h_1$ and set $\hat{n}_U := \widehat{\mathbb{E}}_{h_1}[U \mid \xi]$, $\hat{n}_V := \widehat{\mathbb{E}}_{h_1}[V \mid \xi]$, $\hat{n}_{UV} := \widehat{\mathbb{E}}_{h_1}[UV \mid \xi]$, $\hat{m}_{UU} := \widehat{\mathbb{E}}_{h_1}[UU^\top \mid \xi]$, and $\hat{m}_{V^2} := \widehat{\mathbb{E}}_{h_1}[V^2 \mid \xi]$. Let $\hat{\Omega}_{h_1,h_2}(t,x;\xi)$ be obtained from $\Omega_{\lambda_M}(t,x;\xi)$ by replacing $f(\xi), n_U, n_V, n_{UV}, m_{UU}, m_{V^2}$ with $\hat{f}_1(\xi), \hat{n}_U, \hat{n}_V, \hat{n}_{UV}, \hat{m}_{UU}, \hat{m}_{V^2}$, respectively.

Let $\hat{T} := (\hat{g}_2^\top, \hat{g}_1, \hat{f}_1, \hat{f}_2)^\top$ and $\hat{J}_\Gamma(t,x;\xi) := D\Gamma_t(\hat{T})$. On the event $\{\hat{g}_1\hat{f}_1\hat{f}_2 \neq 0\}$, define

$$\hat{\Sigma}_{h_1,h_2}(t,x;\xi) := \hat{J}_\Gamma(t,x;\xi)\ \hat{\Omega}_{h_1,h_2}(t,x;\xi)\ \hat{J}_\Gamma(t,x;\xi)^\top. \tag{10}$$

All plug-in quantities above may be defined arbitrarily on the complementary event, whose probability tends to zero.

Theorem D.1 (Unequal-bandwidth pointwise inference)**.** *Assume Assumptions 3.1 and 3.2. Let $h_1 \to 0$ and $h_2 \to 0$ satisfy $Mh_1^d \to \infty$, $Mh_2^d \to \infty$, $h_1^\beta\sqrt{Mh_1^d} \to 0$, $h_2^\beta\sqrt{Mh_1^d} \to 0$, and $h_2/h_1 \to \lambda \in (0,\infty)$. Then, for every fixed $(t,x,\xi) \in [s, u-\eta] \times B_R \times \mathrm{int}(B)$,*

$$\sqrt{Mh_1^d}\big(\hat{a}(t,x;\xi) - a^*(t,x;\xi)\big) \Rightarrow \mathcal{N}\big(0, \Sigma_\lambda(t,x;\xi)\big),$$
$$|\hat{\Sigma}_{h_1,h_2}(t,x;\xi) - \Sigma_\lambda(t,x;\xi)|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0.$$

Consequently, for every fixed $v \in \mathbb{R}^d$ such that $v^\top\Sigma_\lambda(t,x;\xi)v > 0$, we have $v^\top\hat{\Sigma}_{h_1,h_2}(t,x;\xi)v > 0$ with probability tending to one and

$$\frac{\sqrt{Mh_1^d}\, v^\top(\hat{a}(t,x;\xi) - a^*(t,x;\xi))}{\sqrt{v^\top\hat{\Sigma}_{h_1,h_2}(t,x;\xi)v}} \Rightarrow \mathcal{N}(0,1).$$

Thus $v^\top\hat{a}(t,x;\xi) \pm z_{1-\alpha/2}\sqrt{v^\top\hat{\Sigma}_{h_1,h_2}(t,x;\xi)v/(Mh_1^d)}$ is an asymptotically valid pointwise $(1-\alpha)$ Wald interval for $v^\top a^*(t,x;\xi)$.

Remark D.2 (Diagonal specialization). If $h_1 = h_2 = h$, then $\lambda_M = \lambda = 1$ and $R_\lambda(K) = R(K)$. A direct simplification gives

$$\Sigma_1(t,x;\xi) = \frac{R(K)\mathrm{Var}(\psi_{t,x,\xi}(X_u) \mid X_s = \xi)}{f(\xi)\Delta(t)^2 D^*(t,x;\xi)^2},$$

where $\psi_{t,x,\xi}(y) := (y - x - \Delta(t)a^*(t,x;\xi))F(t,\xi,x,y)$.

Write $\hat{f}_h := \hat{f}_1 = \hat{f}_2$, $\hat{D}_h := \hat{D}$, $\widehat{\mathbb{E}}_h := \widehat{\mathbb{E}}_{h_1}$, and $\hat{\psi}_h(y) := (y - x - \Delta(t)\hat{a}(t,x;\xi))F(t,\xi,x,y)$, and set $\widehat{\mathrm{Cov}}_h(\phi \mid \xi) := \widehat{\mathbb{E}}_h[\phi\phi^\top \mid \xi] - \widehat{\mathbb{E}}_h[\phi \mid \xi]\widehat{\mathbb{E}}_h[\phi \mid \xi]^\top$. Then the diagonal specialization of (10) is

$$\hat{\Sigma}_h(t,x;\xi) := \hat{\Sigma}_{h,h}(t,x;\xi) = \frac{R(K)\widehat{\mathrm{Cov}}_h(\hat{\psi}_h(X_u) \mid \xi)}{\hat{f}_h(\xi)\Delta(t)^2\hat{D}_h(t,x;\xi)^2}. \tag{11}$$

Hence Theorem D.1 yields all the claims of the main-text Theorem 4.7.

Proof of Theorem D.1. Fix $(t, x, \xi) \in [s, u-\eta] \times B_R \times \mathrm{int}(B)$. Since $\xi \in \mathrm{int}(B)$ and K is compactly supported, for all sufficiently large M we have $\xi + h_j u \in B$ for all $u \in \mathrm{supp}(K)$ and $j \in \{1, 2\}$, so the kernel changes of variables below are valid without boundary effects.

Let $T := \big(g_2(t,x;\xi)^\top, g_1(t,x;\xi), f(\xi), f(\xi)\big)^\top$. By the theorem's assumptions, $\sqrt{Mh_1^d}\, h_1^\beta \to 0$, $\sqrt{Mh_1^d}\, h_2^\beta \to 0$. Since the componentwise bias bounds satisfy

$$|\mathbb{E}\hat{g}_1 - g_1| \lesssim h_1^\beta, \qquad |\mathbb{E}\hat{g}_2 - g_2| \lesssim h_2^\beta, \qquad |\mathbb{E}\hat{f}_j - f| \lesssim h_j^\beta,$$

it follows that $\sqrt{Mh_1^d}\,|\mathbb{E}\hat{T} - T| \to 0$. Since Γ_t is continuously differentiable in a neighborhood of T, it follows that $\sqrt{Mh_1^d}\,\big(\Gamma_t(\mathbb{E}\hat{T}) - \Gamma_t(T)\big) \to 0$.

For each m, define

$$Z_m := \begin{pmatrix} U(X_u^{(m)})\, K_{h_2}(X_s^{(m)} - \xi) \\ V(X_u^{(m)})\, K_{h_1}(X_s^{(m)} - \xi) \\ K_{h_1}(X_s^{(m)} - \xi) \\ K_{h_2}(X_s^{(m)} - \xi) \end{pmatrix} \in \mathbb{R}^{d+3}, \qquad \hat{T} = \frac{1}{M}\sum_{m=1}^{M} Z_m.$$

To identify the covariance limit, write $\widetilde{n}_A(z) := \int_B A(y)\, p(z,y)\, dy$ for bounded measurable $A : B \to \mathbb{R}$, $A : B \to \mathbb{R}^d$, or $A : B \to \mathbb{R}^{d\times d}$. Then $\widetilde{n}_U(\xi) = g_2(t,x;\xi)$, $\widetilde{n}_V(\xi) = g_1(t,x;\xi)$, and

$\widetilde{n}_{UU^\top}(\xi) = f(\xi)\,\mathbb{E}[UU^\top \mid X_s = \xi]$, $\widetilde{n}_{UV}(\xi) = f(\xi)\,\mathbb{E}[UV \mid X_s = \xi]$, $\widetilde{n}_{V^2}(\xi) = f(\xi)\,\mathbb{E}[V^2 \mid X_s = \xi]$.

A standard change of variables gives

$$\begin{aligned} h_1^d\,\mathbb{E}\big[UU^\top K_{h_2}(X_s-\xi)^2\big] &= \frac{h_1^d}{h_2^d}\int_{\mathbb{R}^d} K(u)^2\, \widetilde{n}_{UU^\top}(\xi + h_2 u)\, du \to \lambda^{-d} R(K)\, \widetilde{n}_{UU^\top}(\xi), \\ h_1^d\,\mathbb{E}\big[V^2 K_{h_1}(X_s-\xi)^2\big] &= \int_{\mathbb{R}^d} K(u)^2\, \widetilde{n}_{V^2}(\xi + h_1 u)\, du \to R(K)\, \widetilde{n}_{V^2}(\xi), \\ h_1^d\,\mathbb{E}\big[K_{h_1}(X_s-\xi)^2\big] &\to R(K)\, f(\xi), \qquad h_1^d\,\mathbb{E}\big[K_{h_2}(X_s-\xi)^2\big] \to \lambda^{-d} R(K)\, f(\xi). \end{aligned}$$

Writing $\lambda_M := h_2/h_1$, the mixed terms satisfy

$$h_1^d\,\mathbb{E}\big[UV\, K_{h_2}(X_s-\xi) K_{h_1}(X_s-\xi)\big] = \int_{\mathbb{R}^d} K(u)\, K(\lambda_M u)\, \widetilde{n}_{UV}(\xi + h_2 u)\, du \to R_\lambda(K)\, \widetilde{n}_{UV}(\xi),$$

by dominated convergence. Similarly,

$$\begin{aligned} h_1^d\,\mathbb{E}\big[U\, K_{h_2}(X_s-\xi) K_{h_1}(X_s-\xi)\big] &\to R_\lambda(K)\, \widetilde{n}_U(\xi), \\ h_1^d\,\mathbb{E}\big[V\, K_{h_2}(X_s-\xi) K_{h_1}(X_s-\xi)\big] &\to R_\lambda(K)\, \widetilde{n}_V(\xi), \\ h_1^d\,\mathbb{E}\big[K_{h_2}(X_s-\xi) K_{h_1}(X_s-\xi)\big] &\to R_\lambda(K)\, f(\xi), \\ h_1^d\,\mathbb{E}\big[U\, K_{h_2}(X_s-\xi)^2\big] &\to \lambda^{-d} R(K)\, \widetilde{n}_U(\xi), \\ h_1^d\,\mathbb{E}\big[V\, K_{h_1}(X_s-\xi)^2\big] &\to R(K)\, \widetilde{n}_V(\xi). \end{aligned}$$

Subtracting products of expectations does not change the limit. Indeed, since $\widehat{T} = \frac{1}{M}\sum_{m=1}^M Z_m$, we have $Mh_1^d\,\mathrm{Cov}(\widehat{T}) = h_1^d\Big(\mathbb{E}[Z_m Z_m^\top] - \mathbb{E}Z_m\,\mathbb{E}Z_m^\top\Big)$. The preceding displays identify the limit of $h_1^d\mathbb{E}[Z_m Z_m^\top]$. Moreover, the components of $\mathbb{E}Z_m$ are uniformly bounded, because U, V are bounded on B and the kernel averages converge to the corresponding population quantities. Therefore $h_1^d\,\mathbb{E}Z_m\,\mathbb{E}Z_m^\top = h_1^d O(1) \to 0$. Consequently,

$$Mh_1^d\,\mathrm{Cov}(\widehat{T}) \to \Omega_\lambda(t,x;\xi).$$

Now define the centered triangular-array variables $Y_{m,M} := \sqrt{\frac{h_1^d}{M}}\,(Z_m - \mathbb{E}Z_m)$, $m = 1, \dots, M$. Then $\sqrt{Mh_1^d}\,(\hat{T} - \mathbb{E}\hat{T}) = \sum_{m=1}^M Y_{m,M}$. Because X_s, X_u are supported on the bounded set $B \times B$,

the weights $U(X_u)$ and $V(X_u)$ are bounded, and K is bounded with compact support, there exists $C_Z < \infty$ such that $|Z_m| \le C_Z\big(h_1^{-d} + h_2^{-d}\big)$. Since $h_2/h_1 \to \lambda \in (0,\infty)$, we have $h_2^{-d} \lesssim h_1^{-d}$, and hence

$$|Y_{m,M}| \le C\,\frac{\sqrt{h_1^d/M}}{h_1^d} = \frac{C}{\sqrt{Mh_1^d}} \to 0.$$

Therefore, for every $\varepsilon > 0$, the Lindeberg condition holds: $\sum_{m=1}^{M} \mathbb{E}\big[|Y_{m,M}|^2 \mathbf{1}_{\{|Y_{m,M}|>\varepsilon\}}\big] \to 0$. By the multivariate Lindeberg–Feller theorem, $\sqrt{Mh_1^d}\,(\hat{T} - \mathbb{E}\hat{T}) \;\Rightarrow\; \mathcal{N}\big(0, \Omega_\lambda(t,x;\xi)\big)$.

By definition, $\hat{a}(t,x;\xi) = \Gamma_t(\hat{T})$, $a^*(t,x;\xi) = \Gamma_t(T)$. Since $g_1(t,x;\xi) = f(\xi)D^*(t,x;\xi)$, $f(\xi) \ge f_{\min} > 0$, $D^*(t,x;\xi) \ge D_{\min} > 0$ by Assumption 3.2, the point

$$T = \big(g_2(t,x;\xi)^\top, g_1(t,x;\xi), f(\xi), f(\xi)\big)^\top$$

lies in the open set $\mathbb{R}^d \times (0,\infty) \times (0,\infty) \times (0,\infty)$, on which Γ_t is C^1. Therefore, by the multivariate delta method,

$$\sqrt{Mh_1^d}\,\big(\Gamma_t(\hat{T}) - \Gamma_t(\mathbb{E}\hat{T})\big) \;\Rightarrow\; \mathcal{N}\big(0, \Sigma_\lambda(t,x;\xi)\big),$$

with $\Sigma_\lambda(t,x;\xi) = J_\Gamma(t,x;\xi)\,\Omega_\lambda(t,x;\xi)\,J_\Gamma(t,x;\xi)^\top$.

Finally,

$$\begin{aligned}\sqrt{Mh_1^d}\,\big(\hat{a}(t,x;\xi) - a^*(t,x;\xi)\big) &= \sqrt{Mh_1^d}\,\big(\Gamma_t(\hat{T}) - \Gamma_t(\mathbb{E}\hat{T})\big) \\ &\quad + \sqrt{Mh_1^d}\,\big(\Gamma_t(\mathbb{E}\hat{T}) - \Gamma_t(T)\big).\end{aligned}$$

The first term converges to $\mathcal{N}(0, \Sigma_\lambda(t,x;\xi))$ by the delta-method argument above, and the second vanishes by the bias bound. This proves the unstudentized limit.

It remains to prove covariance consistency. For each $\phi \in \{U, V, UV, UU^\top, V^2\}$, the same change of variables and variance calculation used above give, entrywise,

$$\frac{1}{M}\sum_{m=1}^{M} \phi(X_u^{(m)}) K_{h_1}(X_s^{(m)} - \xi) \xrightarrow{\mathbb{P}} f(\xi)\mathbb{E}[\phi(X_u) \mid X_s = \xi].$$

Indeed, the bias tends to zero by the Hölder regularity in the first coordinate, while the variance is of order $(Mh_1^d)^{-1}$. Together with $\hat{f}_1(\xi) \xrightarrow{\mathbb{P}} f(\xi) > 0$, this yields $\hat{n}_U \xrightarrow{\mathbb{P}} n_U$, $\hat{n}_V \xrightarrow{\mathbb{P}} n_V$, $\hat{n}_{UV} \xrightarrow{\mathbb{P}} n_{UV}$, $\hat{m}_{UU} \xrightarrow{\mathbb{P}} m_{UU}$, and $\hat{m}_{V^2} \xrightarrow{\mathbb{P}} m_{V^2}$.

Since the dimension is fixed, these entrywise convergences imply convergence in operator norm. Moreover, $\lambda_M \to \lambda$ and, by boundedness, compact support, and continuity of K, $R_{\lambda_M}(K) \to R_\lambda(K)$. It follows that

$$|\hat{\Omega}_{h_1,h_2}(t,x;\xi) - \Omega_\lambda(t,x;\xi)|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0.$$

The preceding argument also gives $\hat{T} \xrightarrow{\mathbb{P}} T$. Since $f(\xi) > 0$ and $g_1(t,x;\xi) = f(\xi)D^*(t,x;\xi) > 0$, it follows that

$$\mathbb{P}\Big(\hat{g}_1(t,x;\xi)\hat{f}_1(\xi)\hat{f}_2(\xi) \ne 0\Big) \to 1.$$

Since Γ_t is continuously differentiable near T, $\hat{J}_\Gamma(t,x;\xi) \xrightarrow{\mathbb{P}} J_\Gamma(t,x;\xi)$. Therefore

$$|\hat{\Sigma}_{h_1,h_2}(t,x;\xi) - \Sigma_\lambda(t,x;\xi)|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0$$

by the continuous mapping theorem. The studentized limit follows from the unstudentized CLT and Slutsky's theorem, and the Wald coverage statement follows from continuity of the standard normal distribution. □

E Adaptive selector and oracle inequality

In this Appendix we state and prove an adaptive oracle inequality for the estimator (2) with possibly different bandwidths h_1 and h_2. The main-text Theorem 4.9 and corollary 4.10 are recovered by restricting to the diagonal choice $h_1 = h_2 = h$; see Remark E.2 below.

E.1 Decoupled GL selectors: grids, deviations, and penalties

Let $\mathcal{H}_j = \{h_{j,0}q^k : k = 0, \dots, K_j\} \subset (0, 1]$ be geometric grids for $j \in \{1, 2\}$, with ratio $q \in (0, 1)$, largest element $h_{j,0} \asymp 1$, and smallest element $h_{j,K_j} \asymp M^{-1/d}$. For $h \in \mathcal{H}_1$, let

$$\hat{g}_1{}^h(t, x; \xi) := \frac{1}{M} \sum_{m=1}^{M} F(t, \xi, x, X_u^{(m)})\, K_h(X_s^{(m)} - \xi), \qquad \hat{f}_1{}^h(\xi) := \frac{1}{M} \sum_{m=1}^{M} K_h(X_s^{(m)} - \xi),$$

and for $h \in \mathcal{H}_2$, let

$$\hat{g}_2{}^h(t, x; \xi) := \frac{1}{M} \sum_{m=1}^{M} X_u^{(m)}\, F(t, \xi, x, X_u^{(m)})\, K_h(X_s^{(m)} - \xi), \qquad \hat{f}_2{}^h(\xi) := \frac{1}{M} \sum_{m=1}^{M} K_h(X_s^{(m)} - \xi).$$

Write $\hat{a}_{h_1,h_2}$ for the estimator (2) constructed from $(\hat{g}_1{}^{h_1}, \hat{g}_2{}^{h_2}, \hat{f}_1{}^{h_1}, \hat{f}_2{}^{h_2})$.

For technical completeness, we use the standard denominator-floor convention: on the complement of the event

$$\mathcal{E}(h_1, h_2) := \Big\{ \inf_{\xi \in B_\rho} \hat{f}_1{}^{h_1}(\xi) \geq \tfrac{1}{2} f_{\min},\ \inf_{\xi \in B_\rho} \hat{f}_2{}^{h_2}(\xi) \geq \tfrac{1}{2} f_{\min},\ \inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} \hat{D}_{h_1,h_2}(t, x; \xi) \geq \tfrac{1}{2} D_{\min} \Big\},$$

we set the estimator equal to 0. This convention is inactive on the good event used in the proof below and contributes only to the negligible remainder term in the final risk bound.

Block errors and pairwise deviations. For $h \in \mathcal{H}_1$, define

$$E_1(h) := \max\left\{ \sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} \big|\hat{g}_1{}^h(t, x; \xi) - g_1(t, x; \xi)\big|,\ \sup_{\xi \in B_\rho} \big|\hat{f}_1{}^h(\xi) - f(\xi)\big| \right\},$$

and for $h \in \mathcal{H}_2$, define

$$E_2(h) := \max\left\{ \sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} \big|\hat{g}_2{}^h(t, x; \xi) - g_2(t, x; \xi)\big|,\ \sup_{\xi \in B_\rho} \big|\hat{f}_2{}^h(\xi) - f(\xi)\big| \right\}.$$

For $h' \leq h$ in $\mathcal{H}_j$, define

$$\mathrm{Dev}_j(h', h) := \max\left\{ \sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} \big|\hat{g}_j{}^h(t, x; \xi) - \hat{g}_j{}^{h'}(t, x; \xi)\big|,\ \sup_{\xi \in B_\rho} \big|\hat{f}_j{}^h(\xi) - \hat{f}_j{}^{h'}(\xi)\big| \right\}, \qquad j \in \{1, 2\}.$$

Penalties. Let

$$\Lambda_{\mathcal{H}}(h, \delta) := d \log\left(\frac{C_K}{h}\right) + \log\left(\frac{4(|\mathcal{H}_1| + |\mathcal{H}_2|)}{\delta}\right),$$

where $C_K > 0$ is the entropy constant from Appendix A. Let $A_j, B_{j,1}, B_{j,2} > 0$ be deterministic constants, depending only on the model parameters, such that the uniform deviation bounds of Lemmas A.1 to A.3, together with a union bound over $\mathcal{H}_1 \cup \mathcal{H}_2$, imply that, with probability at least $1 - \delta$,

$$E_j(h) \leq A_j h^\beta + B_{j,1} \sqrt{\frac{\Lambda_{\mathcal{H}}(h, \delta)}{M h^d}} + B_{j,2} \frac{\Lambda_{\mathcal{H}}(h, \delta)}{M h^d}$$

simultaneously for all $h \in \mathcal{H}_j$ and $j \in \{1, 2\}$. Set

$$b_j(h) := A_j h^\beta, \qquad v_j(h) := B_{j,1} \sqrt{\frac{\Lambda_{\mathcal{H}}(h, \delta)}{M h^d}} + B_{j,2} \frac{\Lambda_{\mathcal{H}}(h, \delta)}{M h^d}, \qquad j \in \{1, 2\}$$

and, for calibration constants $\kappa_j, \kappa_j' > 0$ with $\kappa_j' \geq \kappa_j$,

$$\mathrm{pen}_j(h', h) := \kappa_j\, [\, v_j(h) + v_j(h')\,], \qquad \mathrm{pen}_j(h) := \kappa_j' v_j(h), \qquad j \in \{1, 2\}.$$

Bias proxies and decoupled selectors. For each $j \in \{1, 2\}$, define the GL bias proxy

$$B_j(h) := \sup_{h' \leq h} \big\{ \mathrm{Dev}_j(h', h) - \mathrm{pen}_j(h', h) \big\}_+,$$

and select

$$\hat{h}_j \in \arg\min_{h \in \mathcal{H}_j} \Big\{ B_j(h) + \mathrm{pen}_j(h) \Big\}, \qquad j \in \{1, 2\}. \tag{12}$$

Selection is *decoupled*: we use $\hat{h}_1$ for $(\hat{g}_1, \hat{f}_1)$ and $\hat{h}_2$ for $(\hat{g}_2, \hat{f}_2)$. Ties are broken by choosing the largest bandwidth in the argmin.

E.2 Uniform deviations and bias inputs

Let $C_{\mathrm{bias}} := (C_B L_p/\ell_\beta!)\mu_\beta(K)$, where ℓ_β is as in Assumption 3.1. By the proofs of Lemmas A.1 to A.3, there exist deterministic constants $A_j, B_{j,1}, B_{j,2} > 0$, depending only on

$$(\beta, L_K, \mu_\beta(K), \mu_\infty(K), |K|_2, C_f, d, L_p, f_{\min}, D_{\min}, R, \eta, \Delta, \mathrm{diam}(B)),$$

such that the following holds. For every $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for all $h \in \mathcal{H}_j$ and $j \in \{1,2\}$,

$$E_j(h) \le b_j(h) + v_j(h). \tag{13}$$

Moreover, uniformly in $h \in \mathcal{H}_j$ and $j \in \{1,2\}$,

$$\max\left\{ \sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} \left|\mathbb{E}\hat{g}_j^{\,h} - g_j\right|,\ \sup_{\xi\in B_\rho} \left|\mathbb{E}\hat{f}_j^{\,h} - f\right| \right\} \le b_j(h).$$

Theorem E.1 (Unequal-bandwidth oracle inequality)**.** *Under Assumptions 3.1 and 3.2, there exist constants $M_0 \ge 1$ and $C_1, C_2, C_3 > 0$ such that, for all $M \ge M_0$, the decoupled selector $(\hat{h}_1, \hat{h}_2)$ defined by* (12) *satisfies*

$$\sup_{\nu\in\mathfrak{P}(\beta,L_p,f_{\min},B)} \mathcal{R}_\nu\big(\hat{a}_{\hat{h}_1,\hat{h}_2}\big) \le C_1 \inf_{h_1\in\mathcal{H}_1} \left\{ h_1^\beta + \sqrt{\frac{\log M}{M h_1^d}} + \frac{\log M}{M h_1^d} \right\}$$
$$+ C_2 \inf_{h_2\in\mathcal{H}_2} \left\{ h_2^\beta + \sqrt{\frac{\log M}{M h_2^d}} + \frac{\log M}{M h_2^d} \right\} + C_3\, M^{-1}.$$

Remark E.2 (Diagonal specialization). If $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$ and one restricts the selector to the diagonal class $\{(h,h) : h \in \mathcal{H}\}$, then Theorem E.1 reduces, up to a change of constants, to the main-text Theorem 4.9. Likewise, the diagonal specialization of Corollary E.3 below yields the upper-bound part of the main-text Corollary 4.10; the matching lower bound is supplied separately in Appendix F.

Proof of Theorem E.1. Let Ω be the event on which (13) holds simultaneously for all $h \in \mathcal{H}_j$ and $j \in \{1,2\}$. By a union bound over the two geometric grids, after enlarging the constants in v_j if necessary, we may assume $\Pr(\Omega) \ge 1-\delta$.

Fix $j \in \{1,2\}$ and $h' \le h$ in $\mathcal{H}_j$. On Ω, the triangle inequality and (13) give

$$\mathrm{Dev}_j(h',h) \le E_j(h) + E_j(h') \le b_j(h) + b_j(h') + v_j(h) + v_j(h').$$

Since $b_j(\cdot)$ is increasing and $\kappa_j \ge 1$, it follows that $\mathrm{Dev}_j(h',h) - \mathrm{pen}_j(h',h) \le b_j(h) + b_j(h')$, and therefore $B_j(h) \le \sup_{h'\le h}\{b_j(h) + b_j(h')\} = 2b_j(h)$.

Now fix $u \in \mathcal{H}_j$. We claim that on Ω,

$$E_j(\hat{h}_j) \le C\,\big(b_j(u) + v_j(u)\big),$$

for a constant C depending only on κ_j, κ_j'.

If $u \le \hat{h}_j$, then by the triangle inequality,

$$E_j(\hat{h}_j) \le \mathrm{Dev}_j(u,\hat{h}_j) + E_j(u) \le B_j(\hat{h}_j) + \mathrm{pen}_j(u,\hat{h}_j) + E_j(u).$$

Using $\kappa_j' \ge \kappa_j$, $B_j(\hat{h}_j) \ge 0$, and the selector optimality $B_j(\hat{h}_j) + \mathrm{pen}_j(\hat{h}_j) \le B_j(u) + \mathrm{pen}_j(u)$, we obtain $E_j(\hat{h}_j) \le 2\big(B_j(u) + \mathrm{pen}_j(u)\big) + E_j(u)$. If instead $u > \hat{h}_j$, then

$$E_j(\hat{h}_j) \le \mathrm{Dev}_j(\hat{h}_j,u) + E_j(u) \le B_j(u) + \mathrm{pen}_j(\hat{h}_j,u) + E_j(u),$$

and again $\kappa_j' \ge \kappa_j$ together with $\mathrm{pen}_j(\hat{h}_j) \le B_j(\hat{h}_j) + \mathrm{pen}_j(\hat{h}_j) \le B_j(u) + \mathrm{pen}_j(u)$ gives

$$E_j(\hat{h}_j) \le 2\big(B_j(u) + \mathrm{pen}_j(u)\big) + E_j(u).$$

In both cases, using $B_j(u) \le 2b_j(u)$ and (13),

$$E_j(\hat{h}_j) \le C\,\big(b_j(u) + v_j(u)\big).$$

Since $u \in \mathcal{H}_j$ was arbitrary, we conclude that on Ω,

$$E_j(\hat{h}_j) \le C \inf_{u \in \mathcal{H}_j} \left\{ u^\beta + \sqrt{\frac{\Lambda_{\mathcal{H}}(u,\delta)}{Mu^d}} + \frac{\Lambda_{\mathcal{H}}(u,\delta)}{Mu^d} \right\}, \qquad j \in \{1,2\}. \tag{14}$$

We now set $\delta = M^{-2}$. Since the grids are geometric and range from order 1 down to order $M^{-1/d}$, we have $|\mathcal{H}_1| + |\mathcal{H}_2| = O(\log M)$. Therefore, uniformly over $h \in \mathcal{H}_1 \cup \mathcal{H}_2$,

$$\Lambda_{\mathcal{H}}(h, M^{-2}) = d \log\left(\frac{C_K}{h}\right) + \log\big(4M^2(|\mathcal{H}_1| + |\mathcal{H}_2|)\big) \le C \left\{ d \log\left(\frac{C_K}{h}\right) + \log M \right\}$$

for all sufficiently large M. Thus the grid-union factor is absorbed into the displayed $\log M$ rate.

Because the grids span the oracle scale, there exist $\bar{h}_j \in \mathcal{H}_j$ such that $\bar{h}_j \asymp (\log M/M)^{1/(2\beta+d)}$, for $j \in \{1,2\}$. Hence

$$\bar{h}_j^\beta + \sqrt{\frac{\Lambda_{\mathcal{H}}(\bar{h}_j,\delta)}{M\bar{h}_j^d}} \asymp \Big(\frac{\log M}{M}\Big)^{\beta/(2\beta+d)}, \qquad \frac{\Lambda_{\mathcal{H}}(\bar{h}_j,\delta)}{M\bar{h}_j^d} \lesssim \Big(\frac{\log M}{M}\Big)^{2\beta/(2\beta+d)},$$

so for all sufficiently large M,

$$\inf_{u \in \mathcal{H}_j} \left\{ u^\beta + \sqrt{\frac{\Lambda_{\mathcal{H}}(u,\delta)}{Mu^d}} + \frac{\Lambda_{\mathcal{H}}(u,\delta)}{Mu^d} \right\} \lesssim \Big(\frac{\log M}{M}\Big)^{\beta/(2\beta+d)} \to 0.$$

By (14), there exists $M_0 \ge 1$ such that for all $M \ge M_0$, on Ω,

$$E_1(\hat{h}_1) \le \frac{1}{2} \min\left\{ f_{\min}, \frac{D_{\min} f_{\min}^2}{4(B_f + B_{g_1})} \right\}, \qquad E_2(\hat{h}_2) \le \frac{f_{\min}}{2},$$

where $B_f := \sup_{\xi \in B_\rho} f(\xi)$ and $B_{g_1} := \sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} |g_1(t,x;\xi)|$. Therefore, for all $M \ge M_0$, on Ω we have

$$\inf_{\xi \in B_\rho} \hat{f}_1^{\hat{h}_1}(\xi) \ge \frac{f_{\min}}{2}, \qquad \inf_{\xi \in B_\rho} \hat{f}_2^{\hat{h}_2}(\xi) \ge \frac{f_{\min}}{2},$$

and, by Lemma B.1,

$$\inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} \hat{D}_{\hat{h}_1,\hat{h}_2}(t,x;\xi) \ge \frac{D_{\min}}{2}.$$

Thus, for all $M \ge M_0$, the denominator-floor event $\mathcal{E}(\hat{h}_1, \hat{h}_2)$ holds on Ω.

For $M \ge M_0$, on Ω the denominator-floor convention is inactive, and Proposition 4.1, together with the uniform boundedness of f, g_1, g_2, and Q^* on $Q_{R,\eta,\rho}$, yields

$$\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} \big|\hat{a}_{\hat{h}_1,\hat{h}_2}(t,x;\xi) - a^*(t,x;\xi)\big| \le C\big(E_1(\hat{h}_1) + E_2(\hat{h}_2)\big).$$

Using (14), we conclude that on Ω,

$$\begin{aligned}\sup_{(t,x,\xi) \in Q_{R,\eta,\rho}} \big|\hat{a}_{\hat{h}_1,\hat{h}_2}(t,x;\xi) - a^*(t,x;\xi)\big| \le{}& C_1 \inf_{h_1 \in \mathcal{H}_1} \left\{ h_1^\beta + \sqrt{\frac{\Lambda_{\mathcal{H}}(h_1,\delta)}{Mh_1^d}} + \frac{\Lambda_{\mathcal{H}}(h_1,\delta)}{Mh_1^d} \right\} \\ &+ C_2 \inf_{h_2 \in \mathcal{H}_2} \left\{ h_2^\beta + \sqrt{\frac{\Lambda_{\mathcal{H}}(h_2,\delta)}{Mh_2^d}} + \frac{\Lambda_{\mathcal{H}}(h_2,\delta)}{Mh_2^d} \right\}.\end{aligned}$$

With this choice, $\mathbb{P}(\Omega^c) \le cM^{-2}$. On Ω we have the previous oracle bound. It remains to control the contribution of Ω^c to the expected risk.

The denominator-floor convention sets the estimator equal to zero only on the complement of $\mathcal{E}(\widehat{h}_1, \widehat{h}_2)$, not necessarily on all of Ω^c. Nevertheless, a deterministic envelope suffices. If $\mathcal{E}(\widehat{h}_1, \widehat{h}_2)$ fails, then $\widehat{a}_{\widehat{h}_1,\widehat{h}_2} = 0$, and the error is bounded by $\sup_{Q_{R,\eta,\rho}} |a^*| < \infty$. If $\mathcal{E}(\widehat{h}_1, \widehat{h}_2)$ holds, then

$$\widehat{f}_2^{\widehat{h}_2}(\xi) \ge f_{\min}/2, \qquad \widehat{D}^{\widehat{h}_1,\widehat{h}_2}(t,x;\xi) \ge D_{\min}/2.$$

Since the grids satisfy $h_{j,K_j} \gtrsim M^{-1/d}$ and K, F, and y are bounded on the relevant support/query sets,

$$\sup_{Q_{R,\eta,\rho}} |\widehat{g}_2^{\widehat{h}_2}(t,x;\xi)| \le C\widehat{h}_2^{-d} \le CM.$$

Therefore, on $\mathcal{E}(\widehat{h}_1,\widehat{h}_2)$, $\sup_{Q_{R,\eta,\rho}} |\widehat{N}^{\widehat{h}_2}(t,x;\xi)| = \sup_{Q_{R,\eta,\rho}} \left| \frac{\widehat{g}_2^{\widehat{h}_2}(t,x;\xi)}{\widehat{f}_2^{\widehat{h}_2}(\xi)} \right| \le CM$, and the denominator floor for $\widehat{D}$ gives

$$\sup_{Q_{R,\eta,\rho}} |\widehat{a}_{\widehat{h}_1,\widehat{h}_2}(t,x;\xi)| \le C(1+M).$$

Since a^* is uniformly bounded on $Q_{R,\eta,\rho}$,

$$\mathbb{E}\left[\sup_{Q_{R,\eta,\rho}} |\widehat{a}_{\widehat{h}_1,\widehat{h}_2} - a^*| \mathbf{1}_{\Omega^c} \right] \le C(1+M)\mathbb{P}(\Omega^c) \le CM^{-1}.$$

Combining this with the oracle bound on Ω proves the result. □

Corollary E.3 (Unequal-bandwidth adaptive upper bound)**.** *If $\mathcal{H}_j = \{h_{j,0}2^{-k} : k = 0,1,\dots,K_j\}$ with $h_{j,0} \asymp 1$ and $h_{j,K_j} \asymp M^{-1/d}$ for $j \in \{1,2\}$, then, for all sufficiently large M,*

$$\sup_{\nu \in \mathfrak{P}(\beta,L_p,f_{\min},B)} \mathcal{R}_\nu\big(\hat{a}_{\hat{h}_1,\hat{h}_2}\big) \lesssim \left(\tfrac{\log M}{M}\right)^{\beta/(2\beta+d)}.$$

Proof. By Theorem E.1, it suffices to optimize each blockwise term

$$h^\beta + \sqrt{\frac{\log M}{Mh^d}} + \frac{\log M}{Mh^d}$$

over the geometric grid. Since the grid spans from order 1 down to order $M^{-1/d}$, it contains a bandwidth of order $h_j^\star \asymp (\log M/M)^{1/(2\beta+d)}$, $j \in \{1,2\}$, up to a multiplicative constant depending only on the grid ratio. At this scale, the linear term $\log M/(M(h_j^\star)^d)$ is of strictly smaller order than $(h_j^\star)^\beta \asymp \sqrt{\log M/(M(h_j^\star)^d)}$. Therefore both infima in Theorem E.1 are of order $(\log M/M)^{\beta/(2\beta+d)}$. Absorbing constants proves the claim. □

F Local and global minimax lower bounds

Here and throughout Appendix F, we use the convention $\ell_\beta := \max\{m \in \mathbb{N}_0 : m < \beta\}$ and $\theta_\beta := \beta - \ell_\beta \in (0,1]$. The notation $|\cdot|_{C^\beta(B)}$ denotes the Hölder norm from Appendix A, restricted to the set B.

In this Appendix we first prove a pivot-local lower-bound statement and then derive a global minimax lower bound by instantiating it at an explicit uniform pivot. The proof uses a standard two-point Le Cam construction [19, 32, 37] embedded into a shrinking neighborhood of the pivot, intersected with the original global class.

Proposition F.1 (Local minimax lower bound around an interior pivot)**.** *Fix $(t^\circ,x^\circ,\xi^\circ) \in Q_{R,\eta,\rho}$. Let $\nu_0 \in \mathfrak{P}(\beta,L_p,f_{\min},B)$ be a pivot law with joint density p_0. Assume that there exist constants $\delta_f,\delta_D,\delta_H > 0$ such that*

$$\inf_{\xi\in B} f_0(\xi) \ge f_{\min} + 2\delta_f, \qquad \inf_{(t,x,\xi)\in Q_{R,\eta,\rho}} D_0^*(t,x;\xi) \ge D_{\min} + 2\delta_D,$$

and $\sup_{y\in B} |\bar{p}_0(\cdot,y)|_{C^\beta(B)} \le L_p - 2\delta_H$.

For $Q > 0$ and $\varepsilon > 0$, define

$$\mathcal{U}_M(\nu_0;Q,\varepsilon) := \left\{ \nu : p_\nu = p_0 + g,\ \sup_{y\in B} |g(\cdot,y)|_{C^\beta(B)} \le Q,\ |g|_{L^\infty(B\times B)} \le \varepsilon \right\},$$

and the local class inside the global model by

$$\mathfrak{P}^{\cap}_{\mathrm{loc},M}(\nu_0;Q,\varepsilon) := \mathfrak{P}(\beta,L_p,f_{\min},B) \cap \mathcal{U}_M(\nu_0;Q,\varepsilon).$$

Then there exist constants $Q_0 > 0$, $c_{\mathrm{loc}} > 0$, and $c > 0$, depending on ν_0 and the fixed model parameters, such that for all sufficiently large M,

$$\inf_{\tilde{a}} \sup_{\nu \in \mathfrak{P}^{\cap}_{\mathrm{loc},M}\left(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)}\right)} \mathcal{R}_\nu(\tilde{a}) \;\geq\; c\, M^{-\beta/(2\beta+d)}.$$

In particular, combined with Corollary E.3, the selected estimator is locally minimax-optimal around ν_0 for the uniform risk $\mathcal{R}_\nu$, up to logarithmic factors; the lower bound is witnessed by the fixed interior evaluation point $(t^\circ, x^\circ, \xi^\circ)$.

Proof. Write $F^\circ(y) := F(t^\circ, \xi^\circ, x^\circ, y)$, $\mu_0(dy) := p_0(\xi^\circ, y)\, dy$. Since $f_0(\xi^\circ) > 0$, the measure μ_0 is a nonzero finite measure on B, absolutely continuous with respect to Lebesgue measure.

1. Let $S := \mathrm{span}\{1, F^\circ\} \subset L^2(\mu_0)$. For each $r \in \{1, \dots, d\}$ define $g_r(y) := y_r F^\circ(y)$. We claim that there exists r such that

$$g_r \notin S \qquad \text{in } L^2(\mu_0).$$

Suppose otherwise. Then for every r there exist $a_r, b_r \in \mathbb{R}$ such that $y_r F^\circ(y) = a_r + b_r F^\circ(y)$ μ_0-a.e. on B. Equivalently, $h_r(y) := (y_r - b_r)F^\circ(y) - a_r = 0$ μ_0-a.e. on B. Since μ_0 is absolutely continuous and nonzero, h_r vanishes on a set of positive Lebesgue measure. The function h_r is real analytic in y because F° is real analytic. A nontrivial real-analytic function on $\mathbb{R}^d$ cannot vanish on a set of positive Lebesgue measure, so h_r must vanish identically. This is impossible: if $a_r = 0$, then $y_r \equiv b_r$, contradiction; if $a_r \neq 0$, then $F^\circ(y) = a_r/(y_r - b_r)$, but the right-hand side is singular on the hyperplane $y_r = b_r$, whereas F° is smooth and strictly positive. This contradiction proves the claim.

Fix such a coordinate r. Let Π_S denote the orthogonal projection in $L^2(\mu_0)$ onto S, and define $\varphi_0 := g_r - \Pi_S g_r$. Then $\varphi_0 \neq 0$, $\varphi_0 \in S^\perp$, and $\varphi_0 \in L^\infty(B)$ because g_r and F° are bounded on the bounded set B. Set $\varphi := \varphi_0/(2|\varphi_0|_{L^\infty(B)})$. Then $|\varphi|_{L^\infty(B)} \leq 1/2$, and since $\varphi \in S^\perp$,

$$\int_B \varphi(y)\, p_0(\xi^\circ, y)\, dy = 0, \qquad \int_B F^\circ(y)\, \varphi(y)\, p_0(\xi^\circ, y)\, dy = 0. \tag{15}$$

Moreover,

$$\int_B y_r F^\circ(y)\, \varphi(y)\, p_0(\xi^\circ, y)\, dy = \frac{|\varphi_0|^2_{L^2(\mu_0)}}{2|\varphi_0|_{L^\infty(B)}} > 0. \tag{16}$$

Hence the vector $m^\circ := \int_B y\, F^\circ(y)\, \varphi(y)\, p_0(\xi^\circ, y)\, dy$ is nonzero.

2. Choose any nonzero $\psi \in C^\beta_c(\mathbb{R}^d)$ supported in $B(0,1)$ with $\psi(0) = 1$, and set $C_\psi := |\psi|_{C^\beta(\mathbb{R}^d)} < \infty$. Since $\xi^\circ \in \mathrm{int}(B)$, there exists $h_0 > 0$ such that $\mathrm{supp}(\psi_h) \subset B$ for all $0 < h \leq h_0$, where $\psi_h(\xi) := \psi((\xi - \xi^\circ)/h)$.

Set $h := M^{-1/(2\beta+d)}$, $\alpha := a_0 h^\beta$, with $a_0 > 0$ to be chosen small later. For all sufficiently large M, we have $h \leq h_0$. Define $u_h(\xi, y) := \alpha\, \psi_h(\xi)\varphi(y)$, and the unnormalized perturbations $\tilde{p}_\pm(\xi, y) := p_0(\xi, y)\big(1 \pm u_h(\xi, y)\big)$. Since $|\varphi|_{L^\infty(B)} \leq 1/2$, if a_0 is sufficiently small then $|u_h(\xi, y)| \leq \frac{1}{4}$ on $B \times B$, so $\tilde{p}_\pm \geq \frac{3}{4}\, p_0 \geq 0$.

Let $m_h := \iint_{B\times B} p_0(\xi, y)\, u_h(\xi, y)\, d\xi\, dy$. Because p_0 is bounded on $B \times B$, φ is bounded, and ψ_h is supported on a set of volume of order h^d, we have $|m_h| \lesssim \alpha h^d$. Hence $c_\pm := \iint_{B\times B} \tilde{p}_\pm(\xi, y)\, d\xi\, dy = 1 \pm m_h$ satisfy $c_\pm \to 1$. Define the normalized densities

$$p_\pm(\xi, y) := \frac{\tilde{p}_\pm(\xi, y)}{c_\pm} = p_0(\xi, y)\, \frac{1 \pm u_h(\xi, y)}{1 \pm m_h},$$

and let $\nu_\pm$ be the corresponding laws.

3. Define $g_\pm := p_\pm - p_0$. Then $g_\pm(\xi, y) = p_0(\xi, y)\, \frac{\pm(u_h(\xi,y) - m_h)}{1 \pm m_h}$, and therefore

$$|g_\pm|_{L^\infty(B\times B)} \lesssim |u_h|_{L^\infty(B\times B)} + |m_h| \lesssim \alpha \asymp M^{-\beta/(2\beta+d)}.$$

For the C^β control in the first coordinate, write $g_\pm(\cdot, y) = \frac{p_0(\cdot,y)}{1\pm m_h}\Big(\pm\alpha\varphi(y)\psi_h(\cdot) \mp m_h\Big)$. Using the standard product estimate for Hölder spaces on bounded sets, $|fg|_{C^\beta(B)} \le C|f|_{C^\beta(B)}|g|_{C^\beta(B)}$, and the scaling bound $|\psi_h|_{C^\beta(B)} \le CC_\psi h^{-\beta}$, we get, since $\alpha = a_0 h^\beta$, $\alpha|\psi_h|_{C^\beta(B)} \le Ca_0C_\psi$. The m_h term contributes only $O(\alpha h^d)$ in $C^\beta(B)$. Consequently, $\sup_{y\in B}|g_\pm(\cdot,y)|_{C^\beta(B)} \le Ca_0C_\psi$. Choose $Q_0 > 0$ larger than this constant.

4. Because $|f_\pm(\xi) - f_0(\xi)| \le \int_B |g_\pm(\xi, y)|\, dy \le \lambda(B)|g_\pm|_{L^\infty(B\times B)} \lesssim \alpha$, and the assumption of the pivot law $\inf_{\xi\in B} f_0(\xi) \ge f_{\min} + 2\delta_f$, for all sufficiently large M we obtain

$$\inf_{\xi\in B} f_\pm(\xi) \ge f_{\min}.$$

Similarly, define

$$g_{1,\pm}(t,x;\xi) := \int_B F(t,\xi,x,y)\, p_\pm(\xi,y)\, dy, \qquad g_{1,0}(t,x;\xi) := \int_B F(t,\xi,x,y)\, p_0(\xi,y)\, dy.$$

Because F is continuous on the compact set $[s, u-\eta] \times B_R \times B \times B$, there exists

$$C_{F,Q} := \sup_{t\in[s,u-\eta],\, x\in B_R,\, \xi\in B,\, y\in B} F(t,\xi,x,y) < \infty.$$

Hence $|g_{1,\pm}(t,x;\xi) - g_{1,0}(t,x;\xi)| \le C_{F,Q}\lambda(B)|g_\pm|_{L^\infty(B\times B)} \lesssim \alpha$ uniformly on $Q_{R,\eta,\rho}$. Since f_0 and $f_\pm$ are bounded below away from zero for all sufficiently large M, the deterministic ratio inequality gives $\sup_{(t,x,\xi)\in Q_{R,\eta,\rho}} |D^*_\pm(t,x;\xi) - D^*_0(t,x;\xi)| \lesssim \alpha$. As $\inf_{(t,x,\xi)\in Q_{R,\eta,\rho}} D^*_0(t,x;\xi) \ge D_{\min} + 2\delta_D$, for all sufficiently large M we obtain

$$\inf_{(t,x,\xi)\in Q_{R,\eta,\rho}} D^*_\pm(t,x;\xi) \ge D_{\min}.$$

Finally, since $\sup_{y\in B}|p_0(\cdot,y)|_{C^\beta(B)} \le L_p - 2\delta_H$ and $\sup_{y\in B}|g_\pm(\cdot,y)|_{C^\beta(B)} \le Ca_0C_\psi$, choosing a_0 sufficiently small yields $\sup_{y\in B}|p_\pm(\cdot,y)|_{C^\beta(B)} \le L_p - \delta_H < L_p$. Thus $\nu_\pm \in \mathfrak{P}(\beta, L_p, f_{\min}, B)$ for all sufficiently large M.

Consequently, for c_{loc} sufficiently large and all sufficiently large M,

$$\{\nu_+, \nu_-\} \subset \mathfrak{P}^{\cap}_{\mathrm{loc},M}\big(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)}\big). \tag{17}$$

5. At $(t^\circ, x^\circ, \xi^\circ)$, write $g^\circ_{1,\pm} := \int_B F^\circ(y)\, p_\pm(\xi^\circ, y)\, dy$, $g^\circ_{2,\pm} := \int_B y\, F^\circ(y)\, p_\pm(\xi^\circ, y)\, dy$. Since $\psi_h(\xi^\circ) = 1$, using (15) and (16) we obtain

$$g^\circ_{1,\pm} = \frac{1}{1\pm m_h}\int_B F^\circ(y)\, p_0(\xi^\circ, y)\big(1 \pm \alpha\varphi(y)\big)\, dy = \frac{g^\circ_{1,0}}{1\pm m_h},$$

and

$$g^\circ_{2,\pm} = \frac{1}{1\pm m_h}\int_B y\, F^\circ(y)\, p_0(\xi^\circ, y)\big(1 \pm \alpha\varphi(y)\big)\, dy = \frac{g^\circ_{2,0} \pm \alpha m^\circ}{1\pm m_h},$$

where $g^\circ_{1,0} := \int_B F^\circ(y)\, p_0(\xi^\circ, y)\, dy$, $g^\circ_{2,0} := \int_B y\, F^\circ(y)\, p_0(\xi^\circ, y)\, dy$. Therefore the common normalization cancels:

$$Q^\circ_\pm := \frac{g^\circ_{2,\pm}}{g^\circ_{1,\pm}} = \frac{g^\circ_{2,0} \pm \alpha m^\circ}{g^\circ_{1,0}}. \quad \text{Hence} \quad Q^\circ_+ - Q^\circ_- = \frac{2\alpha\, m^\circ}{g^\circ_{1,0}},$$

and therefore

$$\big|a^*_+(t^\circ, x^\circ; \xi^\circ) - a^*_-(t^\circ, x^\circ; \xi^\circ)\big| = \frac{1}{\Delta(t^\circ)}\, |Q^\circ_+ - Q^\circ_-| \ge c_1\, \alpha = c_1 a_0 h^\beta, \tag{18}$$

where $c_1 > 0$ depends only on the fixed model parameters, since $\Delta(t^\circ) \ge \eta$ and

$$g^\circ_{1,0} = f_0(\xi^\circ) D^*_0(t^\circ, x^\circ; \xi^\circ) > 0.$$

6. Using the bound $\mathrm{KL}(\nu_+ \mid \nu_-) \le \chi^2(\nu_+, \nu_-) := \iint_{B\times B} \frac{(p_+ - p_-)^2}{p_-}$, we estimate the one-sample divergence. Since

$$p_+ - p_- = p_0 \left(\frac{1+u_h}{1+m_h} - \frac{1-u_h}{1-m_h} \right) = \frac{2p_0(u_h - m_h)}{1-m_h^2},$$

and since $|u_h| \le 1/4$ and $|m_h| \lesssim \alpha h^d$ for all sufficiently large M, we have

$$p_-(\xi, y) = p_0(\xi, y) \frac{1 - u_h(\xi, y)}{1 - m_h} \ge c\, p_0(\xi, y)$$

for some absolute constant $c > 0$. Therefore $\chi^2(\nu_+, \nu_-) \le C \iint_{B\times B} p_0(\xi, y)\, (u_h(\xi, y) - m_h)^2\, d\xi dy$. Since $m_h = \iint p_0 u_h$ and $\iint p_0 = 1$, Jensen's inequality gives $m_h^2 \le \iint p_0 u_h^2$, and hence

$$\iint p_0 (u_h - m_h)^2 \le 2 \iint p_0 u_h^2 + 2m_h^2 \lesssim \iint p_0 u_h^2.$$

Now p_0 is bounded on $B \times B$, φ is bounded, and ψ_h is supported on a set of ξ-volume of order h^d, so

$$\iint_{B\times B} p_0(\xi, y) u_h(\xi, y)^2\, d\xi dy \lesssim \alpha^2 h^d. \quad \text{Thus} \quad \mathrm{KL}(\nu_+ \mid \nu_-) \lesssim \alpha^2 h^d.$$

Hence for M i.i.d. observations,

$$\mathrm{KL}(\nu_+^{\otimes M} | \nu_-^{\otimes M}) \lesssim M\, \alpha^2 h^d = M\, a_0^2\, h^{d+2\beta}. \tag{19}$$

With $h = M^{-1/(2\beta+d)}$, the right-hand side of (19) is bounded by a universal constant if a_0 is chosen small enough. By Le Cam's two-point lemma,

$$\inf_{\tilde a} \sup_{\nu \in \{\nu_+, \nu_-\}} \mathbb{E}_\nu \big[\big| \tilde a(t^\circ, x^\circ; \xi^\circ) - a^*_\nu(t^\circ, x^\circ; \xi^\circ) \big| \big] \ge c_2 \, \big| a^*_+(t^\circ, x^\circ; \xi^\circ) - a^*_-(t^\circ, x^\circ; \xi^\circ) \big|.$$

Since the risk $\mathcal{R}_\nu$ dominates the pointwise risk at $(t^\circ, x^\circ, \xi^\circ)$,

$$\mathcal{R}_\nu(\tilde a) \ge \mathbb{E}_\nu \big[\big| \tilde a(t^\circ, x^\circ; \xi^\circ) - a^*_\nu(t^\circ, x^\circ; \xi^\circ) \big| \big],$$

and since (17) holds, we conclude from (18) that

$$\inf_{\tilde a} \sup_{\nu \in \mathfrak{P}^{\cap}_{\mathrm{loc}, M}\left(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)}\right)} \mathcal{R}_\nu(\tilde a) \ge \inf_{\tilde a} \sup_{\nu \in \{\nu_+, \nu_-\}} \mathcal{R}_\nu(\tilde a) \gtrsim h^\beta \asymp M^{-\beta/(2\beta+d)},$$

which proves the claim. □

Proposition F.1 is local around a pivot law satisfying the stated slack conditions. The next corollary shows that, under explicit compatibility conditions, a uniform pivot law satisfies those conditions and therefore upgrades the local lower bound to a global minimax lower bound on $\mathfrak{P}(\beta, L_p, f_{\min}, B)$.

Write $|B| := \lambda(B)$. The compatibility conditions $f_{\min} < |B|^{-1}$ and $L_p > |B|^{-2}$ are strict feasibility conditions ensuring that the uniform law lies in the interior of the marginal-density and smoothness constraints of the global class. In particular, $f_{\min} < |B|^{-1}$ is necessary for the existence of any density on B bounded below by $f_{\min}$, since $1 = \int_B f \ge f_{\min}|B|$. The denominator slack is automatic from the deterministic lower bound following Assumption 3.2.

Corollary F.2 (Global minimax lower bound via an explicit uniform pivot)**.** *Let $R_B := \sup_{z\in B} |z| < \infty$ and let $\mathcal{P} := \mathfrak{P}(\beta, L_p, f_{\min}, B)$ be the global model class from Assumption 3.2. Assume further that $f_{\min} < |B|^{-1}$ and $L_p > |B|^{-2}$. Then there exists a constant $c > 0$, depending only on the fixed model parameters, such that for all sufficiently large M,*

$$\inf_{\widetilde a} \sup_{\nu \in \mathcal{P}} \mathcal{R}_\nu(\widetilde a) \ge c\, M^{-\beta/(2\beta+d)}.$$

Consequently, together with Theorem 4.9,

$$\sup_{\nu \in \mathcal{P}} \mathcal{R}_\nu(\hat a_{\widehat h}) \lesssim \left(\frac{\log M}{M} \right)^{\beta/(2\beta+d)},$$

so the adaptive estimator is globally minimax-rate optimal over $\mathcal{P}$ up to logarithmic factors.

Proof. Fix any $(t^\circ, x^\circ, \xi^\circ) \in Q_{R,\eta,\rho}$. Define the uniform pivot law ν_0 on $B \times B$ by the joint density

$$p_0(\xi, y) = \frac{1}{|B|^2} \mathbf{1}_B(\xi) \mathbf{1}_B(y).$$

Then its X_s-marginal density is $f_0(\xi) = \int_B p_0(\xi, y)\, dy = \frac{1}{|B|}$, $\xi \in B$. Since $f_{\min} < |B|^{-1}$, we may choose $\delta_f := \frac{1}{4}\big(|B|^{-1} - f_{\min}\big) > 0$, so that $\inf_{\xi \in B} f_0(\xi) = |B|^{-1} \geq f_{\min} + 2\delta_f$.

Next, for each $y \in B$, the map $\xi \mapsto p_0(\xi, y)$ is constant on B. A valid extension to a neighborhood of B is $\bar{p}_0(z, y) = \frac{1}{|B|^2}$. With the $C^\beta(B)$ norm of Proposition F.1, $\sup_{y \in B} |\bar{p}_0(\cdot, y)|_{C^\beta(B)} = |B|^{-2}$.

Since $L_p > |B|^{-2}$, we choose $\delta_H := \frac{1}{4}\big(L_p - |B|^{-2}\big) > 0$, then $\sup_{y \in B} |\bar{p}_0(\cdot, y)|_{C^\beta(B)} \leq L_p - 2\delta_H$.

Under the uniform pivot,

$$D_0^*(t, x; \xi) = \mathbb{E}_{\nu_0}[F(t, \xi, x, X_u) \mid X_s = \xi] = \frac{1}{|B|} \int_B F(t, \xi, x, y)\, dy.$$

Define $D_{\mathrm{unif}} := \inf_{(t,x,\xi) \in Q_{R,\eta,\rho}} D_0^*(t, x; \xi)$. Moreover, by the automatic denominator bound following Assumption 3.2, $D_{\mathrm{unif}} \geq 2D_{\min}$, so the uniform pivot satisfies the required denominator slack.

It remains to note that $D_{\mathrm{unif}} > 0$. Indeed, on $Q_{R,\eta,\rho}$ we have $\Delta(t) \geq \eta$, and for $x \in B_R$, $y \in B$, $|y - x| \leq |y| + |x| \leq R_B + R$. Since the expanding term in the exponent is nonnegative,

$$F(t, \xi, x, y) = \exp\!\left(-\frac{|y - x|^2}{2\Delta(t)} + \frac{|y - \xi|^2}{2\Delta}\right) \geq \exp\!\left(-\frac{(R_B + R)^2}{2\eta}\right).$$

Hence

$$D_{\mathrm{unif}} \geq \frac{1}{|B|} \int_B \exp\!\left(-\frac{(R_B + R)^2}{2\eta}\right) dy = \exp\!\left(-\frac{(R_B + R)^2}{2\eta}\right) > 0.$$

Thus the explicit uniform pivot ν_0 satisfies the interiority/slack conditions of Proposition F.1. Therefore there exist constants $Q_0 > 0$, $c_{\mathrm{loc}} > 0$, and $c > 0$ such that, for all sufficiently large M,

$$\inf_{\widetilde{a}} \sup_{\nu \in \mathfrak{P}^{\cap}_{\mathrm{loc},M}(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)})} \mathcal{R}_\nu(\widetilde{a}) \geq c\, M^{-\beta/(2\beta+d)}.$$

By construction,

$$\mathfrak{P}^{\cap}_{\mathrm{loc},M}(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)}) \subset \mathcal{P}.$$

Hence

$$\inf_{\widetilde{a}} \sup_{\nu \in \mathcal{P}} \mathcal{R}_\nu(\widetilde{a}) \geq \inf_{\widetilde{a}} \sup_{\nu \in \mathfrak{P}^{\cap}_{\mathrm{loc},M}(\nu_0; Q_0, c_{\mathrm{loc}} M^{-\beta/(2\beta+d)})} \mathcal{R}_\nu(\widetilde{a}) \geq c\, M^{-\beta/(2\beta+d)}.$$

This proves the global minimax lower bound. □

G Experimental Details

This Appendix records the synthetic models, truth computation, bandwidth rules, diagnostics, and reproducibility details used in Section 5. The experiments are theorem-facing: they probe the finite-sample scaling law, the pointwise Gaussian approximation, the terminal singularity, and the finite-sample oracle-inequality behavior of a practical selector. For adaptivity we follow the GL-oracle-inequality viewpoint, so the relevant empirical quantity is a non-unit oracle ratio rather than a target close to one [14, 18].

Benchmark scope and existing generator comparison. We do not re-benchmark SBTS as a time-series generator. Alouadi et al. [1] compare the same finite-history SBTS construction with TSGM-VP/subVP [22], ImagenTime [27], COT-GAN [35], and GAN/VAE baselines including TimeGAN [36] and TimeVAE [9]. Their evaluation covers five-dimensional synthetic data and real datasets with 6 features for Stocks, 13 for Air, and 28 for Energy. For the discriminative score, SBTS obtains 0.010 ± 0.008 on Stocks and 0.036 ± 0.016 on Air, compared with 0.021 ± 0.008

and 0.127 ± 0.010 for TSGM-subVP. On Energy, SBTS obtains 0.356 ± 0.020, while TSGM-subVP and ImagenTime obtain 0.198 ± 0.025 and 0.040 ± 0.004, respectively. For the predictive score on Stocks/Energy/Air, SBTS obtains $0.017/0.072/0.005$, compared with $0.037/0.252/0.005$ for TSGM-subVP and $0.044/0.260/0.024$ for COT-GAN. Thus SBTS is competitive on several moderately multivariate benchmarks, while the Energy discriminative score identifies a harder regime. These end-to-end path metrics complement our synthetic known-drift experiments, which measure drift error, standardized errors, confidence-interval coverage, and adaptive-to-oracle ratios while isolating statistical error from optimization, simulation, discretization, and potential-to-drift recovery errors.

G.1 Synthetic bridge families

All experiments are run on the single interval $[s, u] = [0.2, 1.0]$, so $\Delta = u - s = 0.8$, with support box $B = [-3, 3]^d$. We use $\mathcal{X}_{\text{eval}} = [-2, 2]$ for $d = 1$ and $\mathcal{X}_{\text{eval}} = [-1.5, 1.5]^2$ for $d = 2$; the latter was frozen after a pre-flight denominator-stability check.

We consider two bounded-support synthetic pair laws. In the *Gaussian-to-Gaussian* family, $X_s \sim \text{TruncNormal}(m_s, \Sigma_s; B)$ and $X_u \mid X_s = \xi \sim \text{TruncNormal}(A\xi + b, \Sigma_\varepsilon; B)$. In the *Mixture-to-Mixture* family, $X_s \sim \frac{1}{2}\text{TruncNormal}(m_1, S; B) + \frac{1}{2}\text{TruncNormal}(m_2, S; B)$ and

$$X_u \mid X_s = \xi \sim \pi(\xi)\text{TruncNormal}(A_1\xi + b_1, \Sigma_1; B) + (1 - \pi(\xi))\text{TruncNormal}(A_2\xi + b_2, \Sigma_2; B),$$

with logistic gate $\pi(\xi) = \sigma(\alpha_0 + \alpha^\top \xi)$.

The default parameters are as follows. For GG1 we use $m_s = 0$, $\Sigma_s = (1.0)^2$, $A = 0.7$, $b = 0.3$, and $\Sigma_\varepsilon = (0.35)^2$. For GG2 we use $m_s = (0, 0)$, $\Sigma_s = \left(\begin{smallmatrix} 1.0 & 0 \\ 0 & 0.8 \end{smallmatrix}\right)$, $A = \left(\begin{smallmatrix} 0.75 & 0.15 \\ -0.10 & 0.65 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} 0.25 \\ -0.20 \end{smallmatrix}\right)$, and $\Sigma_\varepsilon = \left(\begin{smallmatrix} 0.14 & 0.03 \\ 0.03 & 0.12 \end{smallmatrix}\right)$. For MM1 we use $m_1 = -1.2$, $m_2 = 1.2$, $S = (0.45)^2$, $A_1 = 0.8$, $b_1 = 0.4$, $\Sigma_1 = (0.25)^2$, $A_2 = -0.5$, $b_2 = -0.3$, $\Sigma_2 = (0.30)^2$, and $\pi(\xi) = \sigma(1.5\xi)$. For MM2 we use $m_1 = (-0.9, 0.9)$, $m_2 = (0.9, -0.9)$, $S = \left(\begin{smallmatrix} 0.16 & 0 \\ 0 & 0.16 \end{smallmatrix}\right)$, $A_1 = \left(\begin{smallmatrix} 0.8 & 0.1 \\ 0 & 0.7 \end{smallmatrix}\right)$, $b_1 = \left(\begin{smallmatrix} 0.3 \\ -0.2 \end{smallmatrix}\right)$, $\Sigma_1 = \left(\begin{smallmatrix} 0.0484 & 0 \\ 0 & 0.0324 \end{smallmatrix}\right)$, $A_2 = \left(\begin{smallmatrix} -0.4 & 0.2 \\ 0.15 & -0.6 \end{smallmatrix}\right)$, $b_2 = \left(\begin{smallmatrix} -0.35 \\ 0.25 \end{smallmatrix}\right)$, $\Sigma_2 = \left(\begin{smallmatrix} 0.08 & 0.02 \\ 0.02 & 0.07 \end{smallmatrix}\right)$, and $\pi(\xi) = \sigma(1.2\xi_1 - 1.0\xi_2)$.

For the rate and adaptivity experiments we use the fixed interior queries GG1: $(t_0, \xi_0) = (0.6, 0)$, GG2: $(0.6, (0, 0))$, MM1: $(0.6, 0.8)$, and MM2: $(0.6, (0.8, -0.8))$. For the one-dimensional CLT experiment we use GG1: $(t_0, x_0, \xi_0) = (0.6, 0.2, 0)$ and MM1: $(0.6, 0.3, 0.8)$. The final four-point rate/adaptivity summaries use $R_M = 50$ for GG1 and MM1 and $R_M = 20$ for GG2 and MM2. The denser adaptivity-figure reruns use the same repetition counts on the seven-point grid $M \in \{10^3, 1500, 2 \times 10^3, 3000, 4 \times 10^3, 6000, 8 \times 10^3\}$. The final one-dimensional CLT runs use $R_M = 300$, while the undersmoothing screen uses $R_M = 150$ at $M \in \{4 \times 10^3, 8 \times 10^3\}$.

G.2 Truth computation and pre-flight checks

The target drift is computed numerically, not by Monte Carlo. For fixed (t, x, ξ) we evaluate

$$D^*(t, x; \xi) = \int_B F(t, \xi, x, y)\, \frac{p(\xi, y)}{f(\xi)}\, dy, \qquad N^*(t, x; \xi) = \int_B y\, F(t, \xi, x, y)\, \frac{p(\xi, y)}{f(\xi)}\, dy,$$

and then

$$a^*(t, x; \xi) = \Delta(t)^{-1} \left(\frac{N^*(t, x; \xi)}{D^*(t, x; \xi)} - x \right).$$

In $d = 1$ we use adaptive Gauss–Legendre quadrature; in $d = 2$ we use a dense tensor-product grid. The truth cache is accepted only after a one-step refinement check showing negligible change at the recorded query points.

Before the main runs we verify constrained sampling, positivity of $f(\xi_0)$ at the chosen interior points, numerical positivity of $\inf_{(t,x,\xi) \in \mathcal{Q}_{\text{eval}}} D^*(t, x; \xi)$ on the actual evaluation domain, and negligible truth-grid refinement error. The pre-flight report is used only to freeze the configurations and evaluation domains.

Global constants and local numerical behavior. The constants in the uniform finite-sample theorem depend on worst-case density and denominator floors over the full query set and on a global envelope for the SB weights. They are sufficient constants and are not claimed to be numerically

Table 3: Pre-flight numerical diagnostics: source density at the chosen conditioning point, minimum source density on the evaluation conditioning grid, minimum population denominator on the evaluation domain, and truth-grid convergence error.

Testbed	$f(\xi_0)$	$\min_{\xi\in\Xi_{\mathrm{grid}}} f(\xi)$	$\min D^*$	truth error
GG1	4.000223×10^{-1}	5.413713×10^{-2}	6.140077×10^{-3}	0
GG2	1.785645×10^{-1}	1.420651×10^{-2}	2.285052×10^{-3}	1.110223×10^{-16}
MM1	2.986354×10^{-1}	2.532522×10^{-2}	2.997735×10^{-1}	0
MM2	4.672258×10^{-1}	4.918337×10^{-9}	8.261432×10^{-3}	2.220446×10^{-16}

sharp. MM2 illustrates the distinction between these uniform quantities and the local conditions at the reported query: the minimum source density on the evaluation conditioning grid is 4.918337×10^{-9}, whereas $f(\xi_0) = 0.4672258$, and the minimum population denominator on the evaluation domain is 8.261432×10^{-3}. Thus the very small grid-wide density value does not represent the local design mass at ξ_0. Consistently, the candidate-level audit in Table 5 reports no activation of either numerical safeguard.

G.3 Grids, losses, and finite-range slope benchmark

For $d = 1$ the evaluation grid uses $|\mathcal{G}_x| = 200$ points on $[-2, 2]$; for $d = 2$ the final runs use a 21×21 grid on $[-1.5, 1.5]^2$. The primary loss is

$$E_{\infty,\mathcal{G}}(t_0, \xi_0; h) := \max_{x\in\mathcal{G}_x} |\hat{a}_h(t_0, x; \xi_0) - a^*(t_0, x; \xi_0)|,$$

with repetition average

$$\overline{E}_{\infty,\mathcal{G}}(M; h) := R_M^{-1} \sum_{r=1}^{R_M} E^{(r)}_{\infty,\mathcal{G}}(t_0, \xi_0; h).$$

As a secondary descriptive metric we also record

$$\mathrm{ISE}(t_0, \xi_0; h) := \left(\int_{\mathcal{X}_{\mathrm{eval}}} |\hat{a}_h(t_0, x; \xi_0) - a^*(t_0, x; \xi_0)|^2 \, dx \right)^{1/2}.$$

The finite-sample theorem predicts $(\log M/M)^{\beta/(2\beta+d)}$. With the product Epanechnikov kernel we use the effective benchmark $\beta = 2$, so $p(d) = 2/(4 + d)$ and the finite-range theoretical secant slope on $[M_1, M_2]$ is

$$s_{\mathrm{theory,sec}}(M_1, M_2; d) = -p(d) + p(d) \frac{\log(\log M_2/\log M_1)}{\log(M_2/M_1)}.$$

For adaptivity we define

$$\Gamma^{(r)}_{\nu,M} := \frac{E^{(r)}_{\infty,\mathcal{G}}(t_0, \xi_0; \hat{h})}{E^{(r)}_{\infty,\mathcal{G}}(t_0, \xi_0; h_{\mathrm{or}})} \geq 1,$$

where $h_{\mathrm{or}} \in \arg\min_{h\in\mathcal{H}_M} E_{\infty,\mathcal{G}}(t_0, \xi_0; h)$. We then report

$$\bar{\Gamma}_\nu(M) := R_M^{-1} \sum_{r=1}^{R_M} \Gamma^{(r)}_{\nu,M}, \qquad \widehat{C}^{\max}_\nu := \max_{M\in\mathcal{M}} \bar{\Gamma}_\nu(M), \qquad \widehat{C}^{\mathrm{avg}}_\nu := |\mathcal{M}|^{-1} \sum_{M\in\mathcal{M}} \bar{\Gamma}_\nu(M),$$

with $\mathcal{M} = \{10^3, 2 \times 10^3, 4 \times 10^3, 8 \times 10^3\}$. This is the empirical oracle-inequality diagnostic. The benchmark is not $\Gamma \approx 1$: GL-type selectors naturally come with non-unit oracle constants, and in the classical kernel setting Goldenshluger–Lepski obtain the factor $1 + 3|K|_1 = 4$ for a nonnegative unit-mass kernel [14]; minimal-penalty results further show that under-penalization may trigger instability [18].

G.4 Bandwidth grid, oracle rule, and practical selector

All rate and adaptive experiments are run on a geometric grid $\mathcal{H}_M = \{h_0 q^k : k = 0, \dots, K_M\}$ with $q = 2^{-1/2}$ and $h_0 = 1.2$. The oracle bandwidth is

$$h_{\mathrm{or}} \in \arg\min_{h \in \mathcal{H}_M} E_{\infty,\mathcal{G}}(t_0, \xi_0; h),$$

with ties broken in favor of the larger bandwidth.

The practical selector used for all quantitative claims in the paper is the raw-max, one-sided GL surrogate

$$|\hat{a}_{h'} - \hat{a}_h|_{\infty,\mathcal{G}} := \max_{x \in \mathcal{G}_x} |\hat{a}_{h'}(t_0, x; \xi_0) - \hat{a}_h(t_0, x; \xi_0)|,$$

$$B(h) := \sup_{h' \le h} \left\{ |\hat{a}_{h'} - \hat{a}_h|_{\infty,\mathcal{G}} - \kappa_{\mathrm{pair}} \sqrt{\frac{\log M}{M(h')^d}} \right\}_+, \qquad \hat{h} := \arg\min_{h \in \mathcal{H}_M} \left\{ B(h) + \kappa_{\mathrm{final}} \sqrt{\frac{\log M}{Mh^d}} \right\},$$

with $\kappa_{\mathrm{pair}} = \kappa_{\mathrm{final}} = 2$.

Why this selector. Write

$$\hat{a}_h := \hat{a}_{h,h}, \qquad v_0(h) := \sqrt{\frac{\log M}{Mh^d}}, \qquad b(h) := Ah^\beta,$$

and $E(h) := |\hat{a}_h - a^*|_{\infty,\mathcal{G}}$. The implemented diagonal selector uses

$$B(h) = \sup_{h' \le h} \{ |\hat{a}_{h'} - \hat{a}_h|_{\infty,\mathcal{G}} - 2v_0(h') \}_+ .$$

On the good event, suppose $E(h) \le b(h) + v_0(h)$. For $h' \le h$,

$$v_0(h') \ge v_0(h), \qquad b(h') \le b(h),$$

and therefore $|\hat{a}_{h'} - \hat{a}_h|_{\infty,\mathcal{G}} \le E(h') + E(h) \le b(h') + b(h) + v_0(h') + v_0(h)$.

Since $2v_0(h') \ge v_0(h') + v_0(h)$,

$$|\hat{a}_{h'} - \hat{a}_h|_{\infty,\mathcal{G}} - 2v_0(h') \le b(h') + b(h) \le 2b(h).$$

Hence $B(h) \le 2b(h)$. By the minimizing property of $\hat{h}$, for every $u \in \mathcal{H}_M$,

$$B(\hat{h}) + 2v_0(\hat{h}) \le B(u) + 2v_0(u).$$

Using the standard one-sided GL comparison,

$$E(\hat{h}) \le 2\{B(u) + 2v_0(u)\} + E(u).$$

Consequently,

$$E(\hat{h}) \le 2\{2b(u) + 2v_0(u)\} + b(u) + v_0(u) = 5\{b(u) + v_0(u)\}.$$

Taking the infimum over $u \in \mathcal{H}_M$, the practical selector tracks the diagonal oracle objective up to the multiplicative constant 5. The appendix oracle inequality uses the full proof-side penalty with constants, grid-union logarithms, and lower-order linear terms; the experiments use this leading-order diagonal proxy.

Why the lower bandwidth floor. The crucial experimental issue is not the selector form itself but the small-h tail of the admissible grid. The stochastic scale entering both the selector penalty and the local denominator estimation is of order $(Mh^d)^{-1/2}$ up to logarithms. If h is too small, then Mh^d is too small, so pairwise discrepancies become excessively noisy and the ratio estimator becomes more fragile because its local denominator is supported by too little effective mass. To prevent the selector from entering this regime, we truncate the grid below at

$$h_{\min}(M, d) = c_{\min} M^{-1/d},$$

with

$$c_{\min} = 81 \quad \text{in } d = 1, \qquad c_{\min} = 9 \quad \text{in } d = 2.$$

Table 4: Bandwidth and oracle-ratio diagnostics for the final adaptive protocol. Here $\mathbb{E}_M[\mathbb{E}(h_{\mathrm{or}} \mid M)]$ and $\mathbb{E}_M[\mathbb{E}(\hat{h} \mid M)]$ are the M-averaged oracle and adaptive selected bandwidths, `BoundaryAvg` is the mean boundary-hit rate, and $\widehat{C}_\nu^{\mathrm{avg}}, \widehat{C}_\nu^{\max}$ are the empirical oracle-ratio summaries.

Testbed	$\mathbb{E}_M[\mathbb{E}(h_{\mathrm{or}} \mid M)]$	$\mathbb{E}_M[\mathbb{E}(\hat{h} \mid M)]$	BoundaryAvg	$\widehat{C}_\nu^{\mathrm{avg}}$	$\widehat{C}_\nu^{\max}$
GG1	0.242760	0.381922	0.0000	1.773613	2.109796
GG2	0.376363	0.383079	0.0000	1.314465	1.431312
MM1	0.238186	0.335428	0.0100	1.960175	2.175424
MM2	0.398902	0.435579	0.0125	1.485740	1.686308

Table 5: Candidate-level numerical-safeguard diagnostics. The last two columns count activations of the corresponding 10^{-12} numerical safeguard.

Testbed	$\min \widehat{f}_h$	$\min_{x \in \mathcal{G}_x} \widehat{D}_h(x)$	Density floor	Denominator floor
GG1	0.2547	2.075×10^{-3}	0	0
GG2	0.1225	8.197×10^{-4}	0	0
MM1	0.1902	3.222×10^{-2}	0	0
MM2	0.1384	2.358×10^{-3}	0	0

These choices are dimensionally matched because in both cases

$$Mh_{\min}(M,d)^d = c_{\min}^d = 81.$$

Hence the entire adaptive search is restricted to

$$Mh^d \geq 81.$$

This imposes a common nominal local-volume restriction across dimensions; it is not a lower bound on the realized kernel effective sample size, which also depends on the empirical kernel weights.In exploratory runs, allowing smaller bandwidths was the dominant source of instability in the hard MM cases; after imposing $Mh^d \geq 81$, the selector became stable on all four testbeds.

Calibration sensitivity. We also evaluated nearby one-sided raw-max calibrations and lower effective-volume restrictions. Across these variants, the four-point mean oracle ratio ranged from 1.69 to 2.13 for GG1, 1.23 to 1.48 for GG2, 1.90 to 2.15 for MM1, and 1.42 to 1.75 for MM2. The canonical calibration $\kappa_{\mathrm{pair}} = \kappa_{\mathrm{final}} = 2$ with $Mh^d \geq 81$ gave 1.77, 1.31, 1.96, and 1.49, respectively. The qualitative ordering and low-single-digit oracle-ratio behavior remained stable when the restriction was varied among $Mh^d \geq 49$, $Mh^d \geq 81$, and $Mh^d \geq 121$. These checks assess the robustness of the practical leading-order calibration and do not estimate the model-dependent constants in the proof-side penalties.

For diagnostics we track the average selected oracle and adaptive bandwidths, the empirical oracle-ratio summaries, and the frequency with which the selected bandwidth lies on the search boundary. In the table below the selected-bandwidth averages are first taken over repetitions at each M and then averaged over $M \in \mathcal{M}$.

Candidate-level numerical-safeguard audit. For each testbed, we recorded the smallest estimated source density and the smallest estimated SB denominator over all sample sizes, repetitions, and candidate bandwidths used by the final adaptive protocol. For the denominator, the minimum is also taken over the prescribed evaluation grid in x.

Across the 6,000 audited candidate evaluations, both activation counts were zero. The smallest observed denominator was more than 8×10^8 times the 10^{-12} numerical floor. Hence the safeguards did not alter the candidate estimators or determine the selected bandwidths. The anonymized supplement contains the complete candidate-level records.

G.5 Rate and adaptivity protocol

For each testbed and each $M \in \mathcal{M}$, we simulate R_M independent datasets and compute $E_{\infty,\mathcal{G}}^{(r)}(t_0, \xi_0; h)$ for every $h \in \mathcal{H}_M$. The oracle-selected bandwidth minimizes this quantity; the

Table 6: Secondary rate results based on the integrated squared error. These are reported only as descriptive diagnostics.

Testbed	d	Finite-range theory	Oracle-h_{or}
GG	1	-0.349379	-0.337109
GG	2	-0.291150	-0.286309
MM	1	-0.349379	-0.383080
MM	2	-0.291150	-0.378696

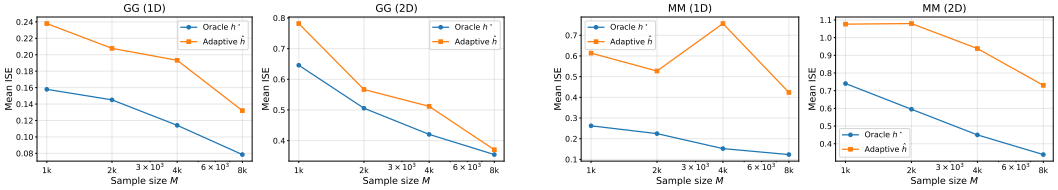


Figure 3: Secondary rate plots based on the integrated squared error.

adaptive bandwidth is obtained from the canonical selector above. The fitted rate slope is then obtained by ordinary least squares on $\log \overline{E}_{\infty,\mathcal{G}}(M)$ versus $\log M$. We also report the slope of the secondary ISE curve as a descriptive diagnostic.

The main adaptivity summaries are $\bar{\Gamma}_\nu(M)$, $\widehat{C}_\nu^{\max}$, and $\widehat{C}_\nu^{\mathrm{avg}}$. Since the theorem is an oracle inequality rather than an oracle-equality statement, these are the right practical diagnostics. The protocol has two distinct theorem-facing ingredients: the selector form is chosen because it is the closest direct implementation of the proof-side GL construction, while the lower truncation $Mh^d \geq 81$ is imposed because the experimental instability came from the too-small-h regime. Once this stabilized admissibility rule is imposed, the canonical raw-max one-sided selector is oracle-competitive on all four testbeds.

The oracle-slope summaries are already reported in Table 1 while the four-point empirical oracle-ratio summaries are reported in Table 4.

Primary sup-grid plots. The primary sup-grid rate plots are shown in the main text in Figure 1. To avoid duplication, we do not repeat them here. This Appendix instead retains the numerical slope Table 1, the oracle-ratio summary and the bandwidth diagnostics Table 4, and the secondary ISE plots in Figure 3.

G.6 Pointwise CLT and confidence intervals

The main-text CLT experiment is one-dimensional. At the fixed interior query (t_0, x_0, ξ_0), we use the plug-in estimator $\hat{\Sigma}_h(t_0, x_0; \xi_0)$ from (11); in this case $\widehat{\mathrm{Cov}}_h$ reduces to the local empirical variance and $\hat{\psi}_h(y) = (y - x_0 - \Delta(t_0)\hat{a}_h(t_0, x_0; \xi_0))F(t_0, \xi_0, x_0, y)$.

The CLT bandwidth must satisfy $Mh_M^d \to \infty$ and $h_M^\beta \sqrt{Mh_M^d} \to 0$. We therefore use $h_M = cM^{-\alpha}$ with $(2\beta+d)^{-1} < \alpha < d^{-1}$. The final reported rules are $(\alpha, c) = (0.22, 1)$ for GG1 and $(0.28, 1)$ for MM1. We also ran a paired audit over $\alpha \in \{0.22, 0.24, 0.26, 0.28\}$ and $c \in \{0.75, 1, 1.25\}$, reusing each simulated sample across all twelve rules. Across the resulting 28,800 estimator evaluations, mean coverage over the four sample sizes ranged from 94.92% to 95.75% for GG1 and from 94.42% to 95.83% for MM1. Mean studentized standard deviations ranged from 0.978 to 0.992 and from 0.973 to 1.003, while the largest absolute standardized means were 0.098 and 0.168. No density, SB-denominator, variance, or standard-deviation floor was activated.

For repetition r, the standardized statistic is

$$Z_M^{(r)} = \frac{\sqrt{Mh_M^d}\big(\hat{a}_{h_M}^{(r)}(t_0, x_0; \xi_0) - a^*(t_0, x_0; \xi_0)\big)}{\sqrt{\hat{\Sigma}^{(r)}(t_0, x_0; \xi_0)}}.$$

Table 7: Final pointwise CLT diagnostics at the fixed interior query point: empirical mean and variance of the standardized statistic and empirical coverage of nominal 95% confidence intervals in the $R_M = 300$ runs.

Testbed	M	$\bar{Z}_M$	$\widehat{\mathrm{Var}}(Z_M)$	Coverage (%)
GG	10^3	-0.029120	1.160508	92.67
GG	2×10^3	-0.083007	0.964855	96.33
GG	4×10^3	0.056018	0.828702	96.67
GG	8×10^3	-0.015311	0.974488	96.67
MM	10^3	0.011673	1.180449	92.33
MM	2×10^3	-0.054668	0.964488	96.67
MM	4×10^3	0.133135	0.879781	96.33
MM	8×10^3	0.041255	0.894031	96.00

Table 8: Formal normality diagnostics for the final pointwise CLT runs. We report the Shapiro–Wilk p-value and the Anderson–Darling statistic together with its 5% critical value. The indicator `reject@5%` equals 1 if the Anderson–Darling statistic exceeds the 5% critical value.

Testbed	M	Shapiro p	Anderson stat	Anderson 5% crit	reject@5%
GG	10^3	0.139036	0.689650	0.750000	0
GG	2×10^3	0.929805	0.253814	0.750000	0
GG	4×10^3	0.180917	0.536864	0.750000	0
GG	8×10^3	0.584810	0.483753	0.750000	0
MM	10^3	0.094434	0.487901	0.750000	0
MM	2×10^3	0.139434	0.500433	0.750000	0
MM	4×10^3	0.905583	0.173670	0.750000	0
MM	8×10^3	0.466130	0.414661	0.750000	0

The final CLT runs use $R_M = 300$ repetitions at each $M \in \{10^3, 2 \times 10^3, 4 \times 10^3, 8 \times 10^3\}$. For each sample size we record the empirical mean $\bar{Z}_M$, empirical variance $\widehat{\mathrm{Var}}(Z_M)$, the empirical coverage of

$$\hat{a}^{(r)}_{h_M}(t_0, x_0; \xi_0) \pm 1.96\sqrt{\hat{\Sigma}^{(r)}(t_0, x_0; \xi_0)/(Mh_M^d)},$$

and formal Shapiro–Wilk / Anderson–Darling diagnostics of the standardized replicates. QQ-plots against $N(0,1)$ are saved for the two largest sample sizes and are used as the main visual CLT diagnostic in the paper.

CLT plots and saved artifacts. The manuscript-level CLT visual is Figure 2. To avoid duplication, we do not repeat it here. This Appendix instead records the numerical diagnostics in Tables 7 and 8. The raw replicate-level statistics are saved as `pointwise_clt.csv` files, and the QQ-plots generated during the final runs are released with the repository under the standard figure directories.

G.7 Terminal-edge experiment

To isolate the terminal singularity, we fix one moderate-to-large sample size and evaluate the estimator at $t \in \{u - 0.40, u - 0.25, u - 0.15, u - 0.10, u - 0.05\}$. The bandwidth is selected once at the interior time $t_0 = 0.6$ and then held fixed. For this terminal-edge experiment only, we abbreviate

$$E_{\infty,\mathcal{G}}(t) := E_{\infty,\mathcal{G}}(t, \xi_0; \hat{h}) = \max_{x \in \mathcal{G}_x} |\hat{a}_{\hat{h}}(t, x; \xi_0) - a^*(t, x; \xi_0)|.$$

The primary diagnostic is this quantity together with its rescaled version $\Delta(t)E_{\infty,\mathcal{G}}(t)$. We also record the corresponding ISE curves as secondary diagnostics.

Terminal-edge plots. Figure 4 reports the raw and rescaled terminal-edge errors. The raw sup-grid error increases as $t \uparrow u$, while the rescaled quantity $\Delta(t)E_{\infty,\mathcal{G}}(t)$ remains substantially flatter. This

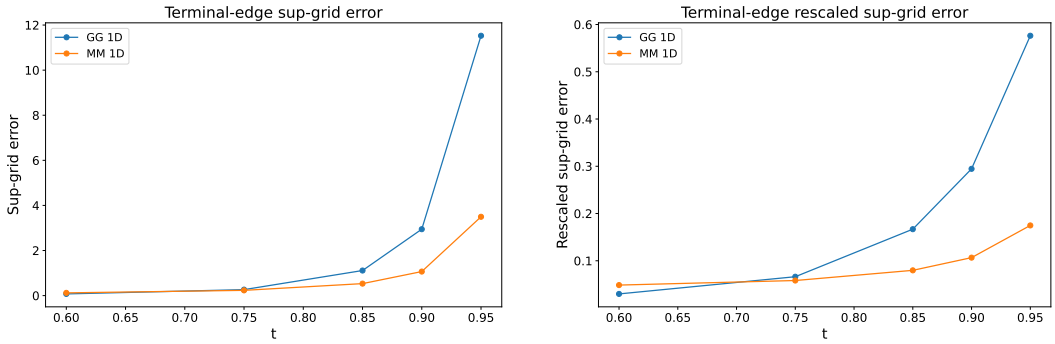


Figure 4: Terminal-edge behavior. Left: raw sup-grid error $E_{\infty,\mathcal{G}}(t)$ as $t \uparrow u$. Right: rescaled error $\Delta(t)E_{\infty,\mathcal{G}}(t)$.

Table 9: Bounded-support stress test at $M_{\text{stress}} = 4000$ and $R_{\text{stress}} = 100$.

Testbed	Setting	$\mathbb{E}[\widehat{\text{Var}}(W^N)]$	$\mathbb{E}[\widehat{\text{Var}}(W^D)]$	$\Pr(\hat{D} \le \tau_D)$	med. $E_{\infty,\mathcal{G}}$
GG1	compact	0.1579	0.7101	0.00	0.1839
GG1	wide	0.1732	0.7843	0.00	0.1570
MM1	compact	0.2638	0.4047	0.00	0.2387
MM1	wide	0.7708	2.1689	0.10	2.4260

supports the interpretation that the observed deterioration is driven primarily by the intrinsic factor $\Delta(t)^{-1}$.

G.8 Bounded-support stress test

To probe the proof role of bounded support, we run a one-dimensional stress test at the fixed interior queries GG1: $(t_0, x_0, \xi_0) = (0.6, 0.2, 0)$ and MM1: $(0.6, 0.3, 0.8)$, with the same canonical selector as in the final adaptive runs (raw-max, one-sided, $\kappa_{\text{pair}} = \kappa_{\text{final}} = 2$, and the same lower-bandwidth floor), at the fixed sample size $M_{\text{stress}} = 4000$ with $R_{\text{stress}} = 100$ repetitions. For GG1, the wide-support variant changes only the truncation box from $[-3, 3]$ to $[-5, 5]$. For MM1, the wide-support variant also inflates the source and conditional variances from $(0.45)^2, (0.25)^2, (0.30)^2$ to $(0.8)^2, (0.9)^2, (1.0)^2$, while keeping the same means, linear maps, and logistic gate.

For each repetition, with selected bandwidth $\hat{h}$, we record the empirical variances of the raw weighted numerator and denominator summands $W_m^N := X_u^{(m)} F(t_0, \xi_0, x_0, X_u^{(m)}) K_{\hat{h}}(X_s^{(m)} - \xi_0)$ and $W_m^D := F(t_0, \xi_0, x_0, X_u^{(m)}) K_{\hat{h}}(X_s^{(m)} - \xi_0)$, the pointwise empirical denominator $\hat{D} := \hat{D}_{\hat{h}}(t_0, x_0; \xi_0)$, and the selected-estimator errors. To quantify denominator fragility, we set

$$\tau_D := \tfrac{1}{4} \ \text{median}_{r,\text{compact}}(\hat{D}^{(r)})$$

separately for each testbed and report $\Pr(\hat{D} \le \tau_D)$.

Interpretation. The stress effect is mild in GG1: the wide-support perturbation raises $\mathbb{E}[\widehat{\text{Var}}(W^N)]$ and $\mathbb{E}[\widehat{\text{Var}}(W^D)]$ only by factors 1.10 and 1.10, while the median sup-grid error and median ISE slightly decrease from 0.1839 to 0.1570 and from 0.1712 to 0.1460. Thus weakening compact support is not automatically catastrophic in an easier Gaussian testbed.

The harder MM1 testbed is qualitatively different. The wide-support perturbation multiplies $\mathbb{E}[\widehat{\text{Var}}(W^N)]$ by 2.92 and $\mathbb{E}[\widehat{\text{Var}}(W^D)]$ by 5.36, raises $\Pr(\hat{D} \le \tau_D)$ from 0 to 0.10, inflates the median sup-grid error from 0.2387 to 2.4260 (a factor 10.16), and inflates the median ISE from 0.2499 to 1.6083 (a factor 6.44). At the same time, the mean oracle ratio rises from 2.09 to 7.12, the mean selected bandwidth drops from 0.3221 to 0.0655, and the boundary-hit rate rises from 0 to 0.10. This is the expected failure mode: the deterministic mean structure is unchanged, but once the

support control is weakened, the weighted empirical terms in the ratio estimator become substantially more variable, denominator stability deteriorates, and the selector is pulled toward pathological small-bandwidth behavior. This is precisely the role bounded support plays in the proof.

Sub-Gaussian extensions. The bounded-support restriction is not intrinsic to every model. For example, if $X_u \mid X_s = \xi \sim \mathcal{N}(m_0 + B\xi, \Sigma)$, completing the square makes the tilted conditional mean, and hence the drift, affine in (x, ξ), so a parametric estimator is available at rate $M^{-1/2}$. For general sub-Gaussian designs, the difficulty is time-dependent. On $t \in [s + \kappa, u - \eta]$, completing the square gives $F(t, \xi, x, y) \le C_{\kappa,\eta,R} e^{-c_{\kappa,\eta}|y|^2}$ for $|x|, |\xi| \le R$, so the positive tilt is controlled by Gaussian damping. At $t = s$, $F(s, \xi, x, y) = e^{\frac{(x-\xi)^\top y}{u-s} + \frac{|\xi|^2 - |x|^2}{2(u-s)}}$. Under a uniform conditional sub-Gaussian condition, $\mathbb{E}[F(s, \xi, x, X_u)^r \mid X_s = \xi] \le e^{Cr^2R^2}$ for $|x|, |\xi| \le R$. Thus all fixed moments exist on a fixed query ball, but the weighted variables have lognormal-type tails and no deterministic envelope, so the present Bernstein/Bousquet and denominator-localization arguments do not extend directly. A radius-R localization leads schematically to

$$e^{CR^2}\left[h^\beta + \sqrt{\frac{\log M}{Mh^d}} + \frac{\log M}{Mh^d}\right] + e^{-cR^2}.$$

For fixed R, the usual polynomial exponent may remain valid with R-dependent constants; for growing query regions, the effective rate deteriorates. A branchwise extension could truncate or robustify the weighted ratio near $t = s$, retain the original estimator in the damped interior, and use a log-gradient representation near the terminal edge.

G.9 Implementation notes and optional selector variants

The implementation is organized around a small set of core modules and driver scripts. The synthetic pair laws are defined in `src/sbdrift/models.py`; deterministic truth evaluation is implemented in `src/sbdrift/truth_engine.py`; the kernel estimator is implemented in `src/sbdrift/estimator.py`; and the kernel definitions are collected in `src/sbdrift/kernels.py`. The experiment drivers are `scripts/00_preflight.py`, `scripts/01_rate.py`, and `scripts/02_clt.py`. The pre-flight script writes numerical checks of the marginal density, denominator floor, truth-convergence probe, and sampling sanity checks. The rate script writes raw per-bandwidth and selected-bandwidth CSV files, processed summaries, and rate plots. The CLT script writes per-repetition pointwise statistics, processed summaries, and per-M QQ-plot files, and also records the undersmoothing-screen runs together with the formal normality diagnostics used to choose the final one-dimensional CLT bandwidths. All experiments are implemented in `Python` using `NumPy`, `SciPy`, `pandas`, and `matplotlib`.

Each run is driven by a YAML configuration file specifying the synthetic family, support and evaluation boxes, query points, bandwidth grid parameters, repetition counts, and random seeds. The rate pipeline caches deterministic truth values on the evaluation grid to avoid recomputation across repeated runs. Outputs are written to the standard experiment directories `results/raw/`, `results/processed/`, and `results/figures/`. The repository state used for the paper contains the processed pre-flight summaries, the final theorem-facing rate and CLT summaries, the undersmoothing-screen and final CLT runs, the final stabilized adaptive-selector summaries, and the figure files corresponding to the saved experiment outputs.

The codebase also retains optional alternative selector implementations, including trimmed-max, two-sided, and ISE-based discrepancy rules. These remain available for users interested in separate comparison studies, but they are not used for the paper's quantitative claims. Accordingly, the reproducibility claim supported by the current codebase is the following: the repository contains the exact synthetic-model definitions, the deterministic truth engine, the kernel estimator implementation, the pre-flight/rate/CLT experiment drivers, the YAML configuration files, and the saved raw, processed, and figure outputs used to assemble the final tables and theorem-facing plots. Exact command lines, random seeds, repetition counts, the selected undersmoothing exponents, the stabilized bandwidth-floor rule, and run-specific parameter choices are recorded in the experiment logs and processed summaries.

H Compute Resources and Broader Impact

Compute resources. All experiments were run on a `Dell5520` machine with an 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60GHz (8 logical cores) and 7.5 GiB RAM; all reported experiments were CPU-only, and no GPU was used. The implementation is in Python using NumPy, SciPy, pandas, and matplotlib, with experiment drivers and YAML configuration files described in Section G. A representative pre-flight run took about 1.9 seconds, a representative four-sample-size rate run with 5 repetitions per sample size took about 24.9 seconds, and a representative four-sample-size CLT run with 20 repetitions per sample size took about 86.0 seconds. Scaling these probes gives roughly 4.2 minutes for a full GG1-style final rate run with 50 repetitions and about 21.5 minutes for a full four-sample-size CLT run with 300 repetitions per sample size. The bounded-support stress-test summary pass, which reconstructs the selected-bandwidth diagnostics over 4×100 saved repetitions, took about 680.8 seconds (≈ 11.3 minutes). The full project also used additional CPU time for pilot undersmoothing screens, selector-calibration checks, confirmation runs, terminal-edge experiments, and superseded exploratory runs. No large-scale model training was performed.

Broader impact. This paper is primarily a theoretical contribution: it studies direct nonparametric estimation of Schrödinger-bridge time-series drifts in the SBTS single-interval setting and evaluates the theory on controlled synthetic experiments. Potential positive impacts include improving the statistical understanding, transparency, and reproducibility of methods related to Schrödinger bridges, which may benefit downstream work in stochastic control, sampling, uncertainty quantification, and time-series modeling.

Potential negative impacts are indirect but real: Schrödinger-bridge methods are connected to modern generative and sequential modeling, so improvements in estimation and inference could eventually contribute to systems that are misused for synthetic-data generation, surveillance-oriented forecasting, or other downstream applications where model errors or distributional mismatch could cause harm. Our work does not release a high-risk dataset or a powerful pretrained generative model, and the paper explicitly emphasizes its narrow scope, assumptions, and limitations; we therefore view the main mitigation here as transparent reporting of assumptions, failure modes, and the non-deployment nature of the present study.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the actual scope of the paper: direct drift-level estimation in the Markov-reduced single-interval setting, a uniform finite-sample bound, a pointwise CLT under genuine undersmoothing, and adaptive near-minimaxity; the experiments are explicitly framed as theorem-targeted diagnostics (Abstract; Sections 1, 4, and 5).

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Answer: [Yes]

   Justification: The paper includes a dedicated "Discussion and Limitations" section that explicitly discusses the bounded-support assumption, curse of dimensionality, terminal singularity, stable-ratio localization, i.i.d. single-interval scope, and the gap between oracle theory and practical selector behavior (Section 6).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumptions are stated explicitly in Section 3.1, the main theoretical results are numbered in Section 4, and the corresponding proofs and specializations are developed in the appendix, including the pointwise CLT, adaptive oracle inequality, and minimax lower bound (Sections 3–4; Appendices A–F).

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper discloses the experimental ingredients needed to reproduce the main diagnostics, including the synthetic families, kernels, sample sizes, query points, repetition counts, bandwidth rules, and the plug-in variance estimator; the exact implementation details are deferred to Appendix G (Section 5; Appendix G).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code, configuration files, and saved experimental outputs are publicly available at `https://github.com/OthmaneAnalytics/sb-drift-experiments`. The repository README provides installation and execution instructions, while Appendix G.9 documents the experiment drivers and implementation details.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Although this is not a train/test benchmark paper in the usual supervised-learning sense, the manuscript specifies the synthetic data-generation setups, evaluation grids, kernels, bandwidth grids and rules, sample sizes, and repetition counts needed to understand the reported results (Section 5; Appendix G).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: For the Monte Carlo diagnostics, the paper reports repetition counts, empirical coverage of nominal 95% confidence intervals, QQ-plots, empirical means and variances of the standardized statistic, and formal normality diagnostics; the adaptive summaries also use repetition-based oracle-ratio diagnostics and interquartile ribbons (Section 5; Appendix G.4–G.6).

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix H reports the hardware environment, approximate wall-clock time for the main experiment groups, and an estimate of additional pilot and exploratory compute beyond the final reported runs; Appendix G already documents the software stack, experiment drivers, configuration files, and saved outputs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work is theoretical and uses controlled synthetic experiments only; it does not involve human subjects, private data, or deceptive deployment claims, and the manuscript does not indicate any deviation from the NeurIPS Code of Ethics.

   Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Appendix H discusses both positive and negative societal impacts. In particular, it explains the theory-first and synthetic nature of the work, its potential benefits for transparent statistical methodology, and possible downstream misuse risks through applications of Schrödinger-bridge methods in generative or sequential modeling.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release a high-risk scraped dataset or a powerful generative model; it studies a statistical estimator and uses synthetic experiments, so special responsible-release safeguards are not applicable here.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The empirical study uses author-generated synthetic data rather than third-party datasets, pretrained models, or borrowed benchmark assets central to the claims. Standard scientific Python packages are used only as implementation dependencies.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The release includes new code and configuration assets for the synthetic experiments, and Appendix G documents the core modules, driver scripts, YAML configuration files, and saved raw/processed/figure outputs used for the reported results (Appendix G.9).

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing, data collection from human participants, or any other human-subject experiment.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The work does not involve human subjects or crowdsourcing, so IRB approval or an equivalent review was not required.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not an important, original, or non-standard component of the proposed methodology or experimental pipeline; the core method is a direct kernel estimator for the Markov-reduced Schrödinger-bridge drift.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.