SYNERGISTIC LEARNING IN DOMAIN ADAPTATION WITH A KNOWN LINK: A MINIMAX UNDERSTANDING

Anonymous authors
Paper under double-blind review

ABSTRACT

When can combining source and target data yield a much better performance than using either one alone? Statistically, we address this issue from a minimax viewpoint: writing $R(n, n_T)$ for the minimax risk attainable from n source and n_T target observations, we ask when $R(n, n_T)$ decays *strictly faster* than both $R(n, 0)$ and $R(0, n_T)$. When this happens, we call it the *Synergistic Learning Phenomenon* (SLP) and quantify it by the *Synergistic Advantage Ratio* $\mathrm{SAR} = \min(R(n, 0), R(0, n_T))/R(n, n_T)$. In this work, we focus on domain adaptation with a *known link*: the source regression function is f, the target regression function is $g = \mathcal{A}f$ for a known operator $\mathcal{A}$, and the two domains may also differ in their designs. Three links—the identity, integration, and differentiation—together with the behaviour of the density ratio $\rho = h_T/h_S$ determine whether synergy is possible, at what rate, and in which sample-size window. For the identity link (covariate shift) with a Beta$(a, 1)$ source and a uniform target, SLP is possible only if $a \geq 2 + 1/(2\beta)$, where β is the smoothness of f (Result I; Zhou & Yang 2025). An integral link $g(x) = \int_0^x f(t)\,dt$ instead can yield the joint rate $(nn_T)^{-1/2}$ even for a bounded ratio (Result II), and under a Beta$(a, 1)$ source it has a phase diagram with cuts at $a = 2$ and $a = 3$: integral SLP for $1 < a < 2$, an extra level-split rate n_T^{-1} for $2 < a < 3$, and covariate-shift-type spatial SLP for $a > 3$ (Result III). A derivative link $g = \dot{f}$ synergizes only under a stronger singularity (Result IV), giving a sharp directional asymmetry under a bounded ratio. Across all three links, synergy arises from *role separation*: each dataset resolves a distinct bottleneck associated with the link and the designs. We characterize the sample-size windows in which combining source and target data improves on both individual minimax rates, and how the SAR varies with the source and target sample sizes. Simulations support the predicted phase structure, and a mirrored integration–differentiation analysis of real vehicle and smartphone signals illustrates the directional asymmetry.

1 INTRODUCTION

Can two datasets together do better than either alone? Labeled target observations are often scarce while a related source domain supplies a far larger sample at lower cost (Pan & Yang, 2010; Weiss et al., 2016; Torrey & Shavlik, 2010). We ask whether the two samples together can achieve a *strictly better* rate of convergence than either achieves alone—a "$1+1>2$" effect rather than a substitution effect. In this work, we consider a domain adaptation setting where the covariate distribution may differ and the two regression functions are related via a known link. We take the source regression to be f and the target to be $g = \mathcal{A}f$ for a known operator $\mathcal{A}$. The two designs—the covariate distributions of the source and the target—may differ with an unbounded density ratio.

Most existing rate improvements in transfer learning arise by imposing structure on an *unknown* discrepancy between the source and target problems, for example through sparse or low-complexity parameter shifts, function proximity, shared representations, or a smoother additive bias; see Section 2 for details. The gain comes from a residual that is easier to estimate than either regression function. Our setting removes this source of gain because the known relation $g = \mathcal{A}f$ determines g from f, leaving no additional regression residual to estimate. Any "$1 + 1 > 2$" effect must instead come from how the known operator interacts with the two designs: the two samples provide com-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

plementary information, rather than an easier unknown offset. Closest to this known-link regime are covariate shift (the identity operator) and classical inverse problems (the same operators observed through a single experiment); both are discussed below.

Under the identity link with a *bounded* density ratio—plain covariate shift—combining the two samples merely matches the faster of the two individual rates (Pathak et al., 2022; Kpotufe & Martinet, 2021). An earlier work (Zhou & Yang, 2025) showed that this limitation can be overcome for suitably singular designs with an *unbounded* density ratio ($\rho(x) \to \infty$ as $x \to x_0$ for some x_0): the combined estimator then strictly outperforms either dataset alone in a sample-size window, the *Synergistic Learning Phenomenon* (SLP), quantified here by the *Synergistic Advantage Ratio* (SAR).

The two non-identity links have not been studied at all to the best of our knowledge. For every $x \in [0,1]$, they take the form

$$g(x) = \int_0^x f(t)\,dt \quad \text{(integral link)} \qquad \text{or} \qquad g(x) = \dot{f}(x) = \frac{d}{dx} f(x) \quad \text{(derivative link)}.$$

Here f has Hölder smoothness β, and the goal is to estimate g under the sampling and Gaussian noise assumptions of Section 3. Both experiments are informative about g, but through different channels and at different smoothness levels, which is what makes the combined rate nonobvious. The classical inverse-problem literature (Cavalier & Tsybakov, 2002; Efromovich & Koltchinskii, 2001) studies the same operators with a *single* dataset, observing either the source or the target regression.

We compare the minimax risk of estimating g from both samples with the risks based on either sample alone. An SLP occurs when the combined risk has a strictly faster rate than both individual risks.

As the source and target sample sizes n and n_T grow jointly, their relative scaling determines which sample alone provides the faster minimax rate and whether combining the samples yields a further rate improvement. We distinguish four types of sample-size regions according to these comparisons. In the *target dominating* (TD) region, the joint risk has the target-only rate; in the *target leading* (TL) region, the target provides the better individual rate and gains from source data; in the *source leading* (SL) region, the source provides the better individual rate and gains from target data; and in the *source dominating* (SD) region, the joint risk has the source-only rate. The SLP holds exactly in TL and SL. Figure 1 illustrates all four regions for representative choices of the functional link and design distributions. In these examples, the TL and SL regions meet at the *equalizing* scale where the individual minimax rates match. The corresponding results and Appendix C give the sample-size windows and SAR formulas for the other parameter settings. In Figure 1(c), the joint minimax rate changes from n_T^{-1} to $(nn_T)^{-1/2}$ within the SL region, at a different sample-size scale from that at which the individual minimax rates match.

We consider both bounded and unbounded density ratios $\rho = h_T/h_S$ on $[0,1]$. In the bounded-ratio setting, both design densities are bounded above and away from zero. In the unbounded-ratio setting, the source follows Beta$(a,1)$ with density $h_S(x) = ax^{a-1}$, $a > 1$, and the target is uniform, so $\rho(x) = a^{-1}x^{1-a} \to \infty$ as $x \to 0$. For these two settings, we characterize the sample-size windows, SAR orders, and mechanisms of synergy under the identity, integral, and derivative links. The two main results are as follows.

- **Integral link (Results II–III).** Under a bounded density ratio, combining an exactly unbiased IPW source pilot with target-based bias correction yields the joint rate $(nn_T)^{-1/2}$ in the synergistic sample-size window. Under a Beta$(a,1)$ source this persists for $1 < a < 2$; $a > 2$ opens a second, level-split SLP at rate n_T^{-1}; $a > 3$ replaces $(nn_T)^{-1/2}$ by a spatial rate; and $a = 2, 3$ carry logarithmic corrections (Theorems 3–5).
- **Derivative link (Result IV).** A joint local polynomial for the derivative coefficient yields an SLP window with fixed polynomial sample-size exponents only when $a > 2 + \frac{3}{2(\beta-1)}$. At equality, SLP remains possible in a subpolynomial boundary layer (Appendix C.11).

These results establish an asymmetry between integration and differentiation: in the bounded-density setting, the integral link admits synergy for suitable relative sample sizes even when the source and target covariate distributions coincide, whereas the derivative link does not. Under Beta-

source designs, the functional link determines the singularity requirements, sample-size windows, and attainable SAR orders. We construct estimators tailored to the functional link and the design distributions, with risks matching the minimax lower bounds up to multiplicative constants. This identifies both the gains available from combining the samples and the regimes in which no estimator can improve on the better individual minimax rate. We also restate the identity-link rates of Zhou & Yang (2025) as Result I, with an improved understanding at $a = 2 + 1/(2\beta)$, so the three links can be compared in one table.

Section 2 surveys related work. Section 3 recalls the known-link model and the SLP. Section 4 gives Results II–IV. Section 5 compares the three links. Section 6 reports simulations and real-data illustrations, and Section 7 concludes. Result I is in Appendix B; notation, deferred details, estimators, and empirical protocols are in Appendices A–F.2. Full proofs are in the companion supplement.

2 Related Work

The identity link: covariate shift. Under a *bounded* ratio, importance-weighted ERM (Cortes et al., 2010; Mansour et al., 2009; Sugiyama et al., 2007; Shimodaira, 2000) and kernel methods (Kpotufe & Martinet, 2021) attain the minimax rate, which equals the faster of the two individual rates (Pathak et al., 2022): no synergy. Under an *unbounded* ratio, Zhou & Yang (2025) showed that a density singularity can produce an SLP, the first rigorous instance of $\mathrm{SAR} \to \infty$. Result I restates their results for the identity link under covariate shift.

Non-identity links and inverse problems. Classical inverse-problem theory studies $g = \mathcal{A}f$ for a compact operator (Cavalier, 2008; Engl et al., 1996), with known single-dataset minimax rates (Cavalier & Tsybakov, 2002; Efromovich & Koltchinskii, 2001) and derivative estimation strictly harder than function estimation (Rice, 1984; Hall & Marron, 1990), but observes either f or g, never both; whether a second dataset on the other side of a known link gives a faster combined rate is what Results II–IV answer.

Unknown source–target relations. A broad transfer-learning literature allows the relation between source and target problems to be unknown but restricts the discrepancy to have additional structure. Examples include sparse coefficient differences (Li et al., 2022), aligned GLM coefficients (Tian & Feng, 2023), shared representations (Tripuraneni et al., 2021; Du et al., 2021), source–target function proximity (Reeve et al., 2021), close class posteriors (Cai & Wei, 2021), small transfer bias (Cai & Pu, 2024), and smoother additive offsets (Lin & Reimherr, 2024). Although the structural assumptions differ, the common mechanism is that an unknown cross-domain discrepancy is restricted to a simpler class, making transfer statistically beneficial. In our model, the target regression function is determined by the source regression function through the known relation $g = \mathcal{A}f$. Hence any improvement from combining the two samples must arise from how the known operator interacts with the two designs, rather than from estimating a simpler unknown source–target discrepancy. Throughout we use minimax rate analysis under Hölder smoothness (Stone, 1982; Tsybakov, 2009; Donoho & Liu, 1991; Nussbaum, 1999), with the SAR of Section 3 measuring when the combined rate strictly dominates either individual rate.

3 Problem Setup and the SLP Definition

We observe two independent samples, each i.i.d. within its domain, with covariates in $[0, 1]$ and regression models

$$Y_i = f(X_i) + \epsilon_i, \quad X_i \sim h_S, \qquad Y_j^T = g(X_j^T) + \epsilon_j^T, \quad X_j^T \sim h_T, \qquad g = \mathcal{A}f, \tag{1}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, n_T$. The operator $\mathcal{A}$ is known, and the goal is to estimate $g = \mathcal{A}f$. Throughout the minimax analysis, the noises are centered Gaussian and independent across observations, across the two samples, and of all covariates. Their source and target variances are fixed, positive, and may differ. The design densities are fixed, with their known or unknown status specified in each result. We write h for h_S when the source design is displayed explicitly. The Hölder class $\mathcal{H}(\beta, \kappa)$ and the comparison notation $\asymp, \preccurlyeq, \ll$ are collected in Appendix A.

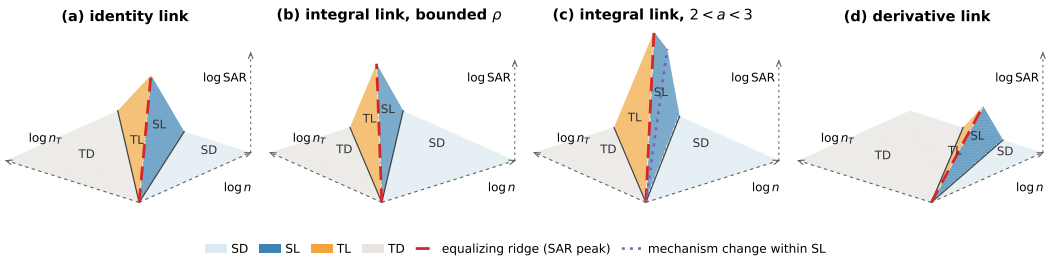

Figure 1: Representative sample-size phase diagrams for four link–design settings: (a) identity/Beta$(a, 1)$, (b) integral/bounded density ratio, (c) integral/Beta$(a, 1)$ with $2 < a < 3$, and (d) derivative/Beta$(a, 1)$. The SLP holds exactly in the SL and TL regions. The red dashed ray marks the equalizing scale where the individual minimax rates match. In (c), the purple dotted ray marks the within-SL transition from n_T^{-1} to $(nn_T)^{-1/2}$. Parameters: $(\beta, a) = (1, 4)$ in (a), $\beta = 1$ in (b), $(1, 2.5)$ in (c), and $(2, 6)$ in (d).

For the linked class $\mathcal{F} = \{(f, \mathcal{A}f) : f \in \mathcal{H}(\beta, \kappa)\}$, define the target $L_2(h_T)$ minimax risk

$$R_{h_T}(\mathcal{F}; n, n_T) = \inf_{\hat{g}} \sup_{f \in \mathcal{H}(\beta,\kappa)} \mathbb{E}_f \int_0^1 \big(\hat{g}(x) - (\mathcal{A}f)(x)\big)^2 h_T(x)\, dx. \tag{2}$$

The infimum is over all estimators based on the available samples, and $\mathbb{E}_f$ is expectation under the specified model. We write $R_{h_T}(g; n, n_T)$ when the linked class is clear. Setting $n_T = 0$ or $n = 0$ gives the source-only risk $R_{h_T}(\mathcal{F}; n, 0)$ or the target-only risk $R_{h_T}(\mathcal{F}; 0, n_T)$, respectively.

Definition 1 (Synergistic Advantage Ratio and SLP)**.** *The* Synergistic Advantage Ratio *(SAR) at* (n, n_T) *is*

$$\mathrm{SAR}(n, n_T) := \frac{\min\big(R_{h_T}(\mathcal{F}; n, 0),\ R_{h_T}(\mathcal{F}; 0, n_T)\big)}{R_{h_T}(\mathcal{F}; n, n_T)}.$$

We also write SAR *when the sample sizes are clear. The* Synergistic Learning Phenomenon *(SLP) occurs along a sample-size sequence if* $\mathrm{SAR}(n, n_T) \to \infty$ *as* $\min(n, n_T) \to \infty$*. Since* $\mathrm{SAR}(n, n_T) \geq 1$*,* $\mathrm{SAR}(n, n_T) = \Theta(1)$ *means no asymptotic rate improvement over the better single-domain estimator, although constant-factor improvements remain possible.*

For the identity link, Result I records when covariate shift produces an SLP. The rates, including a sharper bound at $a = 2 + 1/(2\beta)$, are in Appendix B.

4 Non-identity Links: Integration and Differentiation

Results II–IV below take $\mathcal{A}$ in (1) to be the integral $g(x) = \int_0^x f(t)\, dt$ or the derivative $g(x) = \dot{f}(x)$. The region labels in the rate displays refer to asymptotic sample-size sequences. The SL/TL windows have strict outer endpoints to ensure $\mathrm{SAR} \to \infty$, and share the equalizing scale where the two individual minimax risks are of the same order.

4.1 Result II: Integral Link with Bounded Density Ratio

4.1.1 Setting

The source observes $Y_i = f(X_i) + e_{2i}$ with $X_i \sim h$ and $f \in \mathcal{H}(\beta, \kappa)$; the target observes $Y_i^T = g(X_i^T) + e_{1i}$ with $X_i^T \sim h_T$ and $g(x) = \int_0^x f(t)\, dt \in \mathcal{H}(\beta + 1, \kappa')$ for some $\kappa' > 0$.

Under the conditions of Theorem 1, the individual rates are $R_{h_T}(g; n, 0) \asymp n^{-1}$, parametric because the pilot $\hat{g}_S(x) = n^{-1} \sum_{i \leq n} I(X_i \leq x) Y_i / h(X_i)$ is exactly unbiased for $g(x)$, and $R_{h_T}(g; 0, n_T) \asymp n_T^{-2(\beta+1)/(2\beta+3)}$. Rate optimality in (3) combines this pilot with a boundary-adaptive count-based target residual correction that uses no density estimate (Appendix D.2).

4.1.2 MAIN RESULT

Theorem 1 (Minimax rate for integral link, bounded ratio)**.** *Let the source and target densities satisfy $0 < c_h \le h(x), h_T(x) \le C_h < \infty$ for all $x \in [0,1]$, with constants c_h, C_h independent of the sample sizes. Suppose $g(x) = \int_0^x f(t)\,dt$ and $f \in \mathcal{H}(\beta,\kappa)$ with $\beta > 0$. Then*

$$R_{h_T}(g;\, n, n_T) \asymp \begin{cases} n^{-1}, & n_T \preccurlyeq n \quad \textit{(SD)}, \\ (n\, n_T)^{-1/2}, & n \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2(\beta+1)}} \quad \textit{(SL)}, \\ (n\, n_T)^{-1/2}, & n^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{3}$$

Rate-optimal estimators are in Appendix D.2. Here the source density h is assumed known; the target density h_T need not be. The ratio being upper bounded, there is no covariate-shift singularity and the SLP has a *different* mechanism.

Corollary 2 (SAR for integral link, bounded ratio)**.** *Under the conditions of Theorem 1, the SLP occurs if and only if $n \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$, the SAR being given in Appendix C.3.*

The source IPW estimator already attains the parametric rate n^{-1}, so synergy requires $n \ll n_T$. Within the SLP window, the SL and TL regions meet at $n_T \asymp n^{(2\beta+3)/(2(\beta+1))}$, where the SAR attains its largest order (Appendix C.3).

4.2 RESULT III: INTEGRAL LINK WITH UNBOUNDED DENSITY RATIO

4.2.1 SETTING

For $a > 1$ the source observes $Y_i = f(X_i) + e_{2i}$ with $X_i \sim h(x) = ax^{a-1}$ (Beta$(a,1)$), $f \in \mathcal{H}(\beta,\kappa)$ and anchor $g(0) = 0$, while the target observes $Y_i^T = g(X_i^T) + e_{1i}$ with $X_i^T \sim h_T =$ Uniform and $g(x) = \int_0^x f(t)\,dt$: the integral link of Result II combined with the singular design of Result I. Rate-optimal estimators are given in Appendix D.3.

4.2.2 MILD SINGULARITY: INTEGRAL SLP ONLY

Theorem 3 (Integral SLP for $1 < a < 2$)**.** *Let h and h_T be as defined above, and let $1 < a < 2$. Suppose $g(x) = \int_0^x f(t)\,dt$, $f \in \mathcal{H}(\beta,\kappa)$, $\beta > 0$, and $g(0) = 0$. Then*

$$R_{h_T}(g;\, n, n_T) \asymp \begin{cases} n^{-1}, & n_T \preccurlyeq n \quad \textit{(SD)}, \\ (nn_T)^{-1/2}, & n \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2(\beta+1)}} \quad \textit{(SL)}, \\ (nn_T)^{-1/2}, & n^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{4}$$

Here the source density h and the indices (a,β) are assumed known. Although $\rho \to \infty$, (4) coincides with the rate in Result II. The only synergistic rate is $(nn_T)^{-1/2}$, on $n \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$.

4.2.3 INTERMEDIATE SINGULARITY: LEVEL-SPLIT + INTEGRAL SLP

Theorem 4 (Extra level-split SLP for $2 < a < 3$)**.** *Let h and h_T be as defined above, and let $2 < a < 3$. Suppose $g(x) = \int_0^x f(t)\,dt$, $f \in \mathcal{H}(\beta,\kappa)$, $\beta > 0$, and $g(0) = 0$. Then*

$$R_{h_T}(g;\, n, n_T) \asymp \begin{cases} n^{-\frac{2(\beta+1)}{2\beta+a}}, & n_T \preccurlyeq n^{\frac{2(\beta+1)}{2\beta+a}} \quad \textit{(SD)}, \\ n_T^{-1}, & n^{\frac{2(\beta+1)}{2\beta+a}} \ll n_T \preccurlyeq n \quad \textit{(SL)}, \\ (nn_T)^{-1/2}, & n \preccurlyeq n_T \preccurlyeq n^{\frac{2\beta+3}{2\beta+a}} \quad \textit{(SL)}, \\ (nn_T)^{-1/2}, & n^{\frac{2\beta+3}{2\beta+a}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{5}$$

Here the source density h and the indices (a, β) are assumed known. Appendix C.5 explains why the source-only rate is slower than n^{-1} once $a > 2$ and how the risk splits into irregular running-integral functionals and the remaining structure. The SLP holds on $n^{2(\beta+1)/(2\beta+a)} \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$. Within SL, the rate changes from n_T^{-1} to $(nn_T)^{-1/2}$ at $n_T \asymp n$; the latter rate also holds throughout TL.

4.2.4 Strong singularity: level-split + covariate-shift SLP

Theorem 5 (Spatial singularity SLP for $a > 3$)**.** *Under the setting of Section 4.2.1 with $a > 3$,*

$$R_{h_T}(g;\, n, n_T) \;\asymp\; \begin{cases} n^{-\frac{2(\beta+1)}{2\beta+a}}, & n_T \preccurlyeq n^{\frac{2(\beta+1)}{2\beta+a}} \quad \textit{(SD)}, \\ n_T^{-1}, & n^{\frac{2(\beta+1)}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2\beta+a}} \quad \textit{(SL)}, \\ \left(n_T^{\frac{2\beta+1}{2\beta+3}} n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n^{\frac{2\beta+3}{2\beta+a}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{6}$$

Appendix C.6 explains why the spatial rate replaces $(nn_T)^{-1/2}$. The SLP again holds on $n^{2(\beta+1)/(2\beta+a)} \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$, with rate n_T^{-1} in SL and the spatial rate in TL. The change of rate coincides with the equalizing scale $n_T \asymp n^{(2\beta+3)/(2\beta+a)}$.

4.2.5 The cases $a = 2$ and $a = 3$, phase diagram, and SAR

At $a = 2$, the source-dominated risk becomes $(\log n)n^{-1}$ and the n_T^{-1} window shrinks to $n/\log n \ll n_T \preccurlyeq n$ (Proposition 11). At $a = 3$, the rate $(nn_T)^{-1/2}$ survives only with a logarithmic factor and the polynomial spatial mid-rate does not yet appear (Proposition 12). Why the rates change at these two cuts is discussed in Appendix C.7. The SLP windows, synergistic rates, and SAR values for every range of a are collected in Corollary 13 and Table 2 (Appendices C.8–C.9); in the SD and TD regions, $\mathrm{SAR} = \Theta(1)$. Full proofs are in supplement Section S.2.3.

4.3 Result IV: Derivative Link with Unbounded Density Ratio

4.3.1 Setting

The source observes $Y_i = f(X_i) + e_{2i}$ with $X_i \sim h(x) = ax^{a-1}$, $a > 1$, and $f \in \mathcal{H}(\beta, \kappa)$, $\beta > 1$; the target observes $Y_i^T = \dot{f}(X_i^T) + e_{1i}$ with $X_i^T \sim h_T = \mathrm{Uniform}$ and $\dot{f} \in \mathcal{H}(\beta - 1, \kappa'')$, so differentiation lowers smoothness by 1 while the design carries the Result I singularity.

The target-only rate is $R_T(n_T) \asymp n_T^{-2(\beta-1)/(2\beta-1)}$, while the source-only rate is $R_S(n) \asymp n^{-2(\beta-1)/(2\beta+1)}$ for $a < 2 + \frac{3}{2(\beta-1)}$, carries a $\log n$ factor at equality, and slows to $n^{-(2\beta-1)/(2\beta+a)}$ above it.

4.3.2 Main result

Theorem 6 (Minimax rate for derivative link under Beta$(a, 1)$)**.** *Let $a > 1$, $h(x) = ax^{a-1}$, $h_T \equiv 1$, and $f \in \mathcal{H}(\beta, \kappa)$, $\beta > 1$. The following rates hold.*

- *If $a < 2 + \frac{3}{2(\beta-1)}$:* $R_{h_T}(\dot{f};\, n, n_T) \asymp \min\left(n^{-\frac{2(\beta-1)}{2\beta+1}},\, n_T^{-\frac{2(\beta-1)}{2\beta-1}}\right)$.
- *If $a > 2 + \frac{3}{2(\beta-1)}$:*

$$R_{h_T}(\dot{f};\, n, n_T) \;\asymp\; \begin{cases} n^{-\frac{2\beta-1}{2\beta+a}}, & n_T \preccurlyeq n^{\frac{2\beta-1}{2\beta+a}} \quad \textit{(SD)}, \\ \left(n_T^{\frac{2\beta+1}{2\beta-1}} n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2(\beta-1)}{2\beta-1}}, & n^{\frac{2\beta-1}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{(2\beta-1)^2}{2(\beta-1)(2\beta+a)}} \quad \textit{(SL)}, \\ \left(n_T^{\frac{2\beta+1}{2\beta-1}} n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2(\beta-1)}{2\beta-1}}, & n^{\frac{(2\beta-1)^2}{2(\beta-1)(2\beta+a)}} \preccurlyeq n_T \ll n^{\frac{2\beta-1}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta-1)}{2\beta-1}}, & n_T \succcurlyeq n^{\frac{2\beta-1}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{7}$$

The rate (7) is achieved by the joint local-polynomial estimator (24) (Appendix D.4), whose source residuals enforce $R \approx f$ and target residuals $R' \approx \dot{f}$, with bandwidth set by the Fisher-correct balance (25) (source weight nHt^2, target weight $n_T H_T$). The indices (a, β) are assumed known. If $a < 2 + \frac{3}{2(\beta-1)}$, there is no singularity-driven SLP (same obstruction as Theorem 8), and no parametric IPW fallback: differentiation admits no exact-unbiasedness channel. If $a > 2 + \frac{3}{2(\beta-1)}$, the rates are *not* Result I with $\beta \mapsto \beta - 1$: source levels contribute Fisher weight nHt^2 rather than nH. The case $a = 2 + \frac{3}{2(\beta-1)}$ carries a logarithmic correction and is stated in Appendix C.11. The source-only rate under the Beta design, and why the SLP and SAR conclusions are unaffected by it, are recorded in Appendix C.10.

Corollary 7 (SAR for derivative link under Beta$(a, 1)$)**.** *Under the conditions of Theorem 6, for $a > 2 + \frac{3}{2(\beta-1)}$ the SLP occurs iff $n^{(2\beta-1)/(2\beta+a)} \ll n_T \ll n^{(2\beta-1)/(2\beta+1)}$, and the SAR, given in Appendix C.12, attains its largest order at the SL–TL equalizing scale $n_T \asymp n^{(2\beta-1)^2/(2(\beta-1)(2\beta+a))}$. For $a < 2 + \frac{3}{2(\beta-1)}$, $\mathrm{SAR}(n, n_T) = \Theta(1)$. At $a = 2 + \frac{3}{2(\beta-1)}$, the SLP occurs precisely in the subpolynomial boundary layer specified in Appendix C.11.*

Compared with Results I and III, the threshold $a > 2 + \frac{3}{2(\beta-1)}$ is the *strictest*: estimating $\dot{f}$ from source levels incurs the extra Fisher factor t^2, so a stronger singularity is needed than for a common $\mathcal{H}(\beta)$ regression. At equality the SAR is only subpolynomially large, so there is no fixed polynomial-exponent SLP window; see Appendix C.11. The full proof is in Section S.2.4 of the supplement.

4.4 Asymmetry Under Bounded Ratio

Results II–IV show that the density ratio and the functional link between the source and target regression functions jointly determine whether SLP is possible. To isolate the effect of the functional link, we now restrict to the bounded-density setting of Result II and compare the integral and derivative links. Under these conditions, the integral link admits SLP for suitable relative sample sizes (Result II) but the derivative link does not—a sharp asymmetry between the two links.

For the derivative link, the source observes $Y_i = f(X_i) + e_{2i}$ with $X_i \sim h$ and $f \in \mathcal{H}(\beta, \kappa)$, $\beta > 1$, and the target observes $Y_i^T = \dot{f}(X_i^T) + e_{1i}$ with $X_i^T \sim h_T$ and $\dot{f} \in \mathcal{H}(\beta - 1, \kappa'')$.

Theorem 8 (No SLP for derivative link under bounded ratio)**.** *Let the source and target densities satisfy $0 < c_h \le h(x), h_T(x) \le C_h < \infty$ for all $x \in [0, 1]$, with constants c_h, C_h independent of the sample sizes. Let $\beta > 1$ and use the linked class $\mathcal{F}^{\mathrm{BD}}_{\beta,\kappa} = \{(f, g) : f \in \mathcal{H}(\beta, \kappa),\ g = f'\}$. Write $\mathcal{R}^{\mathrm{G}}_{\mathrm{BD}}(n, n_T) := R_{h_T}(\mathcal{F}^{\mathrm{BD}}_{\beta,\kappa}; n, n_T)$ for its minimax risk under the Gaussian model in Section 3. Then*

$$\mathcal{R}^{\mathrm{G}}_{\mathrm{BD}}(n, n_T) \asymp \min\big(n^{-\frac{2(\beta-1)}{2\beta+1}},\ n_T^{-\frac{2(\beta-1)}{2\beta-1}}\big).$$

The rate is achieved by a joint local-polynomial estimator. $\mathcal{R}^{\mathrm{G}}_{\mathrm{BD}}(n, n_T)$ is the minimum of the two individual risks up to constants, so $\mathrm{SAR}(n, n_T) = \Theta(1)$ for every sample-size sequence, and the SLP does not occur.

The upper bounds for Results II–IV and Theorem 8, in all stated parameter cases, remain valid for independent observations with conditionally mean-zero noises and uniformly bounded conditional second moments.

Remark 1 (Integration–differentiation asymmetry under bounded ratio)**.** *Comparing Theorem 1 with Theorem 8 gives a sharp directional barrier under their common density conditions, with the source density known in both experiments: for $g = \int_0^{\cdot} f$ the SLP occurs on the TL/SL window of Corollary 2 and $\mathrm{SAR}(n, n_T) \to \infty$, whereas for $g = \dot{f}$ it occurs for no (n, n_T) and $\mathrm{SAR}(n, n_T) = \Theta(1)$. The barrier is specific to the bounded-ratio regime. Under the Beta model, the integral link admits SLP for every $a > 1$ (Result III). The derivative link requires $a > 2 + \frac{3}{2(\beta-1)}$ for a polynomial SLP window, with a subpolynomial SLP window at equality (Result IV).*

Integration raises smoothness and admits an exactly unbiased estimator, whereas differentiation lowers smoothness and amplifies source noise (Appendix C.13).

5 Unified Characterization of the SLP

As summarized in Table 1 in Appendix C, the identity link needs a strong singularity, the integral link none, and the derivative link a strictly stronger one: the link sets the price the design pays.

Mechanisms underlying the SLP. Each gain above is a division of labor, not a pooling of samples. Integration averages noise, so the source pilot $\hat{g}_S$ needs no smoothing and its error is pure variance. The target observes g directly and corrects that pilot on a window of width b_T. Balancing b_T/n against $1/(n_T b_T)$ at $b_T = \sqrt{n/n_T}$ gives the $(nn_T)^{-1/2}$ of Result II. A density singularity adds a second division. For $a > 2$, source estimation of the cumulative coefficients of g is limited by the sparse observations near zero; the target estimates this finite-dimensional component at rate n_T^{-1}. If the singularity is strong enough, the two samples also split space: the source carries the bulk and the target the neighborhood of the singularity (Result III). The derivative link reverses the first mechanism. Differentiation amplifies noise, and on a window of scale t a source observation of f informs $\dot{f}$ only after a factor t^2. The source is then a weak channel for the target estimand. Only a much stronger singularity restores synergy (Result IV), and under a bounded ratio nothing does (Theorem 8). Matching lower bounds use alternatives that preserve the functional link and control the joint information from both samples; see Table 1 in the supplement for the corresponding proofs.

Necessary conditions for SLP. Three ingredients recur. (I) *Difficulty asymmetry*: the two problems differ in intrinsic difficulty, in effective sample size (Results I and III) or in smoothness class (Results II and IV). (II) *Complementary information*: each fills a gap the other cannot, spatially near and away from the singularity (Results I and III) or as an unbiased IPW pilot against target bias correction (Result II). (III) *Sample-size window in which SAR diverges*: the SLP holds on the middle wedges of Figure 1, always bracketed by SD and TD, with the SAR attaining its largest order at the equalizing scale in the displayed settings.

Threshold ordering across links. The singularity needed for a polynomial SLP window is ordered by the link: none for the integral link (phase cuts at $a = 2, 3$), $a > 2 + \frac{1}{2\beta}$ for the identity link, and $a > 2 + \frac{3}{2(\beta-1)}$ for the derivative link, the strictest. Condition (I) alone is not enough: the bounded-ratio derivative link satisfies it and still admits no gain (Theorem 8).

6 Numerical and Real-Data Experiments

We use simulations and an analysis of IO-VNBD to examine how the density ratio, functional link, and relative sample sizes affect the gains from combining source and target data. Simulations compare source-only, target-only, and joint estimators under controlled bounded and Beta-source designs. The real-data analysis compares integral and derivative links between two observed signals from the same trajectories. A separate semi-synthetic study examines unbounded covariate shift using constructed designs and functions with empirically calibrated observation errors (Appendix F.2).

Simulations. Figure 2 compares the integral and derivative links under uniform source and target designs, focusing on the effects of relative sample sizes and the contrast between the two functional links. Full experimental details and additional simulations under Beta-source designs are provided in Appendix E.

For the fixed truth in Figure 2, the balanced integral path $\kappa_n = 0.8$ has tail SAR slope 0.112 $[0.065, 0.159]$, versus the theoretical reference 0.100. The fixed-target scan peaks at $\kappa_n = 0.800$, while the derivative control remains near one, ending at 0.998 $[0.989, 1.007]$. Integral SAR uses the minimum of candidate mean risks, not an implemented selector, so its lower bound of one is imposed by construction. The scan's outer baseline follows prescribed switching and does not independently validate the no-SLP regions or their boundaries.

Real-data analysis. The public IO-VNBD benchmark (Onyekpe et al., 2021) records vehicle and smartphone measurements on public-road trajectories. We use vehicle yaw rate recorded through the controller area network (CAN) and smartphone GPS heading, synchronized at 10 Hz. Yaw rate measures the vehicle's turning rate, while GPS heading records its direction of travel, providing

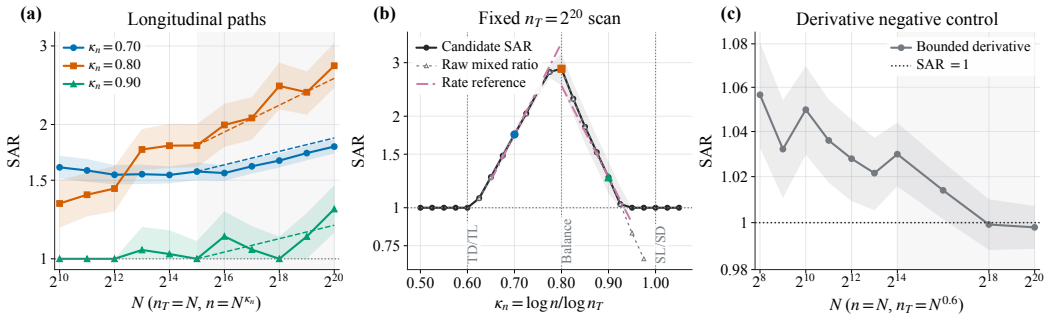

Figure 2: Integral–derivative comparison under uniform designs. (a) Integral candidate SAR for $n_T = N$, $n = N^{\kappa_n}$ with $\kappa_n = 0.7, 0.8, 0.9$ and $\beta = 1$. (b) Fixed-target scan at $n_T = 2^{20}$; gray triangles show the raw mixed ratio only for $0.6 < \kappa_n < 1$. (c) Reversed derivative control for the same function pair, with $\beta = 2$, $n = N$, $n_T = N^{0.6}$, and raw joint SAR. Bands are pointwise paired-bootstrap 95% intervals. Dashed and dash–dot guides show theoretical slopes, vertical dotted lines mark theoretical boundaries and the equalizer, and shaded regions mark fitting windows. Full design and display details are in Appendix E.1.

a physical basis for the functional links studied here. After heading unwrapping, sign-and-scale calibration, and normalization, the heading change g is approximately the integral of the yaw-rate signal f on the audited windows, $g(x) \approx \int_0^x f(u)\,du$. Taking yaw rate as the source response and heading change as the target response gives the integral task; exchanging their roles gives the derivative task. The dataset therefore permits comparison of the two links using the same pair of observed signals, without constructing synthetic responses. Data selection and calibration are detailed in Appendix F.1.

Across 21 audited windows from 11 signal-distinct tracks, the terminal track-geometric SAR is $1.385\ [1.126, 1.699]$ on the $\kappa = 0.8$ integration path and $0.954\ [0.930, 0.976]$ for safeguarded differentiation (Figure 7(b) in Appendix F.1). These exploratory 95% track-bootstrap intervals condition on the selected windows and do not account for window-selection uncertainty. Finite-range rate discrepancies remain, and these illustrations do not establish minimax rates.

7 Discussion and Conclusion

Under a known functional link, synergy arises when the source and target samples provide complementary information about the target regression function, allowing joint estimation to improve on both individual minimax rates for suitable relative sample sizes. When both design densities are bounded above and away from zero, integration admits SLP for suitable relative sample sizes, whereas differentiation does not. Under Beta-source designs, the two links have different singularity requirements and sample-size windows for SLP. We also construct estimators tailored to the functional link and the design distributions, and prove that they attain the corresponding minimax rates.

Limitations. Only one-dimensional regression is studied. The rate-optimal estimators require knowledge of the source density and (a, β), so the adaptation issue remains open, as do extensions to higher-order links and multivariate singularities. Only asymptotic results on convergence rates are provided.

AI USE STATEMENT

In this work, we used a generative AI tool to assist with the generation of numerical simulation code, the preprocessing of data for the empirical analysis and the visual presentation of experimental figures.

Additionally, we used a generative AI tool for grammar checking, language polishing, manuscript organization and typesetting. It shortened prose, adjusted LaTeX formatting, and helped prepare figure scripts. We did not use it to generate research ideas, theorem statements, or proofs. The authors derived and verified every result, checked every AI-assisted edit, and take responsibility for the final content.

REPRODUCIBILITY STATEMENT

The companion supplement provides complete proofs for Results II–IV, the bounded-ratio integration–differentiation comparison. Complete proofs for Result I are taken from Zhou & Yang (2025). Appendix D constructs the estimators that attain each upper bound. The assumptions are stated with the results and collected in Section S.1 of the companion supplement. For Section 6, Appendix E gives the simulation designs, truth libraries, repetition counts, and bootstrap procedures. Appendix F.1 gives the IO-VNBD window audit, synchronization, calibration, and sampling protocol. Appendix F.2 gives the semi-synthetic calibration. The IO-VNBD benchmark (Onyekpe et al., 2021) is public. Anonymized code reproducing every figure and table is submitted as supplementary material.

REFERENCES

T. Tony Cai and Hongming Pu. Transfer learning for nonparametric regression: Non-asymptotic minimax analysis and adaptive procedure. *arXiv preprint arXiv:2401.12272*, 2024.

T Tony Cai and Hongji Wei. Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *The Annals of Statistics*, 49(1):100–128, 2021.

Laurent Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

Laurent Cavalier and Alexandre Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123(3):323–354, 2002. doi: 10.1007/s004400100169.

Corinna Cortes, Yishay Mansour, and Mehryar Mohri. Learning bounds for importance weighting. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/hash/59c33016884a62116be975a9bb8257e3-Abstract.html.

David L Donoho and Richard C Liu. Geometrizing rates of convergence, III. *The Annals of Statistics*, 19(2):668–701, 1991.

Simon S Du, Wei Hu, Sham M Kakade, Jason D Lee, and Qi Lei. Few-shot learning via learning the representation, provably. In *International Conference on Learning Representations*, 2021.

Sam Efromovich and Vladimir Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001. doi: 10.1109/18.959267.

Heinz W Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

Peter Hall and J S Marron. On variance estimation in nonparametric regression. *Biometrika*, 77(2):415–419, 1990.

Samory Kpotufe and Guillaume Martinet. Marginal singularity and the benefits of labels in covariate-shift. *The Annals of Statistics*, 49(6):3299–3323, 2021.

Sai Li, T Tony Cai, and Hongzhe Li. Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B*, 84(1):149–173, 2022.

Haotian Lin and Matthew Reimherr. Smoothness adaptive hypothesis transfer learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30286–30316. PMLR, 2024. URL https://proceedings.mlr.press/v235/lin24q.html.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. *arXiv preprint arXiv:0902.3430*, 2009.

Michael Nussbaum. Minimax risk: Pinsker bound. In *Encyclopedia of Statistical Sciences, Update Volume 3*, pp. 451–460. Wiley, New York, 1999.

Uche Onyekpe, Vasile Palade, Stratis Kanarachos, and Alicja Szkolnik. IO-VNBD: Inertial and odometry benchmark dataset for ground vehicle positioning. *Data in Brief*, 35:106885, 2021. doi: 10.1016/j.dib.2021.106885.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi: 10.1109/TKDE.2009.191.

Reese Pathak, Cong Ma, and Martin Wainwright. A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17517–17530. PMLR, 2022. URL https://proceedings.mlr.press/v162/pathak22a.html.

Henry W. J. Reeve, Timothy I. Cannings, and Richard J. Samworth. Adaptive transfer learning. *The Annals of Statistics*, 49(6):3618–3649, 2021. doi: 10.1214/21-AOS2102.

John A Rice. Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4): 1215–1230, 1984.

Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982. doi: 10.1214/aos/1176345969.

Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul Buenau, and Motoaki Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL https://proceedings.neurips.cc/paper_files/paper/2007/hash/be83ab3ecd0db773eb2dc1b0a17836a1-Abstract.html.

Ye Tian and Yang Feng. Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association*, 118(544):2684–2697, 2023. doi: 10.1080/01621459.2022.2071278.

Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264. IGI Global, 2010. doi: 10.4018/978-1-60566-766-9.ch011.

Nilesh Tripuraneni, Chi Jin, and Michael I Jordan. Provable meta-learning of linear representations. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10434–10443. PMLR, 2021. URL https://proceedings.mlr.press/v139/tripuraneni21a.html.

Alexandre B Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):9, 2016.

Ling Zhou and Yuhong Yang. On a synergistic learning phenomenon in nonparametric domain adaptation. arXiv:2511.17009, 2025. URL https://arxiv.org/abs/2511.17009.

A Function Class and Notation

With $r_\beta = \lceil \beta \rceil - 1$ and $\alpha_\beta = \beta - r_\beta \in (0, 1]$, the Hölder class $\mathcal{H}(\beta, \kappa)$ collects the f with $\max_{j \le r_\beta} \|f^{(j)}\|_\infty \le \kappa$ whose r_β-th derivative is α_β-Hölder with constant κ; integer smoothness m thus means $C^{m-1,1}$, supplying the order-m Taylor remainder used by the local-polynomial proofs. Throughout, b is a bandwidth, the Beta shape a is fixed, and c_j, C_j are constants independent of (n, n_T). For positive sequences, $A \preccurlyeq B$ means $A = O(B)$, $A \succcurlyeq B$ means $B = O(A)$, and $A \asymp B$ means both. We write $A \ll B$ for $A = o(B)$, i.e. $A/B \to 0$, and $A \gg B$ for $B = o(A)$, along the sample-size sequence under consideration. We use $\log_+(u) := \max\{\log u, 0\}$ for $u > 0$.

B The Identity Link: Covariate Shift

With $\mathcal{A} = \mathrm{Id}$ in (1) both domains regress the same function, so the design pair is the only structural difference and $\rho = h_T/h$ is the sole parameter. This cell is settled by Zhou & Yang (2025); we record it in SAR language as the reference point for the non-identity links, and add the boundary case its analysis left open. Take a Beta$(a, 1)$ source, $h(x) = ax^{a-1}$, a uniform target, and $f \in \mathcal{H}(\beta, \kappa)$, so $\rho(x) \to \infty$ as $x \to 0$ once $a > 1$. For $a > 2 + \frac{1}{2\beta}$, the SLP occurs precisely when $n^{(2\beta+1)/(2\beta+a)} \ll n_T \ll n$; there the minimax rate is $(n_T/n)^{1/(a-1)} n_T^{-2\beta/(2\beta+1)}$, faster than either domain alone, the plane splits into the four regions SD/SL/TL/TD, and the SAR attains its largest order at the equalizing scale $n_T \asymp n^{(2\beta+1)^2/(2\beta(2\beta+a))}$. At $a = 2 + \frac{1}{2\beta}$, which is new here, the rate acquires a logarithmic correction and the SLP survives only on the narrow band $n_T/n \to 0$ with $\log(n/n_T) = o(\log n)$. Appendix C.1 gives the full theorem, the SAR formulas, the regions, and the spread-function Nadaraya–Watson estimator that attains the rate.

C Deferred Material for Results I–IV

This appendix collects a summary of Results I–IV, the mechanism discussions, sample-size-region descriptions, rates with logarithmic corrections, and the Result III phase table that are referenced from the main text. The displayed rate and SAR formulas use the same region labels as the main text: SL and TL denote sequences on which the SAR diverges, with strict outer endpoints. The intermediate risk ranges at the identity- and derivative-link thresholds are stated separately from their SLP conditions.

For the repeated Result III comparisons in this appendix, write $n_T^{\mathrm{TD}} \asymp n^{(2\beta+3)/(2\beta+1)}$ and, for $a > 2$, $n_T^{\mathrm{cum}} \asymp n^{2(\beta+1)/(2\beta+a)}$. For $a > 3$, define the spatial-transition scale and rate benchmark by

$$n_T^{\mathrm{sp}} := n^{\frac{2\beta+3}{2\beta+a}}, \qquad R_{\mathrm{spat}}(n, n_T; a, \beta) := \left(n_T^{\frac{2\beta+1}{2\beta+3}} n^{-1} \right)^{\frac{1}{a-1}} n_T^{-\frac{2(\beta+1)}{2\beta+3}}. \tag{8}$$

R_{spat} is a rate benchmark rather than an exact risk. It equals n_T^{-1} at $n_T = n_T^{\mathrm{sp}}$ and matches the target-only rate at $n_T \asymp n_T^{\mathrm{TD}}$.

C.1 Full statement, SAR, and sample-size regions for Result I

The target-only rate is $n_T^{-2\beta/(2\beta+1)}$; the source-only rate is $n^{-2\beta/(2\beta+1)}$ for $a < 2 + \frac{1}{2\beta}$, carries a $\log(en)$ factor at equality, and slows to $n^{-(2\beta+1)/(2\beta+a)}$ above it.

Theorem 9 (Minimax rate under covariate shift)**.** *Let* $h(x) = ax^{a-1}$, $h_T = \mathrm{Uniform}$, *and* $f \in \mathcal{H}(\beta, \kappa)$. *For any* $a > 0$, $\beta > 0$*:*

- *If* $a < 2 + \frac{1}{2\beta}$*:* $R_{h_T}(\mathcal{H}(\beta, \kappa); n, n_T) \asymp (n + n_T)^{-2\beta/(2\beta+1)}$. *No SLP.*
- *If* $a = 2 + \frac{1}{2\beta}$*: the rate is* $\mathfrak{C}_\beta(n, n_T)$ *of* (11)*, stated in Appendix C.2.*

Table 1: Summary of Results I–IV and the bounded-ratio negative result; $\mathfrak{C}_\beta(n, n_T)$ is the identity-link rate at $a = 2 + 1/(2\beta)$ in (11).

Result	Setting	Condition for SLP	Synergistic rate	SAR
I	Identity link (covariate shift), $\rho \to \infty$	$a > 2 + \frac{1}{2\beta}$: polynomial SL/TL window; equality: $\log(n/n_T) = o(\log n)$ and $n_T/n \to 0$	above threshold: $\left(n_T n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2\beta}{2\beta+1}}$; equality: $\mathfrak{C}_\beta(n, n_T)$	Eqs. (10) and (12)
II	Integral link, ρ bounded	$n \ll n_T \ll n^{\frac{2\beta+3}{2\beta+1}}$ (SL/TL; no SLP in SD/TD)	$(nn_T)^{-1/2}$	SL/TL: see (13)
III	Integral link, Beta$(a, 1)$ (phase at $a = 2, 3$)	$1 < a < 2$: as II; $2 < a < 3$: $n_T^{\mathrm{cum}} \ll n_T \ll n_T^{\mathrm{TD}}$; $a > 3$: same window, spatial mid-rate	see (4)–(6); n_T^{-1}, $(nn_T)^{-1/2}$, or $R_{\mathrm{spat}}(n, n_T; a, \beta)$	Cor. 13; Table 2
IV	Derivative link, Beta$(a, 1)$ $(\beta > 1)$	$a > 2 + \frac{3}{2(\beta-1)}$; $n^{\frac{2\beta-1}{2\beta+a}} \ll n_T \ll n^{\frac{2\beta-1}{2\beta+1}}$ (at equality: the subpolynomial layer in Cor. 7; no SLP below the threshold)	above threshold: $\left(n_T^{\frac{2\beta+1}{2\beta-1}} n^{-1}\right)^{\frac{1}{a-1}} \times n_T^{-\frac{2(\beta-1)}{2\beta-1}}$; equality: (19)	see (20)
Neg.	Derivative link, ρ *bounded*	*Never*	$\min(n^{-\frac{2(\beta-1)}{2\beta+1}}, n_T^{-\frac{2(\beta-1)}{2\beta-1}})$	$\Theta(1)$

- *If* $a > 2 + \frac{1}{2\beta}$*:*

$$R_{h_T}(\mathcal{H}(\beta, \kappa);\, n, n_T) \asymp \begin{cases} n^{-\frac{2\beta+1}{2\beta+a}}, & n_T \preccurlyeq n^{\frac{2\beta+1}{2\beta+a}} \quad \textit{(SD)}, \\ \left(n_T n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2\beta}{2\beta+1}}, & n^{\frac{2\beta+1}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{(2\beta+1)^2}{2\beta(2\beta+a)}} \quad \textit{(SL)}, \\ \left(n_T n^{-1}\right)^{\frac{1}{a-1}} n_T^{-\frac{2\beta}{2\beta+1}}, & n^{\frac{(2\beta+1)^2}{2\beta(2\beta+a)}} \preccurlyeq n_T \ll n \quad \textit{(TL)}, \\ n_T^{-\frac{2\beta}{2\beta+1}}, & n_T \succcurlyeq n \quad \textit{(TD)}. \end{cases} \tag{9}$$

The rates for $a \neq 2 + 1/(2\beta)$ are from Zhou & Yang (2025); at $a = 2 + 1/(2\beta)$, we sharpen the bound by specifying its logarithmic factor (see Appendix C.2).

The upper bounds in all stated parameter cases remain valid for independent observations with conditionally mean-zero sub-Gaussian noises and uniformly bounded conditional sub-Gaussian parameters. The conditional-variance argument is given in supplement Section S.2.5.

Corollary 10 (SAR under covariate shift)**.** *For $a > 2 + \frac{1}{2\beta}$, the SLP occurs if and only if $n^{(2\beta+1)/(2\beta+a)} \ll n_T \ll n$, with the SAR given in (10) below. At $a = 2 + 1/(2\beta)$ the SLP occurs exactly when $n_T/n \to 0$ with $\log(n/n_T) = o(\log n)$. For instance, $n_T = n/(\log n)^k$ qualifies while $n_T = n^\tau$, $\frac{2\beta}{2\beta+1} < \tau < 1$, does not. The explicit SAR at equality is given in Appendix C.2.*

In that window,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} \left(n_T\, n^{-\frac{2\beta+1}{2\beta+a}}\right)^{\frac{2\beta}{2\beta+1}-\frac{1}{a-1}}, & n^{\frac{2\beta+1}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{(2\beta+1)^2}{2\beta(2\beta+a)}} \quad \text{(SL)}, \\ \left(nn_T^{-1}\right)^{\frac{1}{a-1}}, & n^{\frac{(2\beta+1)^2}{2\beta(2\beta+a)}} \preccurlyeq n_T \ll n \quad \text{(TL)}. \end{cases} \tag{10}$$

For $a > 2 + 1/(2\beta)$, three boundaries split the (n, n_T)-plane into four sample-size regions (SSR), in which source and target play different roles (Zhou & Yang, 2025):

1. **Source Dominating (SD)** ($n_T \preccurlyeq n^{(2\beta+1)/(2\beta+a)}$): n_T is too small to help near the singularity $x = 0$; the rate equals the source-only rate and $\mathrm{SAR}(n, n_T) = \Theta(1)$.
2. **Source Leading (SL)** ($n^{(2\beta+1)/(2\beta+a)} \ll n_T \preccurlyeq n^{(2\beta+1)^2/(2\beta(2\beta+a))}$): source alone is still better than target alone, but target data assist near $x = 0$ where h vanishes; $\mathrm{SAR}(n, n_T) \to \infty$ and its rate increases with n_T.
3. **Target Leading (TL)** ($n^{(2\beta+1)^2/(2\beta(2\beta+a))} \preccurlyeq n_T \ll n$): target alone is already better; the larger source sample assists away from $x = 0$ where h is not small; $\mathrm{SAR}(n, n_T) \asymp (n/n_T)^{1/(a-1)}$ has a decreasing rate that returns to order one at the TD boundary.

4. **Target Dominating (TD)** ($n_T \succcurlyeq n$): source data no longer improve the rate; $\mathrm{SAR}(n, n_T) = \Theta(1)$.

SLP holds only in SL and TL. The equalizing size $n_T \asymp n^{(2\beta+1)^2/(2\beta(2\beta+a))}$ is where the two individual rates match and the SAR attains its largest order.

The rate (9) is achieved by the spread-function Nadaraya–Watson estimator ($\beta \le 1$) or its boundary-adaptive local-polynomial version ($\beta > 1$); at $a = 2 + 1/(2\beta)$ the same multiscale construction accumulates one $n^{-2\beta/(2\beta+1)}$ contribution per active dyadic shell (Appendix D.1).

C.2 SAR FOR THE IDENTITY LINK AT $a = 2 + 1/(2\beta)$

At $a = 2 + 1/(2\beta)$ the common-function rate $\mathfrak{C}_\beta(n, n_T)$, which gives the *order* of the joint risk when both samples observe the same function, is defined for $n, n_T \ge 1$ by

$$\mathfrak{C}_\beta(n, n_T) := \begin{cases} n^{-\frac{2\beta}{2\beta+1}} \log(en), & n_T \le n^{\frac{2\beta}{2\beta+1}}, \\ n^{-\frac{2\beta}{2\beta+1}} \{1 + \log(n n_T^{-1})\}, & n^{\frac{2\beta}{2\beta+1}} < n_T < n, \\ n_T^{-\frac{2\beta}{2\beta+1}}, & n_T \ge n. \end{cases} \tag{11}$$

$$R_{h_T}(\mathcal{H}(\beta, \kappa);\, n, n_T) \asymp \mathfrak{C}_\beta(n, n_T).$$

In the intermediate risk regime $n^{2\beta/(2\beta+1)} < n_T < n$ the SAR is

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} \dfrac{\log(en)}{1 + \log(n n_T^{-1})}, & n^{\frac{2\beta}{2\beta+1}} < n_T \le n\{\log(en)\}^{-\frac{2\beta+1}{2\beta}}, \\ \dfrac{(n n_T^{-1})^{\frac{2\beta}{2\beta+1}}}{1 + \log(n n_T^{-1})}, & n\{\log(en)\}^{-\frac{2\beta+1}{2\beta}} < n_T < n. \end{cases} \tag{12}$$

The crossover follows by equating the individual rates $n^{-2\beta/(2\beta+1)} \log(en)$ and $n_T^{-2\beta/(2\beta+1)}$. Outside this intermediate regime, $\mathrm{SAR}(n, n_T) \asymp 1$. The intermediate risk regime is not itself an SLP window: at this value of a, SLP occurs exactly when $\log(n/n_T) = o(\log n)$ and $n_T/n \to 0$, or equivalently $z = n/n_T \to \infty$ with $\log z = o(\log n)$. For example, $n_T = n/(\log n)^k$ gives SLP for every fixed $k > 0$, whereas $n_T = n^\tau$ with fixed $\frac{2\beta}{2\beta+1} < \tau < 1$ does not. The derivation is given in Corollary S.2.3.

C.3 SAR AND SAMPLE-SIZE REGIONS FOR RESULT II

In the SLP window of Corollary 2,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} (n_T n^{-1})^{1/2}, & n \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2(\beta+1)}} \quad \text{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{1/2}, & n^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \text{(TL)}. \end{cases} \tag{13}$$

As in Result I, three sample-size boundaries induce four SSR regions. Here synergy requires $n \ll n_T$, because the source IPW estimator already attains the parametric rate n^{-1}. Writing all boundaries in terms of n_T as in (3),

1. **Source Dominating (SD)** ($n_T \preccurlyeq n$): target data do not improve the source-only rate; $\mathrm{SAR}(n, n_T) = \Theta(1)$.
2. **Source Leading (SL)** ($n \ll n_T \preccurlyeq n^{(2\beta+3)/(2(\beta+1))}$): source alone is better; target assists via bias correction, pushing the rate from n^{-1} to $(n n_T)^{-1/2}$; $\mathrm{SAR}(n, n_T) \asymp (n_T n^{-1})^{1/2} \to \infty$.
3. **Target Leading (TL)** ($n^{(2\beta+3)/(2(\beta+1))} \preccurlyeq n_T \ll n^{(2\beta+3)/(2\beta+1)}$): target alone is better; source assists via the exactly unbiased IPW pilot; $\mathrm{SAR}(n, n_T) \asymp (n n_T^{-(2\beta+1)/(2\beta+3)})^{1/2} \to \infty$.
4. **Target Dominating (TD)** ($n_T \succcurlyeq n^{(2\beta+3)/(2\beta+1)}$): source data do not improve the target-only rate; $\mathrm{SAR}(n, n_T) = \Theta(1)$.

SLP holds only in SL and TL. The equalizing size $n_T \asymp n^{(2\beta+3)/(2(\beta+1))}$ is where the two individual rates match and the SAR attains its largest order.

C.4 Boundary bases and the relaxed density assumption for Result III

For $1 < a < 4$, standard boundary wavelets with the stated moments give the weighted primitive bound directly. For $a \geq 4$, the supplement constructs localized, moment-preserving boundary analysis functions with the extra zero-end jets and proves global Riesz stability by an operator-norm perturbation argument. The dual reconstruction basis need not be local; its global Riesz bound suffices for the coefficientwise estimator. Thus the displays below are unconditional minimax equivalences for every fixed $a > 1$.

Remark 2 (Relaxed density assumption for Result III)**.** *Result III uses $h(x) = ax^{a-1}$ with $a > 1$, so $h \to 0$ and $h_T/h \to \infty$ at the origin. The positive lower bound on the source density in Theorem 1 therefore fails. The range $a \leq 1$, where this density-ratio singularity is absent, is not part of Result III.*

C.5 Mechanisms and sample-size regions for Result III, $2 < a < 3$

Remark 3 (Why the source-only rate is slower than n^{-1} for $a > 2$)**.** *Source data observe f, while the target parameter is the running integral $g(x) = \int_0^x f(t)\,dt$. Any nontrivial change in g under the L_2 loss must come from a change in f; under Hölder smoothness, a perturbation of f supported near the origin on an interval of length ε can alter g by at most order $\varepsilon^{\beta+1}$ on a set of positive measure. At the same time, the Beta$(a, 1)$ design places mass only of order ε^a on that interval, so the source sample contains vanishingly little information about such a perturbation as $\varepsilon \to 0$. When $a > 2$, this imbalance is severe enough that the integral functional $f \mapsto g$ is no longer estimable at the parametric rate from source data alone: distinguishing alternatives that differ by $\varepsilon^{\beta+1}$ in g requires a source sample size of order $\varepsilon^{-(2\beta+a)}$, which forces*

$$R_{h_T}(g;\, n, 0) \;\succcurlyeq\; n^{-2(\beta+1)/(2\beta+a)}.$$

For $a > 2$ the right-hand side is strictly slower than n^{-1}. For $1 < a < 2$, this boundary lower bound is weaker than n^{-1}. The parametric floor instead follows from a fixed smooth low-frequency two-point construction (Section S.2.3.2 of the supplement). The matching upper bounds are in Appendix D.3.

Remark 4 (Role split and the two synergistic rates)**.** *Write the risk as the sum of two pieces: the running-integral functionals of g that are irregular for the source when $a > 2$, and the remaining structure of g.*

- ***Irregular running-integral functionals.*** *By Remark 3, source-only estimation of these functionals has rate $n^{-2(\beta+1)/(2\beta+a)}$. The target observes g directly, so the same functionals are ordinary parametric problems with rate n_T^{-1}. Target alone estimating the* whole *of $g \in \mathcal{H}(\beta+1)$ would still cost the nonparametric rate $n_T^{-2(\beta+1)/(2\beta+3)}$; the rate n_T^{-1} applies only to this finite-dimensional piece.*
- ***Remaining structure.*** *For $a < 3$ this part remains parametrically estimable from the source (observations of f), at rate n^{-1}. The Result II-type combination with the target improves it further to $(nn_T)^{-1/2}$, but only when $n_T \gtrsim n$; for $n_T \preccurlyeq n$ one has $(nn_T)^{-1/2} \succcurlyeq n^{-1}$, so the source-only rate n^{-1} is already optimal for this piece.*

Adding the two contributions gives the ordering in (5). In the window $n_T^{\mathrm{cum}} \preccurlyeq n_T \preccurlyeq n$,

$$R_{h_T}(g;\, n, n_T) \;\asymp\; \underbrace{n_T^{-1}}_{\text{irregular functionals (target)}} + \underbrace{n^{-1}}_{\text{remaining structure (source)}} \;\asymp\; n_T^{-1},$$

since $n_T^{-1} \geq n^{-1}$. This is the level-split regime. Once $n_T \gtrsim n$, the remaining structure can be combined at rate $(nn_T)^{-1/2}$, which then dominates n_T^{-1} and yields the integral regime of (5).

Relative to Theorem 3, the case $a > 2$ produces an additional synergistic rate n_T^{-1}, beyond the integral-type rate $(nn_T)^{-1/2}$. The boundaries n_T^{cum}, n, and n_T^{TD} separate four rate regimes (cf. Remarks 3–4). Write $n_T^{\mathrm{eq}} := n^{(2\beta+3)/(2\beta+a)}$ for the size at which the two individual rates $n^{-2(\beta+1)/(2\beta+a)}$ and $n_T^{-2(\beta+1)/(2\beta+3)}$ match; for $2 < a < 3$ one has $n_T^{\mathrm{cum}} < n < n_T^{\mathrm{eq}} < n_T^{\mathrm{TD}}$. The SL–TL boundary occurs at $n_T \asymp n_T^{\mathrm{eq}}$, so the integral-rate regime spans both SL and TL.

1. **Source Dominating (SD)** ($n_T \preccurlyeq n_T^{\mathrm{cum}}$): target sample size is too small to reduce the source-only risk from Remark 3; the combined rate equals $n^{-2(\beta+1)/(2\beta+a)}$ and $\mathrm{SAR} = \Theta(1)$.
2. **Level-split SLP (SL)** ($n_T^{\mathrm{cum}} \ll n_T \preccurlyeq n$): as in Remark 4, the combined risk is $n_T^{-1} + n^{-1} \asymp n_T^{-1}$, so the *binding* term is the target contribution to the irregular functionals, while the source remains the *better individual* estimator (since $n_T \le n < n_T^{\mathrm{eq}}$). Thus $\mathrm{SAR} \asymp n_T / n^{2(\beta+1)/(2\beta+a)} \to \infty$. This regime has no counterpart under $a < 2$.
3. **Integral SLP (SL and TL)** ($n \preccurlyeq n_T \ll n_T^{\mathrm{TD}}$): the remaining structure is combined at rate $(nn_T)^{-1/2}$ (Result II-type), with the level term n_T^{-1} no larger in order. Source is still the better individual on $n \preccurlyeq n_T \preccurlyeq n_T^{\mathrm{eq}}$; target becomes the better individual on $n_T^{\mathrm{eq}} \preccurlyeq n_T \ll n_T^{\mathrm{TD}}$. In both subwindows $\mathrm{SAR} \to \infty$. At $n_T \asymp n_T^{\mathrm{TD}}$, the same risk expression remains valid up to constants, but $\mathrm{SAR} = \Theta(1)$.
4. **Target Dominating (TD)** ($n_T \succcurlyeq n_T^{\mathrm{TD}}$): source data no longer improve upon the target-only Hölder rate $n_T^{-2(\beta+1)/(2\beta+3)}$; $\mathrm{SAR} = \Theta(1)$.

SLP therefore holds on the union of the two middle regimes $n_T^{\mathrm{cum}} \ll n_T \ll n_T^{\mathrm{TD}}$, with two distinct mechanisms: the new rate n_T^{-1}, then the integral rate $(nn_T)^{-1/2}$. (Why no covariate-shift-type rate appears for $2 < a < 3$ is explained after Theorem 5.)

C.6 Mechanism and sample-size regions for Result III, $a > 3$

Remark 5 (Why the mid-rate is covariate-shift type, and why $(nn_T)^{-1/2}$ disappears)**.** *For $a > 3$, even after the irregular running-integral functionals are handled as in Remark 4, the source still cannot estimate arbitrarily fine local features of g near $x = 0$ from observations of f. A feature of width b has source variance of order $n^{-1}b^{3-a}$, which diverges as $b \to 0$. Consequently there is no Result II-type regime in which the remaining structure of g can be combined at the integral rate $(nn_T)^{-1/2}$. The optimal construction must instead use a spatially varying resolution near the singularity—the same geometry as Result I—with source information for levels of g entering through the derivative link $f = \dot{g}$. The resulting mid-rate $R_{\mathrm{spat}}(n, n_T; a, \beta)$ in (8) is therefore a covariate-shift-type synergistic rate, not an integral-type rate.*

By contrast, in Theorem 4 ($2 < a < 3$) one has $n^{-1}b^{3-a} \to 0$ as $b \to 0$, so those local features remain parametrically estimable from the source. The integral combination $(nn_T)^{-1/2}$ is then available, and no covariate-shift-type mid-rate appears. The cut $a = 3$ is exactly the transition between these two behaviours.

Recall from (8) that n_T^{sp} is the onset of the spatial mid-rate in (6). For $a > 3$ one has $n_T^{\mathrm{cum}} < n_T^{\mathrm{sp}} < n < n_T^{\mathrm{TD}}$ (so the level-split window ends before n, unlike Theorem 4). Four sample-size regions arise:

1. **Source Dominating (SD)** ($n_T \preccurlyeq n_T^{\mathrm{cum}}$): combined rate equals the source-only rate $n^{-2(\beta+1)/(2\beta+a)}$; $\mathrm{SAR} = \Theta(1)$.
2. **Level-split SLP (SL)** ($n_T^{\mathrm{cum}} \ll n_T \preccurlyeq n_T^{\mathrm{sp}}$): as in Remark 4, the binding term is the target contribution n_T^{-1} to the irregular functionals, while the source is the better individual estimator; $\mathrm{SAR} \to \infty$.
3. **Covariate-shift SLP (TL)** ($n_T^{\mathrm{sp}} \preccurlyeq n_T \ll n_T^{\mathrm{TD}}$): local features near the singularity are estimated by a spatially adaptive combination of the two samples (Remark 5); the rate is $R_{\mathrm{spat}}(n, n_T; a, \beta)$ from (8), and $\mathrm{SAR} \to \infty$. The target is now the *better individual* estimator, since $n_T \succcurlyeq n_T^{\mathrm{sp}} = n_T^{\mathrm{eq}}$; for $a > 3$ the change of mechanism and the change of leading sample therefore occur at the same size, unlike the case $2 < a < 3$. At $n_T \asymp n_T^{\mathrm{TD}}$, this rate matches the target-only rate and $\mathrm{SAR} = \Theta(1)$. The integral rate $(nn_T)^{-1/2}$ does not appear.
4. **Target Dominating (TD)** ($n_T \succcurlyeq n_T^{\mathrm{TD}}$): source data no longer improve upon the target-only Hölder rate $n_T^{-2(\beta+1)/(2\beta+3)}$; $\mathrm{SAR} = \Theta(1)$.

SLP therefore holds on $n_T^{\mathrm{cum}} \ll n_T \ll n_T^{\mathrm{TD}}$, with two mechanisms: level-split (n_T^{-1}), then covariate-shift (spatial mid-rate).

C.7 THE CASES $a = 2$ AND $a = 3$ FOR RESULT III

Proposition 11 (Minimax rate at $a = 2$)**.** *At $a = 2$,*

$$R_{h_T}(g;\, n, n_T) \asymp \begin{cases} (\log n)\, n^{-1}, & n_T \preccurlyeq n(\log n)^{-1} \quad \textit{(SD)}, \\ n_T^{-1}, & n(\log n)^{-1} \ll n_T \preccurlyeq n \quad \textit{(SL)}, \\ (nn_T)^{-1/2}, & n \preccurlyeq n_T \preccurlyeq \big(n(\log n)^{-1}\big)^{\frac{2\beta+3}{2(\beta+1)}} \quad \textit{(SL)}, \\ (nn_T)^{-1/2}, & \big(n(\log n)^{-1}\big)^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{14}$$

Relative to the limits $a \to 2^{\pm}$, the leading power remains n^{-1}, an extra $\log n$ appears in the source-dominated risk, and the n_T^{-1} window shrinks to $n/\log n \ll n_T \preccurlyeq n$. SLP holds on $n/\log n \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$; SL and TL meet at $n_T \asymp (n/\log n)^{(2\beta+3)/(2(\beta+1))}$.

Proposition 12 (Minimax rate at $a = 3$)**.** *At $a = 3$,*

$$R_{h_T}(g;\, n, n_T) \asymp \begin{cases} n^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \preccurlyeq n^{\frac{2(\beta+1)}{2\beta+3}} \quad \textit{(SD)}, \\ n_T^{-1}, & n^{\frac{2(\beta+1)}{2\beta+3}} \ll n_T \preccurlyeq n(\log n)^{-1} \quad \textit{(SL)}, \\ (\log n)\, n^{-1}, & n(\log n)^{-1} \preccurlyeq n_T \preccurlyeq n \quad \textit{(SL)}, \\ (nn_T)^{-1/2}\Big(1 + \log_+\big(e\, n_T^{-\frac{2\beta+1}{2\beta+3}}\big)\Big), & n \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}, \\ n_T^{-\frac{2(\beta+1)}{2\beta+3}}, & n_T \succcurlyeq n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TD)}. \end{cases} \tag{15}$$

The positive-part logarithm keeps the factor bounded away from zero when $n_T \asymp n^{(2\beta+3)/(2\beta+1)}$, where the last two rate expressions agree up to constants. The rate $(nn_T)^{-1/2}$ survives only with a logarithmic factor; the polynomial spatial mid-rate of Theorem 5 requires $a > 3$. SLP holds on $n^{2(\beta+1)/(2\beta+3)} \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$. The transition at $n_T \asymp n/\log n$ lies within SL, while the equalizing scale $n_T \asymp n$ separates SL and TL.

Remark 6 (Left and right of the cuts $a = 2$ and $a = 3$)**.** *Theorems 3–5 are organised by how well the* source *sample alone can estimate g.*

1. ***Cut** $a = 2$. As explained in Remark 3, when $a > 2$ the source design puts too little mass near the origin for the integral map $f \mapsto g$ to remain parametrically estimable, so $R_{h_T}(g; n, 0)$ is of order $n^{-2(\beta+1)/(2\beta+a)}$ rather than n^{-1}. The target sample observes g directly and attains risk n_T^{-1} for the corresponding finite-dimensional functionals of g. This produces the extra synergistic window of rate n_T^{-1} in Theorem 4.*

2. ***Cut** $a = 3$. Independently of the running-integral functionals above, the source must still estimate local features of g near $x = 0$ from observations of $f = \dot{g}$. For features of width b, the resulting source variance is of order $n^{-1}b^{3-a}$. If $a < 3$ this tends to 0 as $b \to 0$, so arbitrarily fine features remain estimable at parametric source rates and the Result II-type combination $(nn_T)^{-1/2}$ continues to apply. If $a > 3$ the variance blows up as $b \to 0$, so fine features near the singularity cannot be estimated that way; the optimal estimator must use a spatially varying resolution, yielding the mid-rate of Theorem 5 (the same geometry as Result I, but for the integral link).*

Thus unboundedness of $\rho = h_T/h$ (which already holds for all $a > 1$) is not by itself enough to change the SLP structure: the first change occurs when source-only estimation of g is forced below the parametric rate n^{-1} ($a = 2$), and the second when local features of g near 0 cease to be parametrically estimable from the source ($a = 3$).

Remark 7 (Why the rates change exactly at $a = 2$ and $a = 3$)**.** *Propositions 11–12 record the rates at $a = 2$ and $a = 3$. They differ from both neighbouring regimes because each cut is a borderline second-moment condition.*

1. ***At*** $a = 2$***.*** *Truncating the unbiased source estimator of g at scale ε produces variance of order $n^{-1}\int_\varepsilon^1 h^{-1}$. For $a < 2$ the integral stays bounded as $\varepsilon \to 0$; for $a > 2$ it grows like a power ε^{2-a}; exactly at $a = 2$ it grows only logarithmically, $\int_\varepsilon^1 h^{-1} \asymp \log(e/\varepsilon)$. Balancing this variance against the Hölder bias $\varepsilon^{\beta+1}$ yields the source-dominated rate $(\log n)\, n^{-1}$ in (14), rather than n^{-1} or $n^{-2(\beta+1)/(2\beta+a)}$. The level-split window then starts at $n_T \asymp n/\log n$ instead of at a pure power of n.*

2. ***At*** $a = 3$***.*** *Local features of width b near 0 have source variance of order $n^{-1}b^{3-a}$. For $a < 3$ this vanishes as $b \to 0$; for $a > 3$ it diverges; exactly at $a = 3$ it is of order n^{-1} at* every *fine scale. Summing (or optimising over) a logarithmic number of such scales produces the extra $\log n$ factors in (15): a band of rate $(\log n)\, n^{-1}$, and an integral combination of order $(nn_T)^{-1/2}$ multiplied by a slowly varying logarithmic term. The polynomial spatial mid-rate of Theorem 5 appears only for $a > 3$, once the fine-scale source variance truly blows up.*

In both cases the polynomial order matches the limit from the easier side of the cut, while the logarithm is the precise price of sitting on the borderline where the relevant source second moment is neither finite nor power-divergent.

C.8 Phase diagram and SAR for Result III

Corollary 13 (Phase diagram and SAR for Result III)**.** *Table 2 records the SLP windows and synergistic rates. On each such window* $\mathrm{SAR} \to \infty$ *as below; in the SD and TD regions,* $\mathrm{SAR} = \Theta(1)$*. For $1 < a < 2$, the SAR coincides with (13):*

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} (n_T n^{-1})^{1/2}, & n \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2(\beta+1)}} \quad \textit{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{1/2}, & n^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}. \end{cases} \tag{16a}$$

For $2 < a < 3$,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} n_T n^{-\frac{2(\beta+1)}{2\beta+a}}, & n^{\frac{2(\beta+1)}{2\beta+a}} \ll n_T \preccurlyeq n \quad \textit{(SL)}, \\ n^{-\frac{2(\beta+1)}{2\beta+a}+\frac{1}{2}}\, n_T^{1/2}, & n \preccurlyeq n_T \preccurlyeq n^{\frac{2\beta+3}{2\beta+a}} \quad \textit{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{1/2}, & n^{\frac{2\beta+3}{2\beta+a}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}. \end{cases} \tag{16b}$$

For $a > 3$,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} n_T n^{-\frac{2(\beta+1)}{2\beta+a}}, & n^{\frac{2(\beta+1)}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{2\beta+3}{2\beta+a}} \quad \textit{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{\frac{1}{a-1}}, & n^{\frac{2\beta+3}{2\beta+a}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}. \end{cases} \tag{16c}$$

In (16b), the change of rate at $n_T \asymp n$ lies within SL. In (16c), the spatial rate begins at the equalizing scale $n_T \asymp n^{(2\beta+3)/(2\beta+a)}$; the target-only risk is therefore the smaller individual benchmark throughout TL, up to constants at that scale.

At $a = 2$, the individual rates $(\log n)/n$ and $n_T^{-2(\beta+1)/(2\beta+3)}$ cross at $n_T \asymp (n/\log n)^{(2\beta+3)/(2(\beta+1))}$. In the SLP window,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} (\log n)\, n_T n^{-1}, & n(\log n)^{-1} \ll n_T \preccurlyeq n \quad \textit{(SL)}, \\ (\log n)\, (n_T n^{-1})^{1/2}, & n \preccurlyeq n_T \preccurlyeq \left(n(\log n)^{-1}\right)^{\frac{2\beta+3}{2(\beta+1)}} \quad \textit{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{1/2}, & \left(n(\log n)^{-1}\right)^{\frac{2\beta+3}{2(\beta+1)}} \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}. \end{cases} \tag{17}$$

Thus the SLP window is $n/\log n \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$.

Table 2: Result III: SLP windows and synergistic rates (integral link, Beta$(a, 1)$).

Range of a	SLP window	Synergistic rate(s)
$1 < a < 2$	$n \ll n_T \ll n_T^{\mathrm{TD}}$	$(nn_T)^{-1/2}$
$a = 2$	$n/\log n \ll n_T \ll n_T^{\mathrm{TD}}$	n_T^{-1} (thin), then $(nn_T)^{-1/2}$
$2 < a < 3$	$n_T^{\mathrm{cum}} \ll n_T \ll n_T^{\mathrm{TD}}$	n_T^{-1}, then $(nn_T)^{-1/2}$
$a = 3$	$n^{2(\beta+1)/(2\beta+3)} \ll n_T \ll n_T^{\mathrm{TD}}$	n_T^{-1}, then $(\log n)n^{-1}$, then log-corrected $(nn_T)^{-1/2}$
$a > 3$	$n_T^{\mathrm{cum}} \ll n_T \ll n_T^{\mathrm{TD}}$	n_T^{-1}, then $R_{\mathrm{spat}}(n, n_T; a, \beta)$

At $a = 3$, the individual rates are $n^{-2(\beta+1)/(2\beta+3)}$ and $n_T^{-2(\beta+1)/(2\beta+3)}$. In the SLP window,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} n_T n^{-\frac{2(\beta+1)}{2\beta+3}}, & n^{\frac{2(\beta+1)}{2\beta+3}} \ll n_T \preccurlyeq n(\log n)^{-1} \quad \textit{(SL)}, \\ n^{\frac{1}{2\beta+3}}(\log n)^{-1}, & n(\log n)^{-1} \preccurlyeq n_T \preccurlyeq n \quad \textit{(SL)}, \\ \dfrac{\left(n\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)^{1/2}}{1 + \log_+\left(en\, n_T^{-\frac{2\beta+1}{2\beta+3}}\right)}, & n \preccurlyeq n_T \ll n^{\frac{2\beta+3}{2\beta+1}} \quad \textit{(TL)}. \end{cases} \tag{18}$$

The SLP window is $n^{2(\beta+1)/(2\beta+3)} \ll n_T \ll n^{(2\beta+3)/(2\beta+1)}$. At $a = 3$, logarithmic corrections alter the location of the largest SAR order. With $n_T = N$ held fixed, the individual rates match at $n \asymp N$, whereas the SAR attains its largest order at $n \asymp N \log N$. For fixed n, Equation (18) gives a plateau in the SAR rate over $n/\log n \preccurlyeq n_T \preccurlyeq n$. For both $a = 2$ and $a = 3$, $\mathrm{SAR} = \Theta(1)$ in the SD and TD regions. Equations (17)–(18) follow by dividing the smaller individual risk by the joint risk in Propositions 11–12, respectively. Full proofs are in supplement Section S.2.3.

C.9 Phase-diagram table for Result III

C.10 Source-only rate for the derivative link

Remark 8 (Source-only rate under Beta$(a, 1)$ density)**.** *Under bounded source density, differentiating a boundary-adaptive local polynomial gives $n^{-2(\beta-1)/(2\beta+1)}$. Under the Beta source, the same power holds for $a < 2 + 3/\{2(\beta - 1)\}$, acquires a logarithmic factor at equality, and slows to $n^{-(2\beta-1)/(2\beta+a)}$ above it. In the strong-singularity regime the combined rate (7) strictly beats both single-task benchmarks in the synergistic window, so the SLP and SAR conclusions of Corollary 7 remain valid.*

C.11 The case $a = 2 + 3/\{2(\beta - 1)\}$ of Theorem 6

Under the setting of Theorem 6, if $a = 2 + \frac{3}{2(\beta-1)}$, then

$$R_{h_T}(\dot{f}; n, n_T) \asymp \begin{cases} n^{-\frac{2(\beta-1)}{2\beta+1}} \log n, & n_T \preccurlyeq n^{\frac{2(\beta-1)}{2\beta+1}}, \\ n^{-\frac{2(\beta-1)}{2\beta+1}} \left\{1 + \log_+\left(en\, n_T^{-\frac{2\beta+1}{2\beta-1}}\right)\right\}, & n^{\frac{2(\beta-1)}{2\beta+1}} \preccurlyeq n_T \preccurlyeq n^{\frac{2\beta-1}{2\beta+1}}, \\ n_T^{-\frac{2(\beta-1)}{2\beta-1}}, & n_T \succcurlyeq n^{\frac{2\beta-1}{2\beta+1}}. \end{cases} \tag{19}$$

At $n_T \asymp n^{(2\beta-1)/(2\beta+1)}$, the logarithmic factor is of constant order and the middle rate matches the target-only rate up to constants. For the SAR, write the common logarithmic factor as

$$\Lambda_D = 1 + \log_+\left(en\, n_T^{-\frac{2\beta+1}{2\beta-1}}\right).$$

In the intermediate sample-size range,

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} \log n/\Lambda_D, & n^{\frac{2(\beta-1)}{2\beta+1}} \preccurlyeq n_T \preccurlyeq n^{\frac{2\beta-1}{2\beta+1}} (\log n)^{-\frac{2\beta-1}{2(\beta-1)}}, \\ n^{\frac{2(\beta-1)}{2\beta+1}} n_T^{-\frac{2(\beta-1)}{2\beta-1}}/\Lambda_D, & n^{\frac{2\beta-1}{2\beta+1}} (\log n)^{-\frac{2\beta-1}{2(\beta-1)}} \preccurlyeq n_T \preccurlyeq n^{\frac{2\beta-1}{2\beta+1}}. \end{cases}$$

The two branches meet where the individual risks have the same order. This intermediate risk range is not itself an SLP window: the SAR diverges only when $1 \ll n^{(2\beta-1)/(2\beta+1)} n_T^{-1} = n^{o(1)}$. Thus any diverging SAR at equality is subpolynomial, and there is no fixed polynomial-exponent SLP window.

C.12 SAR for Result IV

In the SLP window of Corollary 7:

$$\mathrm{SAR}(n, n_T) \asymp \begin{cases} \left(n_T\, n^{-\frac{2\beta-1}{2\beta+a}}\right)^{\frac{2(\beta-1)}{2\beta-1} - \frac{2\beta+1}{(2\beta-1)(a-1)}}, & n^{\frac{2\beta-1}{2\beta+a}} \ll n_T \preccurlyeq n^{\frac{(2\beta-1)^2}{2(\beta-1)(2\beta+a)}} \quad \text{(SL)}, \\ \left(n\, n_T^{-\frac{2\beta+1}{2\beta-1}}\right)^{\frac{1}{a-1}}, & n^{\frac{(2\beta-1)^2}{2(\beta-1)(2\beta+a)}} \preccurlyeq n_T \ll n^{\frac{2\beta-1}{2\beta+1}} \quad \text{(TL)}. \end{cases} \tag{20}$$

The four SSR regions are SD / SL / TL / TD with sample-size boundaries $n^{(2\beta-1)/(2\beta+a)}$, $n^{(2\beta-1)^2/(2(\beta-1)(2\beta+a))}$, and $n^{(2\beta-1)/(2\beta+1)}$.

C.13 Why the integration–differentiation asymmetry holds

Integration *raises* smoothness ($f \in \mathcal{H}(\beta) \Rightarrow g \in \mathcal{H}(\beta+1)$) and admits an exactly unbiased IPW estimator, enabling a parametric source rate despite g being nonparametric. Differentiation *lowers* smoothness and amplifies source noise. For estimating $g = f'$, the scale-b joint information is of order $nb^3 + n_T b$. Optimizing the matching upper and lower bounds gives the minimum of the two individual rates in Theorem 8. Thus target data can improve upon source-only derivative estimation, but not yield a rate-level gain over the better individual experiment. The full proof is in Section S.2.2.1 of the supplement.

D Estimators

This appendix gives the rate-optimal estimator constructions for Results I–IV. Each subsection provides the full specification omitted from the main text for space.

D.1 Result I: Spread-Function NW Estimator

For each $x \in [0,1]$, the *adaptive bandwidth* $t_n(x) > 0$ is the unique solution to the balance equation

$$t_n(x)^{2\beta} \left[n\, H(x \pm t_n(x)) + n_T\, H_T(x \pm t_n(x)) \right] = 1, \tag{21}$$

where $H(x \pm t) = \int_{\max(x-t,0)}^{\min(x+t,1)} h(y)\, dy$ and $H_T(x \pm t) = \int_{\max(x-t,0)}^{\min(x+t,1)} h_T(y)\, dy$. For $0 < \beta \le 1$, the *spread-function NW estimator* (Zhou & Yang, 2025) is

$$\hat{f}_{n,n_T}(x) = \frac{\displaystyle\sum_{i=1}^{n} Y_i\, K\left(\frac{X_i - x}{t_n(x)}\right) + \sum_{j=1}^{n_T} Y_j^T\, K\left(\frac{X_j^T - x}{t_n(x)}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{X_i - x}{t_n(x)}\right) + \sum_{j=1}^{n_T} K\left(\frac{X_j^T - x}{t_n(x)}\right)}, \tag{22}$$

with a nonnegative compactly supported kernel. For $\beta > 1$, replace the local-constant ratio by the boundary-adaptive local-polynomial fit of degree $p = \lceil \beta \rceil - 1$ using the same spread bandwidth. Below, $\hat{f}_{n,n_T}$ denotes the appropriate one of these two estimators. Set the estimate to zero on empty windows and stabilize the local-polynomial fit using a design-based Gram-matrix safeguard as in Zhou & Yang (2025, Lemma A.10). The rate-optimal estimator by regime:

- *Source-dominated* ($n_T \preccurlyeq n^{(2\beta+1)/(2\beta+a)}$): $\hat{f}_{n,n_T}$ from (22), source data only.
- *Synergistic* ($n^{(2\beta+1)/(2\beta+a)} \preccurlyeq n_T \preccurlyeq n$): $\hat{f}_{n,n_T}$ from (22) with both datasets; $t_n(x)$ automatically interpolates between density regimes.
- *Target-dominated* ($n_T \succcurlyeq n$): Standard NW/local-polynomial estimator $\hat{f}_T$ with bandwidth $b_T \asymp n_T^{-1/(2\beta+1)}$.

D.2 Result II: IPW Estimator with Bias Correction

The rate-optimal estimator by regime:

- *Target-dominated* ($n \preccurlyeq n_T^{(2\beta+1)/(2\beta+3)}$): target-only boundary-adaptive local polynomial of degree $p_g = \lceil \beta + 1 \rceil - 1$ and bandwidth $b_T^* = n_T^{-1/(2\beta+3)}$.
- *Synergistic* ($n_T^{(2\beta+1)/(2\beta+3)} \preccurlyeq n \preccurlyeq n_T$): bias-corrected estimator $\hat{g} = \hat{g}_S - \hat{B}$:

  Step 1. $\hat{g}_S(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x) Y_i / h(X_i)$ (h known). Exactly unbiased: $\mathbb{E}[\hat{g}_S(x)] = g(x)$.

  Step 2. Put $b_T = \sqrt{n/n_T}$, $I_x = [x - b_T, x + b_T] \cap [0,1]$, and $N_x = \sum_{j=1}^{n_T} \mathbf{1}\{X_j^T \in I_x\}$. Define

$$\hat{B}(x) = \frac{1}{N_x \vee 1} \sum_{j=1}^{n_T} \mathbf{1}\{X_j^T \in I_x\} \{\hat{g}_S(X_j^T) - Y_j^T\}.$$

  Step 3. $\hat{g}(x) = \hat{g}_S(x) - \hat{B}(x)$.

  The exact conditional decomposition gives MSE $= O(b_T/n) + O(1/(n_T b_T))$, minimized at $b_T^* = \sqrt{n/n_T}$, yielding rate $(n n_T)^{-1/2}$.
- *Source-dominated* ($n_T \preccurlyeq n$): $\hat{g}_S(x)$ alone; variance $O(n^{-1})$, bias zero.

D.3 Result III: Hierarchical Series Estimator

For every fixed $a > 1$, the minimax upper bounds in Theorems 3–5 and Propositions 11–12 are attained by the following explicit hierarchical construction (no random Gram inverse).

Basis. For $1 < a < 4$, fix the standard interval biorthogonal Riesz system used in the supplement and its shell–scale reindexing; for $a \ge 4$, use the moment-preserving boundary perturbation constructed there. Both have the required global Riesz bounds, polynomial moments, and primitive endpoint behavior. In the perturbed system, the localized analysis functions and the fixed finite root analysis block retain the original index set; the dual reconstruction functions are defined by the bounded inverse operator and need not be locally supported. No modified two-scale identity is required for the coefficient oracle. Write $(\xi_\nu, \widetilde{\xi}_\nu)$ for the finitely many cumulative analysis/reconstruction pairs and $(\psi_\lambda, \widetilde{\psi}_\lambda)$ for the remaining pairs. Then

$$g = \sum_{\nu \in \Lambda_{\mathrm{cum}}} \theta_\nu \widetilde{\xi}_\nu + \sum_{B \in \mathcal{B}} \sum_{\lambda \in B} \theta_\lambda \widetilde{\psi}_\lambda,$$

where Λ_{cum} is a fixed finite set of cumulative (level) indices and $\mathcal{B}$ is a hierarchy of mean-zero (zero-tail) blocks B of length L and mesh b. The corresponding coefficients are $\theta_\nu = \langle g, \xi_\nu \rangle$ and $\theta_\lambda = \langle g, \psi_\lambda \rangle$. For each mean-zero ψ_λ let $\Phi_\lambda(x) = \int_{\inf \operatorname{supp} \psi_\lambda}^{x} \psi_\lambda(t)\, dt$, extended by zero off its support, so $\theta_\lambda = \langle g, \psi_\lambda \rangle = -\langle f, \Phi_\lambda \rangle$.

Step 1 (cumulative / level coefficients). For $\nu \in \Lambda_{\mathrm{cum}}$, denote $K_\nu(t) = \int_t^1 \xi_\nu(x)\, dx$, let $p_\beta = \lceil \beta \rceil - 1$ and $r_\beta(u) = (1, u, \ldots, u^{p_\beta})^\top$, and define

$$M_a = \int_0^1 r_\beta(u) r_\beta(u)^\top a u^{a-1}\, du, \qquad k_{\nu,\varepsilon} = \int_0^1 K_\nu(\varepsilon u) r_\beta(u)\, du.$$

For a cutoff $\varepsilon \in (0, 1]$, estimate the Taylor coefficients of f on $[0, \varepsilon]$ by

$$\widehat{c}_\varepsilon = \frac{M_a^{-1}}{n\varepsilon^a} \sum_{i=1}^{n} \mathbf{1}\{X_i \le \varepsilon\} r_\beta(X_i/\varepsilon) Y_i,$$

and use

$$\widehat{\theta}^{(\varepsilon)}_{\nu,S} = \varepsilon k^\top_{\nu,\varepsilon} \widehat{c}_\varepsilon + \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i > \varepsilon\} \frac{Y_i K_\nu(X_i)}{h(X_i)}, \qquad \widehat{\theta}_{\nu,T} = \frac{1}{n_T} \sum_{j=1}^{n_T} Y_j^T \xi_\nu(X_j^T).$$

The source bias is $O(\varepsilon^{\beta+1})$. Source precision is of order n, $n/\log(e/\varepsilon)$, or $n\varepsilon^{a-2}$ according as $a < 2$, $a = 2$, or $a > 2$. Combine by deterministic inverse-variance weights and shrink to the Hölder/Riesz envelope.

Step 2 (mean-zero block coefficients). For $\lambda \in B$,

$$\hat{\theta}_{\lambda,S} = -\frac{1}{n} \sum_{i=1}^{n} \frac{Y_i \, \Phi_\lambda(X_i)}{h(X_i)}, \qquad \hat{\theta}_{\lambda,T} = \frac{1}{n_T} \sum_{j=1}^{n_T} Y_j^T \, \psi_\lambda(X_j^T).$$

Block informations $I_{B,S} = nL^{a-1}b^{-2}$ and $I_{B,T} = n_T$; combine by inverse-variance and deterministic Hölder shrinkage. Block oracle risk:

$$\min\Big\{ L\, b^{2(\beta+1)},\ \tfrac{L}{nL^{a-1}b^{-1} + n_T b} \Big\}.$$

Step 3 (reconstruct).

$$\hat{g} = \sum_{\nu \in \Lambda_{\mathrm{cum}}} \hat{\theta}_\nu \widetilde{\xi}_\nu + \sum_{B \in \mathcal{B}} \sum_{\lambda \in B} \hat{\theta}_\lambda \widetilde{\psi}_\lambda,$$

projected onto the Hölder envelope of g.

Step 4 (rate-optimal tuning).

- $1 < a < 2$: take $\varepsilon = 0$ (full IPW). For $n_T \preccurlyeq n$ use source coefficients only; for $n \preccurlyeq n_T \preccurlyeq n_T^{\mathrm{TD}}$ use Steps 1–2 with both channels (equivalently the Result II bias-correction path); for $n_T \succcurlyeq n_T^{\mathrm{TD}}$ use a target-only local polynomial of bandwidth $n_T^{-1/(2\beta+3)}$.
- $2 < a < 3$: truncation $\varepsilon \asymp n^{-1/(2\beta+a)}$ when the source owns the level. Level-split window: Step 1 target-dominated for Λ_{cum}, Step 2 source(/target) for mean-zero blocks $\Rightarrow$ rate n_T^{-1}. For $n \preccurlyeq n_T \preccurlyeq n_T^{\mathrm{TD}}$ open both channels $\Rightarrow (nn_T)^{-1/2}$.
- $a > 3$: same Step 1; in Step 2 retain only *boundary-adapted* shells. Equivalently, the local balance

$$n\, H(x \pm t)\, t^{2\beta} + n_T\, H_T(x \pm t)\, t^{2(\beta+1)} = 1 \tag{23}$$

  governs the spatial mid-regime of (6).
- $a = 2$ and $a = 3$: same constructions with the logarithmic replacements of Propositions 11–12.

Inverse-variance weights make the construction unified across regimes; only $(\varepsilon, b_{\min})$ and shell depth depend on a.

D.4 Result IV: Joint Local-Polynomial Estimator (Derivative Link)

Target observes $\dot{f}$ while source observes f. Under $h_T \equiv \mathrm{Uniform}$, a target-only NW has variance depending only on n_T, so $n\,H$ is not an effective sample size for that estimator. Source levels contribute Fisher weight nHt^2 and target derivatives contribute $n_T H_T$.

Estimator. At each x, let $p = \lceil \beta \rceil - 1$ and $R_x(z) = \sum_{k=0}^{p} \frac{a_k}{k!} \big((z-x)/t_n^-(x)\big)^k$ and minimise

$$\sum_{j=1}^{n_T} \big(Y_j^T - R'_x(X_j^T)\big)^2 K\Big(\frac{X_j^T - x}{t_n^-(x)}\Big) + \sum_{i=1}^{n} \big(Y_i - R_x(X_i)\big)^2 K\Big(\frac{X_i - x}{t_n^-(x)}\Big), \tag{24}$$

with bandwidth $t_n^-(x) > 0$ solving

$$\big(t_n^-(x)\big)^{2(\beta-1)} \big(n\, H(x \pm t_n^-(x))\, (t_n^-(x))^2 + n_T\, H_T(x \pm t_n^-(x))\big) = 1. \tag{25}$$

Among the minimizers of (24), choose the coefficient vector with minimum Euclidean norm, using the Moore–Penrose inverse of the normal matrix. Let $M_1 \geq \sup_{f \in \mathcal{H}(\beta,\kappa)} \|f'\|_\infty$ be fixed (one may take $M_1 = \kappa$ under the stated Hölder convention), and set $\hat{g}(x) = \Pi_{[-M_1,M_1]}\{R'_x(x)\}$, where $\Pi_{[-M_1,M_1]}$ denotes projection onto $[-M_1, M_1]$ and $R'_x(x) = a_1/t_n^-(x)$. Source observations of R_x and target observations of R'_x both inform $R'_x(x)$, with Fisher weights $nH(t_n^-)^2$ and $n_T H_T$ respectively (supplement Lemma S.3.10). (H, H_T as in Appendix D.1; a nonnegative box kernel is admissible.)

The projected estimator satisfies the mean-squared-error bound in supplement Lemma S.3.10. Combined with the bandwidth balance (25), this yields (7) for $a > 2 + 3/\{2(\beta-1)\}$ and (19) at equality. Individual rates: $R_{h_T}(\dot{f};\, n, 0) \asymp n^{-2(\beta-1)/(2\beta+1)}$ (bounded-density reference; see Remark 8) and $R_{h_T}(\dot{f};\, 0, n_T) \asymp n_T^{-2(\beta-1)/(2\beta-1)}$. Full construction in supplement Section S.2.4.

E Simulation Illustrations

We illustrate selected transfer scenarios from Sections 4.1–4.3. The experiments examine rate mechanisms, sample-allocation effects, and the limitations of finite-range comparisons with minimax theory.

Common protocol. Design points are sampled independently on the unit cube of the corresponding dimension. All regression experiments use independent Gaussian noise with standard deviation 0.25. Every Monte Carlo configuration in this simulation section and its supplementary diagnostics uses 400 independent repetitions, including regression experiments, surrogate loss simulations, and separate Monte Carlo diagnostics. Analytic-risk calculations have no Monte Carlo repetition count. Unless a subsection specifies otherwise, we approximate integrated risk on a 4096-point grid and construct paired percentile-bootstrap 95% intervals from 5000 resamples; bootstrap resamples are distinct from simulation repetitions. Slopes use the largest four sample sizes unless another tail length is stated. Code for running the simulations and producing the figures is available under `supplementary/simulations`.

For the synthetic regression experiments, β denotes the source-function class index used for tuning and rate references. The target class index is $\beta + 1$ for the integral link, $\beta - 1$ for the derivative link. Individual test functions may be smoother than these class requirements. The surrogate experiments specify their coefficient classes separately.

Risk interpretation. We distinguish raw-regression Monte Carlo, surrogate-model Monte Carlo, and analytic-risk calculations. Monte Carlo is abbreviated as MC. Raw-regression Monte Carlo generates design points and responses and applies the regression estimators. Surrogate-model Monte Carlo samples losses induced by a separately specified observation model and estimator, possibly by exact aggregation of coefficient errors. Analytic-risk calculations evaluate expected losses without Monte Carlo sampling and may concern either a regression estimator or a surrogate. Each experiment specifies its risk target separately. For finite-library comparisons, let $\mathcal{L}$ denote the specified library of target functions and report

$$R_E^{\mathcal{L}} = \max_{g \in \mathcal{L}} \mathbb{E}\|\widehat{g}_E - g\|_2^2, \qquad \frac{\min(R_S^{\mathcal{L}}, R_T^{\mathcal{L}})}{R_C^{\mathcal{L}}},$$

where the second expression is the corresponding SAR and C denotes the combined estimator. The library and any dependence on sample sizes are specified for each experiment. Fixed-truth and finite-library risks concern the specified estimators, not minimax optimization over a function

class. Bootstrap bands quantify pointwise Monte Carlo precision under the stated model and risk definition. They do not measure surrogate-to-regression discrepancy or uncertainty in minimax theory. Finite-range risk stabilization, SAR growth, and scan maxima do not establish minimax exponents or asymptotic window boundaries. Absence of an asymptotic rate improvement requires bounded SAR, not convergence to one.

Display conventions. Sample-size and SAR axes are logarithmic; normalized-risk and relative-risk diagnostics use linear vertical axes. We write $\kappa_n = \log n / \log n_T$ for target-anchored scans and $\kappa_T = \log n_T / \log n$ for source-anchored scans. Markers identify Monte Carlo estimates. Colored dotted curves give expected risks or SAR formed from those risks for the specified estimators, without sampling bands. Rate references instead give theoretical order shapes. Unless stated otherwise, same-color dashed curves give longitudinal references and purple dash–dot curves give scan references. Rate curves are vertically aligned at the first fitted point because the theorems determine rates only up to constants. Light-gray regions mark slope windows. Horizontal gray dotted lines mark SAR$= 1$. Vertical gray dotted lines mark theoretical balances or boundaries, orange dashed lines mark channel crossings computed from the displayed risks, and gray dashed vertical lines mark estimator switches. Experiment-specific reference normalizations are stated below.

E.1 Bounded integral link and the bounded-derivative control

Design. The integral-link experiments use $\beta = 1$, so the target class has smoothness index $\beta+1 = 2$. Let

$$f(x) = 0.7|x - 0.4|, \qquad g(x) = 0.35\{0.4^2 + \operatorname{sgn}(x - 0.4)|x - 0.4|^2\},$$

so that $g' = f$. Both designs are uniform. On the longitudinal paths, $n_T = N$ and $n = \operatorname{round}(N^{\kappa_n})$, with $\kappa_n \in \{0.7, 0.8, 0.9\}$ and $N = 2^{10}, \ldots, 2^{20}$. The source estimator is the empirical IPW integral. The target estimator is degree-one local polynomial regression with bandwidth $1.1N^{-1/5}$. The mixed estimator applies the count-based target correction with $b = 1 \wedge (n/n_T)^{1/2}$.

The transverse scan fixes $n_T = 2^{20}$ and sets $n = \operatorname{round}(n_T^{\kappa_n})$ on the grid $\kappa_n = 0.5, 0.525, \ldots, 1.05$.

Risk diagnostics and prescribed switching. Both integral-link panels report a fixed-truth candidate-risk diagnostic. Let $\widehat{R}_E(g)$ denote estimator E's mean loss over the Monte Carlo replicates at the displayed truth, and define

$$\widehat{R}_{\mathrm{cand}}(g) = \min\{\widehat{R}_S(g), \widehat{R}_T(g), \widehat{R}_B(g)\}, \qquad \widehat{\mathrm{SAR}}_{\mathrm{cand}} = \frac{\widehat{R}_S(g) \wedge \widehat{R}_T(g)}{\widehat{R}_{\mathrm{cand}}(g)}.$$

The candidate minimum is taken *after* averaging each estimator's losses, not separately within each Monte Carlo replicate. It estimates a pointwise minimum over a finite candidate family, not the risk of an implemented data-driven selector or the minimax risk over a function class. The minimum and numerator are recomputed inside each paired bootstrap draw. Nevertheless, $\widehat{\mathrm{SAR}}_{\mathrm{cand}} = \max\{1, (\widehat{R}_S \wedge \widehat{R}_T)/\widehat{R}_B\}$, so its unit lower bound is imposed by construction.

On longitudinal paths, B is the raw mixed estimator. In the fixed-target scan, the third candidate follows the predeclared theoretical branch rule

$$\widehat{g}_B = \begin{cases} \widehat{g}_T, & \kappa_n \le 0.6, \\ \widehat{g}_M, & 0.6 < \kappa_n < 1, \\ \widehat{g}_S, & \kappa_n \ge 1. \end{cases}$$

No unswitched mixed estimator is evaluated outside the open middle interval. Consequently, the scan's outer candidate SAR is identically one by design and does not independently validate the no-SLP regions or locate their boundaries. Gray raw-mixed markers are restricted to $0.6 < \kappa_n < 1$; the complete branch-rule ratios remain in the saved data.

Negative control and fitting protocol. For the bounded-derivative negative control we reverse the operator direction using the same function pair. Source responses now observe the primitive above, and target responses observe its derivative. Thus the derivative-link setting has source index $\beta = 2$ and target index $\beta - 1 = 1$, unlike the integral-link panels. We take $n = N$ and $n_T = \operatorname{round}(N^{0.6})$

for $N = 2^8, \ldots, 2^{14}, 2^{16}, 2^{18}, 2^{20}$, and fit the operator-aware joint local polynomial. This path equalizes the two individual minimax rates for this derivative setting. The negative-control SAR uses the actual joint-estimator risk, without a candidate-minimum floor.

The three integral longitudinal fits use the largest *six* scales, $N = 2^{15}, \ldots, 2^{20}$. The negative control uses four, $N = 2^{14}, 2^{16}, 2^{18}, 2^{20}$. Fixed-target slopes use the six grid points in each of $[0.65, 0.775]$ and $[0.825, 0.95]$ for κ_n, regressing log SAR on $\log n$; the reference slopes $\pm 1/2$ are not slopes with respect to κ_n itself.

Results. Figure 2(a) compares the three longitudinal paths. Their SAR-slope intervals in Table 3 each contain the corresponding theoretical exponent. The largest point estimate occurs on the balanced path, consistent with the theoretical equalizer $\kappa_n = 0.8$. At $\kappa_n = 0.9$, two of the six tail points hit the candidate SAR floor; the raw mixed ratio instead has slope 0.057 $[0.009, 0.109]$ on the same points. The candidate absolute-risk slopes for $\kappa_n = 0.7, 0.8$ lie somewhat above their theoretical decay exponents. Thus SAR compatibility does not establish exact agreement of every absolute-risk exponent.

Figure 2(b) traverses the phase diagram at fixed n_T. Within the actually evaluated mixed region, candidate SAR rises and then falls, with its grid maximum at $\widehat{\kappa}_n = 0.800$ and a gridwise bootstrap interval $[0.775, 0.800]$. The estimated log–log slopes are 0.462 $[0.442, 0.484]$ in the interior of the TL region and -0.501 $[-0.567, -0.421]$ in the SL region, compared with $1/2$ and $-1/2$. The left interval excludes $1/2$; the right estimate includes the candidate floor near $\kappa_n = 0.95$. On the same right-hand points, the raw mixed ratio has slope -0.580 $[-0.671, -0.487]$. The peak SAR is 2.86 $[2.70, 3.23]$, with the maximum recomputed over the grid in each bootstrap draw, rather than a pointwise interval at the estimated peak. The theoretical value $n_T^{1/10} = 4$ is an order benchmark because the rate statement suppresses multiplicative constants. Near $\kappa_n = 1$, the raw mixed estimator is less competitive than source-only in finite samples. The candidate minimum therefore returns to one before the prescribed SD boundary; this is not an empirical estimate of the asymptotic SLP endpoint.

The bounded-derivative control remains near the no-gain baseline in Figure 2(c). Its absolute-risk exponent in Table 3 is close to the theoretical value 0.4. The terminal SAR is 0.998 $[0.989, 1.007]$. The small negative finite-range slope records the unsafeguarded joint estimator settling toward the individual benchmark. Together, the two operator directions separate polynomial synergy for the integral link from bounded gain for the derivative link.

E.2 Unbounded integral link: regression paths and logarithmic mechanisms

Raw-sample estimator for $a = 1.5, 2.5, 4$. For these parameter values, we implement a fixed 64-dimensional version of the hierarchical estimator of Appendix D.3 on fresh regression samples with $\beta = 1$. The analysis family is fixed before any sample size is varied: two anchored cumulative polynomials and 62 compactly supported zero-tail modes form five complete dyadic levels. The population Gram matrix, whose minimum eigenvalue is 0.125 and condition number is 15.15, defines the dual reconstruction family and is never estimated from the data. Source coefficients use the truncated Taylor/IPW and integration-by-parts constructions, target coefficients use direct projection, and deterministic inverse-upper-variance weights combine the two channels. We use five predeclared truths—linear cumulative, boundary cumulative, interior tree, boundary tree, and mixed tree—under one common set of restrictions: $\|g\|_\infty, \|g'\|_\infty \le 0.8$, $\mathrm{Lip}(g') \le L = 64$, and conditional second moments at most one. The analysis-coefficient envelopes have constant one, a separate restriction satisfied by the same library. In particular, the cumulative cutoff minimizes the variance-plus-bias bound using $B_\nu(\epsilon) = L \sup_{[0,\epsilon]} |K_\nu| \epsilon^2$; the joint criterion includes the squared source weight on this bias term. Independent piecewise quadrature checks the actual truncation bias of all five truths against this bound at every candidate cutoff. Loss is computed exactly in the finite span as $(\widehat{\theta} - \theta)^\top G^{-1} (\widehat{\theta} - \theta)$, where G is the population analysis Gram matrix.

Sample-allocation design. To match the notation of Section 4.1, each longitudinal path below uses the target-anchored convention

$$n_T = N, \qquad n = N^{\kappa_n}, \qquad N = 2^8, 2^9, \ldots, 2^{14}.$$

For $a = 1.5$ we take $\kappa_n = 0.70, 0.80, 0.90$; for $a = 2.5$ we take $\kappa_n = 0.70, 0.90, 1.10$; and for $a = 4$ we take $\kappa_n = 1.00, 1.20, 1.35$. At $a = 4$, the three paths lie in the spatial branch, at the spatial–level split and equalizer, and in the level branch, respectively. At $a = 2.5$, the level path lies inside the narrow interval $1 < \kappa_n < 1.125$, with theoretical SAR exponent $1/45$. The path-specific stress truth is fixed before simulation: the interior tree for integral paths, the cumulative-boundary truth for level paths, the boundary tree for the spatial path, and the mixed tree at the $a = 4$ equalizer. These longitudinal curves probe different mechanisms without taking an across-truth maximum; comparisons across different stress truths are not controlled comparisons of κ_n alone.

The lower row instead fixes $n_T = 2^{11}$ and scans $\kappa_n = \log n / \log n_T$ through every theoretical breakpoint. The $a = 2.5$ and $a = 4$ scans use grids ending at $\kappa_n = 1.60$ and 2.20, respectively, to examine the source-rich return. Within each $a = 2.5, 4$ trajectory, source samples are nested across allocations and the same target sample is reused. This preserves each allocation's marginal regression model while coupling the scan points. These scans use the finite-library risks and SAR defined above. The library maximum and channel minimum are recomputed in every bootstrap draw. For these coupled scans, cross-allocation comparisons resample complete trajectories to preserve their dependence. Longitudinal slopes use the largest five values of N.

Regression results for $a = 1.5, 2.5, 4$. Figure 3(a)–(c) distinguishes finite-sample synergy from agreement with the theoretical SAR exponents. At $a = 1.5$, the empirical slopes are 0.106 $[0.070, 0.141]$, 0.148 $[0.102, 0.192]$, and 0.093 $[0.051, 0.135]$, versus references $0.050, 0.100, 0.050$. The equalizer path has the largest point estimate, but the reference exponents lie below the corresponding intervals, so the agreement is qualitative. At $a = 2.5$, the integral-rise and equalizer slopes are 0.065 $[0.032, 0.098]$ and 0.173 $[0.135, 0.211]$, compatible with 0.050 and 0.150. The level path has SAR slope 0.238 $[0.157, 0.317]$, well above $1/45 \simeq 0.022$. Its mixed-risk slope, -1.033 $[-1.111, -0.954]$, is compatible with n_T^{-1}, but target-only risk remains smaller than source-only risk at every sampled scale, unlike the asymptotic ordering on this path. Thus a compatible mixed-risk exponent alone does not validate the level-branch SAR exponent.

At $a = 4$, the spatial, equalizer, and level paths have SAR slopes 0.113 $[0.053, 0.173]$, 0.130 $[0.092, 0.167]$, and 0.227 $[0.164, 0.289]$, versus $0.133, 0.200, 0.100$. Only the spatial path is compatible with its SAR reference; the equalizer grows more slowly and the level path more quickly than predicted. Their normalized mixed-risk slopes, $0.009, 0.032, -0.022$, all have intervals containing zero, but target-only risk is still smaller on the level path at every sampled scale. The design covers both mechanisms without yet reproducing the asymptotic SAR ordering.

The fixed-n_T scans in panels (d)–(f) exhibit a source-rich decline, while retaining finite-sample fluctuations and displaced maxima. For $a = 1.5$, the largest sampled SAR occurs at the theoretical equalizer $\kappa_n = 0.8$, with value 3.84 $[3.46, 4.28]$. For $a = 2.5$, the largest sampled value is 4.13 $[3.61, 4.50]$ at $\kappa_n = 1.20$, falling to 1.23 $[1.19, 1.26]$ at 1.60. For $a = 4$, it is 3.71 $[3.15, 4.16]$ at 1.60, falling to 1.84 $[1.73, 1.94]$ at 2.20. These locations are descriptive grid maxima, not precisely estimated transitions. The predeclared right-side comparisons give SAR ratios $\mathrm{SAR}(1.60)/\mathrm{SAR}(1.20) = 0.297$ $[0.271, 0.341]$ for $a = 2.5$ and $\mathrm{SAR}(2.20)/\mathrm{SAR}(1.90) = 0.509$ $[0.401, 0.644]$ for $a = 4$, using paired trajectory bootstrap intervals. At the two right endpoints, source-only risk is smaller than target-only risk, with source/target ratios 0.085 and 0.273, respectively. The sampled maxima remain to the right of the theoretical equalizers 0.90 and 1.20, and neither curve is monotone throughout each theoretical branch. These fixed-library results support the qualitative transition without establishing exact minimax SLP endpoints or asymptotic boundedness of SAR.

Regression design at $a = 2, 3$. For $a = 2, 3$, we use the random-design regression model with $\beta = 1$ and $g(0) = 0$. The fixed library consists of $f(x) = 0$, $f(x) = 0.3$, and $f(x) = 0.2 + 0.4x$, with $g(x) = \int_0^x f(t)\,dt$. All satisfy $\mathrm{Lip}(f) \le H = 64$ and $\|f\|_\infty, \|g\|_\infty \le 0.8$. A fixed infinite orthonormal piecewise-linear multiresolution system replaces the fixed-dimensional analysis family above. Appendix E.2.1 gives the coefficient construction, risk decompositions, and numerical checks. This is an explicit realization of the coefficient mechanism, not the identical spline construction in the proof. No Gaussian prior is used. Risks and SAR are computed over this three-function library as defined above. Because the libraries and estimators differ from those at $a = 1.5, 2.5, 4$, vertical risk constants should not be compared across the two designs.

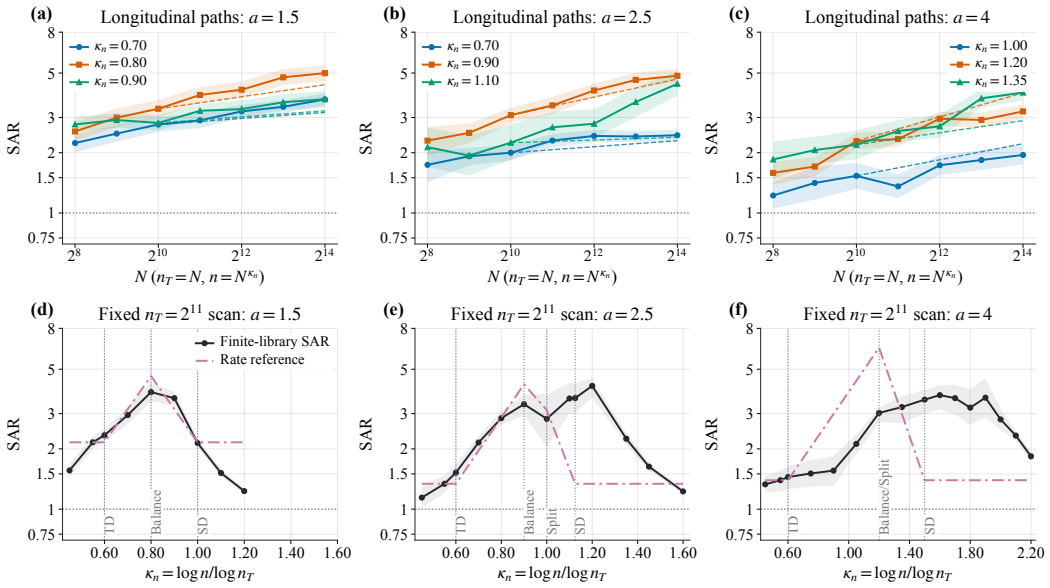

Figure 3: Fixed-dimensional raw-sample diagnostics at $a = 1.5, 2.5, 4$ for Section 4.2, with $\beta = 1$ and cumulative bias bounds using the common smoothness radius $L = 64$. Panels (a)–(c) use $n_T = N$ and $n = N^{\kappa_n}$ and show SAR for predeclared mechanism-specific truths; same-color dashed segments are theoretical tail-rate shapes aligned at the first fitted scale. Panels (d)–(f) fix $n_T = 2^{11}$ and show finite-library SAR against $\kappa_n = \log n / \log n_T$; purple dash–dot curves are aligned theoretical shapes and vertical dotted lines mark the TD boundary, equalizer, mechanism split, and SD boundary when distinct. Panels (e,f) use paired trajectories on grids ending at $\kappa_n = 1.60, 2.20$, respectively. Bands are 95% bootstrap intervals and horizontal dotted lines mark SAR$= 1$.

The longitudinal allocations are $n_T = N$ and $n = \mathrm{round}(N^{\kappa_n})$ for $\kappa_n = 0.70, 0.90, 1.00$ and $N = 2^{12}, 2^{16}, \ldots, 2^{64}$. We also fix $n = 2^{24}, 2^{40}$ and scan $\kappa_T = \log n_T / \log n$ from 0.50 to 1.75, including the polynomial and logarithmic branch boundaries. Sample sizes are rounded to integers. Each fixed-n scan uses one common numerical resolution, so source-only risk is exactly invariant to n_T. All 180 configurations use analytic expectations of the regression estimators, including random-design variance and actual bias. Raw-regression Monte Carlo checks only the six configurations $n = n_T = 2^{12}, 2^{16}, 2^{20}$ at $a = 2, 3$, with 400 repetitions each. SAR intervals use 4000 paired bootstrap resamples, recomputing the library maximum and channel minimum within each draw. Larger-N curves contain no simulated losses or sampling intervals.

Figure 4 reports these longitudinal paths and fixed-source scans together with their theoretical rate references.

Theoretical SAR references. Let $s_a(n, n_T) = \min\{r_{S,a}(n), r_T(n_T)\}/r_{J,a}(n, n_T)$, where $r_T = n_T^{-4/5}$, $r_{S,2} = (\log n)/n$, $r_{S,3} = n^{-4/5}$, and $r_{J,a}$ is the complete unit-constant piecewise expression in (14) or (15) at $\beta = 1$, including its logarithmic factors. Each longitudinal reference is multiplied by a single constant to match the corresponding analytic-risk SAR at $N_0 = 2^{12}$. The scan reference uses $n = 2^{40}$ and is aligned once at $\kappa_T = 1$. No horizontal fitting or separate branch alignment is applied. At a branch interface we use the expression on the left for anchoring and draw the one-sided reference segments separately: the order formulas need not agree in their constants, particularly at $a = 3$. Their aligned constant levels outside the SLP window need not equal one.

Results at $a = 2, 3$. Figure 4 shows increasing longitudinal SAR and hump-shaped allocation scans for these estimators. The reference comparison reveals nonuniform rate agreement. At $a = 2$, the terminal analytic SAR divided by its anchored reference is .856, .126, and .302 for $\kappa_n = .70, .90, 1.00$; the latter two paths grow substantially more slowly over this range. At $a = 3$ the corresponding ratios are 1.894, 1.758, and 2.139, so finite-range drift remains visible even when the

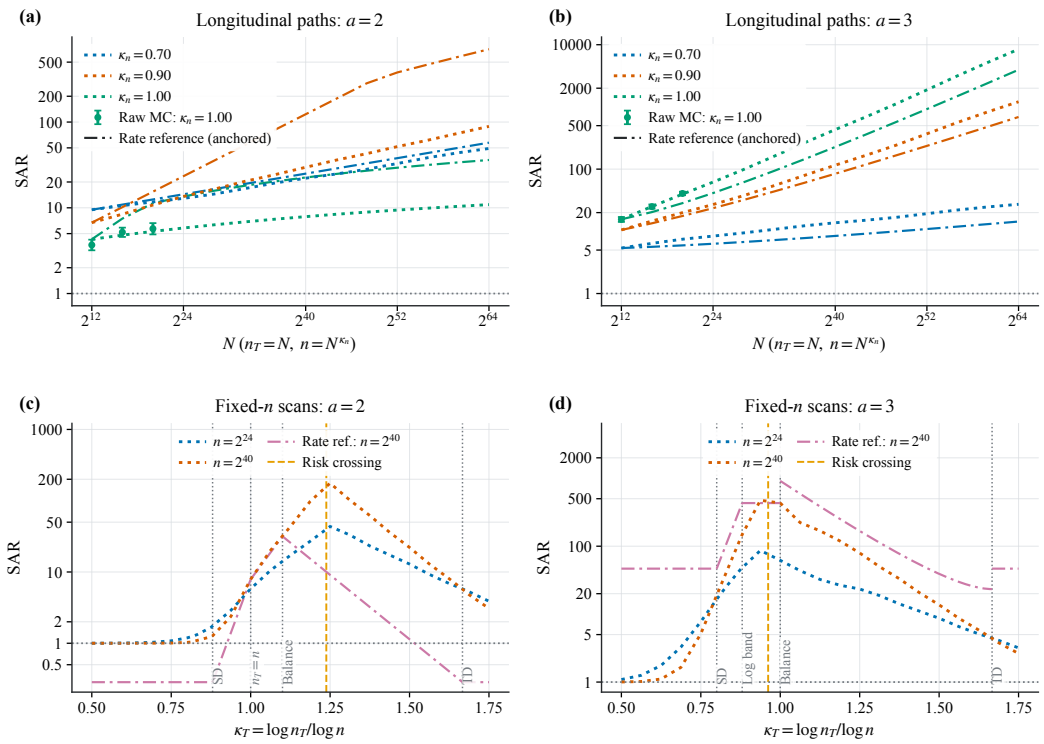

Figure 4: Integral-link SAR at $a = 2, 3$ with $\beta = 1$ on the fixed three-function library. Panels (a,b) use $n_T = N$, $n = N^{\kappa_n}$ for $\kappa_n = .70, .90, 1.00$; dotted curves are analytic-risk SAR and same-color dash–dot curves are theoretical references aligned once at $N_0 = 2^{12}$. Panels (c,d) fix $n = 2^{24}, 2^{40}$ and scan κ_T; purple dash–dot segments give the $n = 2^{40}$ reference aligned once at $\kappa_T = 1$. Gray vertical lines mark theoretical boundaries and balances, while orange dashed lines mark analytic-risk channel crossings. Markers at $\kappa_n = 1$ use 400 raw-regression repetitions with pointwise 95% percentile-bootstrap bars. The horizontal line is SAR$= 1$. These fixed-library diagnostics do not identify exact minimax window boundaries.

broad growth pattern agrees. The scan-grid maxima occur at $\kappa_T = 1.25$ for $a = 2$ and approximately 0.9375 for $a = 3$ at both fixed source sizes. These are finite-library maxima, not estimates of the asymptotic equalizers. At $n = 2^{40}$, interpolating the analytic log risk ratio gives single-channel crossings at 1.238 and $.962$, compared with the unit-constant theoretical balances 1.100 and 1.000. The differing risk constants can shift these crossings and the scan shape; the references are not horizontally adjusted to remove that displacement. At the right endpoint, SAR decreases from 3.89 to 3.08 for $a = 2$ and from 3.18 to 2.64 for $a = 3$ as the fixed source size increases. These constant-sized gains are compatible with no asymptotic rate improvement, but two source sizes do not establish boundedness. All longitudinal paths satisfy $n_T \geq n$. They do not test the narrow $a = 2$ window $n/\log n \lesssim n_T \lesssim n$ or all $a = 3$ branches with $n_T < n$. The fixed scans do not replace longitudinal rate checks in those regimes.

E.2.1 Implementation and risk checks

For $a = 1.5, 2.5, 4$, the normalization checks divide mixed risk by the complete Section 4.2 rate. At the theoretical equalizers $\kappa_n = 0.8, 0.9, 1.2$ for $a = 1.5, 2.5, 4$, the max/min ratios over the five fitted tail scales are $1.10, 1.15, 1.18$ and the normalized-risk slopes are $-0.039, -0.049, 0.032$. The $a = 2.5$ slope interval is $[-0.088, -0.009]$, indicating residual downward drift.

Estimator and risk calculation at $a = 2, 3$. The analysis system for these experiments contains roots 1 and $\sqrt{3}(2x - 1)$, and two orthonormal piecewise-linear wavelets on each dyadic interval. Each wavelet has zero zeroth and first moments. The roots and wavelets form nested piecewise-

polynomial spaces. Orthogonality makes integrated squared error additive without assuming independent coefficient errors. The source statistics use Taylor-compensated IPW for the roots and $-n^{-1}\sum_i Y_i\Psi(X_i)/h(X_i)$ for wavelets, where Ψ is their primitive. Target statistics use direct empirical projection. The fixed restrictions imply a conditional second-moment bound of one and a squared coefficient envelope $(H/8)^2 b^5$ at width b. Combination and shrinkage use these bounds, not truth-specific risks. Root cutoffs minimize a deterministic variance and squared-bias bound, including the source weight on the joint bias.

At $a = 3$, the unregularized boundary wavelets have finite IPW variance but an infinite fourth moment. We therefore apply zeroth-order Taylor compensation below $\delta = (H^2 n)^{-1/5}$, capped at half the wavelet's support width. This rule was fixed before production Monte Carlo. Actual signed boundary bias and modified variance enter the expected risk, while combination and shrinkage use variance plus squared-bias bounds. The bounded weights of each finite-resolution estimator yield finite variance of the sampled losses.

Expected risks include signal-induced design variance, observation noise, boundary and shrinkage bias, and the exact projection tail of the quadratic truths. Numerical resolution grows with the information range to approximate one fixed infinite system, rather than changing the truth class. The bound on the omitted tree contribution and projection tail is at most 0.453% of the corresponding fixed-library risk over the 180 cells. Increasing resolution by two levels at seven prescribed configurations changes the actual risks by at most 0.00478%. The effective block count in the accompanying diagnostics is the participation count $(\sum_d R_d)^2/\sum_d R_d^2$, not a count based on a fitted threshold. The blocks combine orthogonal modes within physical shells and boundary modes assigned by support scale. Their physical supports need not be disjoint, and their sampling errors need not be independent.

These calculations separate cumulative variance at $a = 2$ from oscillatory shell accumulation at $a = 3$. For the first cumulative functional with kernel $K(t) = 1 - t$, the compensated source weight gives noise variance $\sigma^2\{\frac{1}{2}\log(1/b) + \frac{1}{4}\}/n$. At $a = 3$ and $f = 0$, each effective orthogonal shell block contributes approximately $0.0127/N$ to combined oscillatory risk, and the participation count increases from 5.89 at $N = 2^{20}$ to 13.85 at $N = 2^{60}$. At $N = 2^{64}$, the corresponding log-normalized values remain 26.7% and 10.0% above the predicted leading coefficients for $a = 2, 3$. These are analytic-risk checks for the specified estimators, not independent Monte Carlo estimates of asymptotic constants or proofs that the logarithms are necessary for every estimator.

Monte Carlo checks and uncertainty. Common random numbers pair the truths and methods in the raw-regression Monte Carlo checks. The predeclared affine truth $f(x) = 0.2 + 0.4x$ avoids a library-maximization effect in direct mean-risk comparisons. At $a = 3$, $N = 2^{12}$, its 400-repeat source-only mean is 0.966 times the analytic expectation. The first cumulative cutoff contains only 0.074 expected source observations per repetition, making rare boundary observations influential. A separately seeded 400-repeat diagnostic gives ratio 0.975 with relative MC standard error 0.214.

Agreement is not uniform across total-risk checks. At $a = 3$, $N = 2^{20}$, zero-truth combined total risk is 0.909 times its expectation, or -3.14 estimated MC standard errors away. The corresponding affine-truth ratio is 0.963, or -1.16 estimated standard errors away. At $a = 2$, $N = 2^{12}$, the selected cumulative component is -2.84 estimated standard errors below its expectation. The reported errors describe approximate pointwise MC precision, and rare boundary contributions can make normal approximations inaccurate with 400 repetitions. The six component checks do not independently establish the large-N logarithmic limits.

E.3 Unbounded derivative link: regression Monte Carlo and surrogate-risk diagnostics

Design. We set $\beta = 2$, hence $s = 1$ and $a_c = 3.5$. Source covariates have density $h(x) = ax^{a-1}$ and target covariates are uniform. Source responses observe f, whereas target responses observe $\dot{f}$. The left column of Figure 5 reports the raw-sample experiment with

$$a = 6, \quad n = N = 2^{10}, 2^{12}, \ldots, 2^{20}, \quad \kappa_T \in \{.40, .45, .50\}, \quad n_T = \max\{4, \mathrm{round}(N^{\kappa_T})\}.$$

The longitudinal panel isolates the source-limited, equalizer, and target-limited paths inside the theoretical SLP window $.30 < \kappa_T < .60$. The fixed-source panel retains the full sampled range,

including the window boundaries and both outer regimes. Path selection follows the theoretical classification, not the observed SAR or fitted slope. Below-threshold and boundary/outer raw-regression controls are reported in Appendix E.3.2. The complete fixed-source scan for $a = 3$ is provided in the experimental materials. Each cell uses three distinct compact-bump scale stresses and a smooth quadratic control, with a 1025-point integration grid. Risks and SAR use this four-truth library, with mean losses computed before taking the library maximum. The maximum and channel minimum are recomputed in each of 2000 paired bootstrap draws. Horizontal raw scans fix $n = 2^{20}$, the truth library, and the source observations, and report individual risks as well as SAR. The $a = 6$ scan additionally includes $n_T = 2, 3$, corresponding to $\kappa_T = .05$ and $\log_2(3)/20 \approx .07925$, to expose the left side of the finite-sample gain profile. These two scan points are exempt from the longitudinal minimum of four target observations. They use the same estimator and serve as very-small-sample diagnostics.

The experiment at $a = 3.5$ uses surrogate-model Monte Carlo in a separately defined common-data diagonal Gaussian coefficient model, not an equivalent random-design regression experiment. Its basis-derived information constants, common observations, linear estimators, and restricted signal class are specified below. At nominal $N = 2^{24}, 2^{32}, \ldots, 2^{96}$, we follow

$$n = N, \qquad n_T \approx N^{3/5}/(\log N)^c, \qquad c \in \{.5, 1, 1.5, 2\}.$$

These positive-c paths probe the logarithmic SLP mechanism. Their effective sample exponent is

$$\kappa_T(N) = \frac{\log n_T}{\log N} \approx \frac{3}{5} - c\frac{\log\log N}{\log N};$$

unlike the power paths at $a = 6$, they do not hold κ_T fixed. The $c = 0$ bounded-SAR control and fixed-power controls are displayed separately in Appendix E.3.2. Fixed-source scans use $n = 2^{40}, 2^{64}, 2^{96}$. Their Monte Carlo grid has spacing .025 in κ_T; analytic-risk calculations for this surrogate are also evaluated on a fixed .0025 grid across the entire scan range. No information constant is fitted to a desired peak. Three fixed-source sizes diagnose the slow finite-information displacement at $a = 3.5$.

Raw-regression evidence and limitations. In Figure 5(c), the fixed-library source/target crossing for $a = 6$ is approximately .120, obtained by interpolating the log risk ratio, rather than the asymptotic balance .45. The $n_T = 2, 3, 4$ points have SAR $1.347, 1.616, 1.854$, revealing a rising left flank before the displayed grid maximum at $n_T = 16$. The paired SAR ratio from $n_T = 2$ to 3 is 1.200 $[1.128, 1.291]$. Of the 78 longitudinal raw cells, 77 are target-favoured; the sole source-favoured cell is the largest $a = 3, \kappa_T = .30$ configuration. All $a = 6$ longitudinal cells remain target-favoured at the simulated sizes.

At $a = 6$, $\kappa_T = .45$, terminal SAR is 1.166 $[1.133, 1.203]$. Its tail-slope interval in Table 3 excludes zero, but the estimate lies below the asymptotic exponent .05 and the finite-range continuous-rate reference. Thus these raw samples support a finite-range increase, not quantitative confirmation of the predicted equalizer-path rate. The surrogate's corresponding control at $a = 6$ has slope .0460 $[.0448, .0472]$ and terminal SAR 7.99; this supports the mechanism within the restricted coefficient experiment but does not establish the corresponding rate for random-design regression.

Logarithmic paths at $a = 3.5$ and remaining rate uncertainty. Figure 5(b) shows slowly increasing coefficient SAR on the positive-c paths; their empirical terminal SARs for $c = .5, 1, 1.5, 2$ are respectively $3.237, 2.160, 1.593, 1.277$. The asymptotic logarithmic equalizer $c = 1.5$ is not the best of these finite-range paths. At $n = 2^{96}$, the surrogate's analytic channel crossing is near $\kappa_T = .568$; it is reported as a model-specific finite-information crossing, not a match to a unit-constant minimax formula.

Appendix E.3.2 reports the normalization checks and remaining uncertainty about the logarithmic rate using the continuous reference ρ_a defined below.

E.3.1 Implementation details

Regression stress library and mixed fit. Let $\phi(z) = (1 - z^2)^4 \mathbf{1}\{|z| < 1\}$. The three scale stresses use

$$f_N(x) = \frac{b^2}{8}\sum_k (-1)^k \phi\left(\frac{x - c_k}{b}\right), \qquad \dot{f}_N(x) = \frac{b}{8}\sum_k (-1)^k \phi'\left(\frac{x - c_k}{b}\right).$$

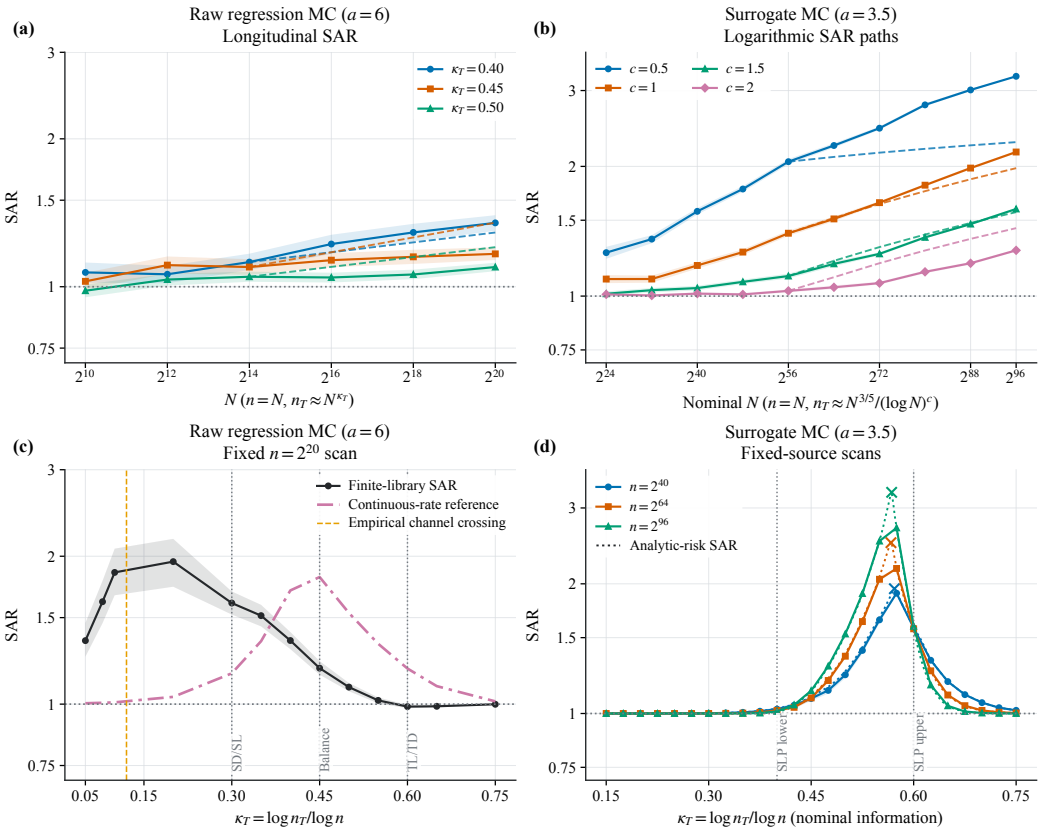

Figure 5: Derivative-link longitudinal SAR and full fixed-source scans with $\beta = 2$. Columns distinguish raw-regression Monte Carlo ($a = 6$, left) from surrogate-model Monte Carlo ($a = 3.5$, right). (a) Finite-library SAR on $\kappa_T = .40, .45, .50$ inside the theoretical SLP window; dashed guides have slopes $1/30, 1/20, 1/30$. (b) Logarithmic SLP paths with $c = .5, 1, 1.5, 2$ and varying effective $\kappa_T(N)$; dashed guides show continuous-rate SAR shapes. Same-color dashed guides in (a) and (b) are aligned at the first fitted scale. (c) The fixed-$n = 2^{20}$ scan with a common truth library and source data, including the $n_T = 2, 3$ small-sample diagnostics. The purple dash–dot curve is the continuous local-rate SAR reference without horizontal calibration; orange dashed lines mark observed channel crossings. Gray dotted vertical lines mark asymptotic balances and window endpoints. (d) Full-range scans at three fixed source-information levels; crosses mark analytic source/target equality and dotted lines mark .4, .6. Colored dotted curves in (d) give exact-risk SAR for the surrogate estimators on a finer scan grid, without sampling bands. Solid lines and markers summarize Monte Carlo estimates; bands are pointwise 95% paired-bootstrap intervals within each model. The joint-estimator SAR is shown without truncation below one. In (b,d), nominal N denotes coefficient information, not raw observations.

Here b is the bump radius. Centers start at $\delta L + b$, have spacing $2.5b$, and end no later than $.92L - b$; the source-, target-, and mixed-scale stresses have $\delta = .12, .20, .16$, respectively. Since the supports are disjoint and $\|\phi''\|_\infty = 8$, each derivative is Lipschitz with constant at most one. Below a_c, the source scale is $(L, b) = (1, .18n^{-1/5})$; above a_c, it is $(L, b) = (n^{-1/(a+4)}, .18n^{-1/(a+4)})$. The target scale is $(1, .18n_T^{-1/3})$. The mixed scale uses $L = 1$ below a_c and $L = 1 \wedge (n_T^{5/3}/n)^{1/(a-1)}$ above it, with $b = .18 \min(n_T^{-1/3}, L)$. Thus a small boundary region is retained by shrinking the radius, not by moving the support to the interior. The fourth truth is $f(x) = x^2/4$. These are explicit scale stresses, not asserted exact least-favourable functions for the regression experiment. For horizontal scans, all four truths are constructed once using $n_T = \text{round}(n^{.45})$ and held fixed.

The degree-one local fit uses an Epanechnikov kernel, exact truncated-window masses, and bandwidth balance $b^{2s}\{nH_a(x, b)b^2 + n_T H_T(x, b)\} = 1$, restricted numerically to $[10^{-4}, .45]$. Source-

only and target-only fits solve the corresponding one-channel balances. A trace-scaled ridge of 10^{-9} stabilizes the normal equations. Projection onto the known derivative bound $[-1, 1]$ is a statistical safeguard, not merely a numerical rounding operation. All methods and truths share the same covariates and Gaussian noises within each repetition. In a fixed-source scan, the source observations and truth library are identical at every n_T; target samples are nested. Released diagnostics separate conditional squared bias and noise risk before projection from the reported projected risk. Independently seeded grid refinement from 1025 to 2049 points changes raw mean risks by at most .0083% on six 400-repeat diagnostic cells.

Surrogate model, Monte Carlo losses, and analytic risks. On shell $[L, 2L]$, $L = 2^{-j}$, use levels with cell width $b = L2^{-\ell}$ and multiplicity $m = 2^{\ell}$, for $1 \le j \le 36$ and $0 \le \ell \le 36 - j$. Set $J_0 = \int_{-1}^{1} \phi^2$, $J_1 = \int_{-1}^{1} (\phi')^2$, $K = J_0/(4J_1) = 7/136$, $A = \sqrt{J_1}/(8\, 2^{3/2})$, and $\sigma = .25$. Here b denotes cell width, so its bump radius is $b/2$. The normalized derivative mode has primitive energy Kb^2 and the coefficient envelope $t_b = Ab^{3/2}$. We explicitly replace the design density by its shell average $\bar{h}_L = (2^a - 1)L^{a-1}$ and use the diagonal Gaussian experiment

$$Z_S = \theta + I_S^{-1/2}\xi_S, \qquad Z_T = \theta + I_T^{-1/2}\xi_T, \qquad I_S = \frac{n\bar{h}_L K b^2}{\sigma^2}, \quad I_T = \frac{n_T}{\sigma^2},$$

with independent standard normal coordinates. Its orthogonal coefficient loss is part of the surrogate definition; neither diagonalization nor shell averaging is asserted equivalent to the random-design regression experiment.

All three estimators use the same coordinates and observations. For information I, let $w(I) = t_b^2 I/(1 + t_b^2 I)$. The source and target fits are $w(I_S)Z_S$ and $w(I_T)Z_T$; the joint fit is

$$\widehat{\theta}_J = w(I_S + I_T)\frac{I_S Z_S + I_T Z_T}{I_S + I_T}.$$

The sequence signal class activates at most one level per shell, with $|\theta_k| \le t_b$ on that level. For each fixed estimator, choose on every shell the level maximizing $mt_b^2(1 - w)^2$, before generating data. The resulting three truths attain the respective worst risks for these estimators on this restricted class; every estimator is evaluated on all three truths, and variance from inactive levels is included. This does not establish minimaxity over all estimators. Monte Carlo losses are computed from common Gaussian signal-direction projections and their orthogonal-noise Wishart sums, an exact aggregation of squared coefficient errors. They are not random draws from a prescribed theoretical risk. We separately evaluate the same estimators' exact expected losses for analytic-risk comparisons. The unresolved signal-energy bound is $4A^2 2^{-72}$, below 3.4×10^{-8} of joint risk. Extending the coefficient resolution from 36 to 40 produces no detectable floating-point change in the analytic risks.

Continuous rate reference. For Section 4.3 normalization, solve the unrestricted exact local-mass balance for $b_*(x)$ and integrate $\rho_a(n, n_T) = \int_0^1 b_*^2(x)\, dx$, including the boundary region. This continuous diagnostic avoids jumps caused by choosing unit constants independently in the branches of an $\asymp$ formula. It is a rate-shape reference, not the exact finite-sample risk of either experiment. The comparison omitting the logarithmic factor at $a = 3.5$ replaces it by $n^{-2/5}$. In the supplementary all-path diagnostic, both normalizations are anchored to one at the first scale, without fitting slopes or channel-information constants. The supplementary representative-path figure reports unrescaled $\widehat{R}_J/\rho_a$, matching the display convention of Section 4.2.

E.3.2 Negative controls and normalization

Figure 6(a,b) shows the six $a = 3 < a_c$ raw-regression paths, $\kappa_T = .30, .45, .50, .55, .60, .75$, and four $a = 6$ boundary and outer paths, $\kappa_T = .15, .30, .60, .75$. Below the threshold, the $\kappa_T = .55, .60, .75$ tail SAR-slope intervals contain zero, whereas $.30, .45, .50$ have positive finite-range slopes .0809 $[.0568, .1060]$, .0193 $[.0083, .0299]$, and .0096 $[.0015, .0180]$. Thus the controls do not support a blanket claim of stabilization over this range. A positive finite-range slope does not by itself contradict an asymptotically bounded SAR. The complete $a = 3$ scan has an observed single-channel crossing near .315, displaced from the asymptotic balance .60.

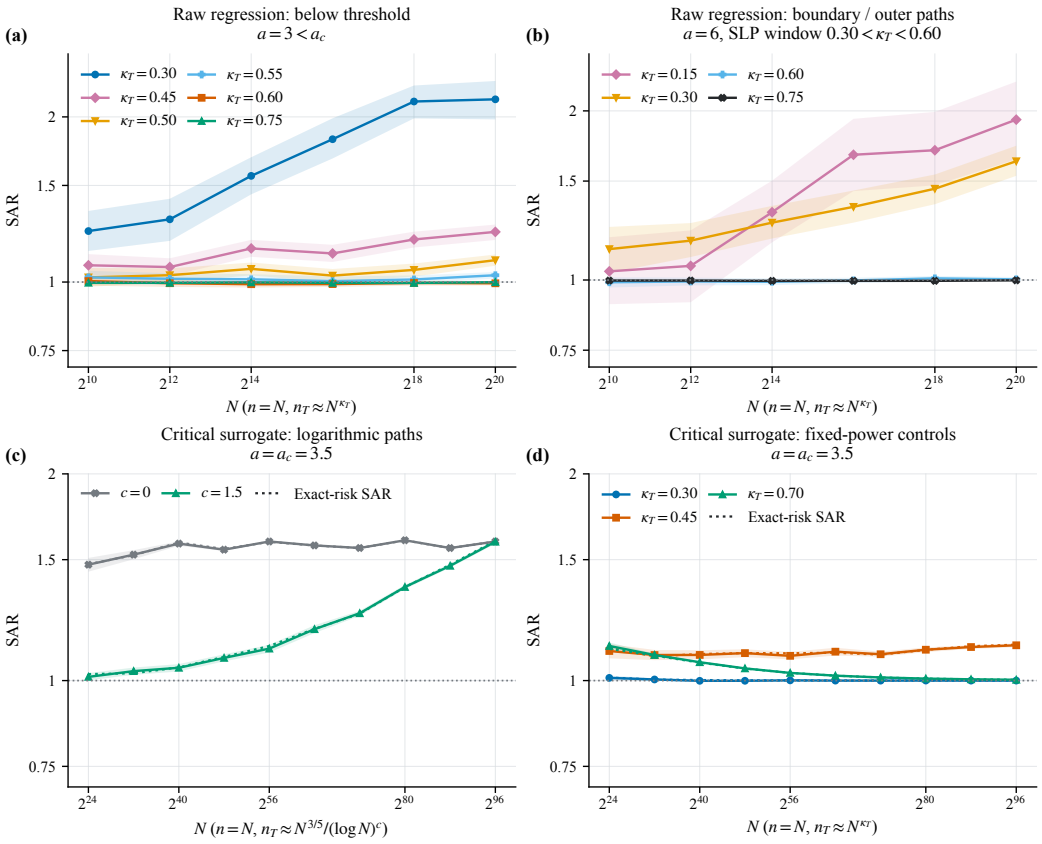

Figure 6: Derivative-link negative controls with $\beta = 2$ and 400 Monte Carlo repetitions per configuration. (a) All six raw-regression paths below the threshold, $a = 3 < a_c = 3.5$. (b) All four raw-regression paths at or outside the $a = 6$ SLP window $.30 < \kappa_T < .60$. (c) Coefficient-surrogate logarithmic paths at $a = 3.5$ with $c = 0$ and $c = 1.5$, where $n_T \approx N^{3/5}/(\log N)^c$. (d) All three fixed-power controls at $a = 3.5$, $n_T \approx N^{\kappa_T}$. Solid curves and markers give MC SAR estimates with pointwise 95% paired-bootstrap bands. Colored dotted curves in (c,d) are exact-risk SARs for the specified surrogate estimators, without sampling bands. Gray horizontal lines mark SAR$= 1$. Below-threshold finite-range growth and SAR below one are shown without truncation. In (c,d), nominal N denotes coefficient information.

At $a = 3.5$, panel (c) pairs the $c = 0$ bounded-SAR control with the theoretical logarithmic equalizer $c = 1.5$. Their terminal SARs are 1.594 and 1.593, respectively. The other positive-c paths remain in Figure 5(b). Panel (d) retains all three fixed-power controls $\kappa_T = .30, .45, .70$, which have no divergent SAR under the rate formula at $a = 3.5$.

The remaining normalization checks use the full-domain reference ρ_a. Over the last six scales at $a = 3.5$, the max/min ratios of $\widehat{R}_J/\rho_a$ for $c = 0, .5, 1, 1.5, 2$ are $1.050, 1.028, 1.049, 1.125, 1.277$. Without the logarithmic factor they are $1.050, 1.095, 1.143, 1.244, 1.426$. The logarithmic reference reduces drift, but the normalized-slope intervals for $c = 1.5, 2$ exclude zero. Thus the exact logarithmic rate remains unresolved in these finite-range diagnostics.

E.4 NUMERICAL SUMMARY

Table 3: Numerical summary of selected regression and coefficient-surrogate rate paths. We report the positive risk-decay exponent $\widehat{\alpha}_R = -d\log(\widehat{R})/d\log N$ and the SAR exponent $\widehat{\alpha}_S = d\log(\widehat{\mathrm{SAR}})/d\log N$; brackets are paired-bootstrap 95% intervals. Section 4.1 integral rows use candidate mean-risk minima with $\beta = 1$ and the six scales $2^{15}, \dots, 2^{20}$; its derivative row uses $\beta = 2$, raw joint risk, and $2^{14}, 2^{16}, 2^{18}, 2^{20}$. Section 4.3 uses $\beta = 2$. Section 4.3 fits use the last four raw scales or six coefficient scales; its references are effective slopes of the continuous local-rate integral and its one-channel analogues, not finite-risk predictions. Other references are asymptotic exponents.

Section	Path	$\widehat{\alpha}_R$ (95% CI)	Reference	$\widehat{\alpha}_S$ (95% CI)	Reference	Terminal SAR
4.1	$\kappa_n = .70$	.881 [.861, .901]	.850	.041 [.023, .057]	.050	1.78
4.1	$\kappa_n = .80$ (equalizer)	.930 [.903, .958]	.900	.112 [.065, .159]	.100	2.71
4.1	$\kappa_n = .90$	.966 [.929, .997]	.950	.051 [.009, .088]	.050	1.29
4.1	derivative control	.414 [.402, .426]	.400	$-.008$ [$-.012$, $-.004$]	0	1.00
4.3	$a = 3 < a_c, \kappa_T = .60$	.412 [.399, .424]	.406	.0012 [$-.0027$, .0048]	$-.0019$	.99
4.3	$a = a_c, c = 1.5$	.3920 [.3917, .3922]	.3962 (log)	.0130 [.0126, .0134]	.0123 (log)	1.59
4.3	$a = 6 > a_c, \kappa_T = .45$	.331 [.315, .345]	.337	.0145 [.0032, .0261]	.0280	1.17

F EMPIRICAL ILLUSTRATIONS WITH IO-VNBD

The IO-VNBD benchmark contains synchronized 10-Hz vehicle CAN and smartphone measurements collected on public-road trajectories (Onyekpe et al., 2021). Yaw rate $\omega(t)$ and unwrapped heading $\theta(t)$ are approximately related by

$$\theta(t) - \theta(0) \approx \int_0^t \omega(s)\, ds, \qquad \omega(t) \approx \theta'(t).$$

We first use observed signals to illustrate operator asymmetry, then calibrate a separate semi-synthetic benchmark under unbounded covariate shift.

F.1 REAL-DATA OPERATOR ASYMMETRY

Data and calibration. The fully observed analysis uses vehicle CAN yaw rate and smartphone GPS heading. After exact-payload deduplication and a prespecified link audit, we retain 21 nonoverlapping 180-second windows from 11 signal-distinct track files. The first 36 seconds form a calibration block and the remaining 144 seconds contain 1441 aligned analysis observations. Window selection uses both calibration and analysis-block link diagnostics. The detailed eligibility thresholds and nonoverlap rules are recorded in `supplementary/applications/preprocessing.md`.

On the calibration block, let $I(t) = \int_0^t \omega(s)\, ds$ and $\Theta(t) = \theta(t) - \theta(0)$. After centering both trajectories, we set

$$\widehat{c} = \frac{\sum_t \{I(t) - \bar{I}\}\{\Theta(t) - \bar{\Theta}\}}{\sum_t \{I(t) - \bar{I}\}^2}.$$

This scale, including its sign, is frozen before the analysis block. Writing $T = 144$ seconds and $x = (t - t_0)/T$, the forward task uses

$$f(x) = \frac{T\widehat{c}}{\pi}\omega(t), \qquad g(x) = \frac{\theta(t) - \theta(t_0)}{\pi}, \qquad g(x) \approx \int_0^x f(u)\, du.$$

The reverse task uses the same normalized signals with source level g and target derivative f. Thus no synthetic responses or artificial response noise are introduced.

Sampling and estimators. For integration we take $n = \max\{8, \mathrm{round}(N^\kappa)\}$, $n_T = N$, with $\kappa \in \{0.5, 0.8, 1\}$; the theory-relevant middle path is $\kappa = 0.8$. For differentiation we take $n = N$

and $n_T = \mathrm{round}(N^{0.6})$. In both directions $N \in \{48, 72, 108, 162, 243, 360, 480\}$ and risk is held-out target MSE. Within each analysis block, five interlaced folds are formed from consecutive five-second dependence blocks. Repetition r holds out fold $r \bmod 5$ and samples disjoint source and target training timestamps by temporal strata. This preserves the quasi-regular recorded time design without using test labels.

For integration, target-only estimation is boundary-truncated local linear regression with Epanechnikov kernel and bandwidth $1.1n_T^{-1/5}$. The source primitive integrates the piecewise-linear interpolant through the sampled yaw labels. The mixed estimator smooths the target residual between that primitive and the observed heading, using correction bandwidth $1 \wedge (n/n_T)^{1/2}$. The three-path sensitivity check uses this same raw mixed estimator for every κ, rather than changing estimators at the theoretical branch boundaries.

For differentiation, the operator-aware degree-one local fit combines source level rows $[1, z]$ with target derivative rows $[0, b^{-1}]$. At evaluation location x, b numerically solves

$$b^2\{nH(x, b)b^2 + n_T H_T(x, b)\} = 1,$$

with the exact mass of the boundary-truncated uniform window. A small ridge 10^{-9} times the local Gram scale and clipping to $[-40, 40]$ are used only for numerical stability. Forty percent of the sampled target labels form an internal validation subset. Starting from source-only, the selector changes to target-only or joint only if the paired validation improvement exceeds one estimated standard error; after selection, the chosen candidate is refitted on all training labels. The held-out fold is never used for this decision.

Risk summaries and results. Each window–budget cell has 400 repetitions. We average first within window and then within track, giving the 11 tracks equal weight. Pointwise 95% track-resampling intervals use 4000 bootstrap resamples of complete track-level paths, preserving cross-budget and cross-estimator dependence. Figure 7 uses the track-geometric quantity $\exp\{11^{-1}\sum_{j=1}^{11}\log(\mathrm{SAR}_j)\}$ for the operator comparison.

As a descriptive sensitivity check, the middle integration path has the largest terminal ratio-of-mean-risks SAR, 1.177, whereas the $\kappa = 0.5$ and $\kappa = 1$ values are 1.064 and 0.998. All three pooled-risk intervals include one, so we do not interpret these points as estimating a phase boundary. On the $\kappa = 0.8$ integration path in Figure 7(b), the terminal track-geometric SAR is 1.385 with 95% interval $[1.126, 1.699]$, and 9 of 11 tracks favor mixing. In the differentiation direction, the corresponding safeguarded SAR is 0.954 $[0.930, 0.976]$, with only 1 of 11 tracks favoring transfer. Eight tracks lie in the joint quadrant "integration gain, differentiation no gain." These analyses give a selection-conditional descriptive illustration consistent with the operator asymmetry in Remark 1. The exploratory track-resampling intervals condition on the audited windows and do not account for window-selection uncertainty. We do not use this analysis to verify the exact minimax exponents.

F.2 Semi-Synthetic Calibration under Unbounded Covariate Shift

The observed time design cannot isolate an unbounded density ratio. The semi-synthetic benchmark therefore uses constructed covariates and latent functions, with observation errors calibrated from vehicle GPS heading and CAN yaw rate.

Residual calibration. We form a separate residual library from audited 600-second windows. On the experiment portion of each window, a cubic spline with nine interior knots is fit to normalized heading. Heading minus this spline gives the level-residual pool, and calibrated yaw rate minus the spline derivative gives the derivative-residual pool. Each pool is centered and standardized to unit root-mean-square magnitude within window. Semi-synthetic responses add independent draws from the relevant pool, scaled by 0.25. This construction retains the empirical marginal error shapes, but not temporal dependence, heteroskedasticity tied to the covariate, or cross-channel coupling.

Design. For integration and for the primary differentiation path, source covariates follow $Q_S = \mathrm{Beta}(6, 1)$ and target covariates are uniform, so $\frac{dQ_T}{dQ_S}(x) = \frac{1}{6x^5}$ is unbounded as $x \downarrow 0$. We take $N = 2^{10}, \ldots, 2^{18}$ source observations and $n_T = \mathrm{round}(N^q)$ target observations. Risks are averaged

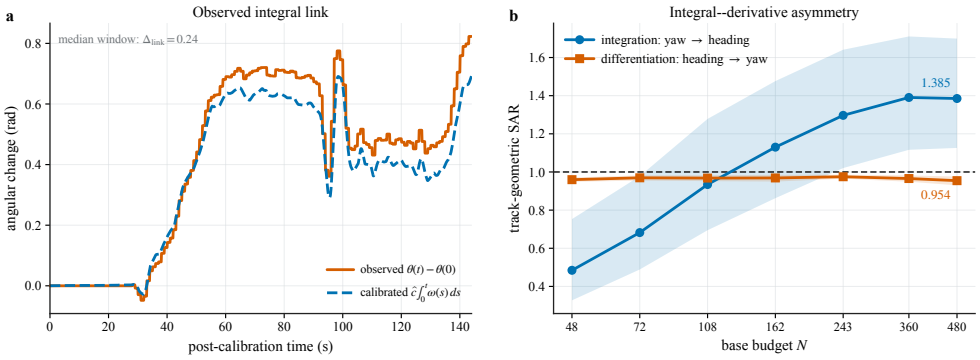

Figure 7: Real-data integral–derivative comparison on IO-VNBD. (a) Observed heading change and the calibrated yaw-rate integral in the deterministically selected median-discrepancy window. (b) Equal-track geometric SAR on the integration middle path and the safeguarded differentiation path. Bands are exploratory 95% track-resampling intervals conditional on the audited windows. The figure supports qualitative operator asymmetry but is not used to estimate an asymptotic exponent.

over repetitions and windows within track before the signal-distinct tracks receive equal weight; the bootstrap preserves each complete sample-size path.

For integration, the source observes g' and the target observes g. We use $q \in \{0.45, 0.60\}$ and a seven-truth library containing the zero truth, signed cumulative and interior alternatives, and their signed sums. Source-only, target-only, and mixed operator-aware ridge estimators use the same finite dictionary within each configuration. Source design rows evaluate dictionary derivatives, and target rows evaluate dictionary function values. The objective sums the squared residuals from both channels, each scaled by 0.25^{-2}, and adds the squared Euclidean norm of the coefficient vector with penalty weight one. Single-channel fits omit the other channel. No rate equivalence with the hierarchical estimator in Appendix D.3 is asserted.

The reported risk is the maximum mean integrated squared error over this library, and the empirical SAR is $\min(R_S^{\mathcal{L}}, R_T^{\mathcal{L}})/R_C^{\mathcal{L}}$. This panel uses 23 windows from 15 signal-distinct tracks, 16 residual draws per window and sample size, and 9999 paired track-bootstrap draws.

For differentiation, the source observes a level function f and the target observes f'. This semi-synthetic panel uses the single-vertex scale-dependent Assouad generator in `supplementary/simulations/derivative_model.py`, whereas Section E.3 uses a four-truth regression library. The ratios in this panel are single-truth diagnostics, not finite-library or minimax SAR. The operator-aware mixed local-polynomial estimator is compared with its two single-channel versions on the primary path $(a, q) = (6, 0.45)$. The controls are the target-dominating path $(6, 0.60)$, the below-threshold path $(3, 0.60)$, and the bounded-ratio path $(1, 0.60)$. This panel uses 33 windows from 20 signal-distinct tracks, 20 residual draws per window and sample size, and 4000 paired track-bootstrap draws.

Results and scope. Figure 8(a) shows the integration SAR paths. On the $q = 0.60$ path, the all-N log-SAR slope is 0.0487 with 95% interval $[0.0309, 0.0665]$, and the terminal SAR is 2.62 $[2.35, 2.92]$. The $q = 0.45$ path has slope 0.0007 $[-0.0116, 0.0163]$ and terminal SAR 1.35 $[1.24, 1.46]$; the paired $q = 0.60$ minus $q = 0.45$ slope contrast is 0.0480 $[0.0220, 0.0708]$.

Figure 8(b) shows the differentiation calibration. On the primary $(a, q) = (6, 0.45)$ path, the tail SAR slope is 0.0323 $[0.0161, 0.0473]$, and the terminal SAR is 1.185 $[1.153, 1.213]$. The primary slope exceeds the largest control slope by 0.0178 $[0.0001, 0.0326]$. The lower endpoint is close to zero and the controls retain small positive finite-range drifts, so the derivative evidence is more tentative than the integration result.

These panels are conditional, estimator-specific finite-range calibrations. They do not estimate the minimax SAR, establish asymptotic divergence, or support a numerical comparison of the two op-

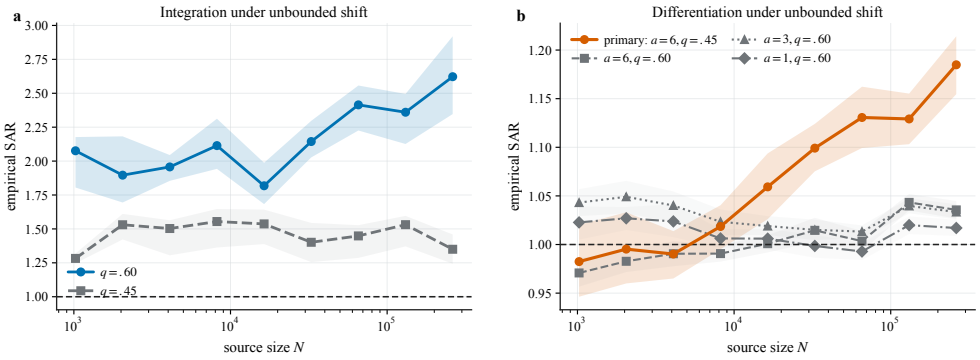

Figure 8: Semi-synthetic unbounded-shift calibration using empirical errors derived from the IO-VNBD signals introduced above. (a) Integration SAR for the $q = .45$ and $q = .60$ paths. (b) Differentiation SAR on the primary path and three controls for the sample-size regime, singularity threshold, and bounded density ratio. Bands are pointwise 95% signal-distinct-track bootstrap intervals, and the horizontal dashed lines mark SAR$= 1$.

erator effects: the two directions use different truth families, estimators, repetition counts, and slope ranges.