Supplementary Material

COBRA: A Combined Regression Strategy
by G. Biau, A. Fischer, B. Guedj and J. D. Malley

A. Proofs

A.1. Proof of Proposition 2.1

We have

$$\begin{aligned}\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2 &= \mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\ &\quad + \mathbb{E}|T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2 \\ &\quad - 2\mathbb{E}[(T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X})))(T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X}))].\end{aligned}$$

As for the double product, notice that

$$\begin{aligned}&\mathbb{E}[(T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X})))(T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X}))] \\ &\quad = \mathbb{E}\left[\mathbb{E}\left[(T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X})))(T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X}))|\mathbf{r}_k(\mathbf{X}), \mathscr{D}_n\right]\right] \\ &\quad = \mathbb{E}\left[(T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X})))\mathbb{E}\left[T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|\mathbf{r}_k(\mathbf{X}), \mathscr{D}_n\right]\right].\end{aligned}$$

But

$$\begin{aligned}\mathbb{E}[r^\star(\mathbf{X})|\mathbf{r}_k(\mathbf{X}), \mathscr{D}_n] &= \mathbb{E}[r^\star(\mathbf{X})|\mathbf{r}_k(\mathbf{X})] \\ &\qquad \text{(by independence of } \mathbf{X} \text{ and } \mathscr{D}_n) \\ &= \mathbb{E}[\mathbb{E}[Y|\mathbf{X}]|\mathbf{r}_k(\mathbf{X})] \\ &= \mathbb{E}[Y|\mathbf{r}_k(\mathbf{X})] \\ &\qquad \text{(since } \sigma(\mathbf{r}_k(\mathbf{X})) \subset \sigma(\mathbf{X})) \\ &= T(\mathbf{r}_k(\mathbf{X})).\end{aligned}$$

Consequently,

$$\mathbb{E}[(T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X})))(T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X}))] = 0$$

and

$$\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2 = \mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 + \mathbb{E}|T(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2.$$

Thus, by definition of the conditional expectation, and using the fact that $T(\mathbf{r}_k(\mathbf{X})) = \mathbb{E}[r^\star(\mathbf{X})|\mathbf{r}_k(\mathbf{X})]$,

$$\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2 \le \mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 + \inf_f \mathbb{E}|f(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2,$$

where the infimum is taken over all square integrable functions of $\mathbf{r}_k(\mathbf{X})$. In particular,

$$\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - r^\star(\mathbf{X})|^2 \\
&\quad \le \min_{m=1,\dots,M} \mathbb{E}|r_{k,m}(\mathbf{X}) - r^\star(\mathbf{X})|^2 + \mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2,
\end{aligned}$$

as desired.

A.2. Proof of Proposition 2.2

Note that the second statement is an immediate consequence of the first statement and Proposition 2.1, therefore we only have to prove that

$$\mathbb{E}\,|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \to 0 \quad \text{as } \ell \to \infty.$$

We start with a technical lemma, whose proof can be found in the monograph by Györfi et al. (2002).

Lemma A.1. *Let $B(n,p)$ be a binomial random variable with parameters $n \ge 1$ and $p > 0$. Then*

$$\mathbb{E}\left[\frac{1}{1+B(n,p)}\right] \le \frac{1}{p(n+1)}$$

and

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{B(n,p)>0\}}}{B(n,p)}\right] \le \frac{2}{p(n+1)}.$$

For all distributions of $(\mathbf{X}, Y)$, using the elementary inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, note that

$$\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\
&\quad = \mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(Y_i - T(\mathbf{r}_k(\mathbf{X}_i)) + T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X})) + T(\mathbf{r}_k(\mathbf{X}))) \right.\\
&\qquad \left. - T(\mathbf{r}_k(\mathbf{X}))\right|^2 \\
&\quad \le 3\mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X})))\right|^2 && \text{(A.1)}\\
&\qquad + 3\mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(Y_i - T(\mathbf{r}_k(\mathbf{X}_i)))\right|^2 && \text{(A.2)}\\
&\qquad + 3\mathbb{E}\left|\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) - 1\right) T(\mathbf{r}_k(\mathbf{X}))\right|^2. && \text{(A.3)}
\end{aligned}$$

Consequently, to prove the proposition, it suffices to establish that (A.1), (A.2) and (A.3) tend to 0 as ℓ tends to infinity. This is done, respectively, in Proposition A.1, Proposition A.2 and Proposition A.3 below.

Proposition A.1. *Under the assumptions of Proposition 2.2,*

$$\lim_{\ell\to\infty} \mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X})))\right|^2 = 0.$$

Proof of Proposition A.1. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X})))\right|^2 \\
&\quad = \mathbb{E}\left|\sum_{i=1}^{\ell} \sqrt{W_{n,i}(\mathbf{X})}\sqrt{W_{n,i}(\mathbf{X})}(T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X})))\right|^2 \\
&\quad \le \mathbb{E}\left[\sum_{j=1}^{\ell} W_{n,j}(\mathbf{X})\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})\,|T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&\quad = \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})\,|T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&\quad := A_n.
\end{aligned}$$

The function T is such that $\mathbb{E}[T^2(\mathbf{r}_k(\mathbf{X}))] < \infty$. Therefore, it can be approximated in an L^2 sense by a continuous function with compact support, say $\tilde{T}$ (see, *e.g.*, Theorem A.1 in Györfi et al., 2002). More precisely, for any $\eta > 0$, there exists a function $\tilde{T}$ such that

$$\mathbb{E}\left|T(\mathbf{r}_k(\mathbf{X})) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))\right|^2 < \eta.$$

Consequently, we obtain

$$\begin{aligned}
A_n &= \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&\le 3\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|T(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{T}(\mathbf{r}_k(\mathbf{X}_i))|^2\right] \\
&\quad + 3\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|\tilde{T}(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&\quad + 3\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|\tilde{T}(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&:= 3A_{n1} + 3A_{n2} + 3A_{n3}.
\end{aligned}$$

Computation of A_{n3}. Thanks to the approximation of T by $\tilde{T}$,

$$\begin{aligned}
A_{n3} &= \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|T(\mathbf{r}_k(\mathbf{X})) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&\leq \mathbb{E}\left|T(\mathbf{r}_k(\mathbf{X})) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))\right|^2 < \eta.
\end{aligned}$$

Computation of A_{n1}. Denote by μ the distribution of $\mathbf{X}$. Then,

$$\begin{aligned}
A_{n1} &= \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|\tilde{T}(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{X}_i))|^2\right] \\
&= \ell\mathbb{E}\left[\frac{\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_1)|\leq\varepsilon_\ell\}}}{\sum_{j=1}^{\ell}\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_j)|\leq\varepsilon_\ell\}}}|\tilde{T}(\mathbf{r}_k(\mathbf{X}_1)) - T(\mathbf{r}_k(\mathbf{X}_1))|^2\right]. \\
&= \ell\mathbb{E}\Bigg\{\int_{\mathbb{R}^d}|\tilde{T}(\mathbf{r}_k(\mathbf{u})) - T(\mathbf{r}_k(\mathbf{u}))|^2 \\
&\times\mathbb{E}\Bigg[\int_{\mathbb{R}^d}\frac{\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{u})|\leq\varepsilon_\ell\}}}{\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{u})|\leq\varepsilon_\ell\}}+\sum_{j=2}^{\ell}\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_j)|\leq\varepsilon_\ell\}}}\mu(\mathrm{d}\mathbf{x}) \\
&\qquad\Bigg|\mathscr{D}_k\Bigg]\mu(\mathrm{d}\mathbf{u})\Bigg\}.
\end{aligned}$$

Letting

$$\begin{aligned}
A'_{n1} = \mathbb{E}\Bigg[\int_{\mathbb{R}^d}&\frac{\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{u})|\leq\varepsilon_\ell\}}}{\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{u})|\leq\varepsilon_\ell\}}+\sum_{j=2}^{\ell}\mathbf{1}_{\bigcap_{m=1}^M\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_j)|\leq\varepsilon_\ell\}}}\mu(\mathrm{d}\mathbf{x}) \\
&\Bigg|\mathscr{D}_k\Bigg],
\end{aligned}$$

let us prove that $A'_{n1} \leq \frac{2^M}{\ell}$. To this aim, observe that

$$\begin{aligned}
A'_{n1} &= \mathbb{E}\left[\int_{\mathbb{R}^d}\frac{\mathbf{1}_{\{\mathbf{x}\in\bigcap_{m=1}^M r_{k,m}^{-1}([r_{k,m}(\mathbf{u})-\varepsilon_\ell,r_{k,m}(\mathbf{u})+\varepsilon_\ell])\}}}{1+\sum_{j=2}^{\ell}\mathbf{1}_{\{\mathbf{X}_j\in\bigcap_{m=1}^M r_{k,m}^{-1}([r_{k,m}(\mathbf{x})-\varepsilon_\ell,r_{k,m}(\mathbf{x})+\varepsilon_\ell])\}}}\mu(\mathrm{d}\mathbf{x})\middle|\mathscr{D}_k\right] \\
&= \mathbb{E}\left[\int_{\mathbb{R}^d}\frac{\mathbf{1}_{\{\mathbf{x}\in\bigcup_{(a_1,\ldots,a_M)\in\{1,2\}^M} r_{k,1}^{-1}(I_{n,1}^{a_1}(\mathbf{u}))\cap\cdots\cap r_{k,M}^{-1}(I_{n,M}^{a_M}(\mathbf{u}))\}}}{1+\sum_{j=2}^{\ell}\mathbf{1}_{\{\mathbf{X}_j\in\bigcap_{m=1}^M r_{k,m}^{-1}([r_{k,m}(\mathbf{x})-\varepsilon_\ell,r_{k,m}(\mathbf{x})+\varepsilon_\ell])\}}}\mu(\mathrm{d}\mathbf{x})\middle|\mathscr{D}_k\right] \\
&\leq \sum_{p=1}^{2^M}\mathbb{E}\left[\int_{\mathbb{R}^d}\frac{\mathbf{1}_{\{\mathbf{x}\in R_n^p(\mathbf{u})\}}}{1+\sum_{j=2}^{\ell}\mathbf{1}_{\{\mathbf{X}_j\in\bigcap_{m=1}^M r_{k,m}^{-1}([r_{k,m}(\mathbf{x})-\varepsilon_\ell,r_{k,m}(\mathbf{x})+\varepsilon_\ell])\}}}\mu(\mathrm{d}\mathbf{x})\middle|\mathscr{D}_k\right].
\end{aligned}$$

Here, $I_{n,m}^1(\mathbf{u}) = [r_{k,m}(\mathbf{u}) - \varepsilon_\ell, r_{k,m}(\mathbf{u})]$, $I_{n,m}^2(\mathbf{u}) = [r_{k,m}(\mathbf{u}), r_{k,m}(\mathbf{u}) + \varepsilon_\ell]$, and $R_n^p(\mathbf{u})$ is the p-th set of the form $r_{k,1}^{-1}(I_{n,1}^{a_1}(\mathbf{u})) \cap \cdots \cap r_{k,M}^{-1}(I_{n,M}^{a_M}(\mathbf{u}))$ assuming that they have been ordered using the lexicographic order of $(a_1, \ldots, a_M)$.

Next, note that

$$\mathbf{x} \in R_n^p(\mathbf{u}) \Rightarrow R_n^p(\mathbf{u}) \subset \bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{x}) - \varepsilon_\ell, r_{k,m}(\mathbf{x}) + \varepsilon_\ell]).$$

To see this, just observe that, for all $m = 1, \ldots, M$, if $r_{k,m}(\mathbf{z}) \in [r_{k,m}(\mathbf{u}) - \varepsilon_\ell, r_{k,m}(\mathbf{u})]$, *i.e.*, $r_{k,m}(\mathbf{u}) - \varepsilon_\ell \le r_{k,m}(\mathbf{z}) \le r_{k,m}(\mathbf{u})$, then, as $r_{k,m}(\mathbf{u}) - \varepsilon_\ell \le r_{k,m}(\mathbf{x}) \le r_{k,m}(\mathbf{u})$, one has $r_{k,m}(\mathbf{x}) - \varepsilon_\ell \le r_{k,m}(\mathbf{z}) \le r_{k,m}(\mathbf{x}) + \varepsilon_\ell$. Similarly, if $r_{k,m}(\mathbf{u}) \le r_{k,m}(\mathbf{z}) \le r_{k,m}(\mathbf{u}) + \varepsilon_\ell$, then $r_{k,m}(\mathbf{u}) \le r_{k,m}(\mathbf{x}) \le r_{k,m}(\mathbf{u}) + \varepsilon_\ell$ implies $r_{k,m}(\mathbf{x}) - \varepsilon_\ell \le r_{k,m}(\mathbf{z}) \le r_{k,m}(\mathbf{x}) + \varepsilon_\ell$. Consequently,

$$\begin{aligned}
A'_{n1} &\le \sum_{p=1}^{2^M} \mathbb{E}\left[\int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\mathbf{x} \in R_n^p(\mathbf{u})\}}}{1 + \sum_{j=2}^{\ell} \mathbf{1}_{\{\mathbf{X}_j \in R_n^p(\mathbf{u})\}}} \mu(\mathrm{d}\mathbf{x}) \middle| \mathscr{D}_k\right] \\
&= \sum_{p=1}^{2^M} \mathbb{E}\left[\frac{\mu\{R_n^p(\mathbf{u})\}}{1 + \sum_{j=2}^{\ell} \mathbf{1}_{\{\mathbf{X}_j \in R_n^p(\mathbf{u})\}}} \middle| \mathscr{D}_k\right] \\
&\le \sum_{p=1}^{2^M} \mathbb{E}\left[\frac{\mu\{R_n^p(\mathbf{u})\}}{\ell \mu\{R_n^p(\mathbf{u})\}} \middle| \mathscr{D}_k\right] \\
&\le \frac{2^M}{\ell}
\end{aligned}$$

(by the first statement of Lemma A.1). Thus, returning to A_{n1}, we obtain

$$A_{n1} \le 2^M \mathbb{E}\left|\tilde{T}(\mathbf{r}_k(\mathbf{X}) - T(\mathbf{r}_k(\mathbf{X})))\right|^2 < 2^M \eta.$$

Computation of A_{n2}. For any $\delta > 0$, write

$$\begin{aligned}
A_{n2} &= \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) |\tilde{T}(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))|^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) |\tilde{T}(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))|^2 \mathbf{1}_{\bigcup_{m=1}^M \{|r_{k,m}(\mathbf{X}) - r_{k,m}(\mathbf{X}_i)| > \delta\}}\right] \\
&\quad + \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) |\tilde{T}(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{T}(\mathbf{r}_k(\mathbf{X}))|^2 \mathbf{1}_{\bigcap_{m=1}^M \{|r_{k,m}(\mathbf{X}) - r_{k,m}(\mathbf{X}_i)| \le \delta\}}\right]
\end{aligned}$$

from which we get that

$$A_{n2} \le 4 \sup_{\mathbf{u} \in \mathbb{R}^d} |\tilde{T}(\mathbf{r}_k(\mathbf{u}))|^2 \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) \mathbf{1}_{\bigcup_{m=1}^M \{|r_{k,m}(\mathbf{X}) - r_{k,m}(\mathbf{X}_i)| > \delta\}}\right] \tag{A.4}$$

$$+ \left(\sup_{\mathbf{u}, \mathbf{v} \in \mathbb{R}^d, \bigcap_{m=1}^M \{|r_{k,m}(\mathbf{u}) - r_{k,m}(\mathbf{v})| \le \delta\}} |\tilde{T}(\mathbf{r}_k(\mathbf{v})) - \tilde{T}(\mathbf{r}_k(\mathbf{u}))|\right)^2. \tag{A.5}$$

With respect to the term (A.4), if $\delta > \varepsilon_\ell$, then

$$\begin{aligned}
&\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})\mathbf{1}_{\bigcup_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|>\delta\}} \\
&\quad = \sum_{i=1}^{\ell} \frac{\mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}}\mathbf{1}_{\bigcup_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|>\delta\}}}{\sum_{j=1}^{\ell}\mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_j)|\le\varepsilon_\ell\}}} \\
&\quad = 0.
\end{aligned}$$

It follows that, for all $\delta > 0$, this term converges to 0 as ℓ tends to infinity. On the other hand, letting $\delta \to 0$, we see that the term (A.5) tends to 0 as well, by uniform continuity of $\tilde{T}$. Hence, A_{n2} tends to 0 as ℓ tends to infinity. Letting finally η go to 0, we conclude that A_n vanishes as ℓ tends to infinity. □

Proposition A.2. *Under the assumptions of Proposition 2.2,*

$$\lim_{\ell\to\infty} \mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(Y_i - T(\mathbf{r}_k(\mathbf{X}_i)))\right|^2 = 0.$$

Proof of Proposition A.2.

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})(Y_i - T(\mathbf{r}_k(\mathbf{X}_i)))\right|^2 \\
&\quad = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \mathbb{E}[W_{n,i}(\mathbf{X})W_{n,j}(\mathbf{X})(Y_i - T(\mathbf{r}_k(\mathbf{X}_i)))(Y_j - T(\mathbf{r}_k(\mathbf{X}_j)))] \\
&\quad = \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})|Y_i - T(\mathbf{r}_k(\mathbf{X}_i))|^2\right] \\
&\quad = \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})\sigma^2(\mathbf{r}_k(\mathbf{X}_i))\right],
\end{aligned}$$

where

$$\sigma^2(\mathbf{r}_k(\mathbf{x})) = \mathbb{E}[|Y - T(\mathbf{r}_k(\mathbf{X}))|^2|\mathbf{r}_k(\mathbf{x})].$$

For any $\eta > 0$, σ^2 can be approximated in an L^1 sense by a continuous function with compact support $\tilde{\sigma}^2$, *i.e.*,

$$\mathbb{E}|\tilde{\sigma}^2(\mathbf{r}_k(\mathbf{X})) - \sigma^2(\mathbf{r}_k(\mathbf{X}))| < \eta.$$

Thus

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})\sigma^2(\mathbf{r}_k(\mathbf{X}_i))\right] \\
&\quad \le \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})\tilde{\sigma}^2(\mathbf{r}_k(\mathbf{X}_i))\right] \\
&\qquad + \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})|\sigma^2(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{\sigma}^2(\mathbf{r}_k(\mathbf{X}_i))|\right] \\
&\quad \le \sup_{\mathbf{u}\in\mathbb{R}^d} |\tilde{\sigma}^2(\mathbf{r}_k(\mathbf{u}))|\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})\right] \\
&\qquad + \mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|\sigma^2(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{\sigma}^2(\mathbf{r}_k(\mathbf{X}_i))|\right].
\end{aligned}
$$

With the same argument as for A_{n1}, we obtain

$$
\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X})|\sigma^2(\mathbf{r}_k(\mathbf{X}_i)) - \tilde{\sigma}^2(\mathbf{r}_k(\mathbf{X}_i))|\right] \le 2^M\eta.
$$

Therefore, it remains to prove that $\mathbb{E}\left[\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X})\right] \to 0$ as $\ell \to \infty$. To this aim, fix $\delta > 0$, and note that

$$
\begin{aligned}
\sum_{i=1}^{\ell} W_{n,i}^2(\mathbf{X}) &= \frac{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}}}{\left(\sum_{j=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_j)|\le\varepsilon_\ell\}}\right)^2} \\
&\le \min\left\{\delta, \frac{1}{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}}}\right\} \\
&\le \delta + \frac{\mathbf{1}_{\left\{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}} > 0\right\}}}{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}}}.
\end{aligned}
$$

To complete the proof, we have to establish that the expectation of the right-hand term tends to 0. Denoting by I a bounded interval on the real line, we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\mathbf{1}\left\{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}>0\right\}}{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}}\right]\\
&\le \mathbb{E}\left[\frac{\mathbf{1}\left\{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}>0\right\}\mathbf{1}_{\left\{\mathbf{X}\in\bigcap_{m=1}^{M}r_{k,m}^{-1}(I)\right\}}}{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}}\right]\\
&\quad +\mu\Big(\bigcup_{m=1}^{M}r_{k,m}^{-1}(I^c)\Big)\\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbf{1}\left\{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}>0\right\}\mathbf{1}_{\left\{\mathbf{X}\in\bigcap_{m=1}^{M}r_{k,m}^{-1}(I)\right\}}}{\sum_{i=1}^{\ell}\mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}}}\right.\right.\\
&\qquad \left.\left.\Big|\mathscr{D}_k,\mathbf{X}\right]\right]+\mu\Big(\bigcup_{m=1}^{M}r_{k,m}^{-1}(I^c)\Big)\\
&\le \frac{2}{(\ell+1)}\mathbb{E}\left[\frac{\mathbf{1}_{\left\{\mathbf{X}\in\bigcap_{m=1}^{M}r_{k,m}^{-1}(I)\right\}}}{\mu(\bigcap_{m=1}^{M}r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell,r_{k,m}(\mathbf{X})+\varepsilon_\ell]))}\right]\\
&\quad +\mu\Big(\bigcup_{m=1}^{M}r_{k,m}^{-1}(I^c)\Big).
\end{aligned}
$$

The last inequality arises from the second statement of Lemma A.1. By an appropriate choice of I, according to the technical statement (2.2), the second term on the right-hand side can be made as small as desired. Regarding the first term, there exists a finite number N_ℓ of points $\mathbf{z}_1,\dots,\mathbf{z}_{N_\ell}$ such that

$$
\bigcap_{m=1}^{M}r_{k,m}^{-1}(I)\subset\bigcup_{(j_1,\dots,j_M)\in\{1,\dots,N_\ell\}^M}r_{k,1}^{-1}(I_{n,1}(\mathbf{z}_{j_1}))\cap\cdots\cap r_{k,M}^{-1}(I_{n,M}(\mathbf{z}_{j_M})),
$$

where $I_{n,m}(\mathbf{z}_j)=[\mathbf{z}_j-\varepsilon_\ell/2,\mathbf{z}_j+\varepsilon_\ell/2]$. Suppose, without loss of generality, that the sets

$$
r_{k,1}^{-1}(I_{n,1}(\mathbf{z}_{j_1}))\cap\cdots\cap r_{k,M}^{-1}(I_{n,M}(\mathbf{z}_{j_M}))
$$

are ordered, and denote by R_n^p the p-th among the $N_\ell^M=(\lceil|I|/\varepsilon_\ell\rceil)^M$ sets. Here $|I|$ denotes the length of the interval I and $\lceil x\rceil$ denotes the smallest integer greater than x. For all p,

$$\mathbf{x} \in R_n^p \Rightarrow R_n^p \subset \bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{x}) - \varepsilon_\ell, r_{k,m}(\mathbf{x}) + \varepsilon_\ell]).$$

Indeed, if $\mathbf{v} \in R_n^p$, then, for all $m = 1, \ldots, M$, there exists $j \in \{1, \ldots, N_\ell\}$ such that $r_{k,m}(\mathbf{v}) \in [\mathbf{z}_j - \varepsilon_\ell/2, \mathbf{z}_j + \varepsilon_\ell/2]$, that is $\mathbf{z}_j - \varepsilon_\ell/2 \le r_{k,m}(\mathbf{v}) \le \mathbf{z}_j + \varepsilon_\ell/2$. Since we also have $\mathbf{z}_j - \varepsilon_\ell/2 \le r_{k,m}(\mathbf{X}) \le \mathbf{z}_j + \varepsilon_\ell/2$, we obtain $r_{k,m}(\mathbf{X}) - \varepsilon_\ell \le r_{k,m}(\mathbf{v}) \le r_{k,m}(\mathbf{X}) + \varepsilon_\ell$. In conclusion,

$$\begin{aligned}
&\mathbb{E}\left[\frac{\mathbf{1}_{\left\{\mathbf{X}\in\bigcap_{m=1}^{M} r_{k,m}^{-1}(I)\right\}}}{\mu(\bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{X}) - \varepsilon_\ell, r_{k,m}(\mathbf{X}) + \varepsilon_\ell]))}\right] \\
&\le \sum_{p=1}^{N_\ell^M} \mathbb{E}\left[\frac{\mathbf{1}_{\{\mathbf{X}\in R_n^p\}}}{\mu(\bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{X}) - \varepsilon_\ell, r_{k,m}(\mathbf{X}) + \varepsilon_\ell]))}\right] \\
&\le \sum_{p=1}^{N_\ell^M} \mathbb{E}\left[\frac{\mathbf{1}_{\{\mathbf{X}\in R_n^p\}}}{\mu(R_n^p)}\right] \\
&= N_\ell^M \\
&= \left\lceil \frac{|I|}{\varepsilon_\ell} \right\rceil^M .
\end{aligned}$$

The result follows from the assumption $\lim_{\ell\to\infty} \ell\varepsilon_\ell^M = \infty$. □

Proposition A.3. *Under the assumptions of Proposition 2.2,*

$$\lim_{\ell\to\infty} \mathbb{E}\left|\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) - 1\right) T(\mathbf{r}_k(\mathbf{X}))\right|^2 = 0.$$

Proof of Proposition A.3. Since $|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) - 1| \le 1$, one has

$$\left|\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) - 1\right) T(\mathbf{r}_k(\mathbf{X}))\right|^2 \le T^2(\mathbf{r}_k(\mathbf{X})).$$

Consequently, by Lebesgue's dominated convergence theorem, to prove the proposition, it suffices to show that $W_{n,i}(\mathbf{X})$ tends to 1 almost surely. Now,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) \neq 1\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{X})-r_{k,m}(\mathbf{X}_i))|\leq\varepsilon_\ell\}} = 0\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{\ell} \mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{X})-\varepsilon_\ell, r_{k,m}(\mathbf{X})+\varepsilon_\ell])\right\}} = 0\right) \\
&= \int_{\mathbb{R}^d} \mathbb{P}\left(\forall i=1,\ldots,\ell, \mathbf{1}_{\left\{\mathbf{X}_i\in\bigcap_{m=1}^{M} r_{k,m}^{-1}([r_{k,m}(\mathbf{x})-\varepsilon_\ell, r_{k,m}(\mathbf{x})+\varepsilon_\ell])\right\}} = 0\right)\mu(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathbb{R}^d} \left[1-\mu(\cap_{m=1}^{M} r_{k,m}^{-1}\left([r_{k,m}(\mathbf{x})-\varepsilon_\ell, r_{k,m}(\mathbf{x})+\varepsilon_\ell]\right))\right]^{\ell} \mu(\mathrm{d}\mathbf{x}).
\end{aligned}$$

Denote by I a bounded interval. Then,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) \neq 1\right) \\
&\leq \int_{\mathbb{R}^d} \exp\left(-\ell\mu(\cap_{m=1}^{M} r_{k,m}^{-1}\left([r_{k,m}(\mathbf{x})-\varepsilon_\ell, r_{k,m}(\mathbf{x})+\varepsilon_\ell]\right))\right) \\
&\quad\times \mathbf{1}_{\{\mathbf{x}\in\bigcap_{m=1}^{M} r_{k,m}^{-1}(I)\}}\mu(\mathrm{d}\mathbf{x}) + \mu\Big(\bigcup_{m=1}^{M} r_{k,m}^{-1}(I^c)\Big) \\
&\leq \max_{\mathbf{u}} \mathbf{u}e^{-\mathbf{u}} \int_{\mathbb{R}^d} \frac{\mathbf{1}_{\{\mathbf{x}\in\bigcap_{m=1}^{M} r_{k,m}^{-1}(I)\}}}{\ell\mu(\cap_{m=1}^{M} r_{k,m}^{-1}\left([r_{k,m}(\mathbf{x})-\varepsilon_\ell, r_{k,m}(\mathbf{x})+\varepsilon_\ell]\right))}\mu(\mathrm{d}\mathbf{x}) \\
&\quad+ \mu\Big(\bigcup_{m=1}^{M} r_{k,m}^{-1}(I^c)\Big).
\end{aligned}$$

Using the same arguments as in the proof of Proposition A.2, the probability $\mathbb{P}\left(\sum_{i=1}^{\ell} W_{n,i}(\mathbf{X}) \neq 1\right)$ is bounded by $\frac{e^{-1}}{\ell}\left\lceil\frac{|I|}{\varepsilon_\ell}\right\rceil^M$. This bound vanishes as n tends to infinity since, by assumption, $\lim_{\ell\to\infty} \ell\varepsilon_\ell^M = \infty$. □

A.3. *Proof of Theorem 2.1*

Choose $\mathbf{x}\in\mathbb{R}^d$. An easy calculation yields that

$$\begin{aligned}
&\mathbb{E}[|T_n(\mathbf{r}_k(\mathbf{x}))-T(\mathbf{r}_k(\mathbf{x}))|^2|\mathbf{r}_k(\mathbf{X}_1),\ldots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k] \\
&= \mathbb{E}\Bigg[\left|T_n(\mathbf{r}_k(\mathbf{x}))-\mathbb{E}[T_n(\mathbf{r}_k(\mathbf{x}))|\mathbf{r}_k(\mathbf{X}_1),\ldots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k]\right|^2 \qquad \text{(A.6)} \\
&\qquad \Bigg|\mathbf{r}_k(\mathbf{X}_1),\ldots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k\Bigg] + \left|\mathbb{E}[T_n(\mathbf{r}_k(\mathbf{x}))|\mathbf{r}_k(\mathbf{X}_1),\ldots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k]-T(\mathbf{r}_k(\mathbf{x}))\right|^2 \\
&:= E_1+E_2. \qquad \text{(A.7)}
\end{aligned}$$

On the one hand, we have

$$E_1 = \mathbb{E}\Big[\big|T_n(\mathbf{r}_k(\mathbf{x})) - \mathbb{E}[T_n(\mathbf{r}_k(\mathbf{x}))|\mathbf{r}_k(\mathbf{X}_1),\dots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k]\big|^2 \Big|\mathbf{r}_k(\mathbf{X}_1),\dots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k\Big]$$

$$= \mathbb{E}\left[\left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{x})(Y_i - \mathbb{E}[Y_i|\mathbf{r}_k(\mathbf{X}_i)])\right|^2 |\mathbf{r}_k(\mathbf{X}_1),\dots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k\right].$$

Developing the square and noticing that $\mathbb{E}\big[Y_j|Y_i,\mathbf{r}_k(\mathbf{X}_1),\dots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k\big] = \mathbb{E}[Y_j|\mathbf{r}_k(\mathbf{X}_j)]$, since Y_j is independent of Y_i and of the $\mathbf{X}_j$'s with $j \neq i$, we have

$$E_1 = \mathbb{E}\Bigg[\frac{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}|Y_i - \mathbb{E}[Y_i|\mathbf{r}_k(\mathbf{X}_i)]|^2}{\left|\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}\right|^2} \Bigg|\mathbf{r}_k(\mathbf{X}_1),\dots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k\Bigg] \tag{A.8}$$

$$= \sum_{i=1}^{\ell} \mathbb{V}(Y_i|\mathbf{r}_k(\mathbf{X}_i))\frac{\mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}}{\left|\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}\right|^2}.$$

Thus,

$$E_1 \leq 4R^2 \frac{\mathbf{1}_{\left\{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}>0\right\}}}{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\leq\varepsilon_\ell\}}}, \tag{A.9}$$

where $\mathbb{V}(Z)$ denotes the variance of a random variable Z. On the other hand, recalling the notation Σ introduced in Section 3, we obtain for the second term E_2:

$$\begin{aligned}
E_2 &= \left|\mathbb{E}[T_n(\mathbf{r}_k(\mathbf{x}))|\mathbf{r}_k(\mathbf{X}_1),\ldots,\mathbf{r}_k(\mathbf{X}_\ell),\mathscr{D}_k] - T(\mathbf{r}_k(\mathbf{x}))\right|^2 \\
&= \left|\sum_{i=1}^{\ell} W_{n,i}(\mathbf{x})\mathbb{E}[Y_i|\mathbf{r}_k(\mathbf{X}_i)] - T(\mathbf{r}_k(\mathbf{x}))\right|^2 \mathbf{1}_{\{\Sigma>0\}} + T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}} \\
&\le \frac{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}} |\mathbb{E}[Y_i|\mathbf{r}_k(\mathbf{X}_i)] - T(\mathbf{r}_k(\mathbf{x}))|^2}{\sum_{j=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_j)|\le\varepsilon_\ell\}}} \mathbf{1}_{\{\Sigma>0\}} \qquad \text{(A.10)} \\
&\quad + T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}} \\
&\quad \text{(by Jensen's inequality)} \\
&= \frac{\sum_{i=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_i)|\le\varepsilon_\ell\}} |T(\mathbf{r}_k(\mathbf{X}_i)) - T(\mathbf{r}_k(\mathbf{x}))|^2}{\sum_{j=1}^{\ell} \mathbf{1}_{\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X}_j)|\le\varepsilon_\ell\}}} \mathbf{1}_{\{\Sigma>0\}} \qquad \text{(A.11)} \\
&\quad + T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}} \\
&\le L^2\varepsilon_\ell^2 + T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}}. \qquad \text{(A.12)}
\end{aligned}$$

Now,

$$\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \le \int_{\mathbb{R}^d} \mathbb{E}|(T_n(\mathbf{r}_k(\mathbf{x})) - T(\mathbf{r}_k(\mathbf{x}))|^2 \mu(d\mathbf{x}).$$

Then, using the decomposition (A.7) and the upper bounds (A.9) and (A.12),

$$\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\
&\quad \le \int_{\mathbb{R}^d} \mathbb{E}\left[\frac{4R^2\mathbf{1}_{\{\Sigma>0\}}}{B}\right]\mu(d\mathbf{x}) + L^2\varepsilon_\ell^2 + \int_{\mathbb{R}^d} \mathbb{E}\left[T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}}\right]\mu(d\mathbf{x}) \\
&\quad \le \int_{\mathbb{R}^d} \mathbb{E}\left\{\mathbb{E}\left[\frac{4R^2\mathbf{1}_{\{\Sigma>0\}}}{B}\Big|\mathscr{D}_k\right]\right\}\mu(d\mathbf{x}) + L^2\varepsilon_\ell^2 \\
&\qquad + \int_{\mathbb{R}^d} \mathbb{E}\left\{\mathbb{E}\left[T^2(\mathbf{r}_k(\mathbf{x}))\mathbf{1}_{\{\Sigma=0\}}|\mathscr{D}_k\right]\right\}\mu(d\mathbf{x}).
\end{aligned}$$

Thus, thanks to Lemma A.1,

$$\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\
&\quad \le \frac{8R^2}{(\ell+1)}\int_{\mathbb{R}^d} \frac{1}{\mu(\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X})|\le\varepsilon_\ell\})}\mu(d\mathbf{x}) + L^2\varepsilon_\ell^2 \\
&\qquad + \int_{\mathbb{R}^d} T^2(\mathbf{r}_k(\mathbf{x}))\left(1-\mu(\bigcap_{m=1}^{M}\{|r_{k,m}(\mathbf{x})-r_{k,m}(\mathbf{X})|\le\varepsilon_\ell\})\right)^\ell \mu(d\mathbf{x}).
\end{aligned}$$

Consequently,

$$
\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\
&\quad\le \frac{8R^2}{(\ell+1)}\int_{\mathbb{R}^d} \frac{1}{\mu(\bigcap_{m=1}^M \{|r_{k,m}(\mathbf{x}) - r_{k,m}(\mathbf{X})| \le \varepsilon_\ell\})}\mu(\mathrm{d}\mathbf{x}) + L^2\varepsilon_\ell^2 \\
&\qquad + \int_{\mathbb{R}^d} T^2(\mathbf{r}_k(\mathbf{x}))\exp\left(-\ell\mu(\bigcap_{m=1}^M \{|r_{k,m}(\mathbf{x}) - r_{k,m}(\mathbf{X})| \le \varepsilon_\ell\})\right)\mu(\mathrm{d}\mathbf{x}) \\
&\quad\le \frac{8R^2}{(\ell+1)}\int_{\mathbb{R}^d} \frac{1}{\mu(\bigcap_{m=1}^M \{|r_{k,m}(\mathbf{x}) - r_{k,m}(\mathbf{X})| \le \varepsilon_\ell\})}\mu(\mathrm{d}\mathbf{x}) + L^2\varepsilon_\ell^2 \\
&\qquad + \Bigg(\sup_{\mathbf{x}\in\mathbb{R}^d} T^2(\mathbf{r}_k(\mathbf{x}))\max_{\mathbf{u}\in\mathbb{R}^+}\mathbf{u}e^{-\mathbf{u}} \\
&\qquad\quad \times \int_{\mathbb{R}^d} \frac{1}{\ell\mu(\bigcap_{m=1}^M \{|r_{k,m}(\mathbf{x}) - r_{k,m}(\mathbf{X})| \le \varepsilon_\ell\})}\mu(\mathrm{d}\mathbf{x})\Bigg).
\end{aligned}
$$

Introducing a bounded interval I as in the proof of Proposition 2.2, we observe that the boundedness of the $\mathbf{r}_k$ yields that

$$
\mu\left(\bigcup_{m=1}^M r_{k,m}^{-1}(I^c)\right) = 0,
$$

as soon as I is sufficiently large, independently of k. Then, proceeding as in the proof of Proposition 2.2, we obtain

$$
\begin{aligned}
&\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \\
&\quad\le 8R^2\left\lceil\frac{|I|}{\varepsilon_\ell}\right\rceil^M \frac{1}{\ell+1} + L^2\varepsilon_\ell^2 + R^2\max_{\mathbf{u}\in\mathbb{R}^+}\mathbf{u}e^{-\mathbf{u}}\left\lceil\frac{|I|}{\varepsilon_\ell}\right\rceil^M\frac{1}{\ell} \\
&\quad\le C_1\frac{R^2}{\ell\varepsilon_\ell^M} + L^2\varepsilon_\ell^2,
\end{aligned}
$$

for some positive constant C_1, independent of k. Hence, for the choice $\varepsilon_\ell \propto \ell^{-\frac{1}{M+2}}$, we obtain

$$
\mathbb{E}|T_n(\mathbf{r}_k(\mathbf{X})) - T(\mathbf{r}_k(\mathbf{X}))|^2 \le C\ell^{-\frac{2}{M+2}},
$$

for some positive constant C depending on L, R and independent of k, as desired.

B. Numerical results

Table 4 (SM): Quadratic errors of the implemented machines and `COBRA` in high-dimensional situations. Means and standard deviations over 200 independent replications.

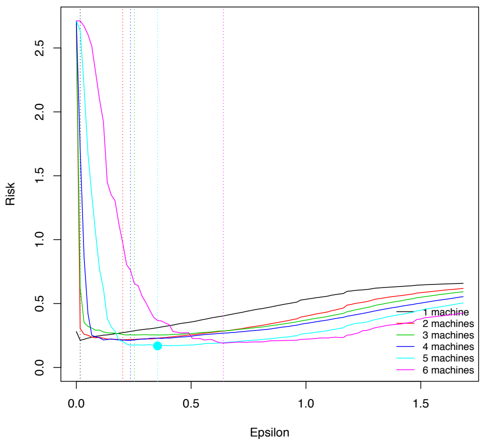
		`lars`	`ridge`	`fnn`	`tree`	`rf`	`COBRA`
Model 9	m.	1.5698	2.9752	3.9285	1.8646	1.5001	**0.9996**
	sd.	0.2357	0.4171	0.5356	0.3751	0.2491	0.1733
Model 10	m.	5.2356	5.1748	6.1395	6.1585	4.8667	**2.7076**
	sd.	0.6885	0.7139	0.9192	0.9298	0.6634	0.3810
Model 11	m.	0.1584	0.1055	0.1363	0.0058	0.0327	**0.0049**
	sd.	0.0199	0.0119	0.0176	0.0010	0.0052	0.0009

Table 5 (SM): Quadratic errors of exponentially weighted aggregate (EWA) and `COBRA`. 200 independent replications.

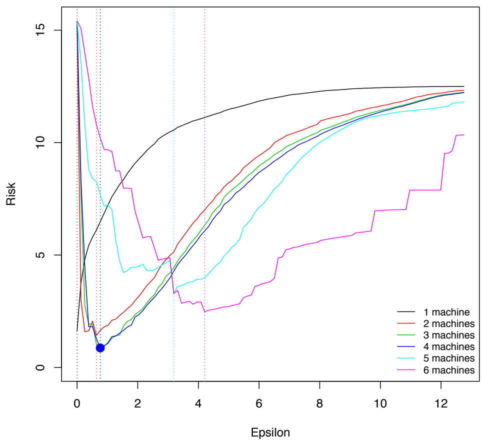
		`EWA`	`COBRA`
Model 9	m.	1.1712	**1.1360**
	sd.	0.2090	0.2468
Model 10	m.	**9.4789**	12.4353
	sd.	5.6275	9.1267
Model 11	m.	0.0244	**0.0128**
	sd.	0.0042	0.0237
Model 12	m.	0.4175	**0.3124**
	sd.	0.0513	0.0884

Figure 2 (SM): Examples of calibration of parameters ε_ℓ and α. The bold point is the minimum.

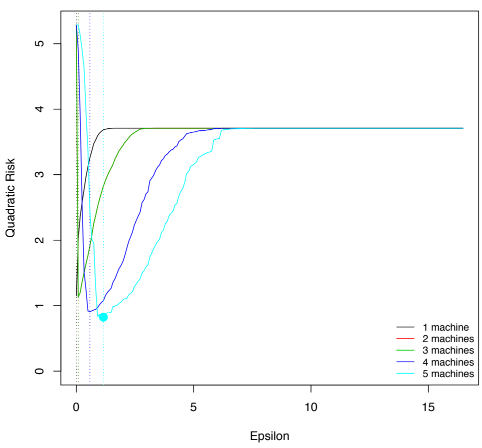
(a) Model 5, uncorrelated design.



(b) Model 5, correlated design.



(c) Model 9.



(d) Model 12.

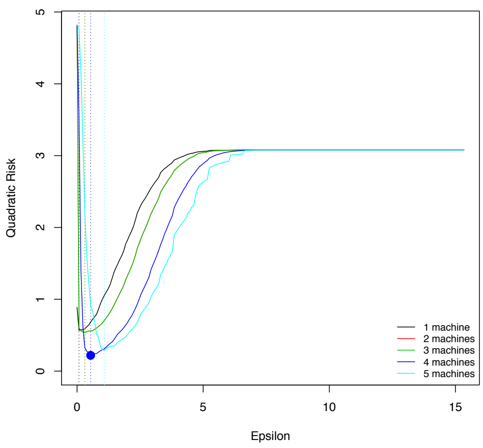
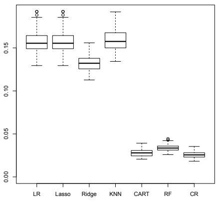
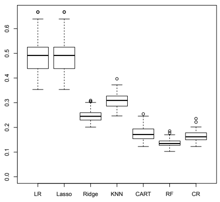


Figure 3 (SM): Boxplots of quadratic errors, uncorrelated design. From left to right: `lars`, `ridge`, `fnn`, `tree`, `randomForest`, `COBRA`.

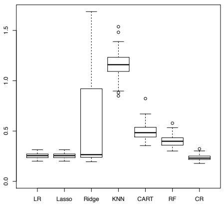
(a) Model 1.



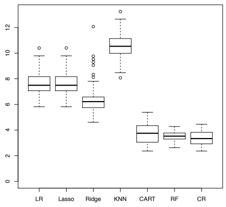
(b) Model 2.



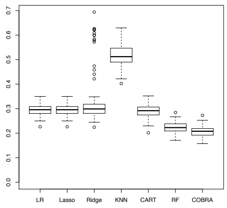
(c) Model 3.



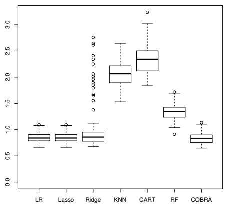
(d) Model 4.



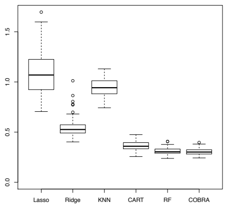
(e) Model 5.



(f) Model 6.



(g) Model 7.



(h) Model 8.

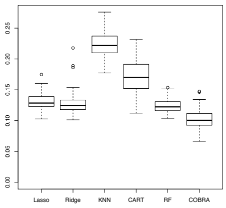
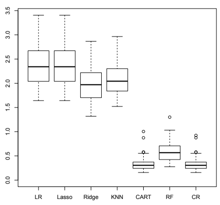
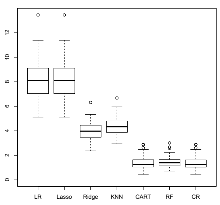


Figure 4 (SM): Boxplots of quadratic errors, correlated design. From left to right: `lars`, `ridge`, `fnn`, `tree`, `randomForest`, `COBRA`.

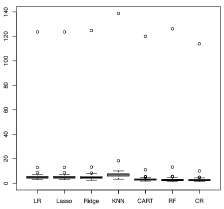
(a) Model 1.



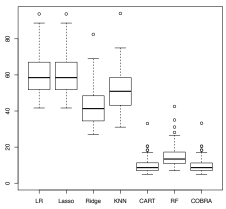
(b) Model 2.



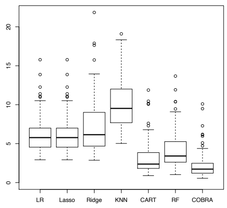
(c) Model 3.



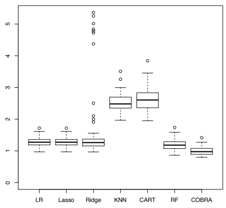
(d) Model 4.



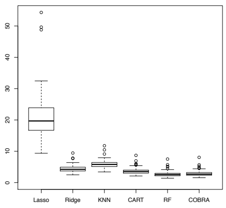
(e) Model 5.



(f) Model 6.



(g) Model 7.



(h) Model 8.

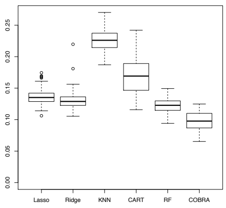
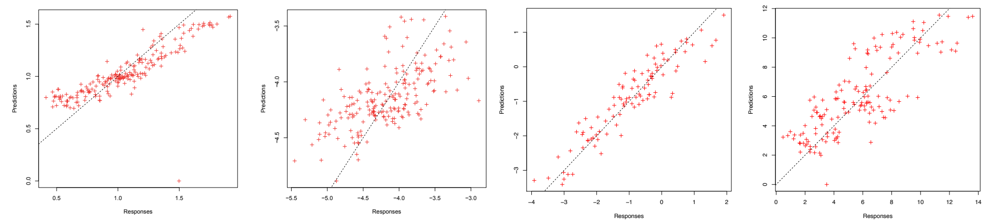


Figure 5 (SM): Prediction over the testing set, uncorrelated design. The more points on the first bissectrix, the better the prediction.

(a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.



(e) Model 5. (f) Model 6. (g) Model 7. (h) Model 8.

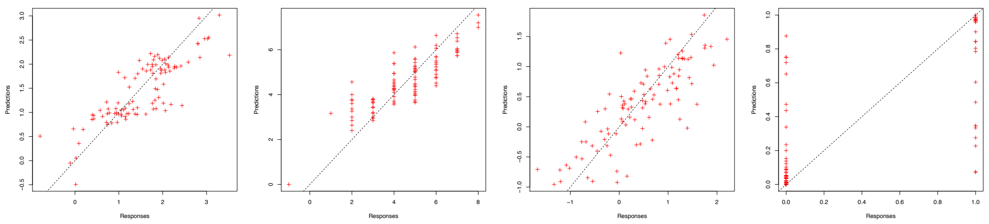
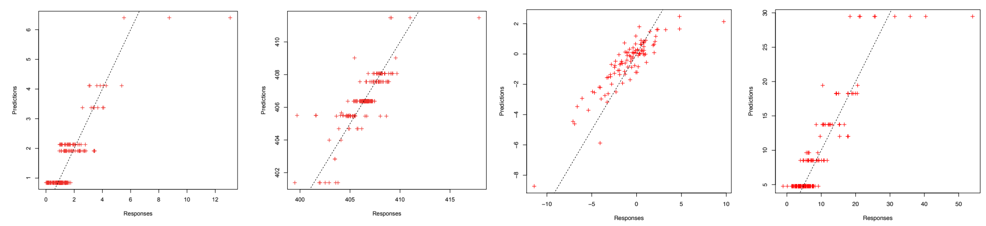


Figure 6 (SM): Prediction over the testing set, correlated design. The more points on the first bissectrix, the better the prediction.

(a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.



(e) Model 5. (f) Model 6. (g) Model 7. (h) Model 8.

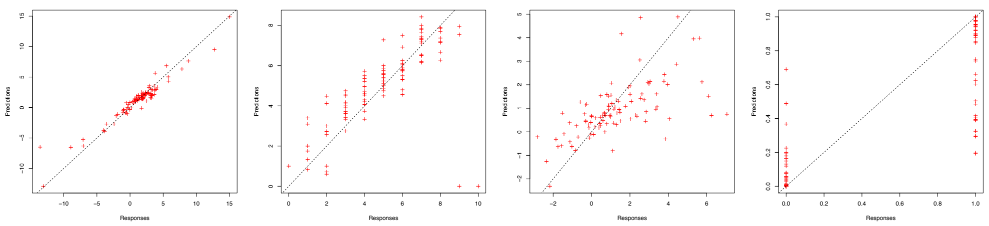
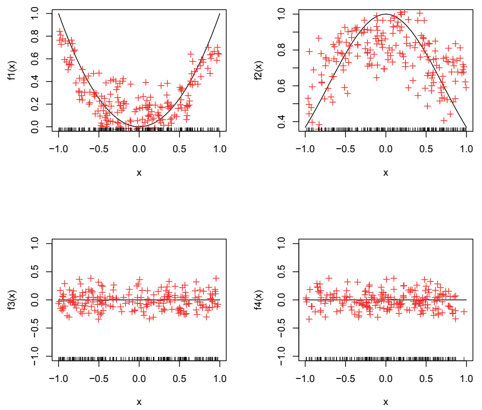
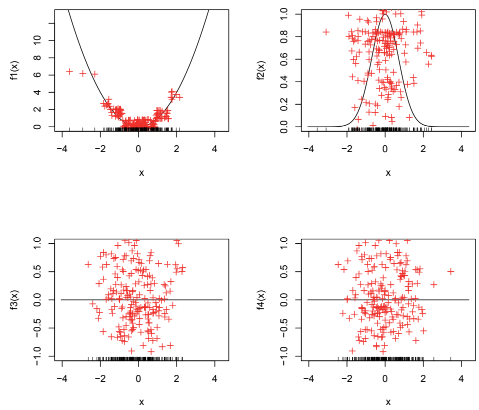


Figure 7 (SM): Examples of reconstruction of the functional dependencies, for covariates 1 to 4.

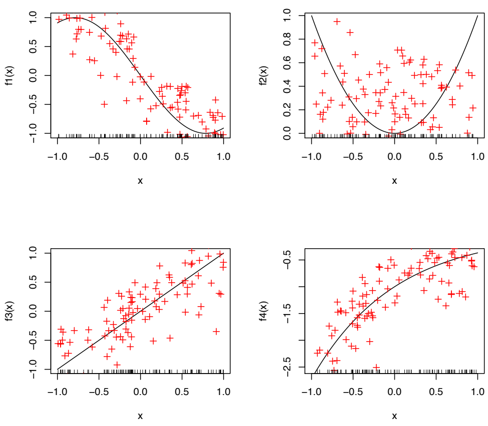
(a) Model 1, uncorrelated design.



(b) Model 1, correlated design.



(c) Model 3, uncorrelated design.



(d) Model 3, correlated design.

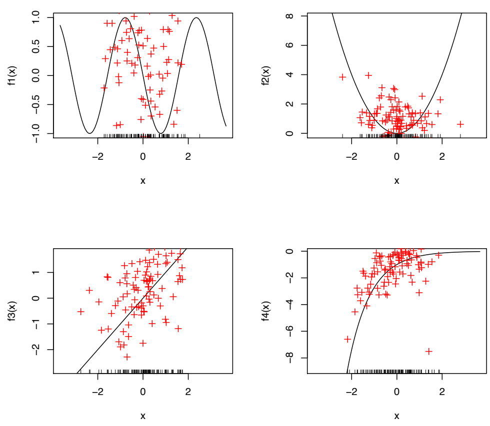


Figure 8 (SM): Boxplot of errors, high-dimensional models.

(a) Model 9 (b) Model 10 (c) Model 11

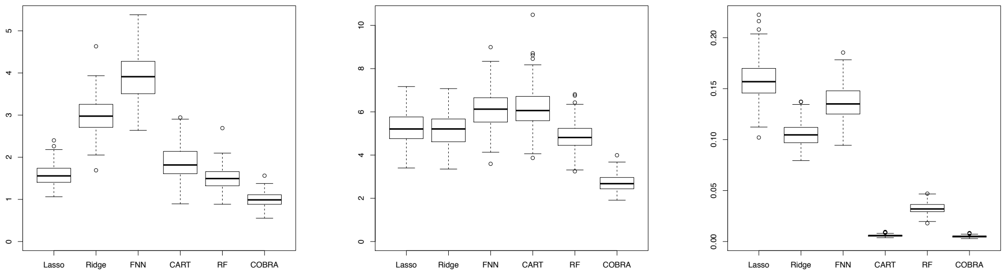


Figure 9 (SM): How stable is COBRA?

(a) Boxplot of errors: Initial sample is randomly cut (1000 replications of Model 12).

(b) Empirical risk with respect to the size of subsample $\mathscr{D}_k$, in Model 12.

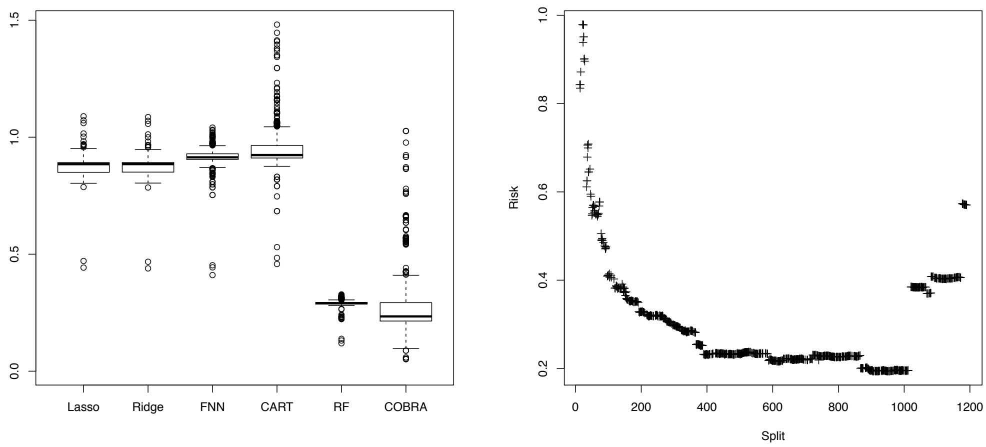


Figure 10 (SM): Boxplot of errors: EWA vs COBRA

(a) Model 9. (b) Model 10. (c) Model 11. (d) Model 12.

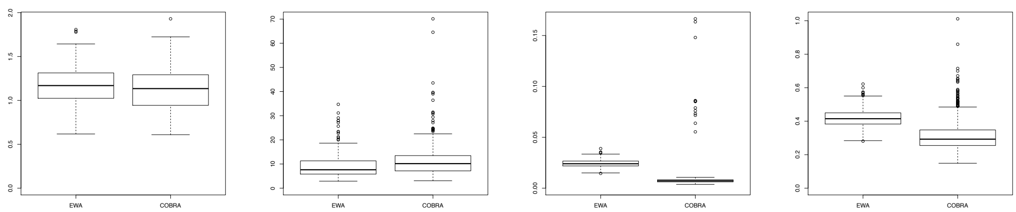
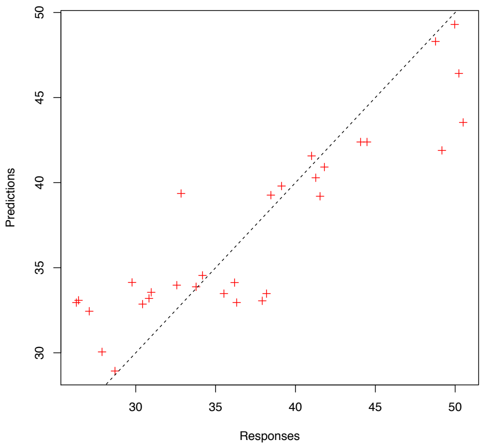
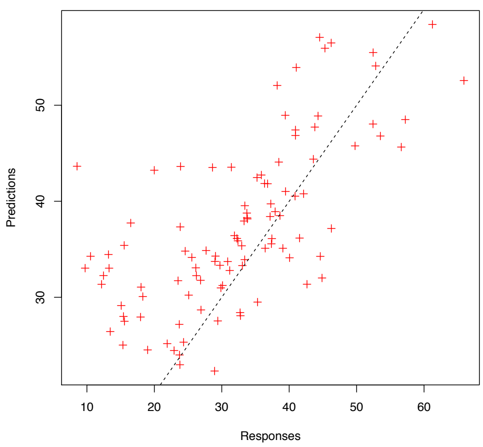


Figure 11 (SM): Prediction over the testing set, real-life data sets.

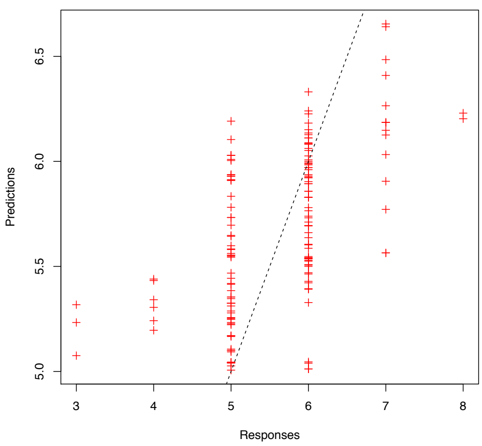
(a) Concrete Slump Test.



(b) Concrete Compressive Strength.



(c) Wine Quality, red wine.



(d) Wine Quality, white wine.

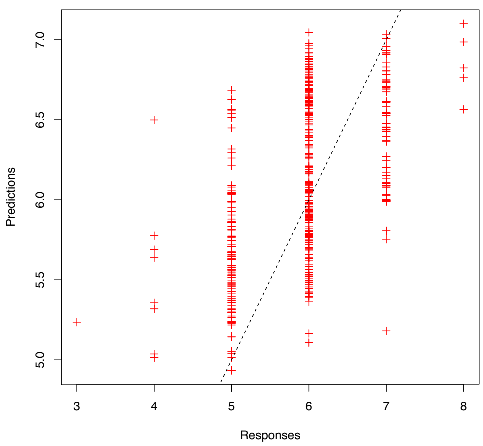


Figure 12 (SM): Boxplot of quadratic errors, real-life data sets.

(a) Concrete Slump Test.

(b) Concrete Compressive Strength.

(c) Wine Quality, red wine.

(d) Wine Quality, white wine.