Linear functional regression with functional output

Gaëlle CHAGNY [(1)], Anouar MEYNAOUI [(1)] and Angelina ROCHE [(2)]

[(1)]Université de Rouen, Laboratoire de Mathématiques Raphaël Salem, France

[(2)]Université Paris-Dauphine, CEREMADE, Paris, France

Monday, August 29, 2022

Table of contents

Table of contents

Regression model
The framework

- Consider the functional linear model:

$$Y = SX + \varepsilon,$$

where X, Y, ε belong to a functional separable Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$.

Regression model
The framework

- Consider the functional linear model:

$$Y = SX + \varepsilon,$$

where X, Y, ε belong to a functional separable Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$. We assume that $\mathbb{E}[X] = \mathbb{E}[\varepsilon] = 0_{\mathbb{H}}$ and $X \perp\!\!\!\perp \varepsilon$.

Regression model
The framework

- Consider the functional linear model:

$$Y = SX + \varepsilon,$$

where X, Y, ε belong to a functional separable Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$. We assume that $\mathbb{E}[X] = \mathbb{E}[\varepsilon] = 0_{\mathbb{H}}$ and $X \perp\!\!\!\perp \varepsilon$.

- In the sequel, $\mathbb{H} = L^2([0,1])$ and

$$\begin{array}{rcl} S : L^2([0,1]) & \longrightarrow & L^2([0,1]) \\ f & \longmapsto & \int_0^1 \mathcal{S}(s,\cdot) f(s) ds, \end{array}$$

where $\mathcal{S} \in L^2([0,1]^2)$ is the kernel function of S.

Regression model
The framework

- Consider the functional linear model:

$$Y = SX + \varepsilon,$$

where X, Y, ε belong to a functional separable Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$. We assume that $\mathbb{E}[X] = \mathbb{E}[\varepsilon] = 0_{\mathbb{H}}$ and $X \perp\!\!\!\perp \varepsilon$.

- In the sequel, $\mathbb{H} = L^2([0,1])$ and

$$\begin{array}{rcl} S : L^2([0,1]) & \longrightarrow & L^2([0,1]) \\ f & \longmapsto & \int_0^1 \mathcal{S}(s,\cdot) f(s) ds, \end{array}$$

where $\mathcal{S} \in L^2([0,1]^2)$ is the kernel function of S.

- Denote $\mathcal{L}(\mathbb{H})$ the space of linear integral operators on $\mathbb{H}$.

Regression model

Useful notations

Regression model
Useful notations

- The covariance operator of X:

$$\begin{array}{rccc} \Gamma : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \mathbb{E}[X \otimes X(f)] = \mathbb{E}[\langle X, f \rangle X]. \end{array},$$

where the tensor product between a and b:

$$\begin{array}{rccc} b \otimes a : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \langle a, f \rangle b. \end{array}$$

Regression model
Useful notations

- The covariance operator of X:

$$\begin{array}{rccl} \Gamma : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \mathbb{E}[X \otimes X(f)] = \mathbb{E}[\langle X, f\rangle X]. \end{array} ,$$

where the tensor product between a and b:

$$\begin{array}{rccl} b \otimes a : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \langle a, f\rangle b. \end{array}$$

- The cross-covariance operator between X and Y,

$$\begin{array}{rccl} \Delta : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \mathbb{E}[Y \otimes X(f)] = \mathbb{E}[\langle X, f\rangle Y]. \end{array}$$

Regression model
Useful notations

- The covariance operator of X:

$$\begin{array}{rccl} \Gamma: & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \mathbb{E}[X \otimes X(f)] = \mathbb{E}[\langle X, f \rangle X]. \end{array},$$

where the tensor product between a and b:

$$\begin{array}{rccl} b \otimes a: & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \langle a, f \rangle b. \end{array}$$

- The cross-covariance operator between X and Y,

$$\begin{array}{rccl} \Delta: & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & f & \longmapsto & \mathbb{E}[Y \otimes X(f)] = \mathbb{E}[\langle X, f \rangle Y]. \end{array}$$

- The empirical versions of Γ and Δ:

$$\Gamma_n : f \mapsto \frac{1}{n} \sum_{i=1}^{n} \langle f, X_i \rangle X_i \quad \text{et} \quad \Delta_n : f \mapsto \frac{1}{n} \sum_{i=1}^{n} \langle f, X_i \rangle Y_i.$$

Regression model

Useful notations

- **Target.** Given an i.i.d. sample $(X_i, Y_i)_{i\in\{1,\ldots,n\}}$ of (X, Y), we aim to estimate S and study the optimality of the estimators.

Regression model

Useful notations

- **Target.** Given an i.i.d. sample $(X_i, Y_i)_{i \in \{1,\ldots,n\}}$ of (X, Y), we aim to estimate S and study the optimality of the estimators.

Output / framework	**Real**	**Functional**
Asymptotic	[Lian, 2015] [Cardot and Johannes, 2010]	[Benatia et al., 2017] [Crambes and Mas, 2013]
Non-asymptotic	[Brunel, Mas and Roche, 2016]	Post-doc Works

Regression model
Useful notations

- **Target.** Given an i.i.d. sample $(X_i, Y_i)_{i\in\{1,\dots,n\}}$ of (X, Y), we aim to estimate S and study the optimality of the estimators.

Output / framework	**Real**	**Functional**
Asymptotic	[Lian, 2015] [Cardot and Johannes, 2010]	[Benatia et al., 2017] [Crambes and Mas, 2013]
Non-asymptotic	[Brunel, Mas and Roche, 2016]	Post-doc Works

- We focus on the Mean Square Prediction Error (MSPE) of an estimator $\hat{S}_n$:

$$\mathrm{MSPE}(\hat{S}_n) = \mathbb{E}\|\hat{S}_n(X_{n+1}) - S(X_{n+1})\|^2,$$

where X_{n+1} is a new observation of X and $X_{n+1} \perp\!\!\!\perp (X_i, \varepsilon_i), i = 1, \dots, n$.

Regression model

Useful notations

- **Target.** Given an i.i.d. sample $(X_i, Y_i)_{i\in\{1,\dots,n\}}$ of (X, Y), we aim to estimate S and study the optimality of the estimators.

Output / framework	**Real**	**Functional**
Asymptotic	[Lian, 2015] [Cardot and Johannes, 2010]	[Benatia et al., 2017] [Crambes and Mas, 2013]
Non-asymptotic	[Brunel, Mas and Roche, 2016]	Post-doc Works

- We focus on the Mean Square Prediction Error (MSPE) of an estimator $\hat{S}_n$:

$$\mathrm{MSPE}(\hat{S}_n) = \mathbb{E}\|\hat{S}_n(X_{n+1}) - S(X_{n+1})\|^2,$$

where X_{n+1} is a new observation of X and $X_{n+1} \perp\!\!\!\perp (X_i, \varepsilon_i), i = 1, \dots, n$.

- We show that

$$\mathrm{MSPE}(\hat{S}_n) = \mathbb{E}\|\hat{S}_n\Gamma^{1/2} - S\Gamma^{1/2}\|_{\mathrm{HS}}^2,$$

where $\Gamma^{1/2} = \sum_{j\geq 1}\sqrt{\lambda_j}\varphi_j \otimes \varphi_j$ with $(\lambda_j, \varphi_j)_{j\geq 1}$ are the eigenelements of Γ.

Regression model

Small reminder

Minimax optimality

Let $\mathcal{C}_\delta$ be a regularity space of S. The non-asymptotic minimax prediction risk over $\mathcal{C}_\delta$ is defined as

$$\underline{\mathcal{R}}_n(C_\delta) = \inf_{\hat{S}_n} \sup_{S \in \mathcal{C}_\delta} \text{MSPE}(\hat{S}_n).$$

Regression model

Small reminder

Minimax optimality

Let $\mathcal{C}_\delta$ be a regularity space of S. The non-asymptotic minimax prediction risk over $\mathcal{C}_\delta$ is defined as

$$\underline{\mathcal{R}}_n(C_\delta) = \inf_{\hat{S}_n} \sup_{S \in \mathcal{C}_\delta} \text{MSPE}(\hat{S}_n).$$

- $\hat{S}_n$ is non-asymptotically minimax over C_δ if

$$\text{MSPE}(\hat{S}_n) \leq C\underline{\mathcal{R}}_n(C_\delta),$$

where $C > 0$.

Regression model
Small reminder

Minimax optimality

Let $\mathcal{C}_\delta$ be a regularity space of S. The non-asymptotic minimax prediction risk over $\mathcal{C}_\delta$ is defined as

$$\underline{\mathcal{R}}_n(C_\delta) = \inf_{\hat{S}_n} \sup_{S \in \mathcal{C}_\delta} \mathrm{MSPE}(\hat{S}_n).$$

- $\hat{S}_n$ is non-asymptotically minimax over C_δ if

$$\mathrm{MSPE}(\hat{S}_n) \leq C \underline{\mathcal{R}}_n(C_\delta),$$

where $C > 0$.

- $\hat{S}_n$ is non-asymptotically adaptive over C_δ if it does not depend on δ.

Regression model

Small reminder

Minimax optimality

Let $\mathcal{C}_\delta$ be a regularity space of S. The non-asymptotic minimax prediction risk over $\mathcal{C}_\delta$ is defined as

$$\underline{\mathcal{R}}_n(C_\delta) = \inf_{\hat{S}_n} \sup_{S \in \mathcal{C}_\delta} \mathrm{MSPE}(\hat{S}_n).$$

- $\hat{S}_n$ is non-asymptotically minimax over C_δ if

$$\mathrm{MSPE}(\hat{S}_n) \le C \underline{\mathcal{R}}_n(C_\delta),$$

  where $C > 0$.

- $\hat{S}_n$ is non-asymptotically adaptive over C_δ if it does not depend on δ.

We aim to construct **minimax/adaptive** estimators of S.

Table of contents

Estimation
Projection estimators

- Define the contrast function γ_n by

$$\gamma_n : \begin{array}{ccc} \mathcal{L}(\mathbb{H}) & \longrightarrow & \mathbb{R}_+ \\ T & \longmapsto & 1/n \sum_{i=1}^{n} \| Y_i - T(X_i) \|^2 . \end{array}$$

Estimation
Projection estimators

- Define the contrast function γ_n by

$$\begin{array}{rccc}\gamma_n : & \mathcal{L}(\mathbb{H}) & \longrightarrow & \mathbb{R}_+ \\ & T & \longmapsto & 1/n\sum_{i=1}^n \|Y_i - T(X_i)\|^2 .\end{array}$$

- Let $\psi = (\psi_j)_{j\in\mathbb{N}^*}$ be an orthonormal basis of $\mathbb{H} = L^2([0,1])$. We introduce the collection of models defined for m_1, m_2 in $\mathbb{N}^*$ by

$$E_{m_1,m_2} = \mathrm{Span}\{\psi_k \otimes \psi_j,\ 1 \leq j \leq m_1, 1 \leq k \leq m_2\}.$$

Estimation
Projection estimators

- Define the contrast function γ_n by

$$\gamma_n : \begin{array}{ccc} \mathcal{L}(\mathbb{H}) & \longrightarrow & \mathbb{R}_+ \\ T & \longmapsto & 1/n \sum_{i=1}^{n} \left\| Y_i - T(X_i) \right\|^2 . \end{array}$$

- Let $\psi = (\psi_j)_{j \in \mathbb{N}^*}$ be an orthonormal basis of $\mathbb{H} = L^2([0,1])$. We introduce the collection of models defined for m_1, m_2 in $\mathbb{N}^*$ by

$$E_{m_1,m_2} = \mathrm{Span}\{\psi_k \otimes \psi_j,\ 1 \leq j \leq m_1, 1 \leq k \leq m_2\}.$$

- We define projection estimators of S by

$$\hat{S}_{m_1,m_2} \in \mathrm{argmin}_{T \in E_{m_1,m_2}} \gamma_n(T).$$

Estimation

Explicit form

- We introduce empirical scalar products defined for g, h in $L^2([0,1])$ by

$$\langle g, h\rangle_n^X = \langle \Gamma_n(g), h\rangle \quad \text{and} \quad \langle g, h\rangle_n^{X,Y} = \langle \Delta_n(g), h\rangle.$$

- Let also A and Y_ψ be the matrices defined by

$$A = \left(\langle \psi_j, \psi_k\rangle_n^X\right)_{\substack{1\leq j\leq m_1\\ 1\leq k\leq m_1}} \quad \text{and} \quad (Y_\psi)_{j,k} = \left(\langle \psi_j, \psi_k\rangle_n^{X,Y}\right)_{\substack{1\leq j\leq m_1\\ 1\leq k\leq m_2}}.$$

Estimation
Explicit form

- We introduce empirical scalar products defined for g, h in $L^2([0,1])$ by

$$\langle g, h\rangle_n^X = \langle \Gamma_n(g), h\rangle \quad \text{and} \quad \langle g, h\rangle_n^{X,Y} = \langle \Delta_n(g), h\rangle.$$

- Let also A and Y_ψ be the matrices defined by

$$A = \left(\langle \psi_j, \psi_k\rangle_n^X\right)_{\substack{1\le j\le m_1 \\ 1\le k\le m_1}} \quad \text{and} \quad (Y_\psi)_{j,k} = \left(\langle \psi_j, \psi_k\rangle_n^{X,Y}\right)_{\substack{1\le j\le m_1 \\ 1\le k\le m_2}}.$$

Proposition (Chagny, Meynaoui and Roche, 2022)

*If A is **invertible**, then $\hat{S}_{m_1,m_2}$ and $\hat{\mathcal{S}}_{m_1,m_2}$ are uniquely defined by*

$$\hat{S}_{m_1,m_2} = \sum_{\substack{1\le j\le m_1 \\ 1\le k\le m_2}} \hat{b}_{j,k}\psi_k \otimes \psi_j \quad \text{and} \quad \hat{\mathcal{S}}_{m_1,m_2} : (s,t) \mapsto \sum_{\substack{1\le j\le m_1 \\ 1\le k\le m_2}} \hat{b}_{j,k}\psi_j(s)\psi_k(t),$$

where $\hat{b}_{j,k} = (A^{-1}Y_\psi)_{j,k}$.

Estimation
Explicit form

- We denote by $(\hat{\lambda}_j, \hat{\varphi}_j)_{j\geq 1}$ the eigenelements of Γ_n, where the $\hat{\lambda}_j$'s are sorted in a decreasing order.

Estimation
Explicit form

- We denote by $(\hat{\lambda}_j, \hat{\varphi}_j)_{j\geq 1}$ the eigenelements of Γ_n, where the $\hat{\lambda}_j$'s are sorted in a decreasing order.

- **Particular case.** If $\psi = (\hat{\varphi}_j)_{j\geq 1}$, then

$$\hat{S}_{m_1,m_2} = \sum_{\substack{1\leq j\leq m_1 \\ 1\leq k\leq m_2}} \frac{1}{\hat{\lambda}_j} \langle \Delta_n(\hat{\varphi}_j), \hat{\varphi}_k \rangle \hat{\varphi}_k \otimes \hat{\varphi}_j.$$

Estimation

Explicit form

- We denote by $(\hat{\lambda}_j, \hat{\varphi}_j)_{j\geq 1}$ the eigenelements of Γ_n, where the $\hat{\lambda}_j$'s are sorted in a decreasing order.

- **Particular case.** If $\psi = (\hat{\varphi}_j)_{j\geq 1}$, then

$$\hat{S}_{m_1,m_2} = \sum_{\substack{1\leq j\leq m_1 \\ 1\leq k\leq m_2}} \frac{1}{\hat{\lambda}_j} \langle \Delta_n(\hat{\varphi}_j), \hat{\varphi}_k \rangle \hat{\varphi}_k \otimes \hat{\varphi}_j.$$

- Define the "pseudo-inverse" of Γ_n as $\Gamma_n^\dagger = \sum_{j=1}^{m_1} 1/\hat{\lambda}_j \hat{\varphi}_j \otimes \hat{\varphi}_j$. Then

$$\hat{S}_{m_1,m_2} = \hat{\Pi}_{m_2} \Delta_n \Gamma_n^\dagger,$$

where $\hat{\Pi}_{m_2}$ is the projection onto $\text{Span}\{\hat{\varphi}_1, \ldots, \hat{\varphi}_{m_2}\}$.

Estimation
Explicit form

- We denote by $(\hat{\lambda}_j, \hat{\varphi}_j)_{j\geq 1}$ the eigenelements of Γ_n, where the $\hat{\lambda}_j$'s are sorted in a decreasing order.
- **Particular case.** If $\psi = (\hat{\varphi}_j)_{j\geq 1}$, then

$$\hat{S}_{m_1,m_2} = \sum_{\substack{1\leq j\leq m_1 \\ 1\leq k\leq m_2}} \frac{1}{\hat{\lambda}_j} \langle \Delta_n(\hat{\varphi}_j), \hat{\varphi}_k \rangle \hat{\varphi}_k \otimes \hat{\varphi}_j.$$

- Define the "pseudo-inverse" of Γ_n as $\Gamma_n^{\dagger} = \sum_{j=1}^{m_1} 1/\hat{\lambda}_j \hat{\varphi}_j \otimes \hat{\varphi}_j$. Then

$$\hat{S}_{m_1,m_2} = \hat{\Pi}_{m_2} \Delta_n \Gamma_n^{\dagger},$$

where $\hat{\Pi}_{m_2}$ is the projection onto $\mathrm{Span}\{\hat{\varphi}_1, \ldots, \hat{\varphi}_{m_2}\}$.

- From $S\Gamma = \Delta \Rightarrow$ Estimators of [Crambes and Mas, 2013]: $\hat{S}_{CM} = \Delta_n \Gamma_n^{\dagger}$.

Table of contents

Upper-bound of the prediction risk
Assumptions

Assumptions on S:

$\mathcal{A}_0$ **:** We assume that S is Hilbert-Schmidt.

Upper-bound of the prediction risk
Assumptions

Assumptions on S:

$\mathcal{A}_0$ **:** We assume that S is Hilbert-Schmidt.

- Consider the regularity space of type ellipsoid (Brunel, Mas and Roche, 2016),

$$\mathcal{W}^R_{\alpha,\beta} = \left\{ T \in \mathcal{L}(\mathbb{H}), \sum_{j=1}^{+\infty} \sum_{r=1}^{+\infty} \eta_\alpha(j)\psi_\beta(r)\langle T(\varphi_j), \varphi_r\rangle^2 \leq R^2 \right\},$$

where $\alpha, \beta, R > 0$ and for all $\gamma > 0$, the functions η_γ and ψ_γ are defined as $u \mapsto u^\gamma$ (polynomial case) or $u \mapsto \exp(u^\gamma)$ (exponential case).

Upper-bound of the prediction risk

Assumptions

Assumptions on S:

$\mathcal{A}_0$ **:** We assume that S is Hilbert-Schmidt.

- Consider the regularity space of type ellipsoid (Brunel, Mas and Roche, 2016),

$$\mathcal{W}_{\alpha,\beta}^{R} = \left\{ T \in \mathcal{L}(\mathbb{H}), \sum_{j=1}^{+\infty} \sum_{r=1}^{+\infty} \eta_\alpha(j)\psi_\beta(r)\langle T(\varphi_j), \varphi_r\rangle^2 \leq R^2 \right\},$$

where $\alpha, \beta, R > 0$ and for all $\gamma > 0$, the functions η_γ and ψ_γ are defined as $u \mapsto u^\gamma$ (polynomial case) or $u \mapsto \exp(u^\gamma)$ (exponential case).

$\mathcal{A}_1$ **:** We assume that $S\Gamma^{1/2} \in \mathcal{W}_{\alpha,\beta}^{R}$ for some $\alpha, \beta, R > 0$.

Upper-bound of the prediction risk

Assumptions

Assumptions on X:

Upper-bound of the prediction risk

Assumptions

Assumptions on X:

- $\mathcal{A}_3$ **:** There exists a convex positive function λ such that, for all j in $\mathbb{N}^*$: $\lambda_j = \lambda(j)$.
- $\mathcal{A}_4$ **:** There exists a constant $\gamma > 0$ for which $\left(j\lambda_j \max\{\log^{1+\gamma}(j), 1\}\right)_{j\geq 1}$ decreases.

Upper-bound of the prediction risk
Assumptions

Assumptions on X:

- $\mathcal{A}_3$ **:** There exists a convex positive function λ such that, for all j in $\mathbb{N}^*$: $\lambda_j = \lambda(j)$.
- $\mathcal{A}_4$ **:** There exists a constant $\gamma > 0$ for which $\left(j\lambda_j \max\{\log^{1+\gamma}(j), 1\}\right)_{j\geq 1}$ decreases.
- $\mathcal{A}_5$ **:** There exists a constant $b > 0$ such that, for all l in $\mathbb{N}^*$,

$$\sup_{j\geq 1} \mathbb{E}\left[\frac{\langle X, \varphi_j\rangle^{2l}}{\lambda_j^l}\right] \leq l! b^{l-1}.$$

Upper-bound of the prediction risk

Assumptions

Assumptions on X***:***

- $\mathcal{A}_3$ **:** There exists a convex positive function λ such that, for all j in $\mathbb{N}^*$: $\lambda_j = \lambda(j)$.
- $\mathcal{A}_4$ **:** There exists a constant $\gamma > 0$ for which $\left(j\lambda_j \max\{\log^{1+\gamma}(j), 1\}\right)_{j\geq 1}$ decreases.
- $\mathcal{A}_5$ **:** There exists a constant $b > 0$ such that, for all l in $\mathbb{N}^*$,

$$\sup_{j\geq 1} \mathbb{E}\left[\frac{\langle X, \varphi_j\rangle^{2l}}{\lambda_j^l}\right] \leq l! b^{l-1}.$$

- $\mathcal{A}_6$ **:** For all $j \neq k$, $\langle X, \varphi_j\rangle$ and $\langle X, \varphi_k\rangle$ are independent.

Upper-bound of the prediction risk

Theoretical result

Theorem (Chagny, Meynaoui and Roche, 2022)

If ψ_β is polynomial with $\beta > 6$ or exponential and under Assumptions $\mathcal{A}_0$ to $\mathcal{A}_6$

$$\inf_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \leq n/\ln^2(n)}} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_{m_1,m_2}) \leq \inf_{\substack{m \in \mathbb{N}^* \\ m_1 \leq n/\ln^2(n)}} \left\{ \sigma_\varepsilon^2 \frac{m}{n} + \frac{3}{\eta_\alpha(m)} \right\} + \frac{c}{n},$$

where $\sigma_\varepsilon^2 = \mathbb{E}\|\varepsilon\|^2$.

- In the polynomial case of η_α, we have for $m_1 \sim n^{1/(1+\alpha)}$ and $m_2 \to +\infty$ that

$$\inf_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \leq n/\ln^2(n)}} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_{m_1,m_2}) \leq C n^{-\alpha/(\alpha+1)}.$$

- In the exponential case of η_α, we show that

$$\inf_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \leq n/\ln^2(n)}} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_{m_1,m_2}) \leq C \frac{\ln(n)^{1/\alpha}}{n}.$$

Table of contents

Lower-bound of the minimax risk
Theoretical result

- **Lower bound** of the minimax prediction risk.

Theorem (Chagny, Meynaoui and Roche, 2022)

Assume that $S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}$, avec $\alpha, \beta, R > 0$. If ε is Gaussian, we have

$$\inf_{\hat{S}_n} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_n) \geq C \inf_{m \in \mathbb{N}^*} \left\{ \sigma_\varepsilon^2 \frac{m}{n} + \frac{3}{\eta_\alpha(m)} \right\},$$

where $C > 0$ and $\sigma_\varepsilon^2 = \mathbb{E}\|\varepsilon\|^2$.

- In the polynomial case of η_α,

$$\inf_{\hat{S}_n} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_n) \geq C n^{-\alpha/(\alpha+1)}.$$

- In the exponential case of η_α,

$$\inf_{\hat{S}_n} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_n) \geq C \frac{\ln(n)^{1/\alpha}}{n}.$$

Model selection

- Fix $m_2 = \infty$, and consider $(\widehat{S}_{m_1,\infty})_{m_1 \in \mathcal{M}_n}$ with
  $\mathcal{M}_n = \{1, \dots, N_n\}, \quad N_n \leq n/\ln^2(n)$, and $\widehat{S}_{m_1,\infty} = \Delta_n \Gamma^{\dagger}_{n,m_1}$.
  How to select the "best" estimator in the collection ?

Model selection

- Fix $m_2 = \infty$, and consider $(\widehat{S}_{m_1,\infty})_{m_1 \in \mathcal{M}_n}$ with $\mathcal{M}_n = \{1, \ldots, N_n\}, \quad N_n \leq n/\ln^2(n)$, and $\widehat{S}_{m_1,\infty} = \Delta_n \Gamma^{\dagger}_{n,m_1}$.
  How to select the "best" estimator in the collection ?

- Contrast penalization method (Barron *et al.*, 1999, Massart, 2007)

$$\widehat{m}_1 = \arg \min_{m_1 \in \mathcal{M}_n} \left(\gamma_n(\widehat{S}_{m_1,\infty}) + \mathrm{pen}(m_1) \right),$$

  where pen is the penalty function defined as

$$\mathrm{pen} : m_1 \mapsto \kappa \sigma_\varepsilon^2 \frac{m_1}{n},$$

  with $\kappa > 0$ a numerical constant that will be tuned in practice.

Model selection

- Fix $m_2 = \infty$, and consider $(\widehat{S}_{m_1,\infty})_{m_1 \in \mathcal{M}_n}$ with $\mathcal{M}_n = \{1, \dots, N_n\}, \quad N_n \leq n/\ln^2(n)$, and $\widehat{S}_{m_1,\infty} = \Delta_n \Gamma^{\dagger}_{n,m_1}$.
  How to select the "best" estimator in the collection ?

- Contrast penalization method (Barron *et al.*, 1999, Massart, 2007)

$$\widehat{m}_1 = \arg \min_{m_1 \in \mathcal{M}_n} \left(\gamma_n(\widehat{S}_{m_1,\infty}) + \mathrm{pen}(m_1) \right),$$

  where pen is the penalty function defined as

$$\mathrm{pen} : m_1 \mapsto \kappa \sigma_\varepsilon^2 \frac{m_1}{n},$$

  with $\kappa > 0$ a numerical constant that will be tuned in practice.

- Additional assumption on ε.

$$\exists p > 6, \quad \tau_p = \mathbb{E}\|\varepsilon\|^p < +\infty.$$

Oracle-type inequality

- **Empirical norm.**

$$\|T\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\|T(X_i)\|^2, \quad T \in \mathcal{L}(\mathbb{H}).$$

Theorem

Under the previous assumption, for all $\zeta > 0$,

$$\mathbb{E}\|S - \widehat{S}_{\widehat{m}_1,\infty}\|_n^2 \leq (1+\zeta)\inf_{m_1\in\mathcal{M}_n}\left\{\mathbb{E}\|S - \widehat{\Pi}^{op}_{m_1,\infty}S\|_n^2 + c(\zeta)\,\mathrm{pen}(m_1)\right\} + \frac{C}{n},$$

where $c(\zeta) = (2+\zeta)/(1+\zeta)$ *and* $\widehat{\Pi}^{op}_{m_1,\infty}$ *is the orthogonal projection onto the closure of* $V_{m_1,\infty} = \mathrm{Span}\{\widehat{\varphi}_k \otimes \widehat{\varphi}_j, 1 \leq j \leq m_1, m_2 \geq 1\}$.

Oracle-type inequality

- **Empirical norm.**

$$\|T\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} \|T(X_i)\|^2, \quad T \in \mathcal{L}(\mathbb{H}).$$

Theorem

Under the previous assumption, for all $\zeta > 0$,

$$\mathbb{E}\|S - \widehat{S}_{\widehat{m}_1,\infty}\|_n^2 \leq (1+\zeta) \inf_{m_1 \in \mathcal{M}_n} \left\{ \mathbb{E}\|S - \widehat{\Pi}^{op}_{m_1,\infty} S\|_n^2 + c(\zeta)\,\mathrm{pen}(m_1) \right\} + \frac{C}{n},$$

where $c(\zeta) = (2+\zeta)/(1+\zeta)$ and $\widehat{\Pi}^{op}_{m_1,\infty}$ is the orthogonal projection onto the closure of $V_{m_1,\infty} = \mathrm{Span}\{\widehat{\varphi}_k \otimes \widehat{\varphi}_j, 1 \leq j \leq m_1, m_2 \geq 1\}$.

$\hookrightarrow$ best bias-variance compromise for $\widehat{S}_{\widehat{m}_1,\infty}$ up to constants.

Oracle-type inequality

- **Empirical norm.**

$$\|T\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\|T(X_i)\|^2, \quad T \in \mathcal{L}(\mathbb{H}).$$

Theorem

Under the previous assumption, for all $\zeta > 0$,

$$\mathbb{E}\|S - \widehat{S}_{\widehat{m}_1,\infty}\|_n^2 \leq (1+\zeta)\inf_{m_1 \in \mathcal{M}_n}\left\{\mathbb{E}\|S - \widehat{\Pi}^{op}_{m_1,\infty}S\|_n^2 + c(\zeta)\,\mathrm{pen}(m_1)\right\} + \frac{C}{n},$$

where $c(\zeta) = (2+\zeta)/(1+\zeta)$ *and* $\widehat{\Pi}^{op}_{m_1,\infty}$ *is the orthogonal projection onto the closure of* $V_{m_1,\infty} = \mathrm{Span}\{\widehat{\varphi}_k \otimes \widehat{\varphi}_j, 1 \leq j \leq m_1, m_2 \geq 1\}$.

$\hookrightarrow$ best bias-variance compromise for $\widehat{S}_{\widehat{m}_1,\infty}$ up to constants.

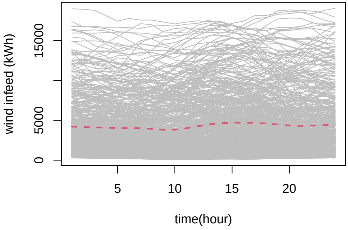
$\hookrightarrow$ similar result possible for the MSPE, with additional technicalities (Brunel *et al.*, 2016).

Table of contents

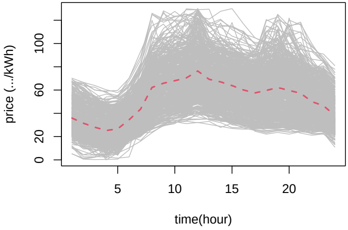
Real data set

Problem : predict the evolution of electricity prices from the wind power in-feed in Germany.

- Liebl (2013), `https://www.dliebl.com`
- $(X_i(t), Y_i(t)) =$(electricity price, wind power in-feed) at time $t \in \{1, \ldots, 24\}$, $i \in \{1, \ldots, n = 516\}$ ($i =$ a day).



Wind power in-feed curves

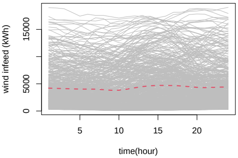


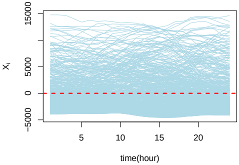
Prices curves

Real data set - Transformation of the data

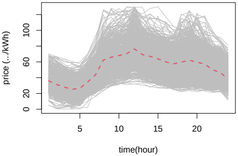
- Remove the outliers
- Re-center the data
- Take the log of the original prices, $+1$

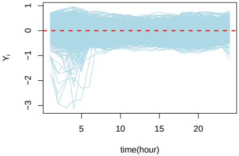
Original data | Transformed data





X



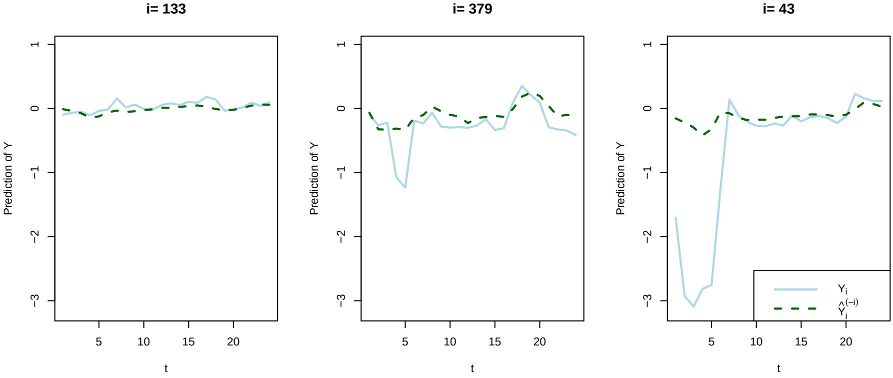


Y

Real data set - prediction results

Cross-validated prediction $\widehat{Y}_i^{-i} = \widehat{S}^{-i}_{\widehat{m}_1^{(-i)},\infty}(X_i)$ versus Y_i for three days.

- $i = 133$: day for which the distance $\|Y_i - \widehat{Y}_i^{-i}\|$ is minimal,
- $i = 379$: median prediction error
- $i = 43$: maximal prediction error.



Conclusion

Summary

Conclusion

Summary

- Estimation of the functional linear regression model by projection.

Conclusion

Summary

- Estimation of the functional linear regression model by projection.
- Upper-bound of the prediction risk w.r.t. the projection dimensions.

Conclusion

Summary

- Estimation of the functional linear regression model by projection.
- Upper-bound of the prediction risk w.r.t. the projection dimensions.
- Lower-bound of the minimax risk over ellipsoidal regularity spaces.

Conclusion

Summary

- Estimation of the functional linear regression model by projection.
- Upper-bound of the prediction risk w.r.t. the projection dimensions.
- Lower-bound of the minimax risk over ellipsoidal regularity spaces.
- Showing the optimality of the adaptive estimators.

Conclusion

Summary

- Estimation of the functional linear regression model by projection.
- Upper-bound of the prediction risk w.r.t. the projection dimensions.
- Lower-bound of the minimax risk over ellipsoidal regularity spaces.
- Showing the optimality of the adaptive estimators.
- Performing simulations on a real data case.

Conclusion

Summary

- Estimation of the functional linear regression model by projection.
- Upper-bound of the prediction risk w.r.t. the projection dimensions.
- Lower-bound of the minimax risk over ellipsoidal regularity spaces.
- Showing the optimality of the adaptive estimators.
- Performing simulations on a real data case.

What's next?

- Proposition of minimax/adaptive kernel estimators for the functional linear regression model.

Thank you for your attention

Appendix

- We can **equivalently** define the contrast function γ_n' by

$$\begin{array}{rccl} \gamma_n' : & L^2([0,1]^2) & \longrightarrow & \mathbb{R}_+ \\ & \mathcal{T} & \longmapsto & 1/n \sum_{i=1}^n \left\| Y_i - \int_{[0,1]} \mathcal{T}(s,\cdot) X_i(s) ds \right\|^2 . \end{array}$$

Appendix

- We can **equivalently** define the contrast function γ'_n by

$$\begin{array}{llll} \gamma'_n : & L^2([0,1]^2) & \longrightarrow & \mathbb{R}_+ \\ & \mathcal{T} & \longmapsto & 1/n \sum_{i=1}^n \left\| Y_i - \int_{[0,1]} \mathcal{T}(s,\cdot) X_i(s) ds \right\|^2 . \end{array}$$

- We introduce the collection of models defined for m_1, m_2 in $\mathbb{N}^*$ by

$$E'_{m_1,m_2} = \mathrm{Span}\{(t,s) \mapsto \psi_j(s)\psi_k(t),\ 1 \le j \le m_1, 1 \le k \le m_2\}.$$

Appendix

- We can **equivalently** define the contrast function γ'_n by

$$\begin{array}{rccl} \gamma'_n : & L^2([0,1]^2) & \longrightarrow & \mathbb{R}_+ \\ & \mathcal{T} & \longmapsto & 1/n\sum_{i=1}^n \left\| Y_i - \int_{[0,1]} \mathcal{T}(s,\cdot)X_i(s)ds \right\|^2. \end{array}$$

- We introduce the collection of models defined for m_1, m_2 in $\mathbb{N}^*$ by

$$E'_{m_1,m_2} = \mathrm{Span}\{(t,s) \mapsto \psi_j(s)\psi_k(t),\ 1 \leq j \leq m_1, 1 \leq k \leq m_2\}.$$

- We define projection estimators of $\mathcal{S}$ by

$$\hat{\mathcal{S}}_{m_1,m_2} \in \mathrm{argmin}_{\mathcal{T}\in E'_{m_1,m_2}} \gamma'_n(\mathcal{T}).$$

Appendix

- We can **equivalently** define the contrast function γ'_n by

$$\begin{array}{llll} \gamma'_n : & L^2([0,1]^2) & \longrightarrow & \mathbb{R}_+ \\ & \mathcal{T} & \longmapsto & 1/n \sum_{i=1}^n \left\| Y_i - \int_{[0,1]} \mathcal{T}(s,\cdot) X_i(s) ds \right\|^2. \end{array}$$

- We introduce the collection of models defined for m_1, m_2 in $\mathbb{N}^*$ by

$$E'_{m_1,m_2} = \mathrm{Span}\{(t,s) \mapsto \psi_j(s)\psi_k(t),\ 1 \le j \le m_1, 1 \le k \le m_2\}.$$

- We define projection estimators of $\mathcal{S}$ by

$$\hat{\mathcal{S}}_{m_1,m_2} \in \mathrm{argmin}_{\mathcal{T} \in E'_{m_1,m_2}} \gamma'_n(\mathcal{T}).$$

If $\hat{S}_{m_1,m_2}$ and $\hat{\mathcal{S}}_{m_1,m_2}$ are **uniquely** defined, then

$$\hat{S}_{m_1,m_2} : f \mapsto \int_0^1 \hat{\mathcal{S}}_{m_1,m_2}(\cdot,s) f(s)\, \mathrm{d}s.$$

Appendix

Sketch of the proof

- We show that

$$\hat{S}_{m_1,m_2} - S = \hat{\Pi}_{m_2} U_n + \hat{\Pi}_{m_2} S \hat{\Pi}_{m_1} - S,$$

where $\hat{\Pi}_m$ is the projection onto $\{\hat{\varphi}_1, \dots, \hat{\varphi}_m\}$ and $U_n = 1/n \sum_{i=1}^n \varepsilon_i \otimes \Gamma_n^\dagger(X_i)$.

Appendix
Sketch of the proof

- We show that
$$\hat{S}_{m_1,m_2} - S = \hat{\Pi}_{m_2} U_n + \hat{\Pi}_{m_2} S \hat{\Pi}_{m_1} - S,$$
where $\hat{\Pi}_m$ is the projection onto $\{\hat{\varphi}_1, \dots, \hat{\varphi}_m\}$ and $U_n = 1/n \sum_{i=1}^n \varepsilon_i \otimes \Gamma_n^{\dagger}(X_i)$.

- Decomposition of prediction risk as a bias-variance trade-off:
$$\mathbb{E}\|\hat{S}_{m_1,m_2}(X_{n+1}) - S(X_{n+1})\|^2 = \mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{HS}^2 + \mathbb{E}\|(S - \hat{\Pi}_{m_2} S \hat{\Pi}_{m_1})\Gamma^{1/2}\|_{HS}^2.$$

Appendix

Sketch of the proof

- We show that

$$\hat{S}_{m_1,m_2} - S = \hat{\Pi}_{m_2} U_n + \hat{\Pi}_{m_2} S \hat{\Pi}_{m_1} - S,$$

where $\hat{\Pi}_m$ is the projection onto $\{\hat{\varphi}_1, \ldots, \hat{\varphi}_m\}$ and $U_n = 1/n \sum_{i=1}^{n} \varepsilon_i \otimes \Gamma_n^{\dagger}(X_i)$.

- Decomposition of prediction risk as a bias-variance trade-off:

$$\mathbb{E}\|\hat{S}_{m_1,m_2}(X_{n+1}) - S(X_{n+1})\|^2 = \mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{\mathrm{HS}}^2 + \mathbb{E}\|(S - \hat{\Pi}_{m_2} S \hat{\Pi}_{m_1})\Gamma^{1/2}\|_{\mathrm{HS}}^2.$$

- The variance term is upper bounded as

$$\mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{\mathrm{HS}}^2 \leq \sigma_\varepsilon^2 \frac{m_1}{n} + A_{n,m_1},$$

where $\sigma_\varepsilon = \mathbb{E}\|\varepsilon\|^2$ and $A_{n,m_1} = \sigma_\varepsilon^2 \frac{C m_1^2 \log^2(m_1)}{n^2}$, with $C > 0$.

Appendix
Sketch of the proof

Sketch of the upper-bound of the variance term

Appendix

Sketch of the proof

Sketch of the upper-bound of the variance term

- We can show that

$$\mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{\mathrm{HS}}^2 \leq \frac{\sigma_\varepsilon^2}{n}\mathbb{E}\left[\mathrm{Tr}\left(\Gamma_n^\dagger \cdot \Gamma\right)\right]$$
$$= \frac{\sigma_\varepsilon^2}{n}\left[\mathrm{Tr}\left(\Gamma^\dagger \cdot \Gamma\right) + \mathrm{Tr}\left(\mathbb{E}\left[\left(\Gamma_n^\dagger - \Gamma^\dagger\right) \cdot \Gamma\right]\right)\right],$$

where $\Gamma^\dagger = \sum_{j=1}^{m_1} 1/\lambda_j \varphi_j \otimes \varphi_j$.

Appendix

Sketch of the proof

Sketch of the upper-bound of the variance term

- We can show that

$$\mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{\mathrm{HS}}^2 \leq \frac{\sigma_\varepsilon^2}{n} \mathbb{E}\left[\mathrm{Tr}\left(\Gamma_n^\dagger \cdot \Gamma\right)\right]$$
$$= \frac{\sigma_\varepsilon^2}{n}\left[\mathrm{Tr}\left(\Gamma^\dagger \cdot \Gamma\right) + \mathrm{Tr}\left(\mathbb{E}\left[\left(\Gamma_n^\dagger - \Gamma^\dagger\right) \cdot \Gamma\right]\right)\right],$$

where $\Gamma^\dagger = \sum_{j=1}^{m_1} 1/\lambda_j \varphi_j \otimes \varphi_j$.

- From (Crambes and Mas, 2013), we have

$$\mathrm{Tr}\left(\mathbb{E}\left[\left(\Gamma_n^\dagger - \Gamma^\dagger\right) \cdot \Gamma\right]\right) \leq \frac{C m_1^2 \log^2(m_1)}{n}.$$

Appendix
Sketch of the proof

Sketch of the upper-bound of the variance term

- We can show that

$$\begin{aligned}\mathbb{E}\|\hat{\Pi}_{m_2} U_n \Gamma^{1/2}\|_{\mathrm{HS}}^2 &\leq \frac{\sigma_\varepsilon^2}{n}\mathbb{E}\left[\mathrm{Tr}\left(\Gamma_n^\dagger \cdot \Gamma\right)\right] \\ &= \frac{\sigma_\varepsilon^2}{n}\left[\mathrm{Tr}\left(\Gamma^\dagger \cdot \Gamma\right) + \mathrm{Tr}\left(\mathbb{E}\left[(\Gamma_n^\dagger - \Gamma^\dagger)\cdot \Gamma\right]\right)\right],\end{aligned}$$

  where $\Gamma^\dagger = \sum_{j=1}^{m_1} 1/\lambda_j \varphi_j \otimes \varphi_j$.

- From (Crambes and Mas, 2013), we have

$$\mathrm{Tr}\left(\mathbb{E}\left[(\Gamma_n^\dagger - \Gamma^\dagger)\cdot \Gamma\right]\right) \leq \frac{C m_1^2 \log^2(m_1)}{n}.$$

- At last, $\mathrm{Tr}\left(\Gamma^\dagger \cdot \Gamma\right) = m_1$.

Appendix

Sketch of the proof

- Upper-bounding of the bias term ⇒ Perturbation theory (Dunford and Schwartz, 1965).

Appendix
Sketch of the proof

- Upper-bounding of the bias term $\Rightarrow$ Perturbation theory (Dunford and Schwartz, 1965).
- Sharp control with a high probability $\hat{\Pi}_m - \Pi_m$,
  where Π_m (resp. $\hat{\Pi}_m$) is the orthogonal projection onto $\text{Span}\{\varphi_1, \dots, \varphi_m\}$ (resp. $\text{Span}\{\hat{\varphi}_1, \dots, \hat{\varphi}_m\}$).

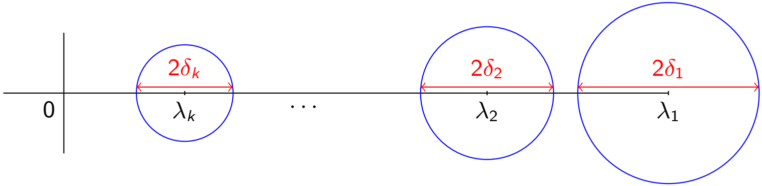


Figure: Contour made of disjoint circles

Appendix
Sketch of the proof

- **Reduction** scheme to a **finite number** of hypothesis (Tsybakov, 2008) :

$$S^\theta = \sum_{j=1}^{m_1} \mu_j \omega_j \varphi_1 \otimes \varphi_j,$$

where $\theta = (\omega_1, \ldots, \omega_{m_1}) \in \Omega_{m_1} = \{0,1\}^{m_1}$ and $\mu_j^2 = R^2/e \times [\lambda_j m_1 \eta_\alpha(m_1)]^{-1}$.

- **First lower-bounding** of the minimax risk :

$$\inf_{\hat{S}_n} \sup_{S\Gamma^{1/2} \in \mathcal{W}^R_{\alpha,\beta}} \mathrm{MSPE}(\hat{S}_n) \geq \frac{R^2}{4e} [m_1 \eta_\alpha(m_1)]^{-1} \inf_{\hat{\theta}} \max_{\theta \in \Omega_{m_1}} \mathbb{E}[\rho(\theta, \hat{\theta})],$$

where ρ is the Hamming distance and the infimum is taken over the estimators with values in Ω_{m_1}.

Lemma (Assouad with Kullback-Leibler version)

Let $\mathbb{P}_\theta^{\otimes n}$ be the distribution of $(X_i, Y_i)_{1 \leq i \leq n}$ under S^θ. Assume for all θ, θ' such that $\rho(\theta, \theta') = 1$, $\mathbf{KL}(\mathbb{P}_\theta^{\otimes n}, \mathbb{P}_{\theta'}^{\otimes n}) \leq \alpha < +\infty$. Then,

$$\inf_{\hat{\theta}} \max_{\theta \in \Omega_m} \mathbb{E}[\rho(\theta, \hat{\theta})] \geq \frac{m}{2} \max\left(\frac{1}{2}\exp(-\alpha), 1 - \sqrt{\alpha/2}\right),$$

where the infimum is taken over the estimators $\hat{\theta}$ with values in Ω_m.

Appendix

Sketch of the proof

Theorem (Cameron–Martin)

Let Z be a random Gaussian centered variable in a Hilbert space $(\mathcal{X}, \langle\cdot,\cdot\rangle, \|\cdot\|)$, with a probability measure P and the covariance operator Γ_Z. Consider the set $H_P \subset \mathcal{X}$ defined as

$$H_P = \left\{ h \in \mathcal{X} \text{ such that } \|\Gamma_Z^{-1/2}(h)\|^2 < +\infty \right\}.$$

For all h in H_P, denote P_h the probability measure of $Z + h$. Then, P_h is absolutely continuous with respect to P and the density $\mathrm{d}P_h/\mathrm{d}P$ given by

$$\mathrm{d}P_h/\mathrm{d}P : x \mapsto \exp\left\{ \langle x, \Gamma_Z^{-1}(h)\rangle - \frac{1}{2}\|\Gamma_Z^{-1/2}(h)\|^2 \right\}.$$

- Theorem of **Cameron–Martin**:

$$\mathbf{KL}(\mathbb{P}_\theta^{\otimes n}, \mathbb{P}_{\theta'}^{\otimes n}) \leq \frac{R^2}{2e\sigma_1} \times \frac{n}{m_1\eta_\alpha(m_1)},$$

  where $1/\sigma_1 = \|\Gamma_\varepsilon^{-1/2}(\varphi_1)\|^2$, with $\Gamma_\varepsilon : f \mapsto \mathbb{E}\left[\langle f, \varepsilon\rangle \varepsilon\right]$.

- Apply **Assouad**'s Lemma with m_1 such that: $m_1\eta_\alpha(m_1) \sim n/c_0$, where $c_0 > 0$ such that $\dfrac{R^2}{2c_0 e\sigma_1} \leq 1/2$.

Appendix

Construction of the estimator

Target. Select optimal m_1 and m_2 **only** based on data.

Appendix

Construction of the estimator

Target. Select optimal m_1 and m_2 **only** based on data.

- Choose $m_2 \to \infty$ and denote $\hat{S}_{m_1,\infty} = \Delta_n \Gamma_n^{\dagger}$.

Appendix

Construction of the estimator

Target. Select optimal m_1 and m_2 **only** based on data.

- Choose $m_2 \to \infty$ and denote $\hat{S}_{m_1,\infty} = \Delta_n \Gamma_n^{\dagger}$.

- Select m_1 based on the method of (Birgé and Massart,1998) from a collection $\mathcal{M}_n = \{1, \ldots, N_n\}$, where $N_n \in \mathbb{N}^*$:

$$\hat{m}_1 = \text{argmin}_{m_1 \in \mathcal{M}_n} \left(\gamma_n(\hat{S}_{m_1,\infty}) + \text{pen}(m_1) \right),$$

where pen : $m_1 \mapsto 8(1+\delta)\sigma_\varepsilon^2 m_1/n$ with $\sigma_\varepsilon^2 = \mathbb{E}\|\varepsilon\|^2$ and $\delta > 0$.

Appendix
A first optimality result

Proposition (Chagny, Meynaoui and Roche, 2022)

Under Assumptions $\mathcal{A}_0$ to $\mathcal{A}_6$, we have

$$\mathbb{E}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \left\{ \mathbb{E}\|S - \hat{S}_{m_1,\infty}\|_n^2 + \text{pen}(m_1) \right\} + \frac{C'}{n},$$

where $C, C' > 0$ and $\|\cdot\|_n$ is defined for all operator T as $\|T\|_n^2 = 1/n \sum_{i=1}^n \|T(X_i)\|^2$.

Appendix
A first optimality result

Proposition (Chagny, Meynaoui and Roche, 2022)

Under Assumptions $\mathcal{A}_0$ *to* $\mathcal{A}_6$, *we have*

$$\mathbb{E}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \left\{ \mathbb{E}\|S - \hat{S}_{m_1,\infty}\|_n^2 + \mathrm{pen}(m_1) \right\} + \frac{C'}{n},$$

where $C, C' > 0$ *and* $\|\cdot\|_n$ *is defined for all operator* T *as* $\|T\|_n^2 = 1/n \sum_{i=1}^n \|T(X_i)\|^2$.

Next steps:

Appendix

A first optimality result

Proposition (Chagny, Meynaoui and Roche, 2022)

Under Assumptions $\mathcal{A}_0$ to $\mathcal{A}_6$, we have

$$\mathbb{E}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \left\{ \mathbb{E}\|S - \hat{S}_{m_1,\infty}\|_n^2 + \mathrm{pen}(m_1) \right\} + \frac{C'}{n},$$

where $C, C' > 0$ and $\|\cdot\|_n$ is defined for all operator T as $\|T\|_n^2 = 1/n \sum_{i=1}^n \|T(X_i)\|^2$.

Next steps:

- Show the same inequality for the prediction risk.

Appendix

A first optimality result

Proposition (Chagny, Meynaoui and Roche, 2022)

Under Assumptions $\mathcal{A}_0$ to $\mathcal{A}_6$, we have

$$\mathbb{E}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \left\{ \mathbb{E}\|S - \hat{S}_{m_1,\infty}\|_n^2 + \mathrm{pen}(m_1) \right\} + \frac{C'}{n},$$

where $C, C' > 0$ and $\|\cdot\|_n$ is defined for all operator T as $\|T\|_n^2 = 1/n \sum_{i=1}^n \|T(X_i)\|^2$.

Next steps:

- Show the same inequality for the prediction risk.
- Well choose the collection $\mathcal{M}_n$ to achieve the minimax risk.

Appendix
Sketch of the proof

- Start with

$$\gamma_n(\hat{S}_{\hat{m}_1,\infty}) - \gamma_n(\hat{S}_{m_1,\infty}) \leq \mathrm{pen}(m_1) - \mathrm{pen}(\hat{m}_1).$$

Appendix

Sketch of the proof

- Start with

$$\gamma_n(\hat{S}_{\hat{m}_1,\infty}) - \gamma_n(\hat{S}_{m_1,\infty}) \leq \mathrm{pen}(m_1) - \mathrm{pen}(\hat{m}_1).$$

- Subsequently,

$$\gamma_n(\hat{S}_{\hat{m}_1,\infty}) - \gamma_n(\hat{S}_{m_1,\infty}) = \|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 - \|S - \hat{S}_{m_1,\infty}\|_n^2 + 2\nu_n(\hat{S}_{\hat{m}_1,\infty} - \hat{S}_{m_1,\infty}),$$

where the empirical process ν_n is defined by $\nu_n : T \mapsto 1/n \sum_{i=1}^n \langle T(X_i), \varepsilon_i \rangle$.

Appendix
Sketch of the proof

- Start with

$$\gamma_n(\hat{S}_{\hat{m}_1,\infty}) - \gamma_n(\hat{S}_{m_1,\infty}) \leq \mathrm{pen}(m_1) - \mathrm{pen}(\hat{m}_1).$$

- Subsequently,

$$\gamma_n(\hat{S}_{\hat{m}_1,\infty}) - \gamma_n(\hat{S}_{m_1,\infty}) = \|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 - \|S - \hat{S}_{m_1,\infty}\|_n^2 + 2\nu_n(\hat{S}_{\hat{m}_1,\infty} - \hat{S}_{m_1,\infty}),$$

where the empirical process ν_n is defined by $\nu_n : T \mapsto 1/n \sum_{i=1}^n \langle T(X_i), \varepsilon_i \rangle$.

- We show that

$$\frac{1}{2}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq \frac{3}{2}\|S - \hat{S}_{m_1,\infty}\|_n^2 + 2\,\mathrm{pen}(m_1) + 4\left(\sup_{\substack{T \in E_{m_1 \vee \hat{m}_1,\infty} \\ \|T\|_n = 1}} \nu_n(T)^2 - p(m_1, \hat{m}_1)\right)_+$$

where for all x in $\mathbb{R}$, $x_+ = x \vee 0$ and $p : (m, m') \mapsto 2(1+\delta)\sigma_\varepsilon^2 \frac{m \vee m'}{n}$. In addition, $E_{m,\infty}$ is the closure of $\mathrm{Span}\{\hat{\varphi}_k \otimes \hat{\varphi}_j,\ 1 \leq j \leq m, k \geq 1\}$.

Appendix

Sketch of the proof

- Inspired from the works of [Brunel, Mas and Roche, 2016], we show the Talagrand-type inequality:

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{\substack{T \in V_{m \vee m', \infty} \\ \|T\|_n = 1}} \nu_n^2(T) - p(m, m')\right)_+\right] \leq \frac{C}{n}.$$

Appendix

Sketch of the proof

- Inspired from the works of [Brunel, Mas and Roche, 2016], we show the Talagrand-type inequality:

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{\substack{T \in V_{m \vee m', \infty} \\ \|T\|_n = 1}} \nu_n^2(T) - p(m, m')\right)_+\right] \leq \frac{C}{n}.$$

- This means that

$$\mathbb{E}\|S - \hat{S}_{\hat{m}_1,\infty}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \left\{\mathbb{E}\|S - \hat{S}_{m_1,\infty}\|_n^2 + \mathrm{pen}(m_1)\right\} + \frac{C'}{n},$$

where $C, C' > 0$.