Biometrika (2021), **1**, 1, *pp*. 1–41
Printed in Great Britain Advance Access publication on 19 March 2021

Scalar-on-function local linear regression and beyond Online supplement

BY F. FERRATY
University of Toulouse, Toulouse Mathematics Institute, 118 route de Narbonne, 31062 Toulouse cedex, France
ferraty@math.univ-toulouse.fr

AND S. NAGY
Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Prague, Czech Republic
nagy@karlin.mff.cuni.cz

SUMMARY

This supplementary file extends the materials from the main paper in two directions. The first part is oriented towards practitioners; it describes additional features of our estimating procedure, the companion `R` package and its implementation for reproducing analyses. The second part details the hypotheses and the proofs of theorems with complementary theoretical developments regarding examples of approximating bases and the original condition (H3).

Some key words: Asymptotics; Functional data; Functional derivative of regression operator; Functional index model; Local linear regression; Scalar-on-function regression.

S.1. PRACTICAL ASPECTS

S.1·1. *Regression operator: complement to prediction quality*

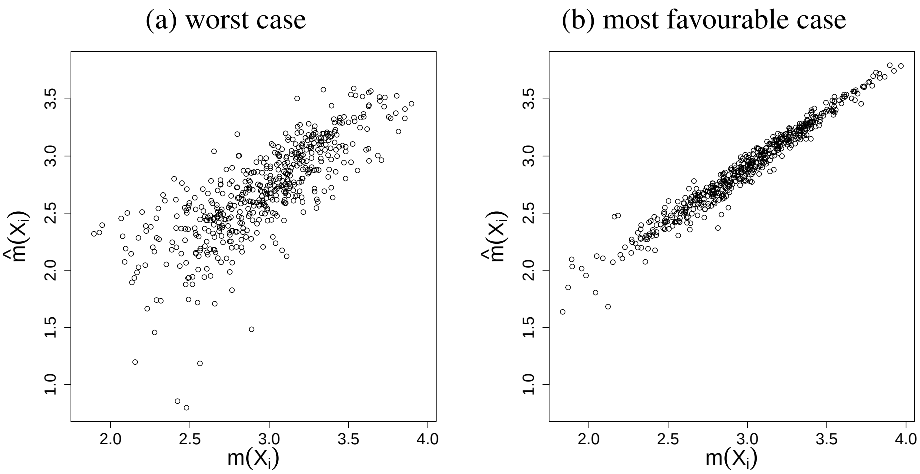

Fig. S.1: (a) $n = 100$ and $nsr = 0.4$, (b) $n = 500$ and $nsr = 0.05$.

Figure S.1 displays the true values $m(X_i)$ against their predictions $\hat{m}(X_i)$ computed from model (M1), considered in Section 4·1 of the main paper, when cross-validation is used. Even in the worst case (small learning sample size and high noise-to-signal ratio), the local linear estimator provides reliable results.

S.1·2. *Bandwidth selection for the functional derivative*

The bootstrap procedure introduces additional randomness into our local linear estimation method. A natural question arises: is the proposed bandwidth selection stable? In other words, if the functional derivatives are estimated several times on the same dataset, are the resulting bandwidths similar? A related issue concerns the number B of bootstrap repetitions set by the user: is the procedure sensitive to this parameter, and how to choose it? To address both these concerns, one dataset is simulated according to (M1). For different B and learning sample sizes n, our estimating algorithm is launched 100 times.

Table S.1 displays the mean and standard deviation (in brackets) of h_{deriv} and $ORMSEP_{deriv}$ in (a) the worst case, and (b) the most favourable case in the simulation study from the main document. The variability of the selected bandwidth is smaller in the most favourable case as expected. The bootstrap bandwidth selection remains stable in both cases despite the additional randomness introduced. In both situations, the variability of $ORMSEP_{deriv}$ is close to 0. The default value ($B = 100$) used in our procedure seems to be large enough to ensure stability as well as accuracy for predictions.

Table S.1: Stability of the bootstrap bandwidth selection.

(a) $n = 100$ and $nsr = 0.4$

B	h_{deriv}	$ORMSEP_{deriv}$
50	36.440 (8.197)	0.231 (0.060)
100	36.600 (7.865)	0.231 (0.058)
500	36.520 (8.093)	0.232 (0.059)
1000	36.520 (7.912)	0.231 (0.059)

(b) $n = 500$ and $nsr = 0.05$

B	h_{deriv}	$ORMSEP_{deriv}$
50	31.590 (7.585)	0.046 (0.004)
100	30.910 (7.190)	0.046 (0.004)
500	30.740 (7.078)	0.046 (0.004)
1000	30.910 (7.190)	0.046 (0.004)

Figure S.2 reproduces the same plot as Figure 2(a) in the main document but $\hat{m}'_{X_i}$ is built with h_{reg}, the bandwidth used for estimating the regression operator with cross-validation, instead of h_{deriv} being the specific bandwidth computed for estimating the functional derivative. In that situation, it is the AIC$_C$ method used for selecting h_{reg} which provides better predictions. Nevertheless, h_{reg} fails drastically, especially for larger sample sizes n; consistency profile is far less obvious than the one displayed in Figure 2(a). The comparison of Figure S.2 with Figure 2(a) confirms the importance of selecting a specific bandwidth for the estimation of functional derivatives.

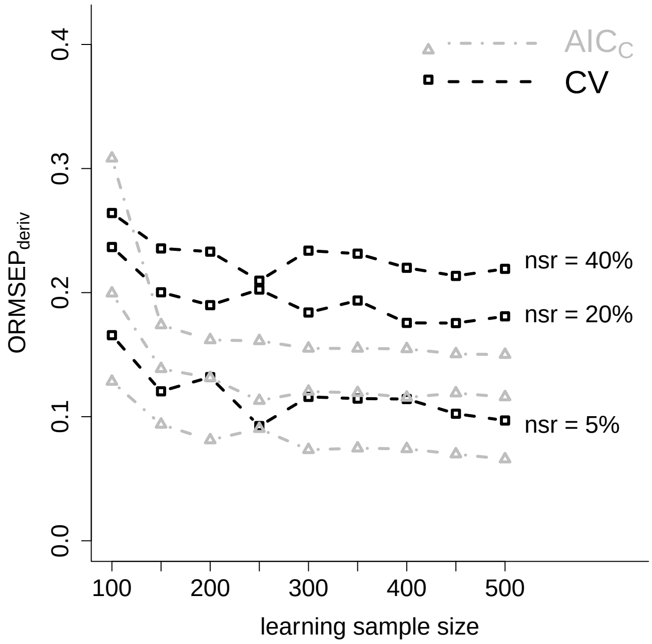

Fig. S.2: Means of $ORMSEP_{deriv}$ over 100 runs in squares and triangles, depending on different noise-to-signal ratios nsr, learning sample sizes, and bandwidth selection methods CV/AIC$_C$ used for computing $\hat{m}$ when $\hat{m}'_{X_i}$ is built with h_{reg}. Smaller values of nsr correspond to lower $ORMSEP_{deriv}$.

The summary statistics of the raw data used for Figure 2(b) in the main paper are provided in Table S.2 which gives the mean and the standard deviation (in brackets) of the computed bandwidths over 100 runs in 27 situations (9 learning sample sizes × 3 noise-to-signal ratios).

Table S.2: Estimated bandwidths.

		h_{reg}	h_{deriv}	h_{deriv}^{oracle}
100	0.05	11.960 (2.309)	18.600 (6.391)	18.060 (2.228)
	0.2	15.200 (3.140)	28.240 (6.352)	22.830 (2.719)
	0.4	18.680 (5.601)	35.960 (8.706)	26.640 (3.606)
150	0.05	12.800 (1.214)	19.650 (5.058)	21.600 (2.211)
	0.2	14.450 (2.492)	30.300 (5.579)	28.430 (2.836)
	0.4	18.550 (4.076)	39.750 (6.867)	33.870 (4.554)
200	0.05	11.350 (1.533)	20.660 (4.749)	24.800 (2.274)
	0.2	16.600 (2.814)	36.340 (5.520)	33.540 (3.480)
	0.4	19.260 (3.047)	43.760 (5.876)	40.080 (3.959)
250	0.05	14.000 (0.000)	24.440 (3.554)	27.260 (2.158)
	0.2	15.530 (3.398)	37.850 (6.051)	37.580 (3.294)
	0.4	20.750 (3.917)	47.930 (5.980)	45.820 (4.423)
300	0.05	12.180 (1.114)	24.150 (5.141)	30.330 (2.857)
	0.2	16.440 (2.066)	40.890 (5.438)	42.540 (3.653)
	0.4	19.740 (2.802)	48.750 (4.961)	51.060 (4.451)
350	0.05	12.580 (1.350)	26.260 (5.346)	33.040 (2.445)
	0.2	16.840 (1.958)	43.150 (4.736)	45.970 (3.802)
	0.4	21.040 (3.101)	53.620 (5.631)	56.020 (4.791)
400	0.05	12.400 (1.206)	29.240 (5.384)	35.680 (2.391)
	0.2	17.760 (2.151)	47.040 (4.727)	50.240 (4.112)
	0.4	21.760 (2.686)	55.800 (5.191)	60.120 (5.004)
450	0.05	13.080 (0.563)	30.880 (4.457)	37.440 (2.660)
	0.2	17.600 (2.000)	48.400 (4.769)	54.120 (4.295)
	0.4	22.440 (3.141)	59.760 (5.677)	64.960 (5.195)
500	0.05	13.760 (0.955)	33.840 (5.083)	40.440 (2.471)
	0.2	18.360 (1.806)	51.360 (3.907)	57.960 (4.000)
	0.4	23.200 (3.191)	62.400 (5.527)	69.600 (4.841)

S.1·3. *Robustness to the model complexity*

We now provide several tables of results to assess the robustness of our estimating procedure with respect to J, the complexity of the simulated model, see Section 4·1 in the main paper. To make the choice of the dimension J more challenging, we add structural perturbation to the functional predictors X. Given the first eight Fourier basis elements $\phi_1, \ldots, \phi_8$, set $X := \sum_{j=1}^{4} U_j \, \phi_j + \eta$ where $\eta := \sum_{j=5}^{8} V_j \, \phi_j$ with U_j and V_j iid uniform random variables defined on $[-1, 1]$ and $[-b, b]$, respectively. The second part η provides a structural noise that is controlled by the ratio $\rho := \mathrm{E}\left(\|\eta\|^2\right) / \mathrm{E}\left(\|X\|^2\right) = b^2/(1+b^2)$. Given any $\rho \in (0,1)$, one can always find a corresponding bound b for simulating the functional predictors. Tables S.3–S.14 display, for J respectively set to 2, 3 and 4, i) the number of times, out of 100, that the dimension is correctly selected, and ii) the corresponding $ORMSEP_{reg}$ and $ORMSEP_{deriv}$ averaged over 100 runs with standard deviation in brackets. For instance, Table S.7 shows how our estimating procedure is robust according to 4 structural noise ratios $\rho = 0.05,\ 0.1,\ 0.2,\ 0.4$ and different learning sample sizes $n = 100, 150, \ldots, 500$ for J set to 4 and nsr set to 0.05. Let us now focus on the coresponding Table S.8. It gives respectively the corresponding $ORMSEP_{reg}$ and $ORMSEP_{deriv}$, each time averaged over 100 runs with standard deviations in brackets. This additional table confirms that our estimating procedure is not too sensitive to perturbations. As usual, the robustness of the selecting procedure degrades with the complexity of the simulated model, that is when J increases.

Table S.3: Number of times, out of 100, that the dimension is correctly selected with $J = 2$, $nsr = 0.05$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	98	100	100	100
150	100	100	100	100
200	99	100	100	100
250	100	100	100	100
300	100	100	100	100
350	100	100	100	100
400	100	100	100	99
450	100	100	100	100
500	100	100	100	99

Table S.4: Average and standard deviation (in brackets) of $ORMSEP$ with $J = 2$, $nsr = 0.05$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.029 (0.009)	0.039 (0.010)	0.057 (0.011)	0.120 (0.019)
	150	0.022 (0.006)	0.029 (0.006)	0.043 (0.007)	0.083 (0.017)
	200	0.017 (0.004)	0.023 (0.004)	0.034 (0.006)	0.068 (0.011)
	250	0.015 (0.004)	0.020 (0.004)	0.029 (0.005)	0.060 (0.009)
	300	0.013 (0.003)	0.017 (0.004)	0.025 (0.004)	0.049 (0.007)
	350	0.011 (0.002)	0.015 (0.003)	0.023 (0.004)	0.042 (0.005)
	400	0.010 (0.002)	0.015 (0.002)	0.021 (0.003)	0.038 (0.005)
	450	0.009 (0.002)	0.013 (0.002)	0.019 (0.003)	0.035 (0.005)
	500	0.009 (0.002)	0.012 (0.002)	0.018 (0.002)	0.033 (0.004)
$ORMSEP_{deriv}$	100	0.032 (0.012)	0.037 (0.010)	0.050 (0.009)	0.110 (0.025)
	150	0.024 (0.004)	0.028 (0.005)	0.038 (0.006)	0.080 (0.016)
	200	0.021 (0.005)	0.023 (0.004)	0.031 (0.004)	0.067 (0.015)
	250	0.018 (0.004)	0.020 (0.003)	0.026 (0.003)	0.058 (0.011)
	300	0.016 (0.003)	0.018 (0.003)	0.024 (0.003)	0.053 (0.013)
	350	0.015 (0.002)	0.017 (0.002)	0.022 (0.003)	0.048 (0.012)
	400	0.014 (0.003)	0.016 (0.002)	0.021 (0.003)	0.046 (0.011)
	450	0.013 (0.002)	0.015 (0.002)	0.020 (0.002)	0.042 (0.011)
	500	0.013 (0.002)	0.014 (0.002)	0.019 (0.002)	0.039 (0.008)

Table S.5: Number of times, out of 100, that the dimension is correctly selected with $J = 3$, $nsr = 0.05$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	99	97	91	79
150	100	100	99	86
200	100	100	99	93
250	100	100	100	100
300	100	100	100	100
350	100	100	100	99
400	100	100	100	100
450	100	100	100	100
500	100	100	100	100

Table S.6: Average and standard deviation (in brackets) of $ORMSEP$ with $J = 3$, $nsr = 0.05$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.130 (0.022)	0.147 (0.026)	0.205 (0.038)	0.380 (0.052)
	150	0.089 (0.014)	0.103 (0.014)	0.150 (0.022)	0.296 (0.042)
	200	0.063 (0.012)	0.081 (0.014)	0.121 (0.019)	0.236 (0.031)
	250	0.055 (0.009)	0.068 (0.010)	0.109 (0.014)	0.215 (0.023)
	300	0.046 (0.008)	0.059 (0.007)	0.093 (0.011)	0.182 (0.018)
	350	0.039 (0.005)	0.053 (0.006)	0.085 (0.010)	0.166 (0.016)
	400	0.036 (0.005)	0.048 (0.006)	0.077 (0.010)	0.154 (0.018)
	450	0.033 (0.005)	0.044 (0.006)	0.071 (0.008)	0.145 (0.016)
	500	0.030 (0.004)	0.041 (0.005)	0.067 (0.007)	0.137 (0.016)
$ORMSEP_{deriv}$	100	0.077 (0.016)	0.097 (0.059)	0.140 (0.086)	0.312 (0.114)
	150	0.054 (0.007)	0.061 (0.008)	0.090 (0.036)	0.237 (0.102)
	200	0.044 (0.005)	0.051 (0.005)	0.073 (0.033)	0.174 (0.074)
	250	0.038 (0.004)	0.044 (0.004)	0.059 (0.006)	0.141 (0.021)
	300	0.034 (0.003)	0.040 (0.004)	0.056 (0.007)	0.124 (0.017)
	350	0.031 (0.004)	0.036 (0.004)	0.051 (0.006)	0.115 (0.035)
	400	0.028 (0.002)	0.033 (0.003)	0.046 (0.004)	0.103 (0.013)
	450	0.027 (0.002)	0.031 (0.003)	0.043 (0.003)	0.097 (0.014)
	500	0.026 (0.002)	0.029 (0.003)	0.041 (0.004)	0.094 (0.012)

Table S.7: Number of times, out of 100, that the dimension is correctly selected with $J = 4$, $nsr = 0.05$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	20	27	18	14
150	41	40	36	9
200	58	71	48	15
250	69	57	58	10
300	89	84	55	15
350	90	90	65	19
400	91	94	78	21
450	97	97	81	20
500	96	96	77	26

Table S.8: Average and standard deviation (in brackets) of $ORMSEP$ with $nsr = 0.05$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.377 (0.053)	0.388 (0.058)	0.435 (0.054)	0.596 (0.056)
	150	0.276 (0.038)	0.287 (0.035)	0.347 (0.042)	0.513 (0.050)
	200	0.218 (0.030)	0.232 (0.030)	0.296 (0.032)	0.460 (0.046)
	250	0.181 (0.026)	0.202 (0.022)	0.260 (0.028)	0.425 (0.042)
	300	0.150 (0.019)	0.171 (0.019)	0.237 (0.027)	0.396 (0.038)
	350	0.134 (0.019)	0.151 (0.017)	0.215 (0.023)	0.371 (0.035)
	400	0.117 (0.016)	0.135 (0.017)	0.196 (0.021)	0.346 (0.029)
	450	0.106 (0.012)	0.124 (0.013)	0.187 (0.019)	0.337 (0.029)
	500	0.099 (0.012)	0.117 (0.013)	0.179 (0.016)	0.318 (0.027)
$ORMSEP_{deriv}$	100	0.379 (0.153)	0.369 (0.157)	0.406 (0.129)	0.585 (0.104)
	150	0.251 (0.141)	0.266 (0.133)	0.320 (0.140)	0.531 (0.110)
	200	0.192 (0.128)	0.169 (0.114)	0.255 (0.119)	0.468 (0.103)
	250	0.147 (0.119)	0.180 (0.121)	0.212 (0.136)	0.441 (0.095)
	300	0.092 (0.076)	0.112 (0.090)	0.213 (0.124)	0.406 (0.092)
	350	0.085 (0.080)	0.093 (0.076)	0.179 (0.116)	0.373 (0.098)
	400	0.078 (0.077)	0.077 (0.060)	0.145 (0.097)	0.359 (0.095)
	450	0.057 (0.044)	0.065 (0.044)	0.129 (0.094)	0.350 (0.095)
	500	0.056 (0.051)	0.063 (0.051)	0.134 (0.102)	0.327 (0.100)

Table S.9: Number of times, out of 100, that the dimension is correctly selected with $J = 2$, $nsr = 0.4$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	91	90	96	88
150	98	98	96	97
200	94	94	97	97
250	99	99	97	98
300	97	99	99	97
350	100	99	100	99
400	99	100	98	99
450	98	100	100	99
500	99	100	100	98

Table S.10: Average and standard deviation (in brackets) of $ORMSEP$ with $J = 2$, $nsr = 0.4$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.107 (0.049)	0.116 (0.045)	0.142 (0.040)	0.238 (0.068)
	150	0.073 (0.025)	0.084 (0.027)	0.111 (0.037)	0.177 (0.042)
	200	0.060 (0.021)	0.072 (0.029)	0.094 (0.025)	0.140 (0.032)
	250	0.049 (0.015)	0.057 (0.017)	0.081 (0.023)	0.125 (0.025)
	300	0.042 (0.013)	0.050 (0.013)	0.070 (0.016)	0.114 (0.023)
	350	0.039 (0.011)	0.047 (0.013)	0.065 (0.013)	0.101 (0.020)
	400	0.034 (0.009)	0.042 (0.011)	0.061 (0.012)	0.096 (0.017)
	450	0.032 (0.009)	0.039 (0.009)	0.055 (0.011)	0.089 (0.018)
	500	0.029 (0.009)	0.037 (0.009)	0.052 (0.012)	0.080 (0.015)
$ORMSEP_{deriv}$	100	0.111 (0.115)	0.114 (0.112)	0.104 (0.056)	0.211 (0.113)
	150	0.070 (0.056)	0.064 (0.020)	0.079 (0.049)	0.145 (0.055)
	200	0.068 (0.095)	0.062 (0.052)	0.073 (0.067)	0.121 (0.052)
	250	0.045 (0.015)	0.045 (0.014)	0.062 (0.054)	0.101 (0.020)
	300	0.049 (0.074)	0.046 (0.052)	0.050 (0.010)	0.091 (0.016)
	350	0.038 (0.011)	0.039 (0.010)	0.045 (0.009)	0.084 (0.019)
	400	0.034 (0.009)	0.035 (0.008)	0.048 (0.047)	0.077 (0.016)
	450	0.033 (0.011)	0.032 (0.007)	0.040 (0.007)	0.071 (0.012)
	500	0.030 (0.009)	0.032 (0.007)	0.038 (0.007)	0.068 (0.014)

Table S.11: Number of times, out of 100, that the dimension is correctly selected with $J = 3$, $nsr = 0.4$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	56	56	57	53
150	64	67	62	56
200	79	81	67	55
250	84	93	77	69
300	90	83	82	70
350	96	89	85	78
400	93	94	81	78
450	96	90	89	85
500	98	91	89	82

Table S.12: Average and standard deviation (in brackets) of $ORMSEP$ with $J = 3$, $nsr = 0.4$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.290 (0.062)	0.310 (0.066)	0.339 (0.065)	0.506 (0.072)
	150	0.216 (0.055)	0.223 (0.049)	0.268 (0.049)	0.408 (0.053)
	200	0.165 (0.034)	0.179 (0.039)	0.228 (0.044)	0.366 (0.054)
	250	0.139 (0.036)	0.152 (0.030)	0.210 (0.034)	0.324 (0.047)
	300	0.116 (0.029)	0.133 (0.028)	0.182 (0.035)	0.294 (0.038)
	350	0.102 (0.021)	0.122 (0.027)	0.165 (0.026)	0.275 (0.034)
	400	0.098 (0.019)	0.116 (0.020)	0.152 (0.026)	0.255 (0.028)
	450	0.087 (0.016)	0.104 (0.019)	0.147 (0.023)	0.243 (0.029)
	500	0.082 (0.017)	0.100 (0.019)	0.140 (0.020)	0.229 (0.029)
$ORMSEP_{deriv}$	100	0.303 (0.203)	0.302 (0.190)	0.313 (0.157)	0.472 (0.145)
	150	0.231 (0.180)	0.230 (0.183)	0.255 (0.143)	0.387 (0.141)
	200	0.154 (0.129)	0.151 (0.120)	0.219 (0.143)	0.355 (0.145)
	250	0.129 (0.128)	0.105 (0.082)	0.178 (0.137)	0.287 (0.126)
	300	0.099 (0.093)	0.125 (0.115)	0.149 (0.119)	0.267 (0.125)
	350	0.077 (0.064)	0.103 (0.097)	0.134 (0.108)	0.228 (0.116)
	400	0.085 (0.084)	0.084 (0.076)	0.138 (0.123)	0.218 (0.117)
	450	0.069 (0.066)	0.092 (0.098)	0.108 (0.096)	0.188 (0.105)
	500	0.063 (0.073)	0.086 (0.094)	0.105 (0.099)	0.193 (0.109)

Table S.13: Number of times, out of 100, that the dimension is correctly selected with $J = 4$, $nsr = 0.4$.

n	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
100	11	8	8	13
150	5	7	4	5
200	6	14	7	5
250	5	13	3	5
300	20	11	10	5
350	20	12	8	3
400	15	16	7	3
450	20	19	8	4
500	21	13	10	2

Table S.14: Average and standard deviation (in brackets) of $ORMSEP$ with $J = 4$, $nsr = 0.4$.

		$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.4$
$ORMSEP_{reg}$	100	0.498 (0.079)	0.515 (0.098)	0.539 (0.090)	0.703 (0.094)
	150	0.390 (0.072)	0.412 (0.061)	0.459 (0.072)	0.605 (0.077)
	200	0.319 (0.051)	0.342 (0.044)	0.390 (0.050)	0.544 (0.054)
	250	0.280 (0.038)	0.304 (0.046)	0.353 (0.042)	0.510 (0.050)
	300	0.264 (0.038)	0.271 (0.042)	0.334 (0.043)	0.472 (0.040)
	350	0.230 (0.031)	0.249 (0.038)	0.300 (0.037)	0.442 (0.043)
	400	0.213 (0.031)	0.234 (0.035)	0.288 (0.039)	0.432 (0.039)
	450	0.199 (0.026)	0.216 (0.031)	0.271 (0.039)	0.410 (0.043)
	500	0.191 (0.027)	0.203 (0.023)	0.257 (0.032)	0.392 (0.035)
$ORMSEP_{deriv}$	100	0.502 (0.160)	0.504 (0.145)	0.527 (0.138)	0.693 (0.112)
	150	0.450 (0.125)	0.460 (0.137)	0.490 (0.128)	0.612 (0.089)
	200	0.411 (0.141)	0.394 (0.153)	0.435 (0.117)	0.582 (0.095)
	250	0.390 (0.123)	0.390 (0.149)	0.428 (0.109)	0.555 (0.104)
	300	0.326 (0.150)	0.359 (0.135)	0.383 (0.117)	0.519 (0.105)
	350	0.311 (0.131)	0.341 (0.125)	0.383 (0.117)	0.488 (0.083)
	400	0.320 (0.133)	0.323 (0.129)	0.365 (0.105)	0.479 (0.087)
	450	0.303 (0.148)	0.295 (0.116)	0.353 (0.110)	0.473 (0.091)
	500	0.279 (0.117)	0.299 (0.100)	0.324 (0.082)	0.466 (0.089)

S.1·4. *Complement to the study from Section* 4·2, $D = 4$

Tables S.19 and S.20 supplement results given in Tables 1 and 2 for two additional perturbation levels and 500 learning functions, and Tables S.15–S.18 provide analogous results for samples of 100 learning functions. The conclusions from the main paper all hold true. Especially in the situation when the learning sample size is small, the method FAM seems to be numerically unstable. Remarkably, the local linear estimation method does not appear to suffer from such drawbacks.

Table S.15: Model (M2) with $D = 4$, $n = 100$, $nsr = 0.05$ and $\rho = 0.05$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.003 (0.002)	0.017 (0.003)	0.120 (0.012)	0.566 (0.045)	1.032 (0.043)
	LC	0.133 (0.028)	0.138 (0.029)	0.177 (0.039)	0.384 (0.054)	0.624 (0.080)
	LL	0.004 (0.002)	0.015 (0.004)	0.066 (0.014)	0.221 (0.045)	0.378 (0.053)
	FAM	0.111 (0.090)	0.127 (0.132)	0.159 (0.070)	0.328 (0.092)	0.458 (0.106)
	GPR	0.018 (0.004)	0.019 (0.005)	0.026 (0.007)	0.056 (0.012)	0.092 (0.018)
Deriv.	L	**1.861 (0.056)**	1.277 (0.456)	1.073 (0.115)	1.044 (0.064)	1.010 (0.029)
	LL	**0.585 (0.230)**	0.931 (0.465)	0.348 (0.181)	0.329 (0.198)	0.317 (0.125)
	FAM	**0.786 (1.021)**	42.419 (254.483)	4.133 (14.675)	0.808 (1.257)	0.427 (0.558)

Table S.16: Model (M2) with $D = 4$, $n = 100$, $nsr = 0.1$ and $\rho = 0.1$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.006 (0.004)	0.020 (0.005)	0.126 (0.015)	0.572 (0.044)	1.031 (0.055)
	LC	0.148 (0.031)	0.154 (0.034)	0.201 (0.041)	0.430 (0.076)	0.637 (0.086)
	LL	0.009 (0.005)	0.022 (0.012)	0.085 (0.021)	0.262 (0.056)	0.404 (0.069)
	FAM	0.127 (0.084)	0.129 (0.090)	0.178 (0.084)	0.354 (0.098)	0.491 (0.122)
	GPR	0.029 (0.007)	0.031 (0.009)	0.039 (0.010)	0.083 (0.016)	0.119 (0.025)
Deriv.	L	**1.864 (0.091)**	1.337 (0.512)	1.094 (0.122)	1.046 (0.070)	1.012 (0.047)
	LL	**0.711 (0.253)**	0.998 (0.467)	0.518 (0.205)	0.384 (0.187)	0.386 (0.137)
	FAM	**1.505 (2.696)**	43.694 (185.737)	5.036 (9.420)	1.476 (4.044)	0.819 (1.465)

Table S.17: Model (M2) with $D = 4$, $n = 100$, $nsr = 0.2$ and $\rho = 0.2$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.014 (0.008)	0.031 (0.011)	0.131 (0.017)	0.575 (0.049)	1.033 (0.058)
	LC	0.190 (0.050)	0.197 (0.049)	0.246 (0.047)	0.480 (0.065)	0.740 (0.091)
	LL	0.023 (0.020)	0.035 (0.015)	0.110 (0.025)	0.318 (0.059)	0.486 (0.068)
	FAM	0.199 (0.138)	0.181 (0.109)	0.249 (0.116)	0.425 (0.170)	0.563 (0.143)
	GPR	0.048 (0.014)	0.050 (0.013)	0.067 (0.016)	0.127 (0.025)	0.194 (0.039)
Deriv.	L	**1.864 (0.063)**	1.622 (0.572)	1.079 (0.090)	1.034 (0.034)	1.010 (0.030)
	LL	**0.875 (0.254)**	1.389 (0.564)	0.725 (0.231)	0.525 (0.216)	0.474 (0.137)
	FAM	**3.140 (5.907)**	65.612 (197.409)	8.354 (14.596)	2.093 (3.349)	1.702 (3.191)

Table S.18: Model (M2) with $D = 4$, $n = 100$, $nsr = 0.4$ and $\rho = 0.4$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.038 (0.018)	0.059 (0.023)	0.161 (0.028)	0.603 (0.054)	1.046 (0.060)
	LC	0.298 (0.066)	0.292 (0.061)	0.367 (0.063)	0.631 (0.079)	0.932 (0.093)
	LL	0.053 (0.029)	0.071 (0.029)	0.149 (0.035)	0.445 (0.100)	0.689 (0.092)
	FAM	0.280 (0.167)	0.289 (0.163)	0.350 (0.176)	0.561 (0.176)	0.732 (0.283)
	GPR	0.079 (0.023)	0.091 (0.026)	0.117 (0.030)	0.241 (0.043)	0.374 (0.058)
Deriv.	L	**1.864 (0.036)**	1.976 (0.465)	1.148 (0.078)	1.042 (0.025)	1.010 (0.016)
	LL	**1.264 (0.218)**	1.973 (0.489)	0.973 (0.135)	0.730 (0.137)	0.685 (0.117)
	FAM	**2.828 (12.855)**	79.267 (170.564)	9.004 (16.088)	1.479 (1.607)	1.533 (6.947)

Table S.19: Model (M2) with $D = 4$, $n = 500$, $nsr = 0.1$ and $\rho = 0.1$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.001 (0.001)	0.015 (0.002)	0.113 (0.009)	0.532 (0.032)	1.008 (0.012)
	LC	0.067 (0.011)	0.070 (0.012)	0.084 (0.015)	0.182 (0.022)	0.301 (0.028)
	LL	0.003 (0.001)	0.011 (0.003)	0.032 (0.005)	0.086 (0.010)	0.139 (0.019)
	FAM	0.034 (0.027)	0.048 (0.025)	0.081 (0.031)	0.187 (0.033)	0.293 (0.048)
	GPR	0.011 (0.002)	0.012 (0.002)	0.016 (0.002)	0.030 (0.004)	0.040 (0.005)
Deriv.	L	**1.851 (0.030)**	1.114 (0.126)	1.024 (0.029)	1.013 (0.016)	1.006 (0.012)
	LL	**0.208 (0.191)**	0.415 (0.229)	0.220 (0.143)	0.123 (0.103)	0.113 (0.104)
	FAM	**0.683 (1.659)**	11.078 (27.439)	1.654 (3.980)	0.451 (0.487)	0.372 (0.910)

Table S.20: Model (M2) with $D = 4$, $n = 500$, $nsr = 0.2$ and $\rho = 0.2$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.003 (0.002)	0.017 (0.002)	0.113 (0.009)	0.535 (0.033)	1.008 (0.011)
	LC	0.091 (0.011)	0.094 (0.014)	0.133 (0.018)	0.247 (0.028)	0.398 (0.036)
	LL	0.006 (0.003)	0.015 (0.004)	0.051 (0.008)	0.140 (0.020)	0.223 (0.024)
	FAM	0.048 (0.028)	0.057 (0.033)	0.089 (0.032)	0.197 (0.042)	0.306 (0.051)
	GPR	0.018 (0.003)	0.019 (0.003)	0.028 (0.004)	0.051 (0.007)	0.070 (0.008)
Deriv.	L	**1.850 (0.023)**	1.131 (0.112)	1.025 (0.024)	1.011 (0.009)	1.003 (0.004)
	LL	**0.501 (0.216)**	0.611 (0.185)	0.330 (0.206)	0.292 (0.134)	0.272 (0.117)
	FAM	**0.687 (0.500)**	19.723 (69.186)	1.906 (2.457)	0.552 (0.504)	0.373 (0.271)

S.1·5. *Complement to the study from Section* 4·2, $D = 15$

Tables S.21–S.28 provide complete results of the extended simulation study for model (M3) and $D = 15$. The conclusions to be drawn match closely our general observations from the situation with $D = 4$.

Table S.21: Model (M2) with $D = 15$, $n = 100$, $nsr = 0.05$ and $\rho = 0.05$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.010 (0.004)	0.027 (0.005)	0.142 (0.017)	0.647 (0.073)	1.036 (0.064)
	LC	0.467 (0.059)	0.483 (0.062)	0.551 (0.067)	0.849 (0.082)	1.013 (0.019)
	LL	0.014 (0.004)	0.030 (0.006)	0.149 (0.020)	0.655 (0.088)	0.979 (0.065)
	FAM	0.425 (0.377)	0.429 (0.225)	0.527 (0.212)	0.790 (0.171)	1.081 (0.241)
	GPR	0.038 (0.012)	0.045 (0.010)	0.085 (0.015)	0.269 (0.037)	0.501 (0.060)
Deriv.	L	**6.972 (0.119)**	1.575 (0.339)	1.146 (0.124)	1.082 (0.074)	1.008 (0.015)
	LL	**6.559 (0.206)**	1.480 (0.356)	1.000 (0.089)	0.987 (0.073)	0.948 (0.029)
	FAM	**12.388 (23.698)**	92.580 (165.531)	9.751 (14.839)	2.206 (1.882)	1.793 (3.438)

Table S.22: Model (M2) with $D = 15$, $n = 100$, $nsr = 0.1$ and $\rho = 0.1$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.025 (0.012)	0.041 (0.013)	0.158 (0.028)	0.675 (0.087)	1.035 (0.057)
	LC	0.469 (0.059)	0.475 (0.059)	0.566 (0.077)	0.864 (0.080)	1.011 (0.025)
	LL	0.031 (0.013)	0.048 (0.014)	0.164 (0.034)	0.684 (0.089)	0.980 (0.057)
	FAM	0.464 (0.213)	0.476 (0.342)	0.517 (0.173)	0.870 (0.284)	1.106 (0.197)
	GPR	0.054 (0.016)	0.059 (0.015)	0.105 (0.020)	0.319 (0.044)	0.558 (0.067)
Deriv.	L	**6.976 (0.097)**	2.215 (1.110)	1.218 (0.247)	1.125 (0.139)	1.007 (0.012)
	LL	**6.631 (0.194)**	2.149 (1.123)	1.105 (0.276)	1.018 (0.100)	0.957 (0.028)
	FAM	**11.009 (16.257)**	83.555 (115.968)	11.405 (16.186)	3.298 (4.952)	1.589 (2.341)

Table S.23: Model (M2) with $D = 15$, $n = 100$, $nsr = 0.2$ and $\rho = 0.2$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.053 (0.021)	0.073 (0.028)	0.184 (0.036)	0.688 (0.068)	1.045 (0.073)
	LC	0.513 (0.065)	0.524 (0.069)	0.586 (0.077)	0.855 (0.082)	1.011 (0.021)
	LL	0.066 (0.024)	0.083 (0.024)	0.200 (0.041)	0.698 (0.086)	1.008 (0.092)
	FAM	0.692 (0.259)	0.744 (0.280)	0.826 (0.290)	1.155 (0.347)	1.405 (0.241)
	GPR	0.083 (0.025)	0.090 (0.021)	0.146 (0.027)	0.399 (0.043)	0.697 (0.096)
Deriv.	L	**6.993 (0.129)**	2.971 (1.441)	1.277 (0.237)	1.093 (0.054)	1.010 (0.017)
	LL	**6.700 (0.185)**	2.802 (1.084)	1.159 (0.216)	1.016 (0.072)	0.967 (0.025)
	FAM	**25.179 (43.463)**	246.828 (251.037)	33.957 (40.713)	8.440 (14.606)	3.636 (6.298)

Table S.24: Model (M2) with $D = 15$, $n = 100$, $nsr = 0.4$ and $\rho = 0.4$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.158 (0.040)	0.196 (0.066)	0.317 (0.073)	0.787 (0.114)	1.046 (0.089)
	LC	0.608 (0.080)	0.615 (0.074)	0.673 (0.066)	0.902 (0.074)	1.014 (0.021)
	LL	0.197 (0.050)	0.237 (0.070)	0.357 (0.077)	0.777 (0.107)	1.052 (0.083)
	FAM	1.053 (0.353)	1.059 (0.351)	1.158 (0.353)	1.478 (0.424)	1.746 (0.406)
	GPR	0.154 (0.042)	0.172 (0.044)	0.248 (0.046)	0.571 (0.070)	1.006 (0.032)
Deriv.	L	**7.000 (0.136)**	5.459 (1.502)	1.587 (0.186)	1.139 (0.050)	1.009 (0.018)
	LL	**6.794 (0.131)**	5.852 (1.636)	1.565 (0.222)	1.086 (0.051)	0.979 (0.017)
	FAM	**30.530 (59.006)**	238.362 (420.575)	32.245 (32.335)	7.476 (9.553)	4.397 (8.501)

Table S.25: Model (M2) with $D = 15$, $n = 500$, $nsr = 0.05$ and $\rho = 0.05$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.002 (0.001)	0.016 (0.002)	0.113 (0.009)	0.540 (0.034)	1.011 (0.019)
	LC	0.310 (0.029)	0.326 (0.029)	0.401 (0.035)	0.751 (0.048)	1.002 (0.007)
	LL	0.005 (0.001)	0.029 (0.005)	0.241 (0.067)	0.539 (0.057)	0.838 (0.038)
	FAM	0.091 (0.049)	0.100 (0.055)	0.164 (0.059)	0.383 (0.050)	0.602 (0.048)
	GPR	0.010 (0.002)	0.012 (0.002)	0.019 (0.003)	0.052 (0.006)	0.087 (0.008)
Deriv.	L	**6.962 (0.064)**	1.326 (0.117)	1.046 (0.026)	1.018 (0.011)	1.004 (0.007)
	LL	**5.919 (0.175)**	0.961 (0.120)	1.522 (0.670)	0.962 (0.066)	0.854 (0.024)
	FAM	**3.854 (3.374)**	20.554 (50.449)	1.896 (3.101)	0.625 (0.417)	0.556 (0.485)

Table S.26: Model (M2) with $D = 15$, $n = 500$, $nsr = 0.1$ and $\rho = 0.1$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.005 (0.001)	0.019 (0.002)	0.119 (0.010)	0.546 (0.032)	1.009 (0.012)
	LC	0.326 (0.030)	0.326 (0.030)	0.415 (0.036)	0.758 (0.052)	1.002 (0.007)
	LL	0.010 (0.002)	0.036 (0.010)	0.240 (0.060)	0.555 (0.057)	0.838 (0.041)
	FAM	0.075 (0.036)	0.084 (0.031)	0.149 (0.037)	0.376 (0.045)	0.600 (0.049)
	GPR	0.017 (0.002)	0.019 (0.003)	0.028 (0.004)	0.068 (0.008)	0.113 (0.011)
Deriv.	L	**6.950 (0.048)**	1.524 (0.261)	1.075 (0.050)	1.029 (0.026)	1.003 (0.002)
	LL	**5.975 (0.184)**	1.129 (0.458)	1.474 (0.612)	0.976 (0.067)	0.862 (0.025)
	FAM	**3.483 (2.655)**	11.804 (14.373)	2.088 (3.507)	0.652 (0.557)	0.503 (0.382)

Table S.27: Model (M2) with $D = 15$, $n = 500$, $nsr = 0.2$ and $\rho = 0.2$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.011 (0.003)	0.025 (0.004)	0.125 (0.012)	0.555 (0.039)	1.014 (0.016)
	LC	0.350 (0.036)	0.361 (0.029)	0.444 (0.035)	0.769 (0.055)	1.004 (0.011)
	LL	0.023 (0.004)	0.046 (0.012)	0.248 (0.075)	0.572 (0.051)	0.851 (0.038)
	FAM	0.120 (0.043)	0.136 (0.051)	0.199 (0.045)	0.439 (0.057)	0.670 (0.062)
	GPR	0.029 (0.005)	0.031 (0.005)	0.047 (0.006)	0.111 (0.012)	0.182 (0.017)
Deriv.	L	**6.950 (0.050)**	1.726 (0.237)	1.093 (0.041)	1.026 (0.016)	1.003 (0.003)
	LL	**6.031 (0.153)**	1.411 (0.414)	1.450 (0.629)	0.982 (0.058)	0.871 (0.022)
	FAM	**6.738 (5.618)**	38.102 (43.432)	6.287 (14.228)	1.787 (3.497)	0.972 (0.809)

Table S.28: Model (M2) with $D = 15$, $n = 500$, $nsr = 0.4$ and $\rho = 0.4$.

		$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
Reg.	L	0.031 (0.008)	0.045 (0.009)	0.145 (0.017)	0.586 (0.041)	1.016 (0.022)
	LC	0.438 (0.032)	0.444 (0.038)	0.515 (0.035)	0.817 (0.057)	1.005 (0.008)
	LL	0.076 (0.026)	0.100 (0.028)	0.262 (0.060)	0.626 (0.063)	0.899 (0.043)
	FAM	0.180 (0.061)	0.189 (0.062)	0.265 (0.070)	0.534 (0.082)	0.825 (0.085)
	GPR	0.047 (0.008)	0.054 (0.009)	0.094 (0.012)	0.249 (0.020)	0.423 (0.035)
Deriv.	L	**6.959 (0.054)**	2.110 (0.287)	1.140 (0.035)	1.041 (0.010)	1.004 (0.004)
	LL	**6.306 (0.128)**	2.637 (1.301)	1.271 (0.441)	1.023 (0.068)	0.910 (0.017)
	FAM	**6.563 (5.674)**	39.583 (47.419)	5.630 (6.619)	1.655 (2.692)	0.947 (0.816)

S.1·6. *Complement to the study from Section* 4·2*: running times*

The next table gathers the average running times of the competing methods over 10 replications of model (M1) with $nsr = 0.4$, estimation of both the regression operator and its functional derivative with testing sample size 500. For the sake of fairness, it is worth noting that method GPR is not optimized for speed; we use the generic numerical optimization method in function `optim` in `R`, whereas the other methods are programmed efficiently in `C++`. Our local linear estimators are competitive with respect to the alternative nonparametric methods, especially FAM and GPR.

Table S.29: Average running times (in s.) of the competing methods, based on 10 replications of the experiment with learning sample size n.

n	L	LC	LL	FAM	GPR
100	0.008	0.075	0.366	0.420	1.061
150	0.014	0.147	0.861	0.909	1.131
200	0.014	0.186	1.121	1.165	2.849
250	0.018	0.228	1.564	1.710	3.683
300	0.025	0.347	1.536	3.723	12.468
350	0.032	0.439	2.344	3.180	14.273
400	0.044	0.525	2.691	4.325	13.966
450	0.048	0.629	4.036	6.002	27.306
500	0.060	0.769	5.428	7.248	31.571

S.1·7. *Complement to the data analysis from Section* 4·3

As explained in Section 4·3, restricting the growth velocity profiles from ages 1–10 to 5–8 does not degrade the quality of estimation. Figure S.3 displays the observed responses versus their

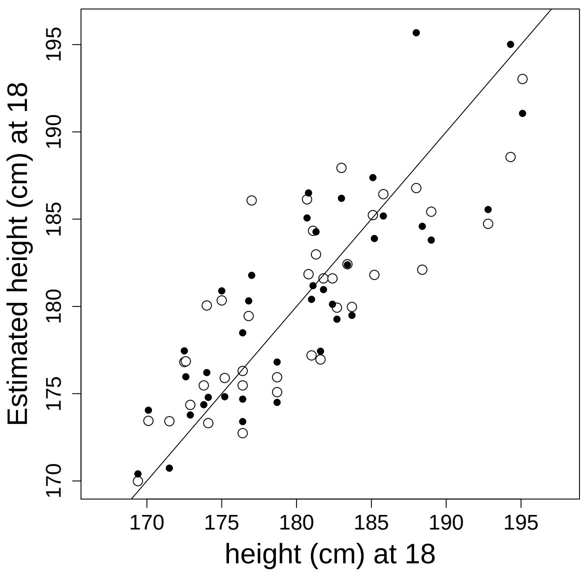

Fig. S.3: Estimates based on the 1–10 growth velocity profiles (black points) and on the 5–8 growth velocity profiles (circles).

estimates when considering the whole growth velocity profile (1–10), or the restricted one (5–8). To quantify the performance of the estimating procedure, we compute the empirical (Pearson's) correlation coefficient between the observations and their estimates in each situation. When the regression model involves the whole trajectory of the growth velocity, the correlation equals 0.847; in the other case where estimates are based on ages 5–8, one gets 0.815. The accuracy of the estimating procedures is almost the same, which confirms that the behaviour of the growth velocity profile outside 5–8 years of age does not influence the adult height at 18.

S.1·8. fllr *R package*

An efficient, fully documented `R` implementation of all the considered estimating procedures, including the automated selection of all their parameters, is freely available as a part of `R` package `fllr`. The package can be downloaded from `https://bitbucket.org/StanislavNagy/fllr`. Using the procedures from `fllr` and the source codes accompanying the present manuscript, all the results from the main paper, and the simulation studies presented in the supplementary material, can be replicated in full.

S.2. PROOFS OF THEOREMS AND THEORETICAL COMPLEMENTS

This section details the proofs of the theorems given in the main paper. Technical lemmas are gathered in Section S.2·5. Throughout this section, we use several conventions. Sums indicated by $\sum_i$ are always meant with i from 1 to n, and sums indicated by $\sum_{j\geq J}$ mean sums with j from J to infinity, for J given; $\mathbf{1}$ stands for a column vector of ones of appropriate dimension; $1_{[0,u]}(t)$ is the indicator of $t \in [0,u]$, i.e. 1 if $t \in [0,u]$ and 0 otherwise. The (j,k)-th element of a matrix Δ can be denoted either by $[\Delta]_{jk}$, or equivalently by Δ_{jk}. At last, to ease the reading of this theoretical part, main notations (even those introduced in the next Section S.2·1) are gathered below:

- $\pi_x(h) := P\left(\|X_i - x\| \leq h\right)$,
- $\pi_x^{-1}(u) := \inf\{h \colon \pi_x(h) \geq u\}$ is the generalized inverse function of $\pi_x(\cdot)$,
- $\gamma_{j_1,\dots,j_M}^{p_1,\dots,p_M}(t) = \mathrm{E}\left(\langle \phi_{j_1}, X_1 - x\rangle^{p_1} \cdots \langle \phi_{j_M}, X_1 - x\rangle^{p_M} | \|X_1 - x\|^{p_1+\dots+p_M} = t\right)$,
- Γ is the $J \times J$ matrix whose (j,k)-th element is defined by $[\Gamma]_{jk} := \gamma_{j,k}^{1,1'}(0)$,
- λ_J is the smallest eigenvalue of the $J \times J$ matrix Γ,
- Λ is the $J \times J$ matrix such that $[\Lambda]_{j,k} := \sqrt{\gamma_{j,k}^{2,2'}(0)}$,
- γ is the J-dimensional vector such that $[\gamma]_j := \gamma_j^{1'}(0)$,
- θ is the J-dimensional vector such that $[\theta]_j := \sqrt{\gamma_j^{2'}(0)}$,
- $\forall u > 0$, $\tau_x(u)$ is the limit of the ratio $\pi_x(h\,u)/\pi_x(h)$ as h goes to 0,
- $b_{x,p,q} := K^q(1) - \int_0^1 \{u^p K^q(u)\}' \tau_x(u) \,\mathrm{d}\, u$.

S.2·1. *Hypotheses*

Let us first focus on the assumptions needed to derive the asymptotic behaviour of $\hat{m}(x)$ and $\hat{m}_x'$.

(H2) The kernel function K is continuously differentiable on its support $(0,1)$ with $K'(s) \leq 0$ for all $s \in (0,1)$ and $K(1) > 0$.

(H3) For any integers $j_1,\dots,j_M,p_1,\dots,p_M \geq 0$ with $M \geq 1$, let us define $\gamma_{j_1,\dots,j_M}^{p_1,\dots,p_M}(t) := \mathrm{E}\left(\langle \phi_{j_1}, X_1 - x\rangle^{p_1} \cdots \langle \phi_{j_M}, X_1 - x\rangle^{p_M} | \|X_1 - x\|^{p_1+\dots+p_M} = t\right)$ and let $\gamma_{j_1,\dots,j_M}^{p_1,\dots,p_M\,'}(t)$ be its derivative at t. The functions $\gamma_{j_1}^{1}$, $\gamma_{j_1,j_2}^{1,1}$, $\dots$, $\gamma_{j_1,\dots,j_4}^{1,\dots,1}$,

$\gamma^2_{j_1}$, $\gamma^{2,1,1}_{j_1,j_2,j_3}$, $\gamma^{2,1,1,1,1}_{j_1,\ldots,j_5}$ and $\gamma^{2,2}_{j_1,j_2}$ are assumed to be continuously differentiable around zero and the smallest eigenvalue λ_J of the $J \times J$ matrix Γ, whose (j,k)-th element is defined by $[\Gamma]_{jk} := {\gamma^{1,1}_{j,k}}'(0)$, is strictly positive.

(H4) $h = h_n$ tends to 0 with n, $J = J_n$ and $n\,\pi_x(h)$ grow to infinity with n so that $h\,J^{1/2} = o(1)$, $h^{-1}\lambda_J^{-1}\{n\,\pi_x(h)\}^{-1/2} = o(1)$ and $h^{-1}(\lambda_J/J)^{-1/2}\{n\,\pi_x(h)\}^{-1/2} = o(1)$, where $h \mapsto \pi_x(h) := P\left(\|X_1 - x\| < h\right)$ is a continuous function in a neighbourhood of zero.

(H5) For all $s > 0$, the ratio $\tau_{x,h}(s) := \dfrac{\pi_x(hs)}{\pi_x(h)}$ tends to $\tau_x(s) \in [0, +\infty]$ as h goes to 0.

(H6) The conditional variance of the error $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ mapping H into $\mathbb{R}$ is a uniformly continuous operator.

The original hypothesis (H3) introduces a particular family of functions. It is shown in LEMMA S.9 that the functions $\gamma^{p_1,\ldots,p_M}_{j_1,\ldots,j_M}$ have interesting properties, in particular the matrix Γ is positive semi-definite. Of course, since the size J of the square matrix Γ tends to infinity with n, it is clear that its smallest eigenvalue λ_J tends to zero with n. Nevertheless, it is not much restrictive to require λ_J to be strictly positive for any fixed J. It is worth noting that the differentiability assumptions imposed on $\gamma^{p_1,\ldots,p_M}_{j_1,\ldots,j_M}$ make this function very useful to approximate $\mathrm{E}\left\{\langle\phi_{j_1}, X_1 - x\rangle^{p_1}\cdots\langle\phi_{j_M}, X_1 - x\rangle^{p_M}\, K^q\left(h^{-1}\|X_1 - x\|\right)\right\}$ that plays a major role in the asymptotic behaviour of the estimators $\hat{m}(x)$ and $\hat{m}'_x$. LEMMA S.15 provides a general situation where (H3) is fulfilled with $\lambda_J > (J+1)^{-1}$, see (S.33). Condition (H5) is a more classical assumption. For standard families of processes an explicit form of the function $\tau_x(s)$ is available, see the discussion just after (H7) in Section S.2·4, and for more details see Ferraty et al., 2007 and references therein.

S.2·2. *Proof of Theorem* 1

PROOF OF THE CONDITIONAL BIAS OF $\hat{m}(x)$. Here only details of the main guidelines of the proof are given. Details and technical lemmas are postponed to Section S.2·5.

For any X_i, the Taylor expansion of m at x can be expressed as

$$m(X_i) \;=\; m(x) + \langle m'_x, X_i - x\rangle + \frac{1}{2}\langle m''_\zeta(X_i - x), X_i - x\rangle \quad \text{a.s.,} \tag{S.1}$$

where $\zeta = x + t(X_i - x)$ with $t \in (0,1)$. By using a basis expansion of m'_x in terms of $\phi_1, \phi_2, \ldots$, we get $m(X_i) = m(x) + \sum_{j\le J}\langle m'_x, \phi_j\rangle\langle\phi_j, X_i - x\rangle + \langle\mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_i - x\rangle + R_{\zeta,x,i}/2$ almost surely with $R_{\zeta,x,i} := \langle m''_\zeta(X_i - x), X_i - x\rangle$. Let $\nabla m_x := [\langle m'_x, \phi_1\rangle, \ldots, \langle m'_x, \phi_J\rangle]^\mathsf{T}$ be the first J coordinates of the gradient of m at x and $R_{\zeta,x} := [R_{\zeta,x,1}, \ldots, R_{\zeta,x,n}]^\mathsf{T}$. Then we can write

$$\begin{bmatrix} m(X_1) \\ \vdots \\ m(X_n) \end{bmatrix} \;=\; \Phi\begin{bmatrix} m(x) \\ \nabla m_x \end{bmatrix} + \begin{bmatrix} \langle\mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_1 - x\rangle \\ \vdots \\ \langle\mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_n - x\rangle \end{bmatrix} + \frac{1}{2}R_{\zeta,x} \quad \text{a.s.} \tag{S.2}$$

According to the definition of $\hat{m}(x)$,

$$\mathrm{E}_{\mathrm{X}}\{\hat{m}(x)\} \;=\; \mathrm{e}^\mathsf{T}\left(\Phi^\mathsf{T}\mathrm{K}\Phi\right)^{-1}\Phi^\mathsf{T}\mathrm{K}\,[m(X_1), \ldots, m(X_n)]^\mathsf{T} \;=\; m(x) + T_1 + \frac{1}{2}T_2 \tag{S.3}$$

where the quantity $T_1 \coloneqq \mathrm{e}^{\mathsf{T}} \left(\Phi^{\mathsf{T}} \mathrm{K} \Phi\right)^{-1} \Phi^{\mathsf{T}} \mathrm{K} \left[\langle \mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x, X_1 - x\rangle, \ldots, \langle \mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x, X_n - x\rangle\right]^{\mathsf{T}}$ and the term $T_2 \coloneqq \mathrm{e}^{\mathsf{T}} \left(\Phi^{\mathsf{T}} \mathrm{K} \Phi\right)^{-1} \Phi^{\mathsf{T}} \mathrm{K} R_{\zeta,x}^{\mathsf{T}}$. The next lemma focuses on the asymptotic behaviour of T_1 and T_2.

LEMMA S.1. *As soon as conditions* (H1)*–*(H5) *are fulfilled,*

(i) $T_1 \;=\; O_P\left(\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|\, h\right)$,

(ii) $T_2 \;=\; O_P\left(h^2\right)$.

Proof of LEMMA S.1-(i) We know that $T_1 = \mathrm{e}^{\mathsf{T}} \left(\Phi^{\mathsf{T}} \widetilde{\mathrm{K}} \Phi\right)^{-1} [A_0, A_1, \ldots, A_J]^{\mathsf{T}}$ with

$$A_0 \coloneqq (n \,\mathrm{E}\, K_1)^{-1} \sum_i K_i \langle \mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x, X_i - x\rangle,$$
$$A_j \coloneqq (n \,\mathrm{E}\, K_1)^{-1} \sum_i K_i \langle \mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x, X_i - x\rangle \langle \phi_j, X_i - x\rangle, \quad \text{for } j = 1, \ldots, J,$$

for $i = 1, \ldots, n$, $K_i \coloneqq K\left(h^{-1}\|X_i - x\|\right)$ and $\widetilde{\mathrm{K}} \coloneqq (n \,\mathrm{E}\, K_1)^{-1}\mathrm{K}$. Set $\delta_0 \coloneqq (n \;\mathrm{E}\, K_1)^{-1} \sum_i K_i$, $\delta_j \coloneqq (n \;\mathrm{E}\, K_1)^{-1} \sum_i \langle \phi_j, X_i - x\rangle K_i$ for $j = 1, \ldots, J$ and let Δ be the $J \times J$ matrix whose the (j,k)-th entry is equal to $\Delta_{jk} \coloneqq (n \;\mathrm{E}\, K_1)^{-1} \sum_i \langle \phi_j, X_i - x\rangle \langle \phi_k, X_i - x\rangle K_i$ for $j, k = 1, \ldots, J$. By using elementary linear algebra,

$$\Phi^{\mathsf{T}} \widetilde{\mathrm{K}} \Phi \;=\; \left[\begin{array}{c|c} \delta_0 & \delta^{\mathsf{T}} \\ \hline \delta & \Delta \end{array}\right],$$

where $\delta \coloneqq [\delta_1, \ldots, \delta_J]^{\mathsf{T}}$. According to standard results with respect to the inverse of a 2×2 block matrix (see for instance Lu & Shiou, 2002),

$$\left(\Phi^{\mathsf{T}} \widetilde{\mathrm{K}} \Phi\right)^{-1} = \left[\begin{array}{c|c} \mu & -\mu\, \delta^{\mathsf{T}} \Delta^{-1} \\ \hline -\mu\, \Delta^{-1}\delta & \Delta^{-1} \;+\; \mu\, \Delta^{-1}\delta\, \delta^{\mathsf{T}} \Delta^{-1} \end{array}\right] \tag{S.4}$$

with $\mu \coloneqq \left(\delta_0 - \delta^{\mathsf{T}} \Delta^{-1} \delta\right)^{-1}$. Then, $T_1 \;=\; \mu \left(A_0 \;-\; \delta^{\mathsf{T}} \Delta^{-1} [A_1, \ldots, A_J]^{\mathsf{T}}\right)$. Before going on, let us focus on Δ^{-1}, the inverse of Δ. From LEMMA S.10, one has $\Delta = \Delta_1 + \Delta_2$ with $\Delta_1 \coloneqq b_{x,0,1}^{-1}\, b_{x,2,1}\, h^2\, \Gamma\, \{1 + o(1)\}$ and $\Delta_2 \coloneqq O_P\left(h^2 \{n\pi_x(h)\}^{-1/2}\right) \Lambda$. As soon as Δ_1^{-1} is invertible, $\Delta^{-1} = \Delta_1^{-1}(\mathrm{I} + \Delta_2 \Delta_1^{-1})^{-1}$. Let $\|\cdot\|_F$ stand for the Frobenius matrix norm and recall that λ_J is the smallest eigenvalue of the $J \times J$ matrix Γ. We have that $\|\Delta_1^{-1}\|_F = O_P\left(\lambda_J^{-1}\, J^{1/2}\, h^{-2}\right)$, and thanks to LEMMA S.9, $\|\Delta_2\|_F = O_P\left(h^2\, \{n\pi_x(h)\}^{-1/2}\right)$ so that the Frobenius norm of $\Delta_2 \Delta_1^{-1}$ is equal to $O_P\left(\lambda_J^{-1}\, J^{1/2}\, \{n\pi_x(h)\}^{-1/2}\right)$. According to (H4), for n large enough, $\|\Delta_2 \Delta_1^{-1}\|_F$ is smaller than 1. Then $\Delta^{-1} = \Delta_1^{-1} \left\{\mathrm{I} + \sum_{k\geq1} (-1)^k \left(\Delta_2 \Delta_1^{-1}\right)^k\right\}$ which results in

$$\Delta^{-1} = b_{x,0,1}\, b_{x,2,1}^{-1}\, h^{-2}\, \Gamma^{-1}\, \{1 + o_P(1)\}\,. \tag{S.5}$$

LEMMAS S.11 and S.12 allow us to write

$$\begin{aligned} T_1 = \mu\, b_{x,0,1}^{-1} b_{x,1,1} \Big\{ &\alpha_{0,x,n}^{bias} h \,\{1 + o(1)\} + O_P\left(\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\| h \{n\pi_x(h)\}^{-1/2}\right) \\ &- \left(\gamma^{\mathsf{T}} + \theta^{\mathsf{T}} O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right) \Gamma^{-1} \alpha_{x,n}^{bias} h \,\{1 + o_P(1)\} \\ &+ \left(\gamma^{\mathsf{T}} + \theta^{\mathsf{T}} O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right) \Gamma^{-1} \sqrt{\alpha_{x,n}^{var}}\, O_P\left(h\, \{n\pi_x(h)\}^{-1/2}\right) \Big\} \end{aligned} \tag{S.6}$$

where $\gamma := [\gamma_1^{1'}(0), \ldots, \gamma_J^{1'}(0)]^\top$, $\theta := \left[\sqrt{\gamma_1^{2'}(0)}, \ldots, \sqrt{\gamma_J^{2'}(0)}\right]^\top$, $\alpha_{x,n}^{bias} := \left[\alpha_{1,x,n}^{bias}, \ldots, \alpha_{J,x,n}^{bias}\right]^\top$ and $\sqrt{\alpha_{x,n}^{var}} := \left[\sqrt{\alpha_{1,x,n}^{var}}, \ldots, \sqrt{\alpha_{J,x,n}^{var}}\right]^\top$. From LEMMA S.12 we know that $\|\alpha_{x,n}^{bias}\|_2 \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|$ and $\|\sqrt{\alpha_{x,n}^{var}}\|_2 \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|$. Here $\|.\|_2$ stands for the Euclidean vector norm. Based on LEMMAS S.9, S.12, and S.13,

$$
\begin{aligned}
\left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\alpha_{x,n}^{bias} &= O_P\left(\lambda_J^{-1}\left\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\right\|\right), \\
\left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\sqrt{\alpha_{x,n}^{var}} &= O_P\left(\lambda_J^{-1}\left\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\right\|\right),
\end{aligned}
\tag{S.7}
$$

where λ_J is the smallest eigenvalue of the $J \times J$ matrix Γ. LEMMA S.9 indicates that the Frobenius norm of Γ is finite and thus λ_J converges to 0 with n. Let us now focus on the term μ. Firstly, one has $\mathrm{E}\,\delta_0 = 1$ and $\mathrm{Var}(\delta_0) = n^{-1}(\mathrm{E}\,K_1)^{-2}\,\mathrm{Var}(K_1)$. Thanks to COROLLARY S.1, it is easy to see that $\mathrm{Var}(\delta_0) = O\left(1/\{n\,\pi_x(h)\}\right)$ which leads to $\delta_0 = 1 + O_P\left(1/\sqrt{n\,\pi_x(h)}\right)$. Secondly, LEMMA S.11 and (S.5) imply that

$$
\begin{aligned}
\delta^\top\Delta^{-1}\delta = {} & b_{x,0,1}^{-1}\, b_{x,1,1}^{2}\, b_{x,2,1}^{-1}\,\gamma^\top\Gamma^{-1}\gamma\,\{1 + o_P(1)\} \\
& + O_P\left(\{n\,\pi_x(h)\}^{-1/2}\right)\theta^\top\Gamma^{-1}\gamma \,+\, O_P\left(\{n\,\pi_x(h)\}^{-1}\right)\theta^\top\Gamma^{-1}\theta.
\end{aligned}
$$

LEMMAS S.9 and S.13 give $\theta^\top\Gamma^{-1}\gamma\,/\,\gamma^\top\Gamma^{-1}\gamma = O(\lambda_J^{-1}) = \theta^\top\Gamma^{-1}\theta\,/\,\gamma^\top\Gamma^{-1}\gamma$. Then (H4) leads to $\{n\,\pi_x(h)\}^{-1/2}\theta^\top\Gamma^{-1}\gamma = o_P(\gamma^\top\Gamma^{-1}\gamma) = \{n\,\pi_x(h)\}^{-1}\theta^\top\Gamma^{-1}\theta$ and finally, the quantity $\delta^\top\Delta^{-1}\delta$ is equal to $b_{x,0,1}^{-1}\, b_{x,1,1}^{2}\, b_{x,2,1}^{-1}\,\gamma^\top\Gamma^{-1}\gamma\,\{1 + o_P(1)\}$. Use again LEMMA S.13 to get

$$
\delta^\top\Delta^{-1}\delta \;=\; O_P\left(\lambda_J^{-1}\right), \text{ which results in } \mu = O_P\left(\lambda_J\right). \tag{S.8}
$$

This last result combined with (S.6) and (S.7) gives $T_1 \;=\; O_P\left(h\,\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|\right)$.

Proof of LEMMA S.1-(ii) We have that $T_2 = T_{21} + T_{22}$ where

- $T_{21} := \mathrm{e}^\top\left(\Phi^\top\widetilde{\mathrm{K}}\Phi\right)^{-1}[B_0, B_1, \ldots, B_J]^\top$ with $B_0 := (n\,\mathrm{E}\,K_1)^{-1}\sum_i K_i R_{x,x,i}$, and for $j = 1, \ldots, J$, $B_j := (n\,\mathrm{E}\,K_1)^{-1}\sum_i K_i R_{x,x,i}\langle\phi_j, X_i - x\rangle$, where $R_{x,x,i}$ is a term involved in (S.2) with ζ replaced by x;
- $T_{22} := \mathrm{e}^\top\left(\Phi^\top\widetilde{\mathrm{K}}\Phi\right)^{-1}[C_0, C_1, \ldots, C_J]^\top$ where $C_0 := (n\,\mathrm{E}\,K_1)^{-1}\sum_i K_i\,(R_{\zeta,x,i} - R_{x,x,i})$, and $C_j := (n\,\mathrm{E}\,K_1)^{-1}\sum_i K_i\,(R_{\zeta,x,i} - R_{x,x,i})\,\langle\phi_j, X_i - x\rangle$ for $j = 1, \ldots, J$.

About T_{21}. LEMMAS S.11 and S.14 allow to write

$$
\begin{aligned}
T_{21} = \mu\, b_{x,0,1}^{-1}\Big\{ & b_{x,2,1}\,\beta_{0,x}^{bias}\,h^2\,\{1 + o(1)\} + O_P\left(h^2\{n\,\pi_x(h)\}^{-1/2}\right) \\
& - b_{x,1,1}b_{x,2,1}^{-1}b_{x,3,1}\left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\beta_x^{bias}\,h^2\,\{1 + o_P(1)\} \\
& + \left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\sqrt{\beta_x^{var}}\,O_P\left(h^2\{n\pi_x(h)\}^{-1/2}\right)\Big\}.
\end{aligned}
\tag{S.9}
$$

Based on LEMMAS S.9, S.13 and S.14,

$$
\begin{aligned}
\left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\beta_x^{bias} &= O_P\left(\lambda_J^{-1}\right), \\
\left(\gamma^\top + \theta^\top O_P\left(\{n\pi_x(h)\}^{-1/2}\right)\right)\Gamma^{-1}\sqrt{\beta_x^{var}} &= O_P\left(\lambda_J^{-1}\right).
\end{aligned}
\tag{S.10}
$$

Now, (S.8), (S.9), and (S.10) result in $T_{21} = O_P\left(h^2\right)$.

About T_{22}. Thanks to (H1), it is easy to show that $C_0 = O_P(h^3)$ and $C_j = O_P(h^4)$ for $j = 1, \ldots, J$. Consequently, $T_{22} = \mathrm{e}^{\mathsf{T}}\left(\Phi^{\mathsf{T}}\widetilde{\mathrm{K}}\Phi\right)^{-1}[1, h, \ldots, h]^{\mathsf{T}}\,O_P(h^3) = \mu\left\{1 - h\,\delta^{\mathsf{T}}\Delta^{-1}\mathbf{1}\right\}O_P(h^3)$. According to (S.8) and LEMMAS S.10, S.11 and S.13, $T_{22} = O_P\left(h^3\sqrt{J}\right)$.

Back to T_2. Because $h\sqrt{J} = o(1)$ thanks to (H4), $T_{22} = o_P(h^2)$ and $T_2 = O_P(h^2)$.

Now, it is enough to plug in LEMMA S.1 with (S.3) to get the claimed conditional bias.

PROOF OF THE CONDITIONAL VARIANCE OF $\hat{m}(x)$. By the assumptions, the covariance matrix of Y given $X_1, \ldots, X_n$ is the diagonal matrix $\operatorname{diag}\{\sigma^2(X_1), \ldots, \sigma^2(X_n)\}$ and the variance $\operatorname{Var}_{\mathrm{X}}\{\hat{m}(x)\} = \mathrm{e}^{\mathsf{T}}\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\Phi^{\mathsf{T}}\mathrm{D}\Phi\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\mathrm{e}$. Here, D is also a diagonal matrix such that $[\mathrm{D}]_{ii} = \sigma^2(X_i)\,[\mathrm{K}]_{ii}^2$. According to (H2), (H4) and (H6), $[\mathrm{D}]_{ii} = \{\sigma^2(x) + o(1)\}\,[\mathrm{K}]_{ii}^2$ and $\operatorname{Var}_{\mathrm{X}}\{\hat{m}(x)\} = \{\sigma^2(x) + o(1)\}\,\mathrm{e}^{\mathsf{T}}\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\Phi^{\mathsf{T}}\,\mathrm{K}^2\,\Phi\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\mathrm{e}$. The remainder of the proof is based on the next lemma which provides a decomposition of the conditional variance with the asymptotic behaviour of each term.

LEMMA S.2. *As soon as conditions* (H1)*–*(H6) *are fulfilled,*

(i) $\operatorname{Var}_X\{\hat{m}(x)\} = \{\sigma^2(x) + o(1)\}\,\mu^2\left\{\widetilde{\delta}_0 - \delta^{\mathsf{T}}\Delta^{-1}\widetilde{\delta} - \widetilde{\delta}^{\mathsf{T}}\Delta^{-1}\delta + \delta^{\mathsf{T}}\Delta^{-1}\widetilde{\Delta}\Delta^{-1}\delta\right\}$, *where* $\widetilde{\delta}_0 := (n\,\mathrm{E}\,K_1)^{-2}\sum_i K_i^2$, $\widetilde{\delta}_j := (n\,\mathrm{E}\,K_1)^{-2}\sum_i\langle\phi_j, X_i - x\rangle\,K_i^2$ *for* $j = 1, \ldots, J$, $\widetilde{\delta} := [\widetilde{\delta}_1, \ldots, \widetilde{\delta}_J]^{\mathsf{T}}$, $\widetilde{\Delta}_{jk} := (n\,\mathrm{E}\,K_1)^{-2}\sum_i\langle\phi_j, X_i - x\rangle\,\langle\phi_k, X_i - x\rangle\,K_i^2$ *for* $j, k = 1, \ldots, J$*, and* $\widetilde{\Delta}$ *is the* $J \times J$ *matrix such that* $[\widetilde{\Delta}]_{jk} := \widetilde{\Delta}_{jk}$*,*

(ii) $\widetilde{\delta}_0 = b_{x,0,1}^{-2}\,b_{x,0,2}\,\{n\,\pi_x(h)\}^{-1}\{1 + o_P(1)\}$, $\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\delta} = O_P\left(\{\lambda_J\,n\,\pi_x(h)\}^{-1}\right)$, *and* $\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\Delta}\Delta^{-1}\delta = O_P\left(\{\lambda_J\,n\,\pi_x(h)\}^{-1}\right)$.

PROOF OF LEMMA S.2-(i). The main term of $\operatorname{Var}_{\mathrm{X}}\{\hat{m}(x)\}$ is given by

$$\mathrm{e}^{\mathsf{T}}\left(\Phi^{\mathsf{T}}\widetilde{\mathrm{K}}\Phi\right)^{-1}\Phi^{\mathsf{T}}\,\widetilde{\mathrm{K}}^2\,\Phi\left(\Phi^{\mathsf{T}}\widetilde{\mathrm{K}}\Phi\right)^{-1}\mathrm{e}$$

where $\Phi^{\mathsf{T}}\widetilde{\mathrm{K}}^2\Phi = \left[\begin{array}{c|c}\widetilde{\delta}_0 & \widetilde{\delta}^{\mathsf{T}} \\ \hline \widetilde{\delta} & \widetilde{\Delta}\end{array}\right]$. Formula (S.4) with elementary linear algebra result in the claimed decomposition.

PROOF OF LEMMA S.2-(ii). By COROLLARY S.1, $\mathrm{E}\,K_1^2 = b_{x,0,2}\,\pi_x(h)\,\{1 + o(1)\}$ and in the same vein of LEMMAS S.10 and S.11, it is easy to see that

$$\begin{aligned}
\widetilde{\delta}_0 &= b_{x,0,1}^{-2}\,b_{x,0,2}\,\{n\,\pi_x(h)\}^{-1}\,\{1 + o_P(1)\}, \\
\widetilde{\delta}_j &= b_{x,0,1}^{-2}\,b_{x,1,2}\,\gamma_j^{1'}(0)\,h\,\{n\,\pi_x(h)\}^{-1}\,\{1 + o_P(1)\} + O_P\left(h\,\{n\,\pi_x(h)\}^{-3/2}\right)\sqrt{\gamma_j^{2'}(0)}, \\
\widetilde{\Delta}_{jk} &= b_{x,0,1}^{-2}\,b_{x,2,2}\,\gamma_{j,k}^{1,1'}(0)\,h^2\,\{n\pi_x(h)\}^{-1}\,\{1 + o_P(1)\} + O_P\left(h^2\,\{n\,\pi_x(h)\}^{-3/2}\right)\sqrt{\gamma_{j,k}^{2,2'}(0)}.
\end{aligned}$$

As a by-product, one has

$$\widetilde{\delta} \,=\, b_{x,0,1}^{-2}\, b_{x,1,2}\, \gamma\, h\, \{n\,\pi_x(h)\}^{-1}\, \{1 + o_P(1)\} \,+\, O_P\left(h\,\{n\,\pi_x(h)\}^{-3/2}\right)\theta, \tag{S.11}$$

and

$$\widetilde{\Delta} \,=\, b_{x,0,1}^{-2}\, b_{x,2,2}\, \Gamma\, h^2\, \{n\pi_x(h)\}^{-1}\, \{1 + o_P(1)\} \,+\, O_P\left(h^2\,\{n\,\pi_x(h)\}^{-3/2}\right)\Lambda. \tag{S.12}$$

These two last equations with (S.5) lead to

$$\begin{aligned}\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\delta} = b_{x,0,1}^{-2}\, b_{x,1,1}\, b_{x,2,1}^{-1}\, b_{x,2,2}\, \gamma^{\mathsf{T}}\Gamma^{-1}\gamma\, \{n\,\pi_x(h)\}^{-1}\{1 + o_P(1)\}&\\ + O_P\left(\{n\,\pi_x(h)\}^{-3/2}\right)\theta^{\mathsf{T}}\Gamma^{-1}\gamma \,+\, O_P\left(\{n\,\pi_x(h)\}^{-2}\right)\theta^{\mathsf{T}}\Gamma^{-1}\theta,&\end{aligned} \tag{S.13}$$

and

$$\begin{aligned}\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\Delta}\Delta^{-1}\delta = b_{x,0,1}^{-2}\, b_{x,1,1}^{2}\, b_{x,2,1}^{-2}\, b_{x,2,2}\, \gamma^{\mathsf{T}}\Gamma^{-1}\gamma\, \{n\,\pi_x(h)\}^{-1}\, \{1 + o_P(1)\}&\\ +\, O_P\left(\{n\,\pi_x(h)\}^{-3/2}\right)\gamma^{\mathsf{T}}\Gamma^{-1}\Lambda\Gamma^{-1}\gamma.&\end{aligned} \tag{S.14}$$

Combination of LEMMAS S.9-(ii) and S.13 gives $\gamma^{\mathsf{T}}\Gamma^{-1}\Lambda\Gamma^{-1}\gamma = O(\lambda_J^{-2})$, $\gamma^{\mathsf{T}}\Gamma^{-1}\gamma = O(\lambda_J^{-1})$ and the claimed assertions hold. Finally, the rate of convergence of the conditional variance is derived from LEMMA S.2 and the fact that $\mu = O_P(\lambda_J)$.

S.2·3. *Proof of Theorem* 2

The asymptotic behaviour of $\|\hat{m}_x' - m_x'\|$ is a direct by-product of the next results which detail the conditional bias and variance of the functional derivative $\hat{m}_x'$.

LEMMA S.3. *As soon as* (H1)*–*(H6) *are fulfilled,*

(i) $\|\,\mathrm{E}_X(\hat{m}_x') - m_x'\| \,=\, O_P\left(\lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x'\|\right) \,+\, O_P\left(\lambda_J^{-1}h\right),$

(ii) and conditionally on $X_1,\dots,X_n$,

$$\|\hat{m}_x' - \mathrm{E}_X(\hat{m}_x')\| \,=\, O_P\left(h^{-1}\{\lambda_J^2 n\pi_x(h)\}^{-1/2}\right) \,+\, O_P\left(h^{-1}\{\lambda_J n\pi_x(h)\}^{-1/2}\sqrt{J}\right).$$

PROOF OF LEMMA S.3-(i). Let us first focus on the conditional bias of $\hat{m}_x'$. By the definition of $\hat{m}_x'$, $\mathrm{E}_X(\hat{m}_x') \,=\, \phi^{\mathsf{T}}\,[0|\mathrm{I}]\,\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\Phi^{\mathsf{T}}\mathrm{K}\,[m(X_1),\dots,m(X_n)]^{\mathsf{T}}$ so that, by (S.1) and (S.2),

$$\mathrm{E}_X(\hat{m}_x') - m_x' \,=\, -\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x' \,+\, Q_1 \,+\, \frac{1}{2}\,Q_2, \tag{S.15}$$

with $Q_1 := \phi^{\mathsf{T}}\,[0|\mathrm{I}]\,\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\Phi^{\mathsf{T}}\mathrm{K}\left[\langle\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x', X_1 - x\rangle,\dots,\langle\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x', X_n - x\rangle\right]^{\mathsf{T}}$, and where $Q_2 := \phi^{\mathsf{T}}\,[0|\mathrm{I}]\,\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\Phi^{\mathsf{T}}\mathrm{K}R_{\zeta,x}^{\mathsf{T}}$.

LEMMA S.4. *As soon as conditions* (H1)*–*(H5) *are fulfilled,*

(i) $\|Q_1\| \,=\, O_P\left(\lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x'\|\right),$

(ii) $\|Q_2\| \,=\, O_P\left(\lambda_J^{-1}\,h\right).$

Proof of LEMMA S.4-(i). With the notations introduced in the proof LEMMA S.1-(i) and (S.4), $Q_1 = -\mu\, A_0\, \phi^{\mathsf{T}}\Delta^{-1}\delta + \phi^{\mathsf{T}}(\Delta^{-1} + \mu\,\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1})\,[A_1,\dots,A_J]^{\mathsf{T}}$. According to LEMMA S.11, LEMMA S.12 and (S.5), $\mu\, A_0\, \phi^{\mathsf{T}}\Delta^{-1}\delta \,=\, O_P\left(\lambda_J\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m_x'\|\,\phi^{\mathsf{T}}\Gamma^{-1}\gamma\right),$

$\phi^{\mathsf{T}}\Delta^{-1}\,[A_1,\ldots,A_J]^{\mathsf{T}} = O_P\left(\phi^{\mathsf{T}}\Gamma^{-1}\alpha_{x,n}^{bias}\right)$ and $\mu\,\phi^{\mathsf{T}}\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1}\,[A_1,\ldots,A_J]^{\mathsf{T}} = O_P\left(\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|\,\phi^{\mathsf{T}}\Gamma^{-1}\gamma\right)$. Set $\mathrm{u} := \lambda_J\Gamma^{-1}\gamma$. For any J, $\|\phi^{\mathsf{T}}\Gamma^{-1}\gamma\| = \lambda_J^{-1}\|\phi^{\mathsf{T}}\mathrm{u}\| = \lambda_J^{-1}\|\mathrm{u}\|_2 \leq \lambda_J^{-1}$. Consequently $\|\phi^{\mathsf{T}}\Gamma^{-1}\gamma\| = O_P\left(\lambda_J^{-1}\right)$. Now, define the vector $\mathrm{v} := \lambda_J\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|^{-1}\Gamma^{-1}\alpha_{x,n}^{bias}$ with $\|\mathrm{v}\|_2 \leq 1$. From the definition of v we have $\|\phi^{\mathsf{T}}\Gamma^{-1}\alpha_{x,n}^{bias}\| = \lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|\,\|\phi^{\mathsf{T}}\mathrm{v}\|$. Because of the orthonormality of the basis, $\|\phi^{\mathsf{T}}\mathrm{v}\| = \|\mathrm{v}\|_2$, $\|\phi^{\mathsf{T}}\Gamma^{-1}\alpha_{x,n}^{bias}\| = \lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|\,\|\mathrm{v}\|_2 \leq \lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|$ which gives $\|\phi^{\mathsf{T}}\Gamma^{-1}\alpha_{x,n}^{bias}\| = O_P\left(\lambda_J^{-1}\|\mathcal{P}_{\mathcal{S}_J^{\perp}} m'_x\|\right)$ and the claimed rate of convergence holds.

Proof of LEMMA S.4-(ii). Similarly to the notations and the proof of LEMMA S.1-(ii), $Q_2 = Q_{21} + Q_{22}$ with

$$Q_{21} := -\mu\,B_0\,\phi^{\mathsf{T}}\Delta^{-1}\delta + \phi^{\mathsf{T}}(\Delta^{-1} + \mu\,\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1})\,[B_1,\ldots,B_J]^{\mathsf{T}}\,,$$
$$Q_{22} := -\mu\,C_0\,\phi^{\mathsf{T}}\Delta^{-1}\delta + \phi^{\mathsf{T}}(\Delta^{-1} + \mu\,\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1})\,[C_1,\ldots,C_J]^{\mathsf{T}}\,.$$

By following the proof of LEMMA S.4-(i) and replacing LEMMA S.12 with LEMMA S.14, $\mu\,B_0\,\|\phi^{\mathsf{T}}\Delta^{-1}\delta\|$ is of order h in probability, $\|\phi^{\mathsf{T}}\Delta^{-1}\,[B_1,\ldots,B_J]^{\mathsf{T}}\,\|$ is of order $\lambda_J^{-1}\,h$ in probability and the quantity $\mu\,\|\phi^{\mathsf{T}}\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1}\,[B_1,\ldots,B_J]^{\mathsf{T}}\,\| = O_P\left(\lambda_J^{-1}\,h\right)$ so that $\|Q_{21}\| = O_P\left(\lambda_J^{-1}\,h\right)$. For studying Q_{22}, recall that $C_0 = O_P(h^3)$ and $C_j = O_P(h^4)$ for $j = 1,\ldots,J$ (see again the proof of LEMMA S.1-(ii)). Then, we obtain that $\mu\,C_0\,\|\phi^{\mathsf{T}}\Delta^{-1}\delta\| = O_P\left(h^2\right)$, $\|\phi^{\mathsf{T}}\Delta^{-1}\,[C_1,\ldots,C_J]^{\mathsf{T}}\,\| = O_P\left(\lambda_J^{-1}\,\sqrt{J}\,h^2\right)$, $\mu\,\|\phi^{\mathsf{T}}\Delta^{-1}\,\delta\,\delta^{\mathsf{T}}\,\Delta^{-1}\,[C_1,\ldots,C_J]^{\mathsf{T}}\,\| = O_P\left(\lambda_J^{-1}\,\sqrt{J}\,h^2\right)$ and the quantity $\|Q_{22}\| = o_P\left(\lambda_J^{-1}\,h\right)$ from (H4). This is enough to get the claimed result.

LEMMA S.4 together with (S.15) provide the claimed expression for the conditional bias.

PROOF OF LEMMA S.3-(ii) Following the proof of the conditional variance of $\hat{m}(x)$

$$\mathrm{E}_{\mathrm{X}}\left(\|\hat{m}'_x - \mathrm{E}_{\mathrm{X}}\left(\hat{m}'_x\right)\|^2\right) = \int_0^1 \mathrm{Var}_{\mathrm{X}}\left\{\hat{m}'_x(t)\right\}\,\mathrm{d}\,t = \sigma^2(x)\,Q\,\{1\,+\,o(1)\},$$

where $Q = \displaystyle\int_0^1 \phi(t)^{\mathsf{T}}\,[0|\mathrm{I}]\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\,\Phi^{\mathsf{T}}\,\mathrm{K}^2\,\Phi\,\left(\Phi^{\mathsf{T}}\mathrm{K}\Phi\right)^{-1}\,[0|\mathrm{I}]^{\mathsf{T}}\phi(t)\,\mathrm{d}\,t$. The remainder of the proof consists in decomposing this conditional variance. Set $\mathrm{M}_0 := \Delta^{-1}\delta\delta^{\mathsf{T}}\Delta^{-1}$. Thanks to (S.4), elementary linear algebra and notations introduced in LEMMA S.2, the following decomposition of the conditional variance

$$\mathrm{E}_{\mathrm{X}}\left(\|\hat{m}'_x - \mathrm{E}_{\mathrm{X}}\left(\hat{m}'_x\right)\|^2\right) = \sigma^2(x)\int_0^1 \phi(t)^{\mathsf{T}}\left\{\sum_{j=1}^{8}\mathrm{M}_j\right\}\phi(t)\,\mathrm{d}\,t\,\{1 + o(1)\}$$

holds with

$$\begin{aligned}
&\mathrm{M}_0 := \Delta^{-1}\delta\delta^{\mathsf{T}}\Delta^{-1}, && \mathrm{M}_1 := \mu^2\widetilde{\delta}_0\mathrm{M}_0, && \mathrm{M}_2 := -\mu\Delta^{-1}\widetilde{\delta}\delta^{\mathsf{T}}\Delta^{-1},\\
&\mathrm{M}_3 := \mathrm{M}_2^{\mathsf{T}}, && \mathrm{M}_4 := -2\mu^2\left(\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\delta}\right)\mathrm{M}_0, && \mathrm{M}_5 := \Delta^{-1}\widetilde{\Delta}\Delta^{-1},\\
&\mathrm{M}_6 := \mu\Delta^{-1}\widetilde{\Delta}\mathrm{M}_0, && \mathrm{M}_7 := \mathrm{M}_6^{\mathsf{T}}, && \mathrm{M}_8 := \mu^2\left(\delta^{\mathsf{T}}\Delta^{-1}\widetilde{\Delta}\Delta^{-1}\delta\right)\mathrm{M}_0.
\end{aligned}$$

According to (S.4), (S.8), (S.11), (S.12), (S.13), and (S.14), LEMMAS S.2, S.10, S.11, and S.13, with the fact that $\phi_1, \phi_2, \ldots$ is an orthonormal basis,

$$\begin{array}{ll} \int_0^1 \phi(t)^\top \mathrm{M}_0 \phi(t)\,\mathrm{d}\,t = O_P\left(\lambda_J^{-2} h^{-2}\right), & \int_0^1 \phi(t)^\top \mathrm{M}_1 \phi(t)\,\mathrm{d}\,t = O_P\left(u_n\right), \\ \int_0^1 \phi(t)^\top \mathrm{M}_2 \phi(t)\,\mathrm{d}\,t = O_P\left(\lambda_J^{-1} u_n\right), & \int_0^1 \phi(t)^\top \mathrm{M}_4 \phi(t)\,\mathrm{d}\,t = O_P\left(\lambda_J^{-1} u_n\right), \\ \int_0^1 \phi(t)^\top \mathrm{M}_5 \phi(t)\,\mathrm{d}\,t = O_P\left(J\lambda_J^{-1} u_n\right), & \int_0^1 \phi(t)^\top \mathrm{M}_6 \phi(t)\,\mathrm{d}\,t = O_P\left(\lambda_J^{-2} u_n\right), \end{array}$$

and $\int_0^1 \phi(t)^\top \mathrm{M}_8 \phi(t)\,\mathrm{d}\,t = O_P\left(\lambda_J^{-1} u_n\right)$, where $u_n := h^{-2}\{n\pi_x(h)\}^{-1}$. By combining these last results, $\mathrm{E}_{\mathrm{X}}\left(\|\hat{m}'_x - \mathrm{E}_{\mathrm{X}}\left(\hat{m}'_x\right)\|^2\right) = O_P\left(\lambda_J^{-2} u_n\right) + O_P\left(J\lambda_J^{-1} u_n\right)$ which is exactly the claimed rate of convergence.

S.2·4. *kNN method and proof of Theorem* 3

Let us first recall the following notations: $\pi_x(h) := P\left(\|X_1 - x\| < h\right)$ and set $\pi_x^{-1}(u) := \inf\{h \colon \pi_x(h) \geq u\}$ the generalized inverse function of $\pi_x(\cdot)$. The consistency of the kNN estimators $\hat{m}_{kNN}$ and $\hat{m}'_{x,\,kNN}$ (see their definition in Section 3) requires the additional assumption:

(H7) $k = k_n$, $J = J_n$ and $k/\log n$ grow to infinity with n so that $k\,/\,n = o(1)$, $\pi_x^{-1}(k/n)\sqrt{J} = o(1)$, $\pi_x^{-1}(k/n)\left(\lambda_J\sqrt{k}\right)^{-1} = o(1)$ and $\pi_x^{-1}(k/n)\sqrt{\lambda_J}/\sqrt{Jk} = o(1)$, where $h \mapsto \pi_x(h)$ is a continuous function in a neighbourhood of zero such that, for all $s > 1$, $\tau_x(s) \geq s$ and $\tau_x(s^{-1}) \leq s^{-1}$.

Condition (H7) is the kNN version of (H4) except that the quantity $\tau_x(s)$ has to fulfil an additional condition. Recall that $\tau_x(s)$ is defined as the limit function of the ratio $\pi_x(hs)/\pi_x(h)$ when h tends to 0. If the functional variable X is a standard Gaussian process (Brownian motion, fractional Brownian motion, integrated Wiener process, Brownian bridge, ordinary Ornstein-Uhlenbeck process, etc.), the ratio $\pi_x(hs)/\pi_x(h)$ is equivalent to $s^\alpha \exp\left\{-C\left(1 - s^\beta\right)/(hs)^\beta\right\}$ when h tends to 0 for some $\alpha > 0$, $\beta > 0$ and $C > 0$ (see Lifshits, 1999; Li & Shao, 2001; Gao et al., 2003, 2004; Nazarov & Nikitin, 2004; Beghin et al., 2005; Nazarov & Nikitin, 2005; Karol et al., 2008). If X takes its values in some d-dimensional subspace of H, $\pi_x(hs)/\pi_x(h) \sim s^d$. In both cases, the assumption on the behaviour of $\tau_x(\cdot)$ is fulfilled.

The proof of THEOREM 3 is based on three main lemmas. LEMMA S.5, inspired by LEMMA 1 in Lian (2011), describes the asymptotic behaviour of the random kNN bandwidth and its related small ball probability (i.e. $\pi_x(H_k) \underset{a.s.}{\sim} k/n$ and $H_k \underset{a.s.}{\sim} \pi_x^{-1}(k/n)$). LEMMA S.6 proposes a kNN version of technical LEMMAS S.8 and S.9. The last results (LEMMA S.7) gather the asymptotic behaviour of the quantities Δ^{kNN}, δ^{kNN}, A_0^{kNN}, $A_1^{kNN}, \ldots, A_J^{kNN}$, B_0^{kNN}, $B_1^{kNN}, \ldots, B_J^{kNN}$, the counterpart of Δ, δ, A_0, $A_1, \ldots, A_J$, B_0, $B_1, \ldots, B_J$ when replacing K_i with $K_i^{kNN} := K\left(H_{k,-i}^{-1}\|X_i - x\|\right)$ for $i = 1, \ldots, n$. Let us first introduce and also recall some notations (in addition to those already mentioned at the beginning of the theoretical part):

- $H_{k,-i} := \inf\left\{h \colon \sum_{\ell=1,\ell\neq i}^n 1_{\{X_\ell \in B(x,h)\}} = k\right\}$,
- $K_i^{kNN} := K\left(H_{k,-i}^{-1}\|X_i - x\|\right)$,
- Δ^{kNN} is the $J \times J$ matrix whose the (j,k)-th entry is equal to $\Delta_{jk} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i \langle\phi_j, X_i - x\rangle\langle\phi_k, X_i - x\rangle K_i^{kNN}$ for $j,k = 1, \ldots, J$,
- $\delta^{kNN} := [\delta_1^{kNN}, \ldots, \delta_J^{kNN}]^\top$ where $\delta_j^{kNN} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i \langle\phi_j, X_i - x\rangle K_i^{kNN}$ for $j = 1, \ldots, J$,
- $A_0^{kNN} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i K_i^{kNN}\langle\mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_i - x\rangle$,
- $A_j^{kNN} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i K_i^{kNN}\langle\mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_i - x\rangle\langle\phi_j, X_i - x\rangle$, for $j = 1, \ldots, J$,

- $B_0^{kNN} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i K_i^{kNN}\langle m_x''(X_i-x), X_i-x\rangle$,
- $B_j^{kNN} := (n\,\mathrm{E}\,K_1^{kNN})^{-1}\sum_i K_i^{kNN}\langle m_x''(X_i-x), X_i-x\rangle\langle\phi_j, X_i-x\rangle$.

LEMMA S.5. *If (H7) is fulfilled, (i)* $\pi_x(H_k) \underset{a.s.}{\sim} k/n$*, and (ii)* $H_k \underset{a.s.}{\sim} \pi_x^{-1}(k/n)$.

PROOF OF LEMMA S.5-*(i)*. For any $\epsilon>0$, set $r_+ := \pi_x^{-1}(s_+)$ where $s_+ := (1+\epsilon)k/n$:

$$\begin{aligned} P\left(H_k \geq r_+\right) &= P\left(\sum_i 1_{\{X_i\in B(x,r_+)\}} \leq k\right)\\ &\leq P\left(\big|\sum_i 1_{\{X_i\in B(x,r_+)\}} - ns_+\big| \geq \epsilon k\right)\\ &\leq 2\exp\left\{-\frac{\epsilon^2k^2}{2ns_+(1-s_+)+2\epsilon k/3}\right\}, \end{aligned}$$

where the last inequality comes from the use of Bernstein's inequality for Bernoulli random variables (see for instance the Appendix in Pollard, 2012). Then, $n\,s_+/k = 1+\epsilon$ and $P\left(H_k\geq r_+\right) \leq 2\exp\left\{-\frac{\epsilon^2k}{2(1+\epsilon)(4/3-s_+)}\right\} \leq 2\exp\left(-(1+\eta)\log n\right)$, where $\eta>0$ for n large enough since $k/\log n$ tends to infinity (see (H7)). Applying the Borel-Cantelli lemma, for any $\epsilon>0$ and for n large enough, $H_k \leq \pi_x^{-1}\left\{(1+\epsilon)k/n\right\}$ almost surely. For any $\epsilon>0$ and for n large enough, a similar reasoning with the replacement of r_+ (resp. s_+) by $r_- := \pi_x^{-1}\left\{(1-\epsilon)k/n\right\}$ (resp. $s_- := (1-\epsilon)k/n$) results in $\pi_x(H_k)\geq(1-\epsilon)k/n$ almost surely. Finally,

$$\forall\epsilon>0,\ \exists n_\epsilon,\ \forall n\geq n_\epsilon,\ \pi_x^{-1}\left\{(1-\epsilon)k/n\right\} \leq H_k \leq \pi_x^{-1}\left\{(1+\epsilon)k/n\right\}\ a.s. \tag{S.16}$$

According to the definition of π_x^{-1}, $\pi_x^{-1}\circ\pi_x(H_k)\leq H_k$ and $\pi_x\circ\pi_x^{-1}(u)\geq u$:

- $\pi_x(H_k)\geq s_+ \ \Rightarrow\ H_k\geq \pi_x^{-1}(s_+)$ results in $P\left\{\pi_x(H_k)\geq s_+\right\} \leq P\left\{H_k\geq\pi_x^{-1}(s_+)\right\}$ and with (S.16),
$$\forall\epsilon>0,\ \exists n_\epsilon,\ \forall n\geq n_\epsilon,\ \pi_x(H_k)\leq(1+\epsilon)k/n,\ a.s. \tag{S.17}$$
- $H_k\geq\pi_x^{-1}(s_-) \ \Rightarrow\ \pi_x(H_k)\geq s_- \Rightarrow\ P\left\{H_k\geq\pi_x^{-1}(s_-)\right\} \leq P\left\{\pi_x(H_k)\geq s_-\right\}$ $\Rightarrow\ P\left\{H_k\leq\pi_x^{-1}(s_-)\right\} \geq P\left\{\pi_x(H_k)\leq s_-\right\}$ and combined with (S.16),
$$\forall\epsilon>0,\ \exists n_\epsilon,\ \forall n\geq n_\epsilon,\ \pi_x(H_k)\geq(1-\epsilon)k/n,\ a.s., \tag{S.18}$$

where (S.17) and (S.18) end the proof of LEMMA S.5-*(i)*.

PROOF OF LEMMA S.5-*(ii)*. For any $\epsilon>0$, set $h_+ := (1+\epsilon)\pi_x^{-1}(k/n)$:

$$\begin{aligned} P\left(H_k\geq h_+\right) &= P\left(\sum_i 1_{\{X_i\in B(x,h_+)\}} \leq k\right)\\ &\leq P\left(\sum_i 1_{\{X_i\in B(x,h_+)\}} - n\,\pi_x(h_+) \leq k - n\,\pi_x(h_+)\right). \end{aligned}$$

The assumption (H7) made on the behaviour of the function $\tau_x(\cdot)$ implies that, for n large enough, $\pi_x(h_+)\geq(1+\epsilon)\,\pi_x\circ\pi_x^{-1}(k/n)\geq(1+\epsilon)\,k\,/\,n$ so that $k\,-\,n\,\pi_x(h_+)\leq-\,\epsilon\,k$: $P\left(H_k\geq h_+\right)\leq P\left(\big|\sum_i 1_{\{X_i\in B(x,h_+)\}} - n\,\pi_x(h_+)\big|\ \geq\ \epsilon\,k\right)$ and by using again a Bernstein's inequality for Bernoulli random variables and the Borel-Cantelli lemma, for any $\epsilon>0$ and for n large enough, $H_k \leq (1+\epsilon)\,\pi_x^{-1}(k/n)$ almost surely. For any $\epsilon > 0$ and for n large enough, a similar reasoning with replacing h_+ by $h_- := (1-\epsilon)\,\pi_x^{-1}(k/n)$ results in $H_k \geq (1-\epsilon)\,\pi_x^{-1}(k/n)$ almost surely. Finally, for any $\epsilon > 0$ and for n large enough, $(1-\epsilon)\,\pi_x^{-1}(k/n) \leq H_k \leq (1+\epsilon)\,\pi_x^{-1}(k/n)$ which is equivalent to $H_k \underset{a.s.}{\sim} \pi_x^{-1}(k/n)$.

LEMMA S.6. *Let $p \geq 0$ and $q > 0$; under (H2), (H3), (H5) and (H7), for n large enough,*

(i) $\forall i,\ \mathrm{E}\,\{(\|X_i - x\|/H_{k,-i})^p K^q(\|X_i - x\|/H_{k,-i})\} \;=\; b_{x,p,q}\, k\, n^{-1}\,\{1 + o(1)\}$,

(ii) $\forall i,\ \mathrm{E}\left\{K_{i,k}^q\, \gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K}\,(\|X_i - x\|^{p_+})\right\} \;=\; b_{x,p_+,a}\, \gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K\,\prime}(0)\, \pi_x^{-1}(k/n)^{p_+}\, \dfrac{k}{n}\,\{1 + o(1)\}$, *with $p_+ = p_1 + \cdots + p_K$,*

where $b_{x,p,q}$ and $b_{x,p_+,a}$ are defined in LEMMA S.8.

PROOF OF LEMMA S.6-*(i)*. Set $K_{i,k} := K(\|X_i - x\|/H_{k,-i})$, $U_i := (\|X_i - x\|/H_{k,-i})^p K_{i,k}^q$ and let $\mathcal{S}_{-i}$ be the set of functional regressors with X_i removed. Similarly to the proof of LEMMA S.8,

$$\mathrm{E}(U_i|\mathcal{S}_{-i}) \;=\; \pi_x(H_{k,-i})\left\{K^q(1) - \int_0^1 \{u^p K^q(u)\}'\,\{\pi_x(H_{k,-i}u)/\pi_x(H_{k,-i})\}\,\mathrm{d}\,u\right\}.$$

Because we only focus on the X_j's such that $K_{j,k} > 0$, the quantity $\|X_i - x\|$ is smaller than $H_{k,-i}$. Then, without loss of generality, one sets $H_{k,-i} = H_{k+1}$ and because $\sup_{s\in[0,1]} |\pi_x(s\,H_{k+1})/\pi_x(H_{k+1}) - \tau_x(s)|$ tends to 0 almost surely when n goes to infinity, $\mathrm{E}(U_i|\mathcal{S}_{-i}) \underset{a.s.}{\sim} b_{x,p,q}\,\pi_x(H_{k+1})$. Thanks to LEMMA S.5-*(i)*, $\mathrm{E}(U_i|\mathcal{S}_{-i}) \underset{a.s.}{\sim} b_{x,p,q}\, k\,/\,n$ which implies that, for any i, $\mathrm{E}(U_i) = b_{x,p,q}\, k\, n^{-1}\,\{1 + o(1)\}$. It remains to state that $\sup_{s\in[0,1]} |\pi_x(s\,H_{k+1})/\pi_x(H_{k+1}) - \tau_x(s)|$ tends to 0 almost surely when n goes to infinity. For any $\epsilon > 0$, let A_ϵ and B_ϵ be the events such that $A_\epsilon := \{\exists n,\, \forall k \geq n,\, H_{k+1} < \epsilon\}$ and $B_\epsilon := \{\exists n,\, \forall k \geq n,\, \sup_s |\pi_x(s\,H_{k+1})/\pi_x(H_{k+1}) - \tau_x(s)| < \epsilon\}$. According to (H5), $\forall \epsilon > 0,\, \exists h_\epsilon,\, \forall h \leq h_\epsilon,\, \sup_s |\pi_x(s\,h)/\pi_x(h) - \tau_x(s)| < \epsilon$. Moreover, $H_{k+1} \leq h_\epsilon \Rightarrow |\pi_x(s\,H_{k+1})/\pi_x(H_{k+1}) - \tau_x(s)| < \epsilon$ results in $A_{h_\epsilon} \subset B_\epsilon$. With LEMMA S.5-*(ii)*, H_{k+1} tends to zero almost surely when n goes to infinity so that $P(A_{h_\epsilon}) = 1$ and hence $P(B_\epsilon) = 1$.

PROOF OF LEMMA S.6-*(ii)*. Set $V_i = K_{i,k}^a\,\gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K}\,(\|X_i - x\|^{p_+})$. Similarly to the proof of LEMMA S.9, $\mathrm{E}(V_i|\mathcal{S}_{-i}) = \gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K\,\prime}(0)\, H_{k,-i}^{p_+}\,\mathrm{E}\left\{(\|X_i - x\|/H_{k,-i})^{p_+} K_{i,k}^a \,|\, \mathcal{S}_{-i}\right\}\{1 + o(1)\}$. LEMMA S.5-*(ii)* and LEMMA S.6-*(i)* provide the claimed result.

The last result (LEMMA S.7) gathers the asymptotic behaviour of the quantities Δ^{kNN}, δ^{kNN}, $A_0^{kNN}, A_1^{kNN}, \dots, A_J^{kNN}$, $B_0^{kNN}, B_1^{kNN}, \dots, B_J^{kNN}$, the counterparts of Δ, δ, $A_0, A_1, \dots, A_J$, $B_0, B_1, \dots, B_J$ when replacing $K_i := K\left(h^{-1}\|X_i - x\|\right)$ with $K_i^{kNN} := K\left(H_{k,-i}^{-1}\|X_i - x\|\right)$ for $i = 1, \dots, n$.

LEMMA S.7. *As soon as (H2), (H3), (H5) and (H7) are fulfilled,*

(i) $\Delta^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,2,1}\,\pi_x^{-1}(k/n)^2\,\Gamma\,\{1 + o_P(1)\} \;+\; O_P\left(\pi_x^{-1}(k/n)^2\, k^{-1/2}\right)\Lambda$,

(ii) $\delta^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,1,1}\,\pi_x^{-1}(k/n)\,\gamma\,\{1 + o_P(1)\} \;+\; O_P\left(\pi_x^{-1}(k/n)\, k^{-1/2}\right)\theta$,

(iii) $A_0^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,1,1}\,\alpha_{0,x,n}^{bias}\,\pi_x^{-1}(k/n)\,\{1 + o(1)\} \;+\; O_P\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|\,\pi_x^{-1}(k/n)\, k^{-1/2}\right)$, *where the sequence $\alpha_{0,x,n}^{bias}$ is upper bounded by $\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|$, and*

$A_j^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,2,1}\, \alpha_{j,x,n}^{bias}\, \pi_x^{-1}(k/n)^2\; \{1+o(1)\} \;+\; O_P\left({\alpha_{j,x,n}^{var}}^{1/2}\, \pi_x^{-1}(k/n)^2\, k^{-1/2}\right)$
for any $j \le J$, *where* $\sum_{j=1}^{J} \left\{\alpha_{j,x,n}^{bias}\right\}^2 \le \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|^2$ *and* $\sum_{j=1}^{J} \alpha_{j,x,n}^{var} \le \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|^2$,

(iv) $B_0^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,2,1}\, \beta_{0,x}^{bias}\, \pi_x^{-1}(k/n)^2\, \{1 + o(1)\} + O_P\left(\pi_x^{-1}(k/n)^2\, k^{-1/2}\right)$, *with* $\beta_{0,x}^{bias} = O(1)$, *and*
$B_j^{kNN} \;=\; b_{x,0,1}^{-1}\, b_{x,3,1}\, \beta_{j,x}^{bias}\, \pi_x^{-1}(k/n)^3\; \{1+o(1)\} \;+\; O_P\left(\beta_{j,x}^{var\,1/2}\, \pi_x^{-1}(k/n)^3\, k^{-1/2}\right)$
for all $j = 1,\dots,J$, *where* $\sum_{j=1}^{J}\left\{\beta_{j,x}^{bias}\right\}^2 = O(1)$ *and* $\sum_{j=1}^{J} \beta_{j,x}^{var} = O(1)$.

The proof of LEMMA S.7 follows exactly the same lines than those given in LEMMAS S.10, S.11, S.12, S.14, but with the adaptation coming from LEMMA S.5 and LEMMA S.6.

Finally, the proof of THEOREM 3 is obtained by grouping the results of LEMMAS S.5, S.6 and S.7, and mimicking the same approach as for THEOREMS 1 and 2.

S.2·5. *Technical lemmas*

LEMMA S.8. *Let* $p \ge 0$ *and* $q > 0$. *As soon as* (H2), (H4) *and* (H5) *are fulfilled,*

$$\int_0^1 t^p K^q(t)\,\mathrm{d}\, P^{\|X_1 - x\|/h}(t) = b_{x,p,q}\,\pi_x(h)\;\{1+o(1)\}\,,$$

where $b_{x,p,q} \coloneqq K^q(1) - \int_0^1 \{u^p K^q(u)\}'\,\tau_x(u)\,\mathrm{d}\,u$.

A useful by-product of this lemma is the following corollary.

COROLLARY S.1. *Under* (H2) *and* (H5), $\mathrm{E}\,K_1^q \;=\; b_{x,0,q}\,\pi_x(h)\;\{1+o(1)\}$ *for any* $q > 0$.

PROOF OF LEMMA S.8. In order to shorten the notations, set $Z_h = \|X_1 - x\|/h$. From the differentiability of K it comes that $t^p K^q(t) = K^q(1) - \int_t^1 \{u^p K^q(u)\}'\,\mathrm{d}\,u$ and

$$\begin{aligned}\int_0^1 t^p K^q(t)\,\mathrm{d}\,P^{Z_h}(t) &= K^q(1)\int_0^1 \mathrm{d}\,P^{Z_h}(t) - \int_0^1\left(\int_t^1 \{u^p K^q(u)\}'\,\mathrm{d}\,u\right)\mathrm{d}\,P^{Z_h}(t)\\ &= K^q(1)\pi_x(h) - \int_0^1\left(\int_0^1 \{u^p K^q(u)\}'\,1_{[0,u]}(t)\,\mathrm{d}\,u\right)\mathrm{d}\,P^{Z_h}(t)\\ &= K^q(1)\pi_x(h) - \int_0^1\left(\int_0^1 1_{[0,u]}(t)\,\mathrm{d}\,P^{Z_h}(t)\right)\{u^p K^q(u)\}'\,\mathrm{d}\,u\\ &= K^q(1)\,\pi_x(h) - \int_0^1 \{u^p K^q(u)\}'\;\pi_x(hu)\,\mathrm{d}\,u.\end{aligned}$$

Finally, $\int_0^1 t^p K^q(t)\,\mathrm{d}\,P^{Z_h}(t) = \pi_x(h)\left\{K^q(1) - \int_0^1 \{u^p K^q(u)\}'\;\{\pi_x(hu)/\pi_x(h)\}\;\mathrm{d}\,u\right\}$ and thanks to (H5) the claimed result holds.

LEMMA S.9. *Under* (H2)–(H5) *one has*

(i) $\mathrm{E}\left\{K_1^a\,\gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K}\left(\|X_1 - x\|^{p_+}\right)\right\} \;=\; b_{x,p_+,a}\,{\gamma_{j_1,\dots,j_K}^{p_1,\dots,p_K}}'(0)\,h^{p_+}\,\pi_x(h)\;\{1+o(1)\}$ *for any* $a > 0$, *where* $p_+ = p_1 + \cdots + p_K$,

(ii) $\sum_{j\geq 1}\left\{\gamma_j^{1'}(0)\right\}^2 \leq 1$, $\sum_{j_1,j_2\geq 1}\left\{\gamma_{j_1,j_2}^{1,1\ '}(0)\right\}^2 \leq 1$, $\sum_{j_1,j_2,j_3\geq 1}\left\{\gamma_{j_1,j_2,j_3}^{1,1,1\ '}(0)\right\}^2 \leq 1$,
$\sum_{j_1,\ldots,j_4\geq 1}\left\{\gamma_{j_1,\ldots,j_4}^{1,\ldots,1\ '}(0)\right\}^2 \leq 1$, $\sum_{j\geq 1}\gamma_j^{2'}(0) \leq 1$, *and* $\sum_{j_1,j_2\geq 1}\gamma_{j_1,j_2}^{2,2\ '}(0) \leq 1$,

(iii) the $J \times J$ matrix Γ is positive semi-definite.

An interesting by-product of this lemma indicates that, for any J, the Frobenius norm of Γ is bounded from above by 1. This results in $\sum_{j=1}^{J}\lambda_j^2 \leq 1$ for any J, where the λ_j's are the eigenvalues of Γ. Consequently, the largest eigenvalue of Γ is smaller than 1, and when n goes to infinity, it exists $\mu > 0$ such that $\lambda_J = O(J^{-1/2-\mu})$.

PROOF OF LEMMA S.9-(i). We can write

$$\begin{aligned}
\mathrm{E}\left\{K_1^a\,\gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K}\left(\|X_1-x\|^{p_+}\right)\right\} &= \int_0^h \gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K}\left(t^{p_+}\right)K^a\left(h^{-1}t\right)\,\mathrm{d}\,P^{\|X_1-x\|}(t)\\
&= \int_0^1 \gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K}\left(h^{p_+}t^{p_+}\right)K^a(t)\,\mathrm{d}\,P^{Z_h}(t)\\
&= \int_0^1 (h\,t)^{p_+}\gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K\ '}(0)\left\{1+\eta(\epsilon)\right\}K^a(t)\,\mathrm{d}\,P^{Z_h}(t),
\end{aligned}$$

where $\eta(\epsilon) = \gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K\ '}(\epsilon)/\gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K\ '}(0) - 1$ with $0 < \epsilon < h^{p_+}t^{p_+}$. Because $\gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K}$ is continuously differentiable, $\sup_\epsilon|\eta(\epsilon)| = o(1)$, which implies

$$\mathrm{E}\left\{K_1^a\,\gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K}\left(\|X_1-x\|^{p_+}\right)\right\} = \gamma_{j_1,\ldots,j_K}^{p_1,\ldots,p_K\ '}(0)\,h^{p_+}\int_0^1 t^{p_+}K^a(t)\,\mathrm{d}\,P^{Z_h}(t)\ \left\{1+o(1)\right\}.$$

Use of LEMMA S.8 gives in the claimed result.

PROOF OF LEMMA S.9-(ii). Let us first remark that

$$\sum_{j\geq 1}\left\{\gamma_j^1(t)\right\}^2 \ \leq\ \sum_{j\geq 1}\mathrm{E}\left(\langle\phi_j, X_1-x\rangle^2 \mid \|X_1-x\| = t\right)$$

by using the definition of $\gamma_j^1(t)$. Because $\sum_{j\geq 1}\langle\phi_j, X_1-x\rangle^2 = \|X_1-x\|^2$, $\sum_{j\geq 1}\left\{\gamma_j^1(t)\right\}^2 \leq \mathrm{E}\left(\|X_1-x\|^2 \mid \|X_1-x\| = t\right) = t^2$. Now, by the definition of the derivative,

$$\sum_{j\geq 1}\left\{\gamma_j^{1'}(0)\right\}^2 = \sum_{j\geq 1}\lim_{t\to 0}\frac{\left\{\gamma_j^1(t)\right\}^2}{t^2}. \tag{S.19}$$

So, for any $d \geq 1$, $|S_d(t) - S_\infty(t)| \leq \mathrm{E}\left(\zeta_d \mid \|X_1-x\| = t\right)$ where we write $S_d(t) := \sum_{j=1}^{d}\left\{\gamma_j^1(t)\right\}^2$ and $\zeta_d := \sum_{j>d}\langle\phi_j, X_1-x\rangle^2$. It is clear that $\{\zeta_d\}_d$ is a non-increasing sequence of random variables that converges almost surely to zero with $d\to\infty$. The monotone convergence theorem implies that for any t, the sequence $\mathrm{E}\left(\zeta_d \mid \|X_1-x\| = t\right)$ is also a non-increasing sequence of random variables converging almost surely to zero with d. Consequently, $\sum_{j\geq 1}t^{-2}\left\{\gamma_j^1(t)\right\}^2$ converges uniformly on $(0,h)$. By remarking that for any j,

$\lim_{t\to 0} t^{-2}\left\{\gamma_j^1(t)\right\}^2 = \left\{\gamma_j^{1'}(0)\right\}^2$, one can exchange the limit with the sum in (S.19)

$$\sum_{j\geq 1}\left\{\gamma_j^{1'}(0)\right\}^2 = \lim_{t\to 0}\sum_{j\geq 1}\frac{\left\{\gamma_j^1(t)\right\}^2}{t^2} \leq \lim_{t\to 0}\frac{1}{t^2}\sum_{j\geq 1}\mathrm{E}\left(\langle\phi_j, X_1 - x\rangle^2 \mid \|X_1 - x\| = t\right).$$

The expression $\sum_{j>d}\langle\phi_j, X_1 - x\rangle^2 = \|\mathcal{P}_{\mathcal{S}_d^\perp}(X_1 - x)\|^2$ converges almost surely to 0 when d tends to infinity, where $\mathrm{E}\left(\|\mathcal{P}_{\mathcal{S}_d^\perp}(X_1 - x)\|^2 \mid \|X_1 - x\| = t\right) \leq t^2$. Thanks to Bourbaki (2004) (see Corollary 2 - INT IV.37), one can put the infinite summation into the expectation

$$\begin{aligned}\sum_{j\geq 1}\left\{\gamma_j^{1'}(0)\right\}^2 &\leq \lim_{t\to 0}\frac{1}{t^2}\,\mathrm{E}\left(\sum_{j\geq 1}\langle\phi_j, X_1 - x\rangle^2 \mid \|X_1 - x\| = t\right)\\ &\leq \lim_{t\to 0}\frac{1}{t^2}\,\mathrm{E}\left(\|X_1 - x\|^2 \mid \|X_1 - x\| = t\right) \leq 1,\end{aligned}$$

which corresponds to the claimed first assertion. The other ones can be obtained by using similar arguments.

PROOF OF LEMMA S.9-(iii). For any J-dimensional vector u and $t \geq 0$,

$$\begin{aligned}\mathrm{u}^\mathsf{T}\,\Gamma\,\mathrm{u} &= \sum_{j,k=1}^{J} u_j u_k \gamma_{j,k}^{1,1'}(0) = \lim_{t\to 0} t^{-1}\sum_{j,k=1}^{J} u_j u_k \gamma_{j,k}^{1,1}(t)\\ &= \lim_{t\to 0} t^{-1}\,\mathrm{E}\left\{\sum_{j,k=1}^{J} u_j u_k\langle\phi_j, X_1 - x\rangle\langle\phi_k, X_1 - x\rangle \mid \|X_1 - x\|^2 = t\right\}\\ &= \lim_{t\to 0} t^{-1}\,\mathrm{E}\left\{\left(\sum_{j=1}^{J} u_j\langle\phi_j, X_1 - x\rangle\right)^2 \mid \|X_1 - x\|^2 = t\right\} \geq 0.\end{aligned}$$

LEMMA S.10. *Under* (H2)*–*(H5)*, one has*

$$\Delta = b_{x,0,1}^{-1}\, b_{x,2,1}\, h^2\,\Gamma\,\{1 + o_P(1)\} + O_P\left(\frac{h^2}{\sqrt{n\pi_x(h)}}\right)\Lambda,$$

where Λ *is the* $J \times J$ *matrix such that* $[\Lambda]_{j,k} := \sqrt{\gamma_{j,k}^{2,2'}(0)}$.

PROOF OF LEMMA S.10. From the definition of Δ_{jk},

$$\begin{aligned}\mathrm{E}\,\Delta_{jk} &= (\mathrm{E}\,K_1)^{-1}\,\mathrm{E}\left(K_1\langle\phi_j, X_1 - x\rangle\langle\phi_k, X_1 - x\rangle\right)\\ &= (\mathrm{E}\,K_1)^{-1}\,\mathrm{E}\left\{K_1\,\gamma_{j,k}^{1,1}\left(\|X_1 - x\|^2\right)\right\}.\end{aligned}$$

Use of LEMMA S.9 and COROLLARY S.1 results in

$$\mathrm{E}\,\Delta_{jk} \;=\; b_{x,0,1}^{-1}\, b_{x,2,1}\,\gamma_{j,k}^{1,1'}(0)\, h^2\,\{1 + o(1)\}. \tag{S.20}$$

One can also derive the asymptotic behaviour of $\mathrm{Var}\,(\Delta_{jk})$:

$$\begin{aligned}\mathrm{Var}\,(\Delta_{jk}) &= n^{-1}(\mathrm{E}\,K_1)^{-2}\,\mathrm{Var}\,(K_1\langle\phi_j, X_1 - x\rangle\langle\phi_k, X_1 - x\rangle)\\ &= n^{-1}(\mathrm{E}\,K_1)^{-2}\,\mathrm{E}\left\{K_1^2\gamma_{j,k}^{2,2}\left(\|X_1 - x\|^4\right)\right\} - n^{-1}(\mathrm{E}\,K_1)^{-2}\left(\mathrm{E}\left\{K_1\,\gamma_{j,k}^{1,1}\left(\|X_1 - x\|^2\right)\right\}\right)^2\\ &= O\left(h^4\{n\,\pi_x(h)\}^{-1}\right)\gamma_{j,k}^{2,2'}(0) \,+\, o\left(\left\{\gamma_{j,k}^{1,1'}(0)\right\}^2 h^4\right),\end{aligned} \tag{S.21}$$

where the last equality comes from LEMMA S.9 and COROLLARY S.1. (S.20) and (S.21) result in

$$\Delta_{jk} \,=\, b_{x,0,1}^{-1}\,b_{x,2,1}\,\gamma_{j,k}^{1,1'}(0)\,h^2\,\{1 + o_P(1)\} \,+\, O_P\left(h^2\{n\,\pi_x(h)\}^{-1/2}\right)\sqrt{\gamma_{j,k}^{2,2'}(0)},$$

which is the element-wise version of the claimed result.

LEMMA S.11. *As soon as (*H2*)–(*H5*) are fulfilled,*

$$\delta = b_{x,0,1}^{-1}\,b_{x,1,1}\,\gamma\,h\,\{1 + o_P(1)\} \,+\, O_P\left(h\,\{n\,\pi_x(h)\}^{-1/2}\right)\theta,$$

where θ is the J-dimensional vector such that $[\theta]_j := \sqrt{\gamma_j^{2'}(0)}$.

PROOF OF LEMMA S.11. This proof is shortened since it follows the same lines as the previous one. From the definition of δ_j, $\mathrm{E}\,\delta_j \,=\, (\mathrm{E}\,K_1)^{-1}\,\mathrm{E}\left\{K_1\,\gamma_j^1\left(\|X - x\|\right)\right\}$ and use of LEMMA S.9 and COROLLARY S.1 results in

$$\mathrm{E}\,\delta_j \,=\, b_{x,0,1}^{-1}\,b_{x,1,1}\,\gamma_j^{1'}(0)\,h\,\{1 + o(1)\}\,. \tag{S.22}$$

Let us now focus on the variance of the δ_j.

$$\begin{aligned}\mathrm{Var}\,(\delta_j) &= n^{-1}(\mathrm{E}\,K_1)^{-2}\,\mathrm{Var}\,(K_1\langle\phi_j, X_1 - x\rangle)\\ &= n^{-1}(\mathrm{E}\,K_1)^{-2}\,\mathrm{E}\left\{K_1^2\,\gamma_j^2\left(\|X - x\|^2\right)\right\} \,-\, n^{-1}(\mathrm{E}\,K_1)^{-2}\,\left(\mathrm{E}\left\{K_1\,\gamma_j^1\left(\|X - x\|\right)\right\}\right)^2\\ &= O\left(h^2\,\{n\,\pi_x(h)\}^{-1}\right)\gamma_j^{2'}(0) \,+\, o\left(\left\{\gamma_j^{1'}(0)\right\}^2 h^2\right),\end{aligned}$$

the last equality resulting from LEMMA S.9 and COROLLARY S.1. By combining this last equation with (S.22),

$$\delta_j \,=\, b_{x,0,1}^{-1}\,b_{x,1,1}\,\gamma_j^{1'}(0)\,h\,\{1 + o_P(1)\} \,+\, O_P\left(h\,\{n\,\pi_x(h)\}^{-1/2}\right)\sqrt{\gamma_j^{2'}(0)},$$

which is the element-wise version of the claimed result.

LEMMA S.12. *Under* H2*)–(*H5*), one has*

(i) $A_0 \,=\, b_{x,0,1}^{-1}\,b_{x,1,1}\,\alpha_{0,x,n}^{bias}\,h\,\{1 + o(1)\} \,+\, O_P\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|\,h\,\{n\,\pi_x(h)\}^{-1/2}\right)$, *where the sequence $\alpha_{0,x,n}^{bias}$ is upper bounded by $\|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|$,*

(ii) $A_j \,=\, b_{x,0,1}^{-1}\,b_{x,2,1}\,\alpha_{j,x,n}^{bias}\,h^2\,\{1 + o(1)\} \,+\, O_P\left(h^2\sqrt{\alpha_{j,x,n}^{var}}\{n\,\pi_x(h)\}^{-1/2}\right)$ *for* $j = 1, \ldots, J$, *where* $\sum_{j=1}^J \left\{\alpha_{j,x,n}^{bias}\right\}^2 \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|^2$ *and* $\sum_{j=1}^J \alpha_{j,x,n}^{var} \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m_x'\|^2$.

PROOF OF LEMMA S.12-(i). We can write

$$\begin{aligned}
\mathrm{E}\, A_0 &= (\mathrm{E}\, K_1)^{-1}\ \mathrm{E}\left(K_1 \langle \mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{j>J} \langle \phi_j, m'_x\rangle\, \mathrm{E}\left(K_1 \langle \phi_j, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{j>J} \langle \phi_j, m'_x\rangle\, \mathrm{E}\left\{K_1\, \gamma_j^1\left(\|X_1 - x\|\right)\right\} \\
&= b_{x,0,1}^{-1}\, b_{x,1,1}\, \alpha_{0,x,n}^{bias}\, h\ \{1 + o(1)\}\,,
\end{aligned}$$

where $\alpha_{0,x,n}^{bias} = \sum_{j>J} \langle \phi_j, m'_x\rangle\, \gamma_j^{1'}(0)$, the last equality coming from LEMMA S.9-(i). Moreover, the Cauchy-Schwartz inequality and LEMMA S.9-(ii) imply that $|\alpha_{0,x,n}^{bias}| \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|$. In the same way, one has

$$\begin{aligned}
\mathrm{Var}(A_0) &= n^{-1}\, (\mathrm{E}\, K_1)^{-2}\ \mathrm{Var}\left\{K_1 \sum_{j>J} \langle \phi_j, m'_x\rangle \langle \phi_j, X_1 - x\rangle\right\} \\
&\leq n^{-1}\, (\mathrm{E}\, K_1)^{-2} \sum_{j>J} \sum_{k>J} \langle \phi_j, m'_x\rangle \langle \phi_k, m'_x\rangle\, \mathrm{E}\left\{K_1^2 \gamma_{j,k}^{1,1}\left(\|X_1 - x\|^2\right)\right\} \\
&= O\left(\alpha_{0,x,n}^{var}\, \frac{h^2}{n\, \pi_x(h)}\right),
\end{aligned}$$

where $\alpha_{0,x,n}^{var} = \sum_{j>J} \sum_{k>J} \langle \phi_j, m'_x\rangle \langle \phi_k, m'_x\rangle\, \gamma_{j,k}^{1,1'}(0)$, the last equality using again LEMMA S.9-(i).

Moreover, $\alpha_{0,x,n}^{var} \leq \left\{\sum_{j>J} \sum_{k>J} \langle \phi_j, m'_x\rangle^2 \langle \phi_k, m'_x\rangle^2\right\}^{1/2} \left\{\sum_{j>J} \sum_{k>J} \left\{\gamma_{j,k}^{1,1'}(0)\right\}^2\right\}^{1/2}$, which results in $\alpha_{0,x,n}^{var} \leq \|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|^2$ (use LEMMA S.9-(ii)). Then $\mathrm{Var}(A_0) = O\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|^2\, h^2 \{n\, \pi_x(h)\}^{-1}\right)$ and the claimed result holds.

PROOF OF LEMMA S.12-(ii). We have

$$\begin{aligned}
\mathrm{E}\, A_j &= (\mathrm{E}\, K_1)^{-1}\ \mathrm{E}\left(K_1 \langle \mathcal{P}_{\mathcal{S}_J^\perp} m'_x, X_1 - x\rangle \langle \phi_j, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{k>J} \langle \phi_k, m'_x\rangle\, \mathrm{E}\left(K_1 \langle \phi_k, X_1 - x\rangle \langle \phi_j, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{k>J} \langle \phi_k, m'_x\rangle\, \mathrm{E}\left\{K_1\, \gamma_{j,k}^{1,1}\left(\|X_1 - x\|^2\right)\right\} \\
&= b_{x,0,1}^{-1}\, b_{x,2,1}\, \alpha_{j,x,n}^{bias}\, h^2\ \{1 + o(1)\}\,,
\end{aligned} \tag{S.23}$$

where $\alpha_{j,x,n}^{bias} = \sum_{k>J} \langle \phi_k, m_x' \rangle \, \gamma_{j,k}^{1,1'}(0)$, the last equality coming from LEMMA S.9-(i). In addition,

$$\begin{aligned}
\sum_{j=1}^{J} \left\{ \alpha_{j,x,n}^{bias} \right\}^2 &= \sum_{j=1}^{J} \sum_{k,\ell>J} \langle \phi_k, m_x' \rangle \langle \phi_\ell, m_x' \rangle \, \gamma_{j,k}^{1,1'}(0) \, \gamma_{j,\ell}^{1,1'}(0) \\
&\leq \left\{ \sum_{k,\ell>J} \langle \phi_k, m_x' \rangle^2 \langle \phi_\ell, m_x' \rangle^2 \right\}^{1/2} \left\{ \sum_{k,\ell>J} \left(\sum_{j=1}^{J} \gamma_{j,k}^{1,1'}(0) \, \gamma_{j,\ell}^{1,1'}(0) \right)^2 \right\}^{1/2} \\
&\leq \| \mathcal{P}_{\mathcal{S}_J^\perp} m_x' \|^2 \left\{ \left(\sum_{j,k\geq 1} \left[\gamma_{j,k}^{1,1'}(0) \right]^2 \right) \left(\sum_{j,\ell \geq 1} \left[\gamma_{j,\ell}^{1,1'}(0) \right]^2 \right) \right\}^{1/2} \\
&\leq \| \mathcal{P}_{\mathcal{S}_J^\perp} m_x' \|^2 .
\end{aligned}$$

To derive the asymptotic behaviour of the variance of A_j, we follow similar arguments:

$$\begin{aligned}
\operatorname{Var}(A_j) &= n^{-1} \, (\mathrm{E}\, K_1)^{-2} \, \operatorname{Var} \left\{ K_1 \left(\sum_{k>J} \langle \phi_k, m_x' \rangle \langle \phi_k, X_1 - x \rangle \right) \langle \phi_j, X_1 - x \rangle \right\} \\
&\leq n^{-1} \, (\mathrm{E}\, K_1)^{-2} \sum_{k>J} \sum_{\ell>J} \langle \phi_k, m_x' \rangle \langle \phi_\ell, m_x' \rangle \, \mathrm{E} \left\{ K_1^2 \gamma_{j,k,\ell}^{2,1,1} \left(\|X_1 - x\|^4 \right) \right\} \qquad \text{(S.24)} \\
&= O\left(\alpha_{j,x,n}^{var} \, h^4 \, \{ n \, \pi_x(h) \}^{-1} \right),
\end{aligned}$$

where $\alpha_{j,x,n}^{var} = \sum_{k>J} \sum_{\ell>J} \langle \phi_k, m_x' \rangle \langle \phi_\ell, m_x' \rangle \, \gamma_{x,j,k,\ell}^{2,1,1\,'}(0)$. Now, by involving arguments similar to those used in LEMMA S.9-(ii) one is able to show $\sum_{j=1}^{J} \alpha_{j,x,n}^{var} \leq \| \mathcal{P}_{\mathcal{S}_J^\perp} m_x' \|^2$. Just combine (S.23) and (S.24) to get the claimed result.

LEMMA S.13. *For any* $u, v \in \mathbb{R}^J$, $\left| u^{\mathsf{T}} \, \Gamma^{-1} \, v \right| \leq \lambda_J^{-1} \, \|u\|_2 \, \|v\|_2$, *where* λ_J *is the smallest eigenvalue of* Γ.

PROOF OF LEMMA S.13. Let us remark that $\lambda_J > 0$ according to (H3). This result involves the Cauchy-Schwartz inequality and the Rayleigh quotient of the inverse of the $J \times J$ matrix Γ:

$$\begin{aligned}
\left| \mathrm{u}^{\mathsf{T}} \, \Gamma^{-1} \, \mathrm{v} \right| &\leq \|\mathrm{u}\|_{\Gamma^{-1}} \, \|\mathrm{v}\|_{\Gamma^{-1}} \\
&\leq \|\mathrm{u}\|_2 \, \|\mathrm{v}\|_2 \left(\frac{\mathrm{u}^{\mathsf{T}} \, \Gamma^{-1} \, \mathrm{u}}{\mathrm{u}^{\mathsf{T}} \mathrm{u}} \right)^{1/2} \left(\frac{\mathrm{v}^{\mathsf{T}} \, \Gamma^{-1} \, \mathrm{v}}{\mathrm{v}^{\mathsf{T}} \mathrm{v}} \right)^{1/2} \\
&\leq \|\mathrm{u}\|_2 \, \|\mathrm{v}\|_2 \, R_1^{1/2} \, R_2^{1/2},
\end{aligned}$$

where R_1 and R_2 are the Rayleigh quotients of Γ^{-1}. Let λ_J be the smallest eigenvalue of Γ. Then λ_J^{-1} is the greatest eigenvalue of Γ^{-1} with $R_1 \leq \lambda_J^{-1}$ and $R_2 \leq \lambda_J^{-1}$, and the claimed result holds.

LEMMA S.14. *Under* (H2)*–*(H5)*, one has*

(i) $B_0 = b_{x,0,1}^{-1} b_{x,2,1} \beta_{0,x}^{bias} \, h^2 \, \{1 + o(1)\} + O_P\left(h^2 \, \{ n \, \pi_x(h) \}^{-1/2} \right)$ *with* $\beta_{0,x}^{bias} = O(1)$,

(ii) $B_j = b_{x,0,1}^{-1}\, b_{x,3,1}\, \beta_{j,x}^{bias}\, h^3\, \{1+o(1)\} + \sqrt{\beta_{j,x}^{var}}\, O_P\left(h^3\, \{n\, \pi_x(h)\}^{-1/2}\right)$ *for all* $j = 1,\ldots,J$, *where* $\sum_{j=1}^J \left\{\beta_{j,x}^{bias}\right\}^2 = O(1)$ *and* $\sum_{j=1}^J \beta_{j,x}^{var} = O(1)$.

PROOF OF LEMMA S.14-(i). A standard expansion of the linear operator m_x'' results in

$$\begin{aligned}
\mathrm{E}\, B_0 &= (\mathrm{E}\, K_1)^{-1} \sum_{j\geq 1}\sum_{k\geq 1} \langle m_x''\phi_j, \phi_k\rangle\, \mathrm{E}\left(K_1 \langle \phi_j, X_1 - x\rangle \langle \phi_k, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{j\geq 1}\sum_{k\geq 1} \langle m_x''\phi_j, \phi_k\rangle\, \mathrm{E}\left\{K_1\, \gamma_{j,k}^{1,1}\left(\|X_1 - x\|^2\right)\right\} \\
&= b_{x,0,1}^{-1}\, b_{x,2,1}\, \beta_{0,x}^{bias}\, h^2\, \{1+o(1)\},
\end{aligned}$$

where $\beta_{0,x}^{bias} := \sum_{j\geq 1}\sum_{k\geq 1}\langle m_x''\phi_j, \phi_k\rangle\, {\gamma_{j,k}^{1,1}}'(0)$, the last equality by LEMMA S.9-(i). Moreover, the Cauchy-Schwartz inequality combined with (H1) and LEMMA S.9-(ii) imply that $\beta_{0,x}^{bias} = O(1)$. In the same way, one has

$$\begin{aligned}
\mathrm{Var}(B_0) &= n^{-1}\, (\mathrm{E}\, K_1)^{-2}\, \mathrm{Var}\left\{K_1 \langle m_x''(X_1 - x), X_1 - x\rangle\right\} \\
&\leq n^{-1}\, (\mathrm{E}\, K_1)^{-2} \sum_{j_1,\ldots,j_4\geq 1} \langle m_x''\phi_{j_1}, \phi_{j_2}\rangle \langle m_x''\phi_{j_3}, \phi_{j_4}\rangle\, \mathrm{E}\left\{K_1^2 \gamma_{j_1,\ldots,j_4}^{1,1,1,1}\left(\|X_1 - x\|^4\right)\right\} \\
&= O\left(\beta_{0,x}^{var}\, h^4\, \{n\, \pi_x(h)\}^{-1}\right),
\end{aligned}$$

where $\beta_{0,x}^{var} := \sum_{j_1,\ldots,j_4\geq 1} \langle m_x''\phi_{j_1}, \phi_{j_2}\rangle \langle m_x''\phi_{j_3}, \phi_{j_4}\rangle\, {\gamma_{j_1,\ldots,j_4}^{1,1,1,1}}'(0)$, the last equality using LEMMA S.9-(i). Moreover,

$$\beta_{0,x}^{var} \leq \left\{\sum_{j_1,\ldots,j_4\geq 1} \langle m_x''\phi_{j_1}, \phi_{j_2}\rangle^2 \langle m_x''\phi_{j_3}, \phi_{j_4}\rangle^2\right\}^{1/2} \left\{\sum_{j_1,\ldots,j_4\geq 1} \left\{{\gamma_{j_1,\ldots,j_4}^{1,1,1,1}}'(0)\right\}^2\right\}^{1/2},$$

which gives that $\beta_{0,x}^{var}$ is a finite quantity (use again (H1) and LEMMA S.9-(ii)). Then $\mathrm{Var}(B_0) = O\left(h^4 \{n\, \pi_x(h)\}^{-1}\right)$ and the claimed result holds.

PROOF OF LEMMA S.14-(ii). Similarly,

$$\begin{aligned}
\mathrm{E}\, B_j &= (\mathrm{E}\, K_1)^{-1} \sum_{k\geq 1}\sum_{\ell\geq 1} \langle m_x''\phi_k, \phi_\ell\rangle\, \mathrm{E}\left(K_1 \langle \phi_j, X_1 - x\rangle \langle \phi_k, X_1 - x\rangle \langle \phi_\ell, X_1 - x\rangle\right) \\
&= (\mathrm{E}\, K_1)^{-1} \sum_{k\geq 1}\sum_{\ell\geq 1} \langle m_x''\phi_k, \phi_\ell\rangle\, \mathrm{E}\left\{K_1\, \gamma_{j,k,\ell}^{1,1,1}\left(\|X_1 - x\|^3\right)\right\} \\
&= b_{x,0,1}^{-1}\, b_{x,3,1}\, \beta_{j,x}^{bias}\, h^3\, \{1+o(1)\},
\end{aligned}$$

where $\beta_{j,x}^{bias} := \sum_{k\geq 1}\sum_{\ell\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\,\gamma_{j,k,\ell}^{1,1,1'}(0)$. Moreover,

$$\begin{aligned}\sum_{j=1}^{J}\left\{\beta_{j,x}^{bias}\right\}^2 &= \sum_{j=1}^{J}\sum_{k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\langle m_x''\phi_p,\phi_q\rangle\gamma_{j,k,\ell}^{1,1,1'}(0)\,\gamma_{j,p,q}^{1,1,1'}(0)\\
&\leq \left\{\sum_{k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle^2\langle m_x''\phi_p,\phi_q\rangle^2\right\}^{1/2}\times\left\{\sum_{k,\ell,p,q\geq 1}\left(\sum_{j=1}^{J}\gamma_{j,k,\ell}^{1,1,1'}(0)\,\gamma_{j,p,q}^{1,1,1'}(0)\right)^2\right\}^{1/2}\\
&\leq \|m_x''\|_{HS}^2\left\{\left(\sum_{j,k,\ell\geq 1}\left[\gamma_{j,k,\ell}^{1,1,1'}(0)\right]^2\right)\left(\sum_{j,p,q\geq 1}\left[\gamma_{j,p,q}^{1,1,1'}(0)\right]^2\right)\right\}^{1/2}\\
&\leq \|m_x''\|_{HS}^2,\end{aligned}$$

where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm of an operator. So, $\sum_{j=1}^{J}\left\{\beta_{j,x}^{bias}\right\}^2 = O(1)$. Let us now focus on the variance of B_j

$$\begin{aligned}\mathrm{Var}(B_j) &= n^{-1}\left(\mathrm{E}\,K_1\right)^{-2}\,\mathrm{Var}\left\{K_1\langle m_x''(X_1-x),X_1-x\rangle\langle\phi_j,X_1-x\rangle\right\}\\
&\leq n^{-1}\left(\mathrm{E}\,K_1\right)^{-2}\sum_{k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\langle m_x''\phi_p,\phi_q\rangle\,\mathrm{E}\left\{K_1^2\gamma_{j,k,\ell,p,q}^{2,1,1,1,1}\left(\|X_1-x\|^6\right)\right\}\\
&= O\left(\beta_{j,x}^{var}\,\frac{h^6}{n\,\pi_x(h)}\right),\end{aligned}$$

where $\beta_{j,x}^{var} := \sum_{k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\langle m_x''\phi_p,\phi_q\rangle\,\gamma_{j,k,\ell,p,q}^{2,1,1,1,1'}(0)$, the last equality using again LEMMA S.9-(i). With similar arguments as those involved to show the second assertion of LEMMA S.9-(ii),

$$\begin{aligned}\sum_{j=1}^{J}\beta_{j,x}^{var} &\leq \sum_{j,k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\langle m_x''\phi_p,\phi_q\rangle\left\{\lim_{t\to 0}t^{-1}\,\gamma_{j,k,\ell,p,q}^{2,1,1,1,1}(t)\right\}\\
&= \sum_{k,\ell,p,q\geq 1}\langle m_x''\phi_k,\phi_\ell\rangle\langle m_x''\phi_p,\phi_q\rangle\left\{\lim_{t\to 0}t^{-2/3}\,\gamma_{k,\ell,p,q}^{1,1,1,1}(t)\right\}\\
&\leq \|m_x''\|_{HS}^2\lim_{t\to 0}t^{-2/3}\left\{\sum_{k,\ell,p,q\geq 1}\left[\gamma_{k,\ell,p,q}^{1,1,1,1}(t)\right]^2\right\}^{1/2}.\end{aligned}$$

Consequently,

$$\sum_{j=1}^{J}\beta_{j,x}^{var} \leq \|m_x''\|_{HS}^2\lim_{t\to 0}t^{-2/3}\left\{\sum_{k,\ell,p,q\geq 1}\gamma_{k,\ell,p,q}^{2,2,2,2}\left(t^{4/3}\right)\right\}^{1/2} \leq \|m_x''\|_{HS}^2,$$

where, of course, $\gamma_{k,\ell,p,q}^{2,2,2,2}\left(t^{4/3}\right)$ is the same as

$$\mathrm{E}\left(\langle\phi_k,X_1-x\rangle^2\langle\phi_\ell,X_1-x\rangle^2\langle\phi_p,X_1-x\rangle^2\langle\phi_q,X_1-x\rangle^2|\|X_1-x\|^6=t\right).$$

S.2·6. *Examples of approximating bases*

The functional local linear estimator depends on the basis $\phi_1, \phi_2, \ldots$ of the space H of square-integrable real functions on $[0,1]$. In this section, we specify the asymptotic behaviour of our estimator when considering particular bases. We start with usual approximating function spaces (cases 1 and 2 below). A more challenging issue consists in replacing the deterministic basis $\phi_1, \phi_2, \ldots$ with a data-driven one. This important question is investigated in case 3.

Case 1: Orthogonal B-spline basis. B-spline basis is a well-known and useful tool for approximating smooth functions. Consider the set of functions that are polynomials of degree q on each interval $[(t-1)/k,\, t/k]$ for $t = 1, \ldots, k$ and are $(q-1)$ times continuously differentiable on $[0,\, 1]$. This $(k+q)$-dimensional subspace of H defines the well-known space of splines. One can derive an orthogonal basis of B-splines of that space $\{B_{k,1}, \ldots, B_{k,k+q}\}$ (see de Boor, 1978, for an overview on spline functions and Redd, 2012, for the orthogonalization of B-spline basis functions). In this situation, one sets $J = k + q$ and for any $j \leq J$, $\phi_j = B_{k,j}/\|B_{k,j}\|$; $k = k_n$ is a sequence that grows to infinity with n. Now, let the Riesz representation of the functional m'_x be smooth enough so that its pth derivative is a Hölder function:

(H8) $|m_x'^{(p)}(u) - m_x'^{(p)}(v)| \leq C|u - v|^\nu$ with $\nu \in [0, 1]$.

From Theorem XII.1 in de Boor (1978) and (H8), $\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\| \leq C\,(J - q)^{-(p+\nu)}$. Because $\|P_{\mathcal{S}_J} m'_x - m'_x\| = \|P_{\mathcal{S}_J^\perp} m'_x\|$, the rate of convergence of the conditional bias of the local linear estimator becomes

$$\mathrm{E}_{\mathrm{X}}\{\hat{m}(x)\} \;=\; m(x) \;+\; O_P(h\, J^{-p-\nu}) \;+\; O_P(h^2). \tag{S.25}$$

Case 2: Fourier basis. When one suspects some periodic features for the functional m'_x, it can be advantageous to expand the functions by means of the Fourier basis $\phi_1(t) = 1$, $\phi_{2j}(t) = \sqrt{2}\sin(2\pi j t)$, and $\phi_{2j+1}(t) = \sqrt{2}\cos(2\pi j t)$ for $j = 1, 2, \ldots$ If one assumes that

(H9) m'_x is a periodic function,

then, according to Zygmund (2002) and assumptions (H8) and (H9), one has $\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\| = O\left(J^{-p-\nu}\right)$ which leads to the same rate of convergence (S.25).

Case 3: Data driven basis. The functional principal components analysis (FPCA) allows to expand a random function X into a basis $\phi_1, \phi_2, \ldots$ in an optimal way (see Karhunen, 1946; Loève, 1946; Rao, 1958; Dauxois et al., 1982 for precursor works and Bosq, 2000; Yao et al., 2005; Hall & Hosseini-Nasab, 2006; Hall et al., 2006 for more recent statistical developments). In this setting, functions ϕ_j are the eigenfunctions of the covariance operator of X and the eigenanalysis of the empirical covariance operator provides a data driven basis $\hat{\phi}_1, \hat{\phi}_2, \ldots$ (by convention, we assume that $\langle \phi_j, \hat{\phi}_j \rangle > 0$). We propose to investigate the asymptotic properties of the functional local linear estimator when replacing $\phi_1, \ldots, \phi_J$ with the data driven basis $\hat{\phi}_1, \ldots, \hat{\phi}_J$. This results in the new estimator $\hat{\hat{m}}(x) \;:=\; \mathrm{e}^\mathsf{T}\left(\hat{\Phi}^\mathsf{T}\mathrm{K}\hat{\Phi}\right)^{-1}\hat{\Phi}^\mathsf{T}\mathrm{K}\mathrm{Y}$ where, for $i = 1, \ldots, n$, $[\hat{\Phi}]_{i1} = 1$ and for $j = 2, \ldots, J$, $[\hat{\Phi}]_{ij} = \langle \hat{\phi}_j,\, X_i - x \rangle$. Thanks to Bosq (2000) and Cardot et al. (1999), as soon as $\mathrm{E}\,\|X\|^4 < \infty$, one has for any $j = 1, 2, \ldots$ that $\|\hat{\phi}_j - \phi_j\| = O_P(a_J^{-1} n^{-1/2})$ with $a_J := \min_{j \leq J}\{\rho_j - \rho_{j+1}, \rho_{j-1} - \rho_j\}$. Here, ρ_j are the eigenvalues of the covariance operator of X (placed in descending order).

THEOREM S.1. *If* $\mathrm{E}\,\|X\|^4 < \infty$, $a_J^{-1} n^{-1/2} = o(1)$ *and* (H1)*–*(H6) *hold,*

(i) $\mathrm{E}_{\mathrm{X}}\left\{\hat{\hat{m}}(x)\right\} = m(x) + O_P\left(J^{1/2}\,\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,h\right) + O_P\left(J^{1/2}\,h^2\right)$
$\qquad + O_P\left(a_J^{-1}\,n^{-1/2}\,J^{1/2}\,h\right),$

(ii) $\mathrm{Var}_{\mathrm{X}}\left\{\hat{\hat{m}}(x)\right\} \;=\; O_P\left(\{n\,\pi_x(h)\}^{-1}\right).$

Consideration of the data-driven basis degrades slightly the conditional bias by introducing in both original terms the quantity $J^{1/2}$ and by adding a third term $O_P\left(a_J^{-1}\,n^{-1/2}\,J^{1/2}\,h\right)$. However, the conditional variance is not sensitive to the introduction of the data-driven basis.

What about kNN estimators? Thanks to the theoretical developments involved in the proof of THEOREM 3, it is easy to derive the rates of convergence of the kNN estimator in each of the three cases (orthogonal B-spline basis, Fourier basis, and data driven basis) by replacing the global bandwidth h with $\pi_x^{-1}(k/n)$ and the quantity $n\,\pi_x(h)$ with k. In cases 1 and 2, $\mathrm{E}_{\mathrm{X}}\left\{\hat{m}_{kNN}(x)\right\} = m(x) + O_P\left\{\pi_x^{-1}(k/n)\,J^{-p-\nu}\right\} + O_P\left\{\pi_x^{-1}(k/n)^2\right\}$. For case 3, let $\hat{\hat{m}}_{kNN}(x) := \mathrm{e}^{\mathsf{T}}\left(\hat{\Phi}^{\mathsf{T}}\mathrm{K}_{kNN}\hat{\Phi}\right)^{-1}\hat{\Phi}^{\mathsf{T}}\mathrm{K}_{kNN}\mathrm{Y}$ be the kNN version of $\hat{\hat{m}}(x)$; $\mathrm{E}_{\mathrm{X}}\left\{\hat{\hat{m}}_{kNN}(x)\right\} = m(x) + O_P\left\{J^{1/2}\,\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,\pi_x^{-1}(k/n)\right\} + O_P\left\{J^{1/2}\,\pi_x^{-1}(k/n)^2\right\}$ $+ O_P\left\{a_J^{-1}\,n^{-1/2}\,J^{1/2}\,\pi_x^{-1}(k/n)\right\}$ and $\mathrm{Var}_{\mathrm{X}}\left\{\hat{\hat{m}}_{kNN}(x)\right\} \;=\; O_P\left\{k^{-1}\right\}$.

PROOF OF THEOREM S.1 Let us start with the conditional bias. Similarly to (S.3), $\mathrm{E}_{\mathrm{X}}\left\{\hat{\hat{m}}(x)\right\} = m(x) + \hat{T}_1 + \hat{T}_2 + \hat{T}_3$ where $\hat{T}_1$ (resp. $\hat{T}_2$) corresponds to T_1 (resp. T_2) when replacing Φ with $\hat{\Phi}$, and $\hat{T}_3 := \mathrm{e}^{\mathsf{T}}\left(\hat{\Phi}^{\mathsf{T}}\mathrm{K}\hat{\Phi}\right)^{-1}\hat{\Phi}^{\mathsf{T}}\mathrm{K}(\Phi - \hat{\Phi})\left[m(x)|\nabla m_x{}^{\mathsf{T}}\right]^{\mathsf{T}}$.

About $\hat{T}_1$. Similarly to the proof of LEMMA S.1-(i), $\hat{T}_1 = \hat{\mu}\left(A_0 - \hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}[\hat{A}_1 \ldots \hat{A}_J]^{\mathsf{T}}\right)$ where $\hat{\delta}$ (resp. $\hat{\Delta}$, $\hat{A}_1,\ldots,\hat{A}_J$) are defined as δ (resp. Δ, $A_1,\ldots,A_J$) but $\phi_1,\ldots,\phi_J$ are replaced with $\hat{\phi}_1,\ldots,\hat{\phi}_J$ and where $\hat{\mu} := \left(\delta_0 - \hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}\hat{\delta}\right)^{-1}$. Because $A_0 = O_P\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,h\right)$ and $\hat{A}_j = O_P\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,h^2\right)$ for $j = 1,\ldots,J$,

$$\hat{T}_1 = \hat{\mu}\left(1 - h\,\hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}\mathbf{1}\right)O_P\left(\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,h\right). \tag{S.26}$$

According to the definitions of $\hat{\delta}$ and $\hat{\Delta}$, it is easy to see that $\hat{\delta} = \delta\,(1 + U_n)$ and $\hat{\Delta}^{-1} = \Delta^{-1}\,(1 + U_n)$ with $U_n := O_P\left(a_J^{-1}\,n^{-1/2}\right)$ being the rate of convergence of $\max_{j\le J}\|\phi_j - \hat{\phi}_j\|$. On one side, $\hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}\hat{\delta} = \delta^{\mathsf{T}}\Delta^{-1}\delta\,(1 + U_n)$, and on the other side, $\hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}\mathbf{1} = \delta^{\mathsf{T}}\Delta^{-1}\mathbf{1}\,(1 + U_n)$. LEMMAS S.10, S.11 and S.13 give $\hat{\delta}^{\mathsf{T}}\hat{\Delta}^{-1}\mathbf{1} = O_P\left(h^{-1}\,\lambda_J^{-1}\,J^{1/2}\right)$ and $\delta^{\mathsf{T}}\Delta^{-1}\delta = O_P\left(\lambda_J^{-1}\right)$ thanks to (S.8). Finally, $\hat{\mu} = O_P\left(\lambda_J\right)$, and with (S.26)

$$\hat{T}_1 \;=\; O_P\left(J^{1/2}\,\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\|\,h\right). \tag{S.27}$$

About $\hat{T}_2$. Following the guidelines in the proof of LEMMA S.1-(ii), $\hat{T}_2 = \hat{T}_{21} + \hat{T}_{22}$ where $\hat{T}_{21} := \mathrm{e}^{\mathsf{T}}\left(\hat{\Phi}^{\mathsf{T}}\widetilde{\mathrm{K}}\hat{\Phi}\right)^{-1}\left[B_0, \hat{B}_1, \ldots, \hat{B}_J\right]^{\mathsf{T}}$ and $\hat{T}_{22} := \mathrm{e}^{\mathsf{T}}\left(\hat{\Phi}^{\mathsf{T}}\widetilde{\mathrm{K}}\hat{\Phi}\right)^{-1}\left[C_0, \hat{C}_1, \ldots, \hat{C}_J\right]^{\mathsf{T}}$ with $\hat{B}_1,\ldots,\hat{B}_J$ (resp. $\hat{C}_1,\ldots,\hat{C}_J$) defined as $B_1,\ldots,B_J$ (resp. $C_1,\ldots,C_J$) but $\phi_1,\ldots,\phi_J$ are replaced with $\hat{\phi}_1,\ldots,\hat{\phi}_J$. One has $B_0 = O_P(h^2)$, $\hat{B}_j = O_P(h^3)$ for $j = 1,\ldots,J$, and with (H1),

$C_0 = O_P(h^3)$ and $\hat{B}_j = O_P(h^4)$ for $j = 1, \ldots, J$. Consequently, $\hat{T}_{22}$ is negligible with respect to $\hat{T}_{21}$. Similarly to $\hat{T}_1$, $\hat{T}_{21} = \hat{\mu}\left(1 - h\,\hat{\delta}^\mathsf{T}\hat{\Delta}^{-1}\mathbf{1}\right) O_P\left(h^2\right)$ so that

$$\hat{T}_2 \;=\; O_P\left(J^{1/2}\,h^2\right). \tag{S.28}$$

About $\hat{T}_3$. By the definition of $\hat{T}_3$ and since $\max_{j\leq J}\|\phi_j - \hat{\phi}_j\| = O_P\left(a_J^{-1}\,n^{-1/2}\right)$,

$$\begin{aligned}\hat{T}_3 &= \mathrm{e}^\mathsf{T}\left(\hat{\Phi}^\mathsf{T}\widetilde{\mathrm{K}}\hat{\Phi}\right)^{-1}\hat{\Phi}^\mathsf{T}\widetilde{\mathrm{K}}\left[\delta_0\,|\,\hat{\delta}^\mathsf{T}\right]^\mathsf{T} O_P\left(a_J^{-1}\,n^{-1/2}\,J^{1/2}h\right)\\ &= \hat{\mu}\left(\delta_0 - \hat{\delta}^\mathsf{T}\hat{\Delta}^{-1}\hat{\delta}\right) O_P\left(a_J^{-1}\,n^{-1/2}\,J^{1/2}h\right).\end{aligned}$$

Since $\hat{\mu}\left(\delta_0 - \hat{\delta}^\mathsf{T}\hat{\Delta}^{-1}\hat{\delta}\right) = 1$, we have that

$$\hat{T}_3 \;=\; O_P\left(a_J^{-1}\,n^{-1/2}\,J^{1/2}\,h^2\right). \tag{S.29}$$

Now it is enough to combine the decomposition of $\mathrm{E}_\mathrm{X}\left\{\hat{\hat{m}}(x)\right\}$ with (S.27), (S.28), and (S.29) to get the claimed conditional bias.

Let us now focus on the conditional variance. Similarly to LEMMA S.2-(i),

$$\mathrm{Var}_\mathrm{X}\left\{\hat{\hat{m}}(x)\right\} = \{\sigma^2(x) \,+\, o(1)\}\,\hat{\mu}^2\left(\widetilde{\delta}_0 - \hat{\delta}^\mathsf{T}\hat{\Delta}^{-1}\hat{\widetilde{\delta}} - \hat{\widetilde{\delta}}^\mathsf{T}\hat{\Delta}^{-1}\hat{\delta} + \hat{\delta}^\mathsf{T}\hat{\Delta}^{-1}\hat{\widetilde{\Delta}}\hat{\Delta}^{-1}\hat{\delta}\right)$$

where $\hat{\widetilde{\delta}}$ (resp. $\hat{\widetilde{\Delta}}$) is defined as $\widetilde{\delta}$ (resp. $\widetilde{\Delta}$) but $\phi_1, \ldots, \phi_J$ are replaced with $\hat{\phi}_1, \ldots, \hat{\phi}_J$. From the definition of $\hat{\widetilde{\delta}}$ and $\hat{\widetilde{\Delta}}$, it is easy to state that $\hat{\widetilde{\delta}} = \widetilde{\delta}\,(1 + U_n)$ and $\hat{\widetilde{\Delta}} = \widetilde{\Delta}\,(1 + U_n)$ which results in

$$\begin{aligned}\mathrm{Var}_\mathrm{X}\left\{\hat{\hat{m}}(x)\right\} &= \mathrm{Var}_\mathrm{X}\left\{\hat{m}(x)\right\}(1 + U_n)\\ &\quad + \{\sigma^2(x) \,+\, o(1)\}\,\hat{\mu}^2\left(-\delta^\mathsf{T}\Delta^{-1}\widetilde{\delta} - \widetilde{\delta}^\mathsf{T}\Delta^{-1}\delta + \delta^\mathsf{T}\Delta^{-1}\widetilde{\Delta}\Delta^{-1}\delta\right) U_n\\ &= O_P\left(\{n\,\pi_x(h)\}^{-1}\right) + O_P\left(a_J^{-1}\,n^{-1/2}\,\lambda_J\{n\,\pi_x(h)\}^{-1}\right),\end{aligned}$$

the last equality coming from LEMMA S.2-(ii) and the fact that $\hat{\mu} = O_P(\lambda_J)$. Since $a_J^{-1}\,n^{-1/2}$ is equal to $o(1)$, the claimed expression for the conditional variance holds.

S.2·7. *Curse of dimensionality issue*

Case 1: X $\mathcal{S}_D$-valued random function. Let X be a random function valued in some D-dimensional subspace (i.e. $D = J$ does not depend on n).

COROLLARY S.2. *Let X be an $\mathcal{S}_D$-valued random function with $D \geq 1$ a fixed integer (i.e. $D = J$ does not depend on n). If the probability density function of $[\langle X,\,\phi_1\rangle, \cdots, \langle X,\,\phi_D\rangle]^\mathsf{T}$ exists and is continuous in a neighbourhood of $[\langle x,\,\phi_1\rangle, \cdots, \langle x,\,\phi_D\rangle]^\mathsf{T}$, then $\hat{m}(x) - m(x) = O_P\left(n^{-2/(D+4)}\right)$, and $\|\hat{m}'_x - m'_x\| = O_P\left(n^{-1/(D+4)}\right)$.*

Proof of Corollary S.2. If X is an $\mathcal{S}_D$-valued random function with D fixed (i.e. $D = J$ does not depend on n), then the smallest eigenvalue λ_D of the $D \times D$ matrix Γ is a non-zero positive constant (i.e. does not depend on n). The regression operator m maps $\mathcal{S}_D$ into $\mathbb{R}$, and m'_x belongs to $\mathcal{S}_D$ so that $\|\mathcal{P}_{\mathcal{S}_D^\perp} m'_x\| = 0$. For any $u \in H$, set $u_D := [\langle X,\,\phi_1\rangle, \cdots, \langle X,\,\phi_D\rangle]^\mathsf{T}$;

because the probability density function f of X_D exists and is continuous in a neighbourhood of x_D, and since $\mathcal{S}_D$ is isomorphic to $\mathbb{R}^D$, $\pi_x(h) = h^D\, f(x_D)\, V_D\, \{1 + o(1)\}$ where V_D is the volume of the unit ball in $\mathbb{R}^D$. Conditionally on $X_1, \ldots, X_n$, THEOREM 1 results in $\hat{m}(x) - m(x) = O_P\left(h^2\right) + O_P\left(1/\sqrt{n\, h^D}\right)$, and THEOREM 2 leads to $\|\hat{m}'_x - m'_x\| = O_P\left(h\right) + O_P\left(1/\sqrt{n\, h^{D+2}}\right)$. When balancing the leading terms, the claimed rates of convergence hold.

Case 2: X MIPS-valued random function. Let us now consider the recently introduced concept of the mixture inner product space (MIPS) and the notion of a MIPS-valued random variable (Lin et al., 2018). For $k = 1, 2, \ldots$, let $S_k := \{\sum_{j=1}^k a_j\, \phi_j \colon a_k \neq 0\}$ be a k-dimensional subset of H and let $S := \cup_{k=1}^\infty S_k$ be an infinite-dimensional linear subspace of H. Lin et al. (2018) proved that S is dense in H, and, if one sets $w_k := P(X \in S_k)$ and $\forall u \in S$, $f_k(u) := f(\langle u, \phi_1\rangle, \ldots, \langle u, \phi_k\rangle | X \in S_k)$ is the conditional density of $[\langle u, \phi_1\rangle, \ldots, \langle u, \phi_k\rangle]^\mathsf{T}$ given $X \in S_k$, then the mapping $f(u) := \sum_{j=1}^\infty w_j\, f_j(u)\, 1_{S_j}(u)$ is a well defined mixture density on S as soon as, for all $k = 1, 2, \ldots$, the conditional densities f_k exist.

COROLLARY S.3. *Set $D := \min\{k \colon w_k > 0\}$. If X is an S-valued random function, the conditional density f_D is continuous in a neighbourhood of x, m'_x is a Lipschitz function, $\phi_1, \phi_2, \ldots$ is an orthonormal B-spline basis (for more details, see Case 1 in Section S.2·6), then $\hat{m}(x) - m(x) = O_P\left(n^{-2/(D+4)}\right)$, and if λ_J is asymptotically equivalent to $J^{-\mu}$ with $1/2 < \mu < 1$, $\|\hat{m}'_x - m'_x\| = O_P\left(n^{(-1+\mu)/(D+5-\mu)}\right)$.*

Proof of Corollary S.3. Thanks to Case 1 of Section S.2·6, if m'_x is a Lipschitz function and $\phi_1, \phi_2, \ldots$ an orthonormal B-spline basis, then $\|\mathcal{P}_{\mathcal{S}_J^\perp} m'_x\| = O(J^{-1})$. With the definition of the mixture density, $\pi_x(h) = \sum_{k=1}^\infty w_k \int_{B(x,h)\cap S_k} f_k(u)\ \mathrm{d}\, Leb_k(u)$, where, for $k = 1, 2, \ldots$, by Leb_k we mean the Lebesgue measure on the k-dimensional subspace S_k. Then, $\pi_x(h) \geq w_D \int_{B(x,h)\cap S_D} f_D(u)\ \mathrm{d}\, Leb_D(u)$ where $D := \inf\{k \colon w_k > 0\}$, and the continuity of f_D in a neighbourhood of x results in $\int_{B(x,h)\cap S_D} f_D(u)\, dLeb_D(u) = h^D f_D(x) V_D\{1 + o(1)\}$ so that $\pi_x(h)^{-1} = O(h^{-D})$. Set $h \sim J^{-1}$ to equalize both bias terms in THEOREMS 1 and 2; just balance the bias and variance terms in THEOREM 1 to get the claimed rate of convergence for $\hat{m}(x)$ (it does not change when comparing with the finite-dimensional situation). Given $\lambda_J \sim J^{-\mu}$ with $1/2 < \mu < 1$, $\lambda_J^{-1} < \sqrt{J/\lambda_J}$ and $h^{-1}\{\lambda_J\, n\, \pi_x(h)\}^{-1/2}\sqrt{J}$, the term corresponding to variance is the leading term in the asymptotic behaviour of $\hat{m}'_x$ and THEOREM 2 leads to $\|\hat{m}'_x - m'_x\| = O_P\left(h^{1-\mu}\right) + O_P\left((n\, h^{D+3+\mu})^{-1/2}\right)$. Again, when balancing the bias and variance terms (in THEOREM 2), one gets the claimed rate of convergence for $\hat{m}'_x$.

These rates of convergence are obtained when λ_J is assumed asymptotically equivalent to $J^{-\mu}$ with $1/2 < \mu < 1$. First, LEMMA S.9 and the comment just after result in $\lambda_J = O(J^{-\mu})$ with $\mu > 1/2$. Second, the discussion on the original hypothesis (H3) indicates that $\lambda_J > (J+1)^{-1}$ in a quite general situation (see (S.33)). Therefore, the assumption on λ_J seems reasonable.

S.2·8. *Theoretical complement on (H3)*

In this section, we investigate hypothesis (H3) that requires regularity of the family of functions

$$\gamma_{j_1,\ldots,j_M}^{p_1,\ldots,p_M}(t) = \mathrm{E}\left(\langle \phi_{j_1}, X_1 - x\rangle^{p_1} \cdots \langle \phi_{j_M}, X_1 - x\rangle^{p_M} | \|X_1 - x\|^{p_1 + \cdots + p_M} = t\right).$$

The next lemma provides a general condition on the functional predictor X in order to fulfil (H3).

LEMMA S.15. *Suppose that for some $x \in H$ the random vector*

$$\left[\langle \phi_1, X_1 - x\rangle, \dots, \langle \phi_J, X_1 - x\rangle, \|X_1 - x\| - \sqrt{\sum_{i=1}^{J} \langle \phi_i, X_1 - x\rangle^2}\right]^{\mathsf{T}}$$

is absolutely continuous with a density in $\mathbb{R}^{J+1}$ that is positive at the origin, continuous at the origin in its first J coordinates, and continuous at the origin from the right in its last coordinate. Then condition (H3) is satisfied for all functions in H.

PROOF OF LEMMA S.15. For $x \in H$ fixed and $j = 1, \dots, J$ denote $Z_j = \langle \phi_j, X_1 - x\rangle$. We want to establish that the matrix

$$\Gamma = \left[\gamma_{j,k}^{1,1\prime}(0)\right]_{j,k=1}^{J} = \lim_{t \to 0} \frac{1}{t} \left[\mathrm{E}\left(Z_j Z_k \middle| \|X_1 - x\|^2 = t\right)\right]_{j,k=1}^{J}$$

is positive definite. That is equivalent with the fact that for any $\mathrm{u} = [u_1, \dots, u_J]^{\mathsf{T}} \in \mathbb{R}^J$ and $\|\mathrm{u}\| = 1$

$$\begin{aligned} 0 < \mathrm{u}^{\mathsf{T}} \Gamma \mathrm{u} &= \lim_{t \to 0} \frac{1}{t} \sum_{j=1}^{J} \sum_{k=1}^{J} \mathrm{E}\left(u_j Z_j Z_k u_k \middle| \|X_1 - x\|^2 = t\right) \\ &= \lim_{t \to 0} \frac{1}{t} \mathrm{E}\left(\left(\sum_{j=1}^{J} u_j Z_j\right)^2 \middle| \|X_1 - x\|^2 = t\right). \end{aligned}$$

For any univariate random variable Z with finite variance and $z \in \mathbb{R}$ we know that

$$\mathrm{E}\,(Z - z)^2 = \mathrm{E}\,(Z - \mathrm{E}\,Z)^2 + (\mathrm{E}\,Z - z)^2 \geq \mathrm{Var}\,Z, \tag{S.30}$$

with equality if and only if $z = \mathrm{E}\,Z$. Use the conditional version of this inequality to obtain

$$\begin{aligned} &\mathrm{E}\left(\left(\sum_{j=1}^{J} u_j Z_j\right)^2 \middle| \|X_1 - x\|^2 = t\right) \\ &= \mathrm{E}\left(\left(\sum_{j=1}^{J} u_j Z_j - \mathrm{E}\left(\sum_{j=1}^{J} u_j Z_j \middle| \|X_1 - x\|^2 = t\right)\right)^2 \middle| \|X_1 - x\|^2 = t\right) \\ &\quad + \mathrm{E}\left(\left(\mathrm{E}\left(\sum_{j=1}^{J} u_j Z_j \middle| \|X_1 - x\|^2 = t\right)\right)^2 \middle| \|X_1 - x\|^2 = t\right) \\ &\geq \mathrm{E}\left(\left(\sum_{j=1}^{J} u_j Z_j - \mathrm{E}\left(\sum_{j=1}^{J} u_j Z_j \middle| \|X_1 - x\|^2 = t\right)\right)^2 \middle| \|X_1 - x\|^2 = t\right) \\ &= \mathrm{Var}\left(\sum_{j=1}^{J} u_j Z_j \middle| \|X_1 - x\|^2 = t\right) = \mathrm{Var}\left(\sum_{j=1}^{J} u_j Z_j \middle| \sum_{i=1}^{\infty} Z_i^2 = t\right). \end{aligned} \tag{S.31}$$

Therefore, it suffices to show that the conditional variance of no projection of the vector $[Z_1, \ldots, Z_J]^\mathsf{T}$ into a line spanned by a unit vector is of order $o(t)$ with $t \to 0$.

We assume that the random vector $\mathrm{Z} := \left[Z_1, \ldots, Z_J, \sqrt{\sum_{i=J+1}^\infty Z_i^2}\right]^\mathsf{T}$ is absolutely continuous in $\mathbb{R}^{J+1}$. For an independent Rademacher random variable R, i.e. $P(R=1) = P(R=-1) = 1/2$, define $\widetilde{\mathrm{Z}} := \left[Z_1, \ldots, Z_J, R\sqrt{\sum_{i=J+1}^\infty Z_i^2}\right]^\mathsf{T}$. This random vector is absolutely continuous, with density f_J positive and continuous at the origin. It differs from the original random vector Z only in its last coordinate, and $\mathrm{Z}^\mathsf{T}\mathrm{Z}$ has the same distribution as $\widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}}$. The conditional density of $\widetilde{\mathrm{Z}}$ given $\sum_{i=1}^\infty Z_i^2 = \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t$ takes the form

$$\frac{f_J(\mathrm{z})\, 1\left[t = \mathrm{z}^\mathsf{T}\mathrm{z}\right]}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}} \quad \text{for } \mathrm{z} \in \mathbb{R}^{J+1}, \tag{S.32}$$

where $1\left[t = \mathrm{z}^\mathsf{T}\mathrm{z}\right]$ is 1 if $t = \mathrm{z}^\mathsf{T}\mathrm{z}$, 0 otherwise. The integral in (S.32), and in analogous expressions below, is taken with respect to the Hausdorff measure on an appropriate sphere in $\mathbb{R}^{J+1}$. By our assumptions, f_J is positive and continuous in the neighbourhood of the origin. Then, for t small enough, $c_{J,t} = \inf_{\{\mathrm{z}^\mathsf{T}\mathrm{z}=t\}} f_J(\mathrm{z})$ must be positive. Using (S.30) again, we can therefore write

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{j=1}^J u_j Z_j \middle| \sum_{i=1}^\infty Z_i^2 = t\right) \\
&= \int_{\mathbb{R}^{J+1}} \left(\sum_{j=1}^J u_j z_j - \mathrm{E}\left(\sum_{j=1}^J u_j Z_j \middle| \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t\right)\right)^2 \frac{f_J(\mathrm{z})\, 1\left[t = \mathrm{z}^\mathsf{T}\mathrm{z}\right]}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}}\,\mathrm{d}\,\mathrm{z} \\
&= \int_{\{\mathrm{z}^\mathsf{T}\mathrm{z}=t\}} \left(\sum_{j=1}^J u_j z_j - \mathrm{E}\left(\sum_{j=1}^J u_j Z_j \middle| \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t\right)\right)^2 \frac{f_J(\mathrm{z})}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}}\,\mathrm{d}\,\mathrm{z} \\
&\geq \int_{\{\mathrm{z}^\mathsf{T}\mathrm{z}=t\}} \left(\sum_{j=1}^J u_j z_j - \mathrm{E}\left(\sum_{j=1}^J u_j Z_j \middle| \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t\right)\right)^2 \frac{c_{J,t}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}}\,\mathrm{d}\,\mathrm{z} \\
&= \frac{c_{J,t}\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1\,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}} \int_{\{\mathrm{z}^\mathsf{T}\mathrm{z}=t\}} \left(\sum_{j=1}^J u_j z_j - \mathrm{E}\left(\sum_{j=1}^J u_j Z_j \middle| \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t\right)\right)^2 g_t(\mathrm{z})\,\mathrm{d}\,\mathrm{z} \\
&= \frac{c_{J,t}\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1\,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}}\,\mathrm{E}\left(\sqrt{t}\sum_{j=1}^J u_j U_j - \mathrm{E}\left(\sum_{j=1}^J u_j Z_j \middle| \widetilde{\mathrm{Z}}^\mathsf{T}\widetilde{\mathrm{Z}} = t\right)\right)^2 \\
&\geq \frac{c_{J,t}\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1\,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}} \operatorname{Var}\left(\sqrt{t}\sum_{j=1}^J u_j U_j\right) \\
&= \frac{c_{J,t}\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1\,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}}\, t\operatorname{Var}(U_1) = \frac{c_{J,t}\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1\,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v})\,\mathrm{d}\,\mathrm{v}} \frac{t}{J+1}.
\end{aligned}$$

Here, $g_t(\mathrm{z}) := \left(\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1 \,\mathrm{d}\,\mathrm{v}\right)^{-1}$ is the reciprocal of the Hausdorff measure of a sphere, and $\mathrm{U} := [U_1, \dots, U_{J+1}]^\mathsf{T}$ is a random vector distributed uniformly on the unit sphere in $\mathbb{R}^{J+1}$. The second inequality is from (S.30). The first equality on the last line in the formula above follows from the spherical symmetry of the vector U — any projection of U onto a line has the same distribution as U_1. The final equality follows from $\operatorname{Var}\mathrm{U} = \mathrm{I}/(J+1)$, for I the $(J+1) \times (J+1)$ identity matrix.

From the continuity of f_J around the origin it follows that

$$\lim_{t\to 0} \frac{c_{J,t} \int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1 \,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v}) \,\mathrm{d}\,\mathrm{v}} = \lim_{t\to 0} \frac{f_J(0) \int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} 1 \,\mathrm{d}\,\mathrm{v}}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(0) \,\mathrm{d}\,\mathrm{v}} = 1.$$

Therefore, we obtain

$$\lim_{t\to 0} \frac{1}{t} \operatorname{Var}\left(\sum_{j=1}^{J} u_j Z_j \middle| \sum_{i=1}^{\infty} Z_i^2 = t \right) \geq \frac{1}{J+1},$$

and also the desired

$$\mathrm{u}^\mathsf{T}\Gamma\mathrm{u} \geq \frac{1}{J+1} > 0. \tag{S.33}$$

Note that the uniformity in u is assured by the spherical symmetry of U utilized above.

Finally, to see that, for instance, function $\gamma_{j,k}^{1,1}$ is continuously differentiable at the origin, note that from (S.32) we get

$$\gamma_{j,k}^{1,1}(t) = \int_{\{\mathrm{z}^\mathsf{T}\mathrm{z}=t\}} \frac{z_j z_k f_J(\mathrm{z})}{\int_{\{\mathrm{v}^\mathsf{T}\mathrm{v}=t\}} f_J(\mathrm{v}) \,\mathrm{d}\,\mathrm{v}} \,\mathrm{d}\,\mathrm{z},$$

for z_j and z_k the jth and kth elements of the vector $\mathrm{z} \in \mathbb{R}^{J+1}$, respectively. The assertion follows from the continuity of f_J, and both integrands above in the formula for $\gamma_{j,k}^{1,1}(t)$, around the origin and the fundamental theorem of calculus.

For a non-degenerate Gaussian process X we know that, in the notation from the proof above, each Z_i has a univariate, non-degenerate normal distribution. For the special case of $\{\phi_i\}_i$ being the eigenbasis of the covariance operator of X, $\{Z_i\}_i$ is a sequence of independent random variables with non-degenerate normal distributions. Therefore, in this case the elements of the random vector $\widetilde{\mathrm{Z}}$ are independent, and absolutely continuous random variables. Absolute continuity of the last element $\sqrt{\sum_{i=J+1}^{\infty} \langle \phi_i, X_1 - x \rangle^2}$ follows from the orthonormality of the basis $\{\phi_i\}_i$, independence and absolute continuity of the terms of the sequence $\{Z_i\}_i = \{\langle \phi_i, X_1 - x \rangle\}_i$, and absolute continuity of $\|X_1 - x\|$. All the marginal densities of $\widetilde{\mathrm{Z}}$ are positive and continuous at the origin. Therefore, for any non-degenerate Gaussian process with $\{\phi_i\}_i$ its collection of eigenfunctions, the assumptions of LEMMA S.15 are satisfied, and X satisfies (H3).

REFERENCES

BEGHIN, L., NIKITIN, Y. & ORSINGHER, E. (2005). Exact small ball constants for some Gaussian processes under the L 2-norm. *Journal of Mathematical Sciences* **128**, 2493–2502.

BOSQ, D. (2000). *Linear processes in function spaces*, vol. 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

BOURBAKI, N. (2004). *Integration I. Chapters 1–6*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin.

CARDOT, H., FERRATY, F. & SARDA, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11–22.

DAUXOIS, J., POUSSE, A. & ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.* **12**, 136–154.

DE BOOR, C. (1978). *A practical guide to splines*, vol. 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin.

FERRATY, F., MAS, A. & VIEU, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* **49**, 267–286.

GAO, F., HANNIG, J., LEE, T.-Y. & TORCASO, F. (2004). Exact L 2 small balls of Gaussian processes. *J. Theor. Probab.* **17**, 503–520.

GAO, F., HANNIG, J. & TORCASO, F. (2003). Integrated brownian motions and exact L 2-small balls. *Ann. Probab.* **31**, 1320–1337.

HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 109–126.

HALL, P., MÜLLER, H.-G. & WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493–1517.

KARHUNEN, K. (1946). Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* **34**, 1–7.

KAROL, A., NAZAROV, A. & NIKITIN, Y. (2008). Small ball probabilities for Gaussian random fields and tensor products of compact operators. *Trans. Amer. Math. Soc.* **360**, 1443–1474.

LI, W. V. & SHAO, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. *Handbook of Statistics* **19**, 533–597.

LIAN, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electron. J. Stat.* **5**, 31–40.

LIFSHITS, M. (1999). Asymptotic behavior of small ball probabilities. In *Probab. Theory and Math. Statist. Proc. VII International Vilnius Conference*.

LIN, Z., MÜLLER, H.-G. & YAO, F. (2018). Mixture inner product spaces and their application to functional data analysis. *Ann. Statist.* **46**, 370–400.

LOÈVE, M. (1946). Fonctions aléatoires à décomposition orthogonale exponentielle. *Revue Sci.* **84**, 159–162.

LU, T.-T. & SHIOU, S.-H. (2002). Inverses of 2×2 block matrices. *Comput. Math. Appl.* **43**, 119–129.

NAZAROV, A. I. & NIKITIN, Y. Y. (2004). Exact L 2-small ball behavior of integrated Gaussian processes and spectral asymptotics of boundary value problems. *Probab. Theory Relat. Fields* **129**, 469–494.

NAZAROV, A. I. & NIKITIN, Y. Y. (2005). Logarithmic L 2-small ball asymptotics for some fractional Gaussian processes. *Theor. Probab. Appl.* **49**, 645–658.

POLLARD, D. (2012). *Convergence of stochastic processes*. Springer Science & Business Media.

RAO, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14**, 1–17.

REDD, A. (2012). A comment on the orthogonalization of B-spline basis functions and their derivatives. *Stat. Comput.* **22**, 251–257.

YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577–590.

ZYGMUND, A. (2002). *Trigonometric series. Vol. I, II*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 3rd ed.

[*Received* 10 *October* 2019. *Editorial decision on* 19 *March* 2021]