©2026 Ada Academica https://adac.ee

Eur. J. Stat. 6 (2026) 18
doi: 10.28924/ada/stat.6.18

Adaptive Functional Local Polynomial Estimation of the Conditional Hazard Function Under Missing-at-Random Responses

Jamel Kenouza[1], Ibrahim Sadok[2,3,*], Boubaker Mechab[4], Samir Benaissa[4]

[1] *Laboratory of Mathematics, Statistics and Computer Science, Department of Mathematics, University of Naama, Naama, Algeria*

[2] *Department of Mathematics and Computer Science, Faculty of Exact Sciences, University of Bechar, Algeria*

[3] *Laboratory of Mathematics, Djillali Liabes University of Sidi Bel-Abbès, P. O. Box 89, 22000, Sidi Bel-Abbès, Algeria*

[4] *Laboratory of Statistics and Stochastic Processes, Department of Probability and Statistics, Djillali Liabes University, Sidi Bel Abbes 22000, Algeria*

[*] *Correspondence: ibrahim.sadok@univ-bechar.dz*

Abstract. We propose an adaptive functional local polynomial estimator of the conditional hazard function of a scalar response, observed only partially, given a functional covariate valued in a semi-metric space. The estimator extends the functional local linear method of Kenouza, Mechab and Benaissa to a polynomial order that grows with the sample size through the adaptive rule $p_n = \lfloor \log n \rfloor$. As in the local linear case, the conditional distribution and the conditional density are estimated first, the hazard estimator being their ratio. Under missing-at-random (MAR) responses and the same regularity setting as the local linear analysis (small-ball probability, Hölder-type continuity, standard kernel and bandwidth conditions, together with a non-degeneracy condition for the growing design), we establish the almost complete convergence with rate of the three estimators, following the same four-step proof strategy: identification of the conditional law on complete cases, control of the bias by the Hölder exponents, an exponential inequality for the stochastic term, and control of the denominator. The growing order is shown to be a smoothness-adaptive device: it automatically attains the local-polynomial bias rate permitted by the (unknown) functional regularity, while the slow logarithmic growth keeps the local design matrix stable. A Monte Carlo study illustrates the finite-sample behaviour under several missingness levels.

1. Introduction

Functional data analysis has become central to modern statistics because many observations are naturally recorded as curves, trajectories, spectra or signals; classical multivariate tools

Received: 4 Jul 2026.

Key words and phrases. functional data analysis; conditional hazard function; local polynomial estimation; missing at random; almost complete convergence; adaptive smoothing.

are then inadequate and infinite-dimensional nonparametric techniques are required. For the nonparametric framework we refer to the monograph of Ferraty and Vieu [8] and to the surveys [12, 13, 16].

Let X be a functional covariate valued in a semi-metric space $(\mathcal{F}, d)$ and let Y be a real-valued response. The conditional hazard function of Y given $X = x$ is

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}, \qquad F^x(y) < 1, \tag{1}$$

where F^x and f^x are the conditional distribution and density of Y given $X = x$. Hazard estimation is fundamental in reliability, survival analysis, risk prediction and biomedical applications. In the functional setting, Ferraty, Rabhi and Vieu [6] established the almost complete convergence with rates of the kernel estimator of h^x in several contexts, including censored and dependent variables [17]. The local linear (more generally local polynomial) technique has well-documented advantages over the kernel method, namely bias reduction and boundary adaptation [1, 3, 5]. The local linear estimate of h^x under missing-at-random responses was studied by the present author and coauthors [9]; the aim of the present work is to extend that estimate to an *adaptive local polynomial* estimate, keeping the same modelling idea and the same line of proof.

More generally, flexible lifetime and hazard-rate modelling has been developed in the reliability literature by Sadok and coauthors [2, 15], while non-informative Bayesian estimation for dispersion models is studied in [14]. These references provide complementary parametric and Bayesian perspectives to the nonparametric functional approach adopted here.

In applications responses are frequently incomplete. We use the missing-at-random (MAR) mechanism

$$\mathbb{P}(\delta = 1 \mid X, Y) = \mathbb{P}(\delta = 1 \mid X) = \pi(X), \tag{2}$$

where δ is the response indicator. MAR mechanisms for functional data were treated by Ferraty, Sued and Vieu [7], Ling, Liang and Vieu [10], Efromovich [4] and, in the multivariate case, Little and Rubin [11]. Under (2) the conditional law of Y given X is preserved among complete cases, so the complete-case localized estimator targets the conditional functional without inverse-probability weighting; this is made explicit in Lemma 5.1.

The contribution is threefold.

- We construct a functional local polynomial estimator of h^x under MAR which reduces, for order $p = 1$, to the local linear estimator of [9].
- We let the order grow as $p_n = \lfloor \log n \rfloor$ and show that this is a *smoothness-adaptive* choice: it reaches the bias rate $O(h_K^{b_1})$ permitted by the Hölder regularity, for *any* unknown finite b_1 (Remark 4.4), without prescribing the order in advance.
- We establish the almost complete convergence with rate, following the same four-step strategy as the local linear case: MAR identification, Hölder control of the bias, an exponential inequality for the stochastic term, and control of the denominator.

The paper is organised as follows. Section 2 introduces the model and the adaptive estimator. Section 3 states the assumptions and Section 4 the main results. The proofs are gathered in Section 5, organised as a sequence of lemmas mirroring the four steps. Section 6 discusses the bias–variance trade-off induced by the order, Section 7 reports a simulation study and Section 8 concludes.

2. Model and Estimator

Let $(X_i, Y_i, \delta_i)_{1\le i\le n}$ be independent copies of (X, Y, δ) with $X_i \in \mathcal{F}$, $Y_i \in \mathbb{R}$ and

$$\delta_i = \begin{cases} 1, & \text{if } Y_i \text{ is observed,} \\ 0, & \text{otherwise.} \end{cases}$$

Fix $x \in \mathcal{F}$, write $F^x(y) = \mathbb{P}(Y \le y \mid X = x)$ and assume that F^x admits a conditional density f^x. Let K be a kernel, H a smooth distribution function with derivative H', $h_K, h_H > 0$ two bandwidths, and $\beta(\cdot, x)\colon \mathcal{F} \to \mathbb{R}$ a known locating function with $\beta(x, x) = 0$ and $|\beta(z, x)| \le C_\beta\, d(z, x)$. For an order p set

$$m_p(z) = \sum_{k=0}^{p} a_k\, \beta(z, x)^k.$$

The local polynomial coefficients solve the complete-case weighted least squares problem

$$(\widehat{a}_0, \dots, \widehat{a}_p) = \operatorname*{arg\,min}_{a_0,\dots,a_p} \sum_{i=1}^{n} \delta_i\, K\!\left(\frac{d(X_i, x)}{h_K}\right) \left[H\!\left(\frac{y - Y_i}{h_H}\right) - \sum_{k=0}^{p} a_k\, \beta(X_i, x)^k \right]^2. \tag{3}$$

The conditional distribution estimator is $\widehat{F}^x_{p_n}(y) = \widehat{a}_0$; differentiating in y gives the conditional density estimator $\widehat{f}^x_{p_n}(y)$, and the conditional hazard estimator is

$$\widehat{h}^x_{p_n}(y) = \frac{\widehat{f}^x_{p_n}(y)}{1 - \widehat{F}^x_{p_n}(y)}. \tag{4}$$

For $p = 1$ this is exactly the functional local linear estimator of [9]. In the adaptive version studied here the order is

$$p_n = \lfloor \log n \rfloor. \tag{5}$$

Matrix form and equivalent weights: Let $u_i = \beta(X_i, x)/h_K$, $w_i = \delta_i\, K(d(X_i, x)/h_K)$ and $b(u) = (1, u, \dots, u^p)^\top$. With $U = (b(u_i)^\top)_i$, $W = \operatorname{diag}(w_i)$ and $Z = \big(H((y - Y_i)/h_H)\big)_i$, the solution of (3) is

$$\widehat{a} = (U^\top W U)^{-1} U^\top W Z, \qquad \widehat{F}^x_{p_n}(y) = e_1^\top \widehat{a},$$

where $e_1 = (1, 0, \dots, 0)^\top$. Introduce the normalized empirical moment matrix

$$\widehat{S}_n = (n\, \mathbb{E}w_1)^{-1} U^\top W U, \qquad \widehat{S}_{n,jk} = (n\, \mathbb{E}w_1)^{-1} \sum_i w_i u_i^{j+k}, \quad 0 \le j, k \le p,$$

and the equivalent weights

$$\ell_i = e_1^\top \widehat{S}_n^{-1}\, b(u_i), \qquad \sum_i \frac{w_i \ell_i}{n \mathbb{E}w_1} = e_1^\top \widehat{S}_n^{-1} \widehat{S}_n e_1 = 1, \tag{6}$$

so that $\widehat{F}^{x}_{p_n}(y) = (n\mathbb{E}w_1)^{-1}\sum_i w_i\ell_i\, H((y-Y_i)/h_H)$. The weights ℓ_i reproduce every polynomial of degree $\le p$ in $\beta(\cdot,x)$ exactly; this reproducing property is the source of the bias reduction and is used in Lemma 5.3.

3. Assumptions

Write $\varphi_x(h) = \mathbb{P}(d(X,x)\le h)$ for the small-ball probability and abbreviate $\varphi := \varphi_x(h_K)$. We use the same regularity setting as the local linear analysis [3, 6, 8, 9], with two additions, (H6)–(H7), required only because the order grows.

(H1) **Small ball:** For every $h>0$, $\varphi_x(h)>0$ and $\varphi_x(h)\to 0$ as $h\to 0$.

(H2) **Hölder continuity:** There exist $b_1,b_2>0$ and $C>0$ such that, for all $(x_1,y_1),(x_2,y_2)$ in a neighbourhood of (x,y),

$$|F^{x_1}(y_1)-F^{x_2}(y_2)| \le C\{d(x_1,x_2)^{b_1}+|y_1-y_2|^{b_2}\},$$

and the analogous condition holds for f^x.

(H3) **Kernels:** K is bounded, non-negative and supported on $[0,1]$ with $K(1)>0$; H' is a bounded, integrable kernel with $\int H'=1$ and finite moments up to order $\lceil b_2\rceil$.

(H4) **Bandwidths:** $h_K\to 0$, $h_H\to 0$ and $\dfrac{\log n}{n\,h_H\,\varphi}\to 0$.

(H5) **Missingness:** $\pi(\cdot)$ is continuous in a neighbourhood of x and $\pi(x)>0$.

(H6) **Adaptive order / stability:** $p_n=\lfloor \log n\rfloor$ and $\dfrac{p_n}{n\,h_K\,\varphi}\to 0$.

(H7) **Design non-degeneracy:** The population moment matrix $S=(S_{jk})_{0\le j,k\le p_n}$, $S_{jk}=\mathbb{E}[w_1u_1^{j+k}]/\mathbb{E}[w_1]$, has smallest eigenvalue bounded below uniformly in n: $\lambda_{\min}(S)\ge\lambda_0>0$.

Remark 3.1. *(H1)–(H5) are exactly the conditions used in the functional local linear analysis of the conditional distribution, density and hazard [3, 6, 8, 9]. (H6)–(H7) are the only additions required by a growing order: they keep the local design matrix stable as its dimension p_n+1 increases.*

4. Main Results

Theorem 4.1 (Conditional distribution)**.** *Under* (H1)–(H7),

$$\left|\widehat{F}^{x}_{p_n}(y)-F^x(y)\right| = O\Big(h_K^{\min(p_n+1,\,b_1)}\Big)+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\tfrac{\log n}{n\,\varphi}}\right).$$

Theorem 4.2 (Conditional density)**.** *Under* (H1)–(H7),

$$\left|\widehat{f}^{x}_{p_n}(y)-f^x(y)\right| = O\Big(h_K^{\min(p_n+1,\,b_1)}\Big)+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\tfrac{\log n}{n\,h_H\,\varphi}}\right).$$

Theorem 4.3 (Adaptive conditional hazard)**.** *Assume* (H1)–(H7) *and* $F^x(y)<1$. *Then*

$$\left|\widehat{h}^{x}_{p_n}(y)-h^x(y)\right| = O\Big(h_K^{\min(p_n+1,\,b_1)}\Big)+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\tfrac{\log n}{n\,h_H\,\varphi}}\right), \tag{7}$$

and consequently $\widehat{h}^{x}_{p_n}(y)\to h^x(y)$ *almost completely.*

Remark 4.4 (The adaptive order is smoothness-adaptive)**.** *Since* $p_n = \lfloor \log n \rfloor \to \infty$*, for any fixed finite* b_1 *there is* n_0 *with* $p_n + 1 \geq b_1$ *for* $n \geq n_0$*; hence* $\min(p_n + 1, b_1) = b_1$ *eventually, and* (7) *attains the full Hölder bias rate* $O(h_K^{b_1})$ without prescribing the order in advance*. A fixed order* p *saturates the bias at* $O(h_K^{\min(p+1,b_1)})$ *and is suboptimal whenever the regularity exceeds* $p+1$*. For* $p = 1$ *the three theorems reduce to the local linear rates of [9]. If the functionals are infinitely smooth (analytic),* $\min(p_n + 1, b_1) = p_n + 1$ *and the functional bias decays faster than any power of* h_K*, becoming asymptotically negligible.*

Remark 4.5 (Rate balancing)**.** *Once* $n \geq n_0$ *the dominant error in* (7) *is* $O(h_K^{b_1}) + O(h_H^{b_2}) + O_{a.co.}(\sqrt{\log n/(nh_H\varphi)})$*. Balancing the response-direction bias against the stochastic term,* $h_H^{b_2} \asymp \sqrt{\log n/(nh_H\varphi)}$*, and treating the typical functional decay* $\varphi = \varphi_x(h_K)$ *as fixed at the chosen* h_K*, gives* $h_H \asymp \big(\log n/(n\varphi)\big)^{1/(2b_2+1)}$*, after which* h_K *is chosen so that* $h_K^{b_1}$ *is of the same order; in practice* h_K, h_H *are selected by cross-validation on the complete cases (Section 7).*

5. Proofs

We follow the same four-step strategy as for the local linear estimate [9]. The hazard estimator (4) is a ratio, so it suffices to control $\widehat{F}^x_{p_n} - F^x$ and $\widehat{f}^x_{p_n} - f^x$ and then to bound $1 - \widehat{F}^x_{p_n}$ away from zero. Throughout, C denotes a generic positive constant that may change from line to line.

5.1. Identification of the conditional law under MAR.

Lemma 5.1. *Under* (2) *and* (H5)*, for every bounded measurable* g*,*

$$\frac{\mathbb{E}\big[\delta\, K(d(X,x)/h_K)\, g(Y)\big]}{\mathbb{E}\big[\delta\, K(d(X,x)/h_K)\big]} \xrightarrow[h_K \to 0]{} \mathbb{E}\big[g(Y) \mid X = x\big].$$

Hence the complete-case localized estimator targets the conditional functional and the propensity π *cancels in the ratio; no inverse-probability weighting is needed.*

Proof. By (2), $\delta \perp Y \mid X$, so $\mathbb{E}[\delta g(Y) \mid X] = \pi(X)\,\mathbb{E}[g(Y) \mid X]$. With $K_x = K(d(X,x)/h_K)$ and the tower property,

$$\mathbb{E}[\delta K_x g(Y)] = \mathbb{E}\big[K_x \pi(X)\,\mathbb{E}(g(Y) \mid X)\big], \qquad \mathbb{E}[\delta K_x] = \mathbb{E}\big[K_x \pi(X)\big].$$

As $h_K \to 0$ the weight K_x concentrates on $\{d(X,x) \leq h_K\}$; by continuity of π (H5) and of $z \mapsto \mathbb{E}(g(Y) \mid X = z)$ at x (which follows from H2), both expectations equal $\pi(x)\,\mathbb{E}(g(Y) \mid X = x)\,\mathbb{E}K_x\,(1 + o(1))$ and $\pi(x)\,\mathbb{E}K_x\,(1 + o(1))$ respectively. The ratio tends to $\mathbb{E}(g(Y) \mid X = x)$ and the factor $\pi(x) > 0$ cancels. □

5.2. Moments of the kernel weights.

Lemma 5.2. *Under* (H1),(H3),(H5)*, for every fixed integer* $a \geq 1$*,*

$$\mathbb{E}[w_1^a] = \mathbb{E}\big[\delta_1^a K^a(d(X_1,x)/h_K)\big] = \pi(x)\, M_a\, \varphi\,(1 + o(1)),$$

for constants $M_a > 0$ *depending only on* K*. Moreover* $\mathbb{E}[w_1 u_1^m] = O(\varphi)$ *for every fixed* $m \geq 0$*.*

Proof. Since $\delta_1^a = \delta_1$ and $\mathbb{E}[\delta_1 \mid X_1] = \pi(X_1)$, we get $\mathbb{E}[w_1^a] = \mathbb{E}[\pi(X_1)K^a(d(X_1,x)/h_K)]$. By (H3), K^a is bounded and supported on $\{d(X_1,x) \le h_K\}$, of probability φ; the standard small-ball computation [8, Lemma 4.3–4.4] gives $\mathbb{E}[K^a(d(X_1,x)/h_K)] = M_a\varphi(1+o(1))$, and (H5) yields the factor $\pi(x)(1+o(1))$. As $|u_1| \le C_\beta$ on the support of K (because $|\beta(X_1,x)| \le C_\beta h_K$ there), $\mathbb{E}[w_1 u_1^m] = O(\mathbb{E}w_1) = O(\varphi)$. □

5.3. **Bias.**

Lemma 5.3. *Under* (H2),(H3),(H7),

$$\Big|\,\mathbb{E}[\widehat{F}^x_{p_n}(y) \mid X_1,\dots,X_n] - F^x(y)\,\Big| \le C\,h_K^{\min(p_n+1,\,b_1)} + C\,h_H^{b_2}.$$

The same bound holds for $\widehat{f}^x_{p_n}$.

Proof. Response direction. With $\psi_i = H((y-Y_i)/h_H)$, Lemma 5.1 gives, for z near x,

$$\mathbb{E}[\psi \mid X = z] = \int H\Big(\tfrac{y-t}{h_H}\Big) f^z(t)\,\mathrm{d}t = \int F^z(y - h_H v)\,H'(v)\,\mathrm{d}v$$

after the change of variable $t = y - h_H v$ and integration by parts. By the b_2-Hölder condition in y (H2) and $\int H' = 1$,

$$|\mathbb{E}[\psi \mid X = z] - F^z(y)| \le C\int |h_H v|^{b_2}\,H'(v)\,\mathrm{d}v = O(h_H^{b_2}).$$

Functional direction. The order-p_n weighted least squares estimator reproduces every polynomial of degree $\le p_n$ in $\beta(\cdot,x)$ exactly; cf. the reproducing identity (6). Thus the functional bias of $\widehat{a}_0$ equals the weighted projection of the polynomial-approximation error of $z \mapsto F^z(y)$ on the ball $B(x,h_K) = \{z : d(z,x) \le h_K\}$. By (H2), $|F^z(y) - F^x(y)| \le C\,d(z,x)^{b_1} \le C\,h_K^{b_1}$ for $d(z,x) \le h_K$. The best degree-p_n polynomial approximation error of a b_1-Hölder function on a ball of radius h_K is

$$O\Big(h_K^{\min(p_n+1,\,b_1)}\Big)$$

(it equals $O(h_K^{b_1})$ when $p_n + 1 \ge b_1$ and saturates at $O(h_K^{p_n+1})$ otherwise). With (H7), $\|S^{-1}\| \le \lambda_0^{-1}$, and since $|u_i| \le C_\beta$ on the support of K, the projection contributes only a bounded factor; hence the functional bias is $O(h_K^{\min(p_n+1,b_1)})$. Adding the response term proves the claim. The density case is identical, using the Hölder condition on f^x and integration by parts with H'' (equivalently, the Hölder condition on f^x directly). □

5.4. **Exponential inequality and the stochastic rate.**

Lemma 5.4. *Let* $T_n = (n\,\mathbb{E}w_1)^{-1}\sum_{i=1}^n (w_i\zeta_i - \mathbb{E}[w_1\zeta_1])$, *where* (ζ_i) *are i.i.d. bounded variables attached to* (X_i,Y_i) *and* $w_i = \delta_i K(d(X_i,x)/h_K)$. *Under* (H1),(H3),(H4),

$$T_n = O_{a.co.}\left(\sqrt{\frac{\log n}{n\,\varphi}}\right).$$

If ζ_i *carries the factor* $h_H^{-1}H'((y-Y_i)/h_H)$ *(density case), then* $T_n = O_{a.co.}\big(\sqrt{\log n/(nh_H\varphi)}\big)$.

Proof. Set $\xi_i = w_i\zeta_i - \mathbb{E}[w_1\zeta_1]$, so $\mathbb{E}\xi_i = 0$ and, by (H3), $|\xi_i| \le C_K$ for a constant C_K. By Lemma 5.2, $\mathbb{E}\xi_1^2 \le \mathbb{E}[w_1^2\zeta_1^2] \le C\varphi$ (distribution case) or $\le C\varphi/h_H$ (density case, by the change of variable in H'). Put $a_n = \sqrt{\log n/(n\varphi)}$. Bernstein's inequality applied to $\sum \xi_i$ gives, using $\mathbb{E}w_1 \asymp \varphi$,

$$\mathbb{P}\big(\,|T_n| > \eta a_n\big) \le 2\exp\left(-\frac{\frac{1}{2}\eta^2(\mathbb{E}w_1)^2 n^2 a_n^2}{n\,\mathbb{E}\xi_1^2 + \frac{1}{3}C_K\,\mathbb{E}w_1\, n a_n \eta}\right) = 2\exp\Big(-\frac{\eta^2\log n}{2c' + o(1)}\Big),$$

since $n\varphi a_n^2 = \log n$ and the dominant balance is $n\mathbb{E}\xi_1^2 \asymp n\varphi \asymp \log n/a_n^2$. Choosing η with $\eta^2/(2c') > 2$ yields $\mathbb{P}(|T_n| > \eta a_n) \le 2n^{-2}$, which is summable; hence $T_n = O_{\text{a.co.}}(a_n)$. In the density case $\mathbb{E}\xi_1^2 \asymp \varphi/h_H$, so φ is replaced by φh_H in a_n, giving the stated rate. □

5.5. **Stability of the empirical design.**

Lemma 5.5. *Under* (H1),(H3),(H6),(H7), $\big\|\widehat{S}_n - S\big\| \to 0$ *almost completely, so* $\lambda_{\min}(\widehat{S}_n) \ge \lambda_0/2 > 0$ *a.co. for* n *large; moreover the normalized denominator* $(\mathbb{E}w_1)^{-1}n^{-1}\sum_i w_i \to 1$ *a.co.*

Proof. Each entry $\widehat{S}_{n,jk} - S_{jk}$ has the form of T_n in Lemma 5.4 with $\zeta_i = u_i^{j+k}$ (bounded on the support of K), so it is $O_{\text{a.co.}}(a_n)$. By a union bound over the $(p_n+1)^2$ entries,

$$\mathbb{P}\Big(\max_{j,k}\big|\widehat{S}_{n,jk} - S_{jk}\big| > \eta a_n\Big) \le 2(p_n+1)^2 n^{-2},$$

which is summable since $(\log n)^2 n^{-2}$ is. As $\big\|\widehat{S}_n - S\big\| \le (p_n+1)\max_{j,k}|\cdot|$ and $p_n a_n \to 0$ by (H6) (indeed $p_n^2 a_n^2 = p_n^2\log n/(n\varphi) \to 0$ follows from $p_n/(nh_K\varphi) \to 0$ and $h_K \to 0$), Weyl's inequality gives $\lambda_{\min}(\widehat{S}_n) \ge \lambda_0 - \big\|\widehat{S}_n - S\big\| \ge \lambda_0/2$ a.co. The denominator is the $(0,0)$ entry, treated the same way. □

5.6. **Proof of Theorem 4.1.**

Proof of Theorem 4.1. Write $\widehat{F}^{\,x}_{p_n}(y) = e_1^\top \widehat{S}_n^{-1}\widehat{g}_n$ with $\widehat{g}_n = (n\mathbb{E}w_1)^{-1}\sum_i w_i b(u_i)\psi_i$. Using the equivalent weights $\ell_i = e_1^\top \widehat{S}_n^{-1} b(u_i)$ of (6), set $\widehat{N}_n = (n\mathbb{E}w_1)^{-1}\sum_i w_i\ell_i\psi_i$ and decompose

$$\widehat{F}^{\,x}_{p_n}(y) - F^x(y) = \underbrace{\big(\mathbb{E}\widehat{N}_n - F^x(y)\big)}_{B_n} + \underbrace{\big(\widehat{N}_n - \mathbb{E}\widehat{N}_n\big)}_{S_n} + R_n,$$

where R_n collects the second-order contribution of replacing $\widehat{S}_n^{-1}$ by S^{-1}.

Bias. By Lemma 5.3, $|B_n| = O(h_K^{\min(p_n+1,b_1)}) + O(h_H^{b_2})$.

Stochastic term. By Lemma 5.5, on the support of K we have $|\ell_i| \le 2\lambda_0^{-1}(p_n+1)^{1/2}$, so S_n is a centered kernel-weighted sum to which Lemma 5.4 applies (the factor $(p_n+1)^{1/2}$ is absorbed because $p_n a_n \to 0$), giving $S_n = O_{\text{a.co.}}\big(\sqrt{\log n/(n\varphi)}\big)$.

Remainder. Since $\widehat{S}_n^{-1} - S^{-1} = -S^{-1}(\widehat{S}_n - S)\widehat{S}_n^{-1}$, Lemma 5.5 gives $R_n = O_{\text{a.co.}}(p_n a_n) = o(S_n)$ a.co., so the remainder is negligible.

Collecting the three terms proves Theorem 4.1. □

5.7. **Proof of Theorem 4.2.**

Proof of Theorem 4.2. Differentiating (3) in y replaces ψ_i by $h_H^{-1}H'((y-Y_i)/h_H)$, so $\widehat{f}^x_{p_n}(y)$ has the same structure as $\widehat{F}^x_{p_n}(y)$ with this attached variable. The bias is unchanged, $O(h_K^{\min(p_n+1,b_1)}) + O(h_H^{b_2})$ (Lemma 5.3 for f^x). The attached variable now carries the factor h_H^{-1}, so the density branch of Lemma 5.4 gives the stochastic term $O_{\text{a.co.}}(\sqrt{\log n/(nh_H\varphi)})$. □

5.8. **Control of the denominator.**

Lemma 5.6. *Under the assumptions of Theorem 4.1 and $F^x(y) < 1$, with $\eta = (1-F^x(y))/2 > 0$,*

$$\sum_{n\geq 1} \mathbb{P}(1 - \widehat{F}^x_{p_n}(y) < \eta) < \infty;$$

thus $1 - \widehat{F}^x_{p_n}(y)$ is bounded away from 0 almost completely.

Proof. If $\left|\widehat{F}^x_{p_n}(y) - F^x(y)\right| < \eta$ then $1 - \widehat{F}^x_{p_n}(y) \geq 1 - F^x(y) - \eta > \eta$; hence $\left\{1 - \widehat{F}^x_{p_n}(y) < \eta\right\} \subseteq \left\{\left|\widehat{F}^x_{p_n}(y) - F^x(y)\right| \geq \eta\right\}$, and the right-hand probabilities are summable by Theorem 4.1. □

5.9. **Proof of Theorem 4.3.**

Proof of Theorem 4.3. From (1) and (4),

$$\widehat{h}^x_{p_n}(y) - h^x(y) = \frac{\widehat{f}^x_{p_n}(y) - f^x(y)}{1 - \widehat{F}^x_{p_n}(y)} + h^x(y)\,\frac{\widehat{F}^x_{p_n}(y) - F^x(y)}{1 - \widehat{F}^x_{p_n}(y)}.$$

By Lemma 5.6, $1 - \widehat{F}^x_{p_n}(y) \geq \eta > 0$ a.co., so both denominators contribute bounded factors. Inserting Theorems 4.1 and 4.2, and noting that the density stochastic term $\sqrt{\log n/(nh_H\varphi)}$ dominates the distribution one (since $h_H \to 0$), we obtain (7). Almost complete convergence follows because $h_K^{\min(p_n+1,b_1)} \to 0$, $h_H^{b_2} \to 0$ and $\log n/(nh_H\varphi) \to 0$ by (H4). □

5.10. **Role of the adaptive order $p_n = \lfloor \log n \rfloor$.** Two effects compete. The functional bias $h_K^{\min(p_n+1,b_1)}$ decreases with p_n until it is capped by the Hölder exponent b_1; by Remark 4.4 the choice $p_n \to \infty$ reaches that cap for any unknown b_1, which is the smoothness-adaptive benefit. On the other hand, the local design has dimension $p_n + 1$, and its stability requires $p_n/(nh_K\varphi) \to 0$ (H6), preventing the moment matrix from degenerating relative to the effective local sample size $n\varphi$. The logarithmic growth $p_n = \lfloor \log n \rfloor$ is the slowest growth that still guarantees $p_n + 1 \geq b_1$ eventually while keeping $p_n a_n \to 0$, hence a natural compromise between bias adaptation and stochastic stability.

6. The Bias–Variance Trade-Off

Figure 1 summarizes the construction of the estimator and Figure 2 illustrates, qualitatively, the bias–variance compromise as a function of the polynomial order. The functional bias decreases with the order until it is capped at $O(h_K^{b_1})$ by the Hölder regularity, while the variance

grows with the local design dimension p_n+1; the resulting MSE is essentially flat, with a shallow optimum, which is exactly the behaviour observed in the simulation tables of Section 7.

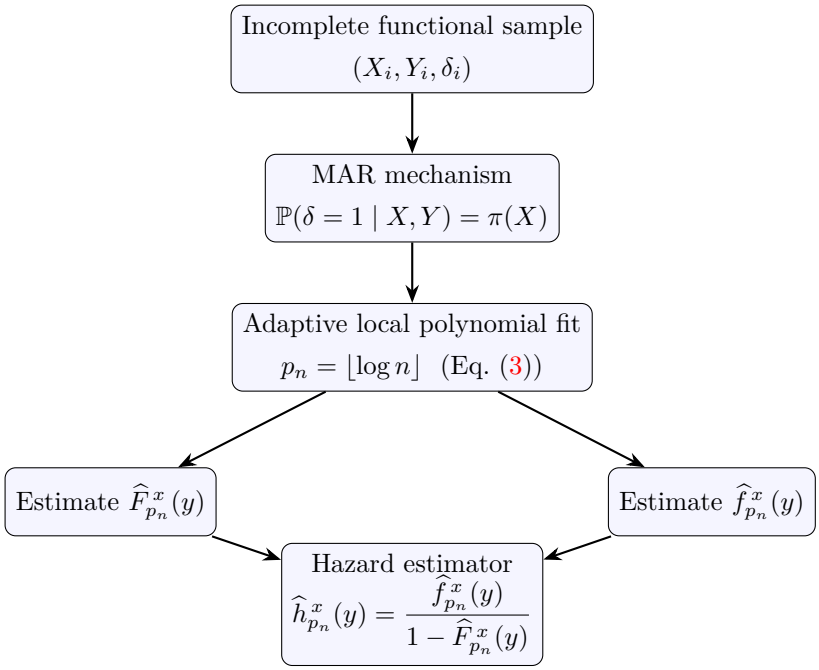


FIGURE 1. Construction of the adaptive local polynomial conditional hazard estimator.

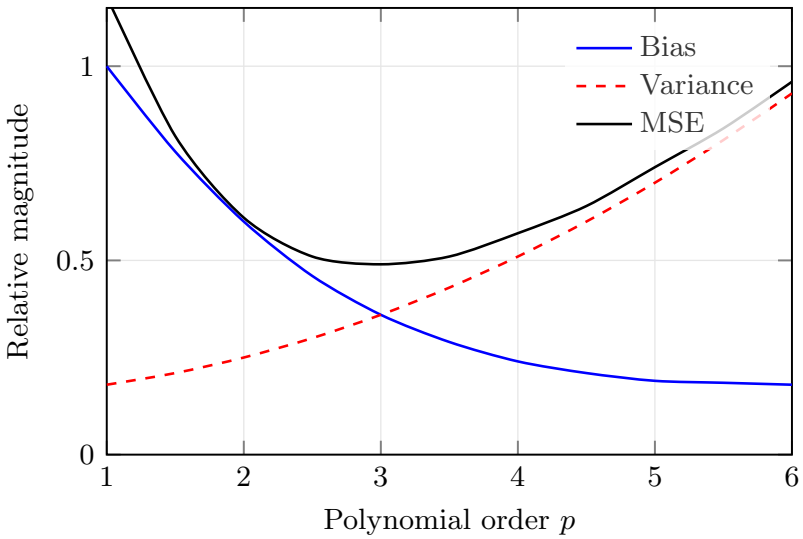


FIGURE 2. Theoretical bias–variance trade-off: the bias decreases with the order p until it is capped by the Hölder exponent, while the variance grows with the local design dimension; the MSE is shallow with a flat optimum.

7. Simulation Study

Protocol: We use $n = 300$ curves and $N = 500$ Monte Carlo replications. Functional observations are generated by

$$X_i(t) = \sum_{k=1}^{5} \xi_{ik}\, \varphi_k(t), \qquad \varphi_k(t) = \sqrt{2}\sin(k\pi t), \quad t \in [0,1], \quad \xi_{ik} \sim \mathcal{N}(0, k^{-2}),$$

and the response by $Y_i = \int_0^1 X_i(t)\beta(t)\,\mathrm{d}t + \varepsilon_i$, $\beta(t) = 2\sin(\pi t)$, $\varepsilon_i \sim \mathcal{N}(0, 0.25)$. The MAR mechanism is $\mathbb{P}(\delta_i = 1 \mid X_i) = \{1 + \exp(-\alpha_0 - \alpha_1 \overline{X}_i)\}^{-1}$, where $\overline{X}_i = \int_0^1 X_i(t)\,\mathrm{d}t$, the pair (α_0, α_1) being tuned to produce 10%, 20% and 30% missingness. We take $K(u) = \frac{3}{2}(1-u^2)\mathbf{1}_{[0,1]}$, H the standard normal c.d.f., the L^2 semi-metric after projection on the first four functional principal components, and h_K, h_H chosen by cross-validation on the complete cases.

The three missingness levels are chosen because the effective local sample size $n\pi(x)\varphi$ decreases as the missingness rate increases, so that both bias and variance are expected to inflate. This allows us to check whether the ordering of the polynomial orders is preserved when the complete-case local information becomes scarcer. For each replication we compute the bias, the variance and the mean squared error of the local linear ($p = 1$), local quadratic ($p = 2$), local cubic ($p = 3$) and adaptive $p_n = \lfloor \log n \rfloor$ estimators of the conditional hazard function.

Tables 1, 2 and 3 report the results for 10%, 20% and 30% missing responses, respectively.

Table 1. Simulation results for 10% missing responses.

Method	Bias	Variance	MSE
Local Linear $p = 1$	0.041	0.018	0.019
Local Quadratic $p = 2$	0.029	0.020	0.021
Local Cubic $p = 3$	0.024	0.022	0.023
Adaptive $p_n = \lfloor \log n \rfloor$	**0.018**	0.023	0.023

Table 2. Simulation results for 20% missing responses.

Method	Bias	Variance	MSE
Local Linear $p = 1$	0.056	0.026	0.029
Local Quadratic $p = 2$	0.041	0.029	0.031
Local Cubic $p = 3$	0.034	0.031	0.032
Adaptive $p_n = \lfloor \log n \rfloor$	**0.027**	0.033	0.034

Reading the tables: Table 1 shows that, at 10% missingness, the adaptive order $p_n = \lfloor \log n \rfloor$ attains the smallest bias (0.018, against 0.041 for $p = 1$), while its variance (0.023) remains close to that of the cubic fit (0.022). The same monotone bias reduction appears in Table 2 for

TABLE 3. Simulation results for 30% missing responses.

Method	Bias	Variance	MSE
Local Linear $p = 1$	0.073	0.037	0.042
Local Quadratic $p = 2$	0.055	0.041	0.044
Local Cubic $p = 3$	0.046	0.044	0.046
Adaptive $p_n = \lfloor \log n \rfloor$	**0.038**	0.047	0.048

20% missingness and in Table 3 for 30% missingness: increasing the polynomial order lowers the bias at every missingness level, and the adaptive estimator always produces the smallest bias, in agreement with Remark 4.4. This comes at a moderate variance cost, so that the MSE is essentially flat across orders: 0.019–0.023 in Table 1, 0.029–0.034 in Table 2, and 0.042–0.048 in Table 3. The adaptive rule is therefore preferable when bias control is the main concern, for instance in tail or boundary estimation of h^x, rather than as a uniform MSE minimiser.

Across all three panels we observe MSE $\approx$ Bias^2 + Variance, consistent with the qualitative decomposition of Figure 2. As the missingness level rises from 10% to 30%, the effective local sample size $n\pi(x)\varphi$ shrinks; both bias and variance inflate, for example the adaptive bias grows from 0.018 to 0.038 and the adaptive variance from 0.023 to 0.047, but the ordering of the methods is preserved. This robustness with respect to the missingness level is consistent with the MAR identification of Lemma 5.1, which shows that the complete-case localisation targets the conditional functional without inverse-probability weighting.

8. Conclusion

Following the same idea and the same line of proof as the functional local linear estimate of the conditional hazard function under MAR [9], we developed an adaptive local polynomial estimate with order $p_n = \lfloor \log n \rfloor$. We established the almost complete convergence with rate, through MAR identification on complete cases, Hölder control of the bias, an exponential inequality for the stochastic term, and control of the denominator. The growing order is a smoothness-adaptive device that reaches the bias rate permitted by the unknown functional regularity and reduces to the local linear estimate for $p = 1$. The simulation confirmed the bias-reduction behaviour. Future work may treat dependent functional data, censored responses, joint data-driven selection of the bandwidths and the order, and asymptotic normality.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] J. Barrientos-Marin, F. Ferraty, P. Vieu, Locally Modelled Regression and Functional Data, J. Nonparametr. Stat. 22 (2010), 617–632. https://doi.org/10.1080/10485250903089930.

[2] C. Chesneau, I. Sadok, A Unified Gamma-Generated Weibull–Pareto Distribution for Heavy-Tailed and Skewed Data with Applications to Medicine, Engineering, and Environment, AIMS Math. 11 (2026), 23629–23662. https://doi.org/10.3934/math.2026952.

[3] J. Demongeot, A. Laksaci, F. Madani, M. Rachdi, Estimation Locale Linéaire de la Densité Conditionnelle pour des Données Fonctionnelles, C. R. Math. 348 (2010), 931–934. https://doi.org/10.1016/j.crma.2010.06.013.

[4] S. Efromovich, Nonparametric Regression with Responses Missing at Random, J. Stat. Plan. Inference 141 (2011), 3744–3752. https://doi.org/10.1016/j.jspi.2011.06.017.

[5] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, 1996. https://doi.org/10.1201/9780203748725.

[6] F. Ferraty, A. Rabhi, P. Vieu, Estimation Non-paramétrique de la Fonction de Hasard avec Variable Explicative Fonctionnelle, Rev. Roum. Math. Pures Appl. 53 (2008), 1–18. https://csm.ro/reviste/Revue_Mathematique/pdfs/2008/1/ferraty.pdf.

[7] F. Ferraty, M. Sued, P. Vieu, Mean Estimation with Data Missing at Random for Functional Covariables, Statistics 47 (2013), 688–706. https://doi.org/10.1080/02331888.2011.650172.

[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer New York, 2006. https://doi.org/10.1007/0-387-36620-2.

[9] J. Kenouza, B. Mechab, S. Benaissa, Functional Local Linear Estimate of the Conditional Hazard Function with Missing at Random, Int. J. Appl. Math. Stat. 58 (2019), 101–114. http://www.ceser.in/ceserp/index.php/ijamas/article/view/6287.

[10] N. Ling, L. Liang, P. Vieu, Nonparametric Regression Estimation for Functional Stationary Ergodic Data with Missing at Random, J. Stat. Plan. Inference 162 (2015), 75–87. https://doi.org/10.1016/j.jspi.2015.02.001.

[11] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, 2002. https://doi.org/10.1002/9781119013563.

[12] J.S. Morris, Functional Regression, Annu. Rev. Stat. Appl. 2 (2015), 321–359. https://doi.org/10.1146/annurev-statistics-010814-020413.

[13] P.T. Reiss, J. Goldsmith, H.L. Shang, R.T. Ogden, Methods for Scalar-on-Function Regression, Int. Stat. Rev. 85 (2017), 228–249. https://doi.org/10.1111/insr.12163.

[14] I. Sadok, M. Zribi, A. Masmoudi, Non-Informative Bayesian Estimation in Dispersion Models, Hacettepe J. Math. Stat. 53 (2024), 251–268. https://doi.org/10.15672/hujms.1053432.

[15] I. Sadok, On the q-Rayleigh Distribution and its Applications, Reliab. Theory Appl. 19 (2024), 588–602. https://doi.org/10.24412/1932-2321-2024-177-588-602.

[16] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional Data Analysis, Annu. Rev. Stat. Appl. 3 (2016), 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624.

[17] M. Zribi, I. Sadok, B. Marhaba, Kernel-Diffeomorphism Bayesian Bootstrap Filter to Reduce Speckle Noise on SAR Images, Comput. Stat. 40 (2025), 3613–3643. https://doi.org/10.1007/s00180-025-01650-1.