Adaptive estimation in the linear random coefficients model when regressors have limited variation

CHRISTOPHE GAILLAC and ERIC GAUTIER

Toulouse School of Economics, 1 esplanade de l’université, 31000 Toulouse, France
E-mail: christophe.gaillac@tse-fr.eu; eric.gautier@tse-fr.eu

We consider a linear model where the coefficients - intercept and slopes - are random and independent from the regressors. The law of the coefficients is nonparametric. Without further restriction, nonparametric identification requires the regressors to have a support which is the whole space. This is hardly ever the case in practice. It is possible to handle regressors with limited variation when the coefficients can have a compact support. This is not compatible with unbounded error terms as usual in regression models. In this paper, the regressors can have a support which is a proper subset but the slopes do not have heavy-tails. Lower bounds on the minimax risk for the estimation of the joint density of the random coefficients density are obtained for a wide range of smoothness. Some allow for polynomial and nearly parametric rates of convergence. We present a minimax optimal estimator and a data-driven rule for adaptive estimation. A R package is available to implement this estimator.

Keywords: Adaptation, Inverse Problem, Minimax, Random Coefficients.

1. Introduction

Inferring causal effects from a data set is of great importance for applied researchers. This paper assumes that the explanatory variables are determined outside the model (e.g., a treatment is randomly assigned) and addresses the question of the heterogeneity of the effects. We can argue that the effects are heterogeneous in nearly all applications of linear regression. In such a case, the coefficients of the linear regression capture average effects. For example, it is well understood that the effect of the income of the parents or the class size on pupils’ achievements differ across pupils. A second example consists of models of consumer demand such as Engel curves. These are models of the effect of the total budget an household can spend on consumption goods on the budget share spent on a particular one (e.g., food). It is also well known that there is a great deal of heterogeneity in consumer demand (see, e.g., [29]). Quantifying the heterogeneity of the effects can have useful policy implications. For example, it could be worthwhile to implement a treatment on the population subgroups for which it has the highest impact.
Taking into account the heterogeneity has implications on prediction intervals (see [3]), welfare measures, and counterfactual effects.

The linear regression with random coefficients is a continuous mixture of linear regressions. Maintaining parametric assumptions on the mixture density is open to criticism because these assumptions can drive the results (see [27]). For this reason, this paper considers a nonparametric setup. Unfortunately, most of the estimation theory for this model has relied on assumptions which are almost never satisfied. This is probably the reason why, up to now, applied researchers have preferred to use the quantile regression to account for heterogeneous effects. There, the conditional quantiles of an outcome given the regressors are linear in the regressors. When the conditional quantiles are strictly increasing then the quantile regression defines the same data generating process as a linear random coefficients model where the coefficients are functions of a scalar uniform distribution. However, it can be hard to argue for such degeneracy of the random coefficients distribution or for the linearity of the conditional quantiles. The unobserved scalar uniform variable can be interpreted as a ranking variable. In the education and demand examples, one can argue that such ranking can be based on a more complex multidimensional vector of unobserved attributes. Restricting unobserved heterogeneity to a scalar can have undesirable implications such as monotonicity in the literature on policy evaluation (see [22]).

Formally, for a random variable $\alpha$ and random vectors $X$ and $\beta$ of dimension $p$, the linear random coefficients model is

$$Y = \alpha + \beta^\top X,$$

$$\alpha, \beta^\top \text{ and } X \text{ are independent.}$$

The researcher has at her disposal $n$ observations $(Y_i, X_i^\top)_{i=1}^n$ of $(Y, X^\top)$ but does not observe the realizations $(\alpha_i, \beta_i^\top)_{i=1}^n$ of $(\alpha, \beta^\top)$. $\alpha$ subsumes the intercept and error term and the vector of slope coefficients $\beta$ is heterogeneous (i.e., varies across $i$). $(\alpha, \beta^\top)$ corresponds to multidimensional unobserved heterogeneity and $X$ to observed heterogeneity. Other random coefficients models have been analyzed recently in econometrics (see, e.g., [9, 23, 28, 38] and references therein). This paper focuses on the most simple model but addresses important practical issues.

Estimation of the density of the random coefficients $f_{\alpha, \beta}$ has similarities with tomography problems involving the Radon transform (see [4, 5, 29]). Indeed, the density of $Y/\sqrt{1 + |X|^2_2}$ given $S = (1, X^\top)/\sqrt{1 + |X|^2_2}$, where $|\cdot|_2$ is the Euclidian norm, at point $u$ given $s$ is the integral of $f_{\alpha, \beta}$ on the affine hyperplane defined via the pair $(u, s)$. It could be tempting to analyze the related tomography problem with an additive Gaussian white noise. But the random coefficients model (1)-(2) has it specificities. Treating it requires: (1) that $(\alpha, \beta^\top)$ has a noncompact support to allow for usual unbounded errors, (2) to allow the dimension to be larger than in tomography due to more than one or two regressors, and (3) the directions to have an unknown but estimable density.

To obtain rates of convergence, [29] assumes the density of $S$ is bounded from below. When $p = 1$, this holds when $X$ has tails at least as fat as the ones of the Cauchy distribution. Recently, [18] motivates testing large features of the density by the possible
Random coefficients with bounded regressors

slow rates of convergence of density estimation and [31] obtains rates of convergence for density estimation with less heavy tails on $X$. But assuming the support of $X$ is $\mathbb{R}^p$ is unrealistic for nearly all applications. In the motivating examples, the income of the parents, the class size, and the total budget have limited variation.

Limited angle tomography considers the case where $p = 1$, $S$ has a support which is a known cap (i.e., the support of the angle is an interval), and the object has support in a ball. [19] proposes a soft-thresholded curvelet regularization for the problem with an additive bounded noise but does not obtain results for the statistical problem (e.g., consistency). Importantly, [30] shows that the rate of the minimax risk in Sobolev type ellipsoids relative to the right-singular functions of the Radon transform is logarithmic. It shows that projection estimators are adaptive. It gives the analogy with a random coefficients model where $p = 1$, $(\alpha, \beta^\top)$ has support in the unit ball, for some known densities of the regressors. It concludes that a lot remains to be done to handle $p > 1$ and estimable densities of the regressors. The random coefficients model when the support of $X$ can be a proper (i.e. strict) subset is considered in [5]. When $p = 1$ and $(\alpha, \beta^\top)$ has compact support, it is shown that a minimum distance estimator is consistent. Section 2 in the online appendix of [28] proposes another consistent estimator in this case.

This paper is directly applicable to (1)-(2). It allows for the essential feature that $\alpha$ can have a noncompact support, that the researcher does not need to have knowledge on the support of $\beta$ if compact, and that the latter can also be noncompact. It also allows for estimable density of the regressors and arbitrary $p$. We assume the marginals of $\beta$ (but not of $\alpha$) do not have heavy tails. This allows for many parametric families which are used in mixture modelling, while leaving unspecified the parametric family. We do not rely on the Radon transform but on the truncated Fourier transform (see, e.g., [1]). Due to (2), the conditional characteristic function of $Y$ given $X = x$ at $t$ is the Fourier transform of $f_{\alpha, \beta}$ at $(t, tx^\top)^\top$. Hence, the family of conditional characteristic functions indexed by $x$ in the support of $X$ gives access to the Fourier transform of $f_{\alpha, \beta}$ on a double cone of axis $(1, 0, \ldots, 0) \in \mathbb{R}^{p+1}$ and apex 0. When $\alpha = 0$ and the supports of $\beta$ and $X$ are compact with nonempty interior, this is the problem of out-of-band extrapolation or super-resolution (see, e.g., [6]). Because we do not restrict $\alpha$ and the support of $\beta$ can be noncompact, we generalize this approach.

A related problem is extrapolation. It is used in [39] to perform deconvolution of compactly supported densities, allowing the characteristic function of the error to vanish. This paper does not use extrapolation nor assume densities are analytic. Rather, the operator of the inverse problem is a composition of two operators based on partial Fourier transforms. One involves a truncated Fourier transform and we make use of properties of the singular value decomposition.

Unlike [5, 28], we go beyond consistency and provide a full analysis of the general case. Similar to [24, 30, 31], we study minimax optimality. However, we obtain lower bounds under a wide variety of assumptions. We show that polynomial and nearly parametric rates can be attained. Hence, we can lose little in terms of rates of convergence from going from a parametric model to a nonparametric one. This contrasts with the pessimistic logarithmic rates in [30] (also mentioned in [28]) and the message to avoid estimating densities in [18]. We present an estimator involving: series based estimation of the partial
Fourier transform of the density with respect to the first variable, interpolation around zero, and inversion of the partial Fourier transform. The orthonormal system is a tensor products of the Prolate Spheroidal Wave Functions (see [42]) when the law of $\beta$ has a support included in a known bounded set. Else, it is composed of singular functions of an operator first studied in [41, 47]. The relevant results on these systems are given in the appendix. They can also be used in a wide range of applications such as for stable analytic continuation by Hilbert space techniques (see [20]). We use a Goldenshluger-Lepski type method to obtain data-driven estimators. We consider estimation of the marginal $f_\beta$ in Appendix C.

The adaptive estimator is implemented in the R package RandomCoefficients. In contrast with the EM algorithm for parametric continuous mixtures of regressions, the estimator has the advantage of being robust to misspecification of the parametric family. It also avoids possible non convergence issues of the EM algorithm. The proposed estimator relies on the computation of a SVD which we obtain once and for all by a numerically efficient method, on two simple sums, and a one dimensional Fourier transform carried by FFT. Additional practical and computational details are available in [21].

The paper is organized as follows. Section 2 gives preliminaries. It introduces the baseline assumption on $f_{\alpha,\beta}$ and the variation of the regressors. It frames the recovery of $f_{\alpha,\beta}$ as an inverse problem involving a composition at the basis of the estimation procedure. The section also introduces a related Gaussian white noise model and the main elements on interpolation, which is key to obtain an optimal estimator. Finally, it presents the sets of smooth and integrable functions for the minimax analysis and the risk. Section 3 provides the lower bounds. Section 4 describes the estimator and its rates of convergence assuming knowledge of the parameters of the sets of functions. Section 4.3 provides a data-driven estimator and presents its nearly minimax rates of convergence. Section 5 concludes with details on the numerical implementation and simulations illustrating the finite sample performances of the data-driven estimator.

2. Preliminaries

2.1. Notations

The notations $\cdot$, $\cdot_1$, $\cdot_2$, $\star$ are used to denote variables in a function. $a \wedge b$ (resp. $a \vee b$) is used for the minimum (resp. maximum) between $a$ and $b$, $(\cdot)_+$ for $0 \vee \cdot$, and $1 \{A\}$ for the indicator function of set $A$. $\mathbb{N}$ and $\mathbb{N}_0$ stand for the positive and nonnegative integers. Bold letters are used for vectors. For all $r \in \mathbb{R}$, $r$ is the vector, which dimension will be clear from the text, where each entry is $r$. For $x \geq 1$ we denote by $\ln_2(x) = \ln(\ln(x))$. $W$ is the inverse of $x \in [0, \infty) \mapsto xe^x$. $|\cdot|_q$ for $q \in [1, \infty]$ stands for the $\ell_q$ norm of a vector or sequence. For all $\beta \in \mathbb{C}^d$, $(f_m)_{m \in \mathbb{N}_0}$ functions with values in $\mathbb{C}$, and $m \in \mathbb{N}_0^d$, denote by $\beta^m = \prod_{k=1}^d \beta_k^{m_k}$, $|\beta|^m = \prod_{k=1}^d |\beta_k|^{m_k}$, and $f_m = \prod_{k=1}^d f_{m_k}$. For a function $f$ of real variables, $\text{supp}(f)$ denotes its support. The inverse of a mapping $f$, when it exists, is denoted by $f^{-1}$. We denote the interior of $\mathcal{S} \subseteq \mathbb{R}^d$ by $\mathring{\mathcal{S}}$. 
When \( \mathcal{S} \) is a Borel set and \( h \) a nonnegative function from \( \mathcal{S} \) to \([0, \infty] \), \( L^2(h) \) is the space of complex-valued square integrable functions equipped with \( \langle f, g \rangle_{L^2(h)} = \int_{\mathcal{S}} f(x) \overline{g(x)} h(x) \, dx \). This is denoted by \( L^2(\mathcal{S}) \) when \( h = 1 \). For a Borel set \( \mathcal{S} \subseteq \mathbb{R}^d \), we denote by \( i_{\mathcal{S}} = \mathbb{1}(\mathcal{S}) + \infty \mathbb{1}(\mathcal{S}^c) \). We have \( L^2(i_{\mathcal{S}}) = \{ f \in L^2(\mathbb{R}^d) : \text{supp}(f) \subseteq \mathcal{S} \} \) and \( \langle f, g \rangle_{L^2(i_{\mathcal{S}})} = \int_{\mathcal{S}} f(x) \overline{g(x)} h(x) \, dx \). Denote by \( \mathcal{D} \) the set of densities and by \( \otimes \) the product of functions (e.g., \( h^{\otimes d}(b) = \prod_{j=1}^d h(b_j) \)) or measures.

The Fourier transform of \( f \in L^1(\mathbb{R}^d) \) is \( \mathcal{F}[f](x) = \int_{\mathbb{R}^d} e^{ib^T x} f(b) \, db \) and \( \mathcal{F}[f] \) is also the Fourier transform in \( L^2(\mathbb{R}^d) \). For all \( c > 0 \), denote the Paley-Wiener space by \( PW(c) := \{ f \in L^2(\mathbb{R}) : \text{supp}(\mathcal{F}[f]) \subseteq [-c, c] \} \), by \( \mathcal{P}_c \) the projector from \( L^2(\mathbb{R}) \) to \( PW(c) \) \( (\mathcal{P}_c[f] = \mathcal{F}^T \mathbb{1}([-c, c]) \mathcal{F}[f]) \). For all \( f \in L^1(\mathbb{R}^d) \), \( \mathcal{F}_{\text{int}}[f](t, \cdot) \) denotes the partial Fourier transform of \( f \) with respect to the first variable.

Now, \( W \) is a function which can be either \( i_{[-R,R]} \) or \( \cosh(\cdot / R) \) and which depends on a given \( R > 0 \). The truncated Fourier transform \( \mathcal{F}_c \) is defined by

\[
\forall c \neq 0, \quad \mathcal{F}_c : \quad L^2(W^{\otimes d}) \to L^2([-1,1]^d) \quad f \to \mathcal{F}[f](c \cdot).
\]

For a random vector \( X \), \( \mathbb{P}_X \) is its law, \( f_X \) its density, \( f_{X|X} \) the truncated density of \( X \) given \( X \in \mathcal{X}, \mathbb{S}_X \) its support, and \( f_{Y|X=x} \) the conditional density.

### 2.2. Baseline assumption

Assumption 1 below imposes restrictions on the integrability of the density of the random coefficients and on the variation of the regressors ensuring nonparametric identification.

**Assumption 1 (H1.1)** \( f_X \) and \( f_{\alpha,\beta} \) exist;

(H1.2) There exist \( w \geq 1 \) and \( R > 0 \) such that \( f_{\alpha,\beta} \in L^2(\mathbb{1}_{[-R,R]} \otimes W^{\otimes p}) \), where \( W \) is either \( i_{[-R,R]} \) or \( \cosh(\cdot / R) \);

(H1.3) There exists \( x_0 > 0 \) and \( \mathcal{X} = [-x_0, x_0]^p \subseteq \mathbb{S}_X \).

The assumption (H1.2) is important to rely on Hilbert-space methods. It is satisfied when \( W = \cosh(\cdot / R) \) if

\[
\mathbb{E} \left[ w(\alpha) \prod_{k=1}^p e^{l|\beta_k|/R} \right] < \infty \quad \text{and} \quad \|f_{\alpha,\beta}\|_{L^\infty(\mathbb{R}^{p+1})} < \infty
\]

hold. The first condition is a condition on the tails of the density of the random coefficients. The weights are such that \( L^2(\mathbb{1}_{[-R,R]} \otimes W^{\otimes p}) \subseteq L^2(\mathbb{1}_{[-R,R]} \otimes (e^{l|\cdot|/R})^{\otimes p}) \). When \( W = i_{[-R,R]} \); \( f_{\alpha,\beta} \in L^2(\mathbb{1}_{[-R,R]} \otimes W^{\otimes p}) \) implies that \( \mathcal{S}_\beta \subseteq [-R,R]^p \). If \( w^{-1} \in L^1(\mathbb{R}) \), (H1.2) implies that the slopes of \( \beta \) do not have heavy tails. This means that their tails are not heavier than that of the Laplace distribution (i.e., the Laplace transform of their absolute value is finite near 0). Indeed, we have, for all \( \epsilon \in (0, 1) \) and \( k = 1, \ldots, p \), for \( \lambda = (1 - \epsilon)/(2R) \), by the Cauchy-Schwarz inequality,

\[
\mathbb{E} \left[ e^{l|\beta_k|} \right] \leq \|f_{\alpha,\beta}\|_{L^2(\mathbb{1}_{[-R,R]} \otimes W^{\otimes p})} \|w^{-1}\|_{L^1(\mathbb{R})}^{1/2} (2R/e)^{p/2} < \infty.
\]
(H1.2) allows for slopes which marginal distributions are Gaussian, inverse Gaussian, logistic, and all distributions with compact support, among others. But it rules out
the lognormal distribution. The case \( w = 1 \) corresponds to mild integrability in the
first variable. The condition that the support of the regressors contains an hypercube
\( X = [-x_0, x_0]^p \) in (H1.3) is not restrictive because \( Y = \alpha + \beta \top \varpi + \beta \top M^{-1} M(X - \varpi) \),
we can take \( \varpi, M \) an invertible \( p \times p \) matrix, and \( x_0 \) such that \( X \subseteq \mathbb{S}_M(X - \varpi) \), and there
is a one-to-one mapping between \( f_{\alpha+\beta \top \varpi}((M^{-1}) \top \beta) \) and \( f_{\alpha, \beta} \).

2.3. Inverse problem in Hilbert spaces

Estimation of \( f_{\alpha, \beta} \) corresponds to solving a statistical ill-posed inverse problem. Indeed,
we can relate \( f_{\alpha, \beta} \) to the conditional characteristic function \( f_{Y|X=x} \) known over a subset
of \( \mathbb{R}^{p+1} \), namely a double cone. This can be formalized via the introduction of the
operator: for all \( t \in \mathbb{R} \) and \( u \in [-1, 1]^p \),

\[
\mathcal{K} f_{\alpha, \beta}(t, u) = \mathcal{F} \left[ f_{Y|X=x \cap u} \right] (t) x_0 |t|^{p/2},
\]

where

\[
\mathcal{K} : L^2(w \otimes W^{\otimes p}) \rightarrow L^2(\mathbb{R} \times [-1, 1]^p)
\]

\[
f \rightarrow \ (t, u) \mapsto \mathcal{F} [f](t, x_0 t u) x_0 |t|^{p/2}.
\]

**Proposition 1** \( L^2(w \otimes W^{\otimes p}) \) is continuously embedded into \( L^2(\mathbb{R}^{p+1}) \). Moreover, \( \mathcal{K} \) is
injective and continuous, and not compact if \( w = 1 \).

Thus, when \( w = 1 \), the SVD of \( \mathcal{K} \) does not exist. This makes it difficult to prove rates
of convergence even for estimators which do not rely explicitly on the SVD such as the
Tikhonov and Landweber method (Gerchberg algorithm in out-of-band extrapolation, see
[6]). Rather than working with \( \mathcal{K} \) directly, we use that \( \mathcal{K} \) is the composition of operators
which are easier to analyze

\[
\forall \ t \in \mathbb{R}, \ \mathcal{F}[f](t, *) = \mathcal{F}_{tx_0} \left[ \mathcal{F}_{1st} [f] (t, 2) \right] (r) x_0 |t|^{p/2} \text{ in } L^2([-1, 1]^p).
\]

For all \( f \in L^2(w \otimes W^{\otimes p}) \) and \( t \in \mathbb{R} \), \( \mathcal{F}_{1st} [f] (t, 2) \) belongs to \( L^2(\mathbb{W}^{\otimes p}) \) and, for
\( c \neq 0 \), \( \mathcal{F}_c : L^2(\mathbb{W}^{\otimes p}) \rightarrow L^2([-1, 1]^p) \) admits a SVD, where both orthonormal
systems are complete. This is a tensor product of the SVD when \( p = 1 \) that we denote by
\( (\sigma^W, \varphi^W, g^W)_{m \in \mathbb{N}_0} \), where \( (\sigma^W)_{m \in \mathbb{N}_0} \in (0, \infty)^{\mathbb{N}_0} \) is in decreasing order repeated
according to multiplicity. Note that (6) allows to rely on the property of the unidimensional
truncated Fourier transform to analyze the problem of the multidimensional truncated
Fourier transform on a double cone. This is a natural way to exploit the geometry of the
problem.

**Proposition 2** For all \( c \neq 0 \), \( (g^W)_{m \in \mathbb{N}_0} \) and \( (\varphi^W)_{m \in \mathbb{N}_0} \) are bases of, respectively,
\( L^2([-1, 1]) \) and \( L^2(W) \).
The SVD when $W = [i_{-1,1}]$ is well studied. The singular functions $(g_m^{i_{-1,1},c})_{m \in \mathbb{N}_0}$ are the Prolate Spheroidal Wave Functions (henceforth PSWF). They can be extended as entire functions in $L^2(\mathbb{R})$ and form a complete orthogonal system of $PW(c)$ for which we use the same notation. They are useful to carry interpolation and extrapolation (see, e.g., [36]) with Hilbert techniques. Nonasymptotic upper bounds on the singular values show that the latter decay exponentially with $m$ faster than $e^{-m \ln(4(m+3/2)/c|c|)}$ (see, e.g., Lemma B.4. in Appendix B and [8]).

The singular functions $(g_m^{\cosh(\cdot/R),c})_{m \in \mathbb{N}_0}$ allow to carry extrapolation of nonbandlimited functions (see [20]). This is useful even if $S_\beta$ is compact when the researcher does not know a superset containing $S_\beta$. Extending the work of [47], [20] also proves nonasymptotic lower and upper bounds on the singular values, which show that, for $c < 1$, the latter decay with $m$ faster than $e^{-m \ln(1/|c|)/\sqrt{2m + 1}}$ (see Theorem 7 in [20]). Useful results for the present paper on the corresponding SVD are in Appendix B. More properties and a numerical algorithm to compute it are in [20].

Nonasymptotic upper bounds on $\|g_m^{W,c}\|_{L^\infty(-1,1)}$ are proved in [8] for the PSWF and in [20] when $W = \cosh(\cdot/R)$. They are recalled in Proposition B.2. They take the form $\|g_m^{W,c}\|_{L^\infty(-1,1)} \leq H(c)\sqrt{m + 1/2}$ for a given $H(c)$. We use the explicit dependence of $H$ in $c$ for all $m \in \mathbb{N}_0$ to prove the results in Section 4.3.

### 2.4. Related Gaussian white noise model

The next idealized model is related to (1)-(2) when $f_X$ is known:

$$dZ(t) = K[f](t, \cdot, 2) dt + \frac{1}{\sqrt{n}}dG(t), \quad t \in \mathbb{R},$$

where $f$ plays the role of $f_{n, \beta}$ and $(G(t))_{t \in \mathbb{R}}$ is a complex two-sided cylindrical Gaussian process on $L^2([-1,1]^p)$ (see [16]). $Z$ and $G$ are function-valued processes and those functions can take complex values. The partial derivative in the sense of distributions with respect to time of $G$ is the space time white noise in $L^2(\mathbb{R} \times [-1,1]^p)$. $G$ is the partial derivative in the sense of distributions obtained by differentiating once with respect to each space variable the Brownian sheet.

By taking the inner product of both sides of (7) with $g_m^{W,x_0,t}$ for all $m \in \mathbb{N}_0$, we get the system of independent equations

$$Z_m(t) = \int_0^t \sigma_m^{W,x_0,s} b_m(s) ds + \frac{1}{\sqrt{n}} B_m(t), \quad t \in \mathbb{R},$$

where $Z_m := \langle Z(\cdot), g_m^{W,x_0,t} \rangle_{L^2([-1,1]^p)}$, $B_m(t) = B_{m}^R(t) + iB_{m}^I(t)$, and $(B_{m}^R(t))_{t \in \mathbb{R}}$ and $(B_{m}^I(t))_{t \in \mathbb{R}}$ are independent two-sided Brownian motions.

It is common in statistics of inverse problems to analyze an idealized Gaussian white noise model (see, e.g., [30]). This paper sometimes refer to model (8) to develop intuitions. Some of the minimax lower bounds are proved only in this context for simplicity.
Asymptotic equivalence with white noise models has been proved in some cases such as regression models (see, e.g., [32, 43]). These results do not apply in the current random coefficients model because, among other things, it is an inverse problem. Asymptotic equivalence results for some inverse problems have been proved (see, e.g., [40] for the functional linear regression). Proving such an equivalence in our context is out of the scope of this paper.

2.5. Interpolation

This paper relies on interpolation when the variance of an initial estimator \( \hat{f}^0(t) \) of \( f(t) \) is large when \( t \) is close to 0 but \( \| f - \hat{f}^0 \|_{L^2(\mathbb{R} \setminus (-\epsilon, \epsilon))}^2 \) is small. In (8), because \( \sigma^W_{m, \epsilon, \alpha} \) is small when \( |m|_q \) is large or \( t \) is small (see Lemma B.4), the estimator of Section 4.1 truncates large values of \( |m|_q \) and does not rely on small values of \( |t| \) but uses interpolation. Specifically, we work with

\[
\hat{f}(t) = \hat{f}^0(t) \mathbb{1}_{\{ |t| \geq \epsilon \}} + \mathcal{I}_{\mathbb{R}, \epsilon}[\hat{f}^0](t) \mathbb{1}_{\{ |t| < \epsilon \}},
\]

where, for all \( \alpha, \epsilon > 0 \), \( \mathcal{I}_{\mathbb{R}, \epsilon} \) is the interpolation operator on \( L^2(\mathbb{R}) \) with domain \( PW(\alpha) \),

\[
\mathcal{I}_{\mathbb{R}, \epsilon}[f] := \sum_{m \in \mathbb{N}_0} \frac{\rho_{m, \epsilon}^{[-1,1]} \cdot \hat{h}}{\left( 1 - \rho_{m, \epsilon}^{[-1,1]} \cdot \hat{h} \right) \epsilon} f_{m, \epsilon}^{[-1,1]} \mathcal{P}_{\mathbb{R}, \epsilon}(\cdot) + \mathcal{P}_{\mathbb{R}, \epsilon}[f],
\]

and, for all \( m \in \mathbb{N}_0 \) and \( c \neq 0 \), \( \rho_{m, \epsilon}^{W, c} := 2\pi(\sigma_{m, \epsilon}^{W, c})^2 / |c| \). Then, (11) below yields

\[
\| f - \hat{f} \|_{L^2(\mathbb{R})}^2 \leq (1 + 2C_0(\epsilon \alpha)) \| f - \hat{f}^0 \|_{L^2(\mathbb{R} \setminus (-\epsilon, \epsilon))}^2 + 2(1 + C_0(\epsilon \alpha)) \| f - \mathcal{P}_{\mathbb{R}, \epsilon}[f] \|_{L^2(\mathbb{R})}^2,
\]

where \( C_0 := 4 \cdot (\pi(1 - \rho_{m, \epsilon}^{[-1,1]} \cdot \hat{h}))^2 \), which bounds the error made on \( \mathbb{R} \) by the interpolated estimator \( \hat{f} \) by the sum of the error of the initial estimator \( \hat{f}^0 \) on \( \mathbb{R} \setminus (-\epsilon, \epsilon) \) and a term related to the distance of \( f \) to its projection on \( PW(\alpha) \).

**Proposition 3** For all \( \alpha, \epsilon > 0 \), \( \mathcal{I}_{\mathbb{R}, \epsilon}(L^2(\mathbb{R})) \subseteq L^2([\epsilon, \epsilon]) \) and, for all \( g \in PW(\alpha) \), \( \mathcal{I}_{\mathbb{R}, \epsilon}[g] = g \) in \( L^2(\mathbb{R}) \), and, for all \( f, h \in L^2(\mathbb{R}) \),

\[
\| f - \mathcal{I}_{\mathbb{R}, \epsilon}[h] \|_{L^2([\epsilon, \epsilon])}^2 \leq (2 + C_0(\epsilon \alpha)) \| f - \mathcal{P}_{\mathbb{R}, \epsilon}[f] \|_{L^2(\mathbb{R})}^2 + 2(1 + C_0(\epsilon \alpha)) \| f - \mathcal{P}_{\mathbb{R}, \epsilon}[f] \|_{L^2(\mathbb{R})}^2.
\]

If \( f \in PW(\alpha) \), \( \mathcal{I}_{\mathbb{R}, \epsilon}[f] \) only relies on \( f \mathbb{1}_{\mathbb{R} \setminus (-\epsilon, \epsilon)} \) and \( \mathcal{I}_{\mathbb{R}, \epsilon}[f] = f \) on \( \mathbb{R} \setminus (-\epsilon, \epsilon) \), so (9) provides an analytic formula to carry interpolation on \( [-\epsilon, \epsilon] \) of functions in \( PW(\alpha) \). Else, (11) provides an upper bound on the error made by approximating \( f \) by \( \mathcal{I}_{\mathbb{R}, \epsilon}[h] \) on \( [-\epsilon, \epsilon] \) when \( h \) approximates \( f \) outside \( [-\epsilon, \epsilon] \). When \( \operatorname{supp}(\mathcal{F}[f]) \) is compact, \( \alpha \) is taken such that \( \operatorname{supp}(\mathcal{F}[f]) \subseteq [-\alpha, \alpha] \). Else, \( \alpha \) goes to infinity so the second term in (10) goes to 0. \( \epsilon \) is taken such that \( \alpha \epsilon \) is constant because, due to (3.87) in [42], \( \lim_{t \to \infty} C_0(t) = \infty \) and (11) and (10) become useless. We set \( C = 2(1 + C_0(\epsilon \alpha)) \).
2.6. Sets of smooth and integrable functions

Define, for \( q \in \{1, \infty\} \),

\[
b_m(t) := \langle \mathcal{F}_{\text{1st}} [f] (t, \cdot, 2), \varphi^a m \rangle_{L^2(W \otimes \rho)}^{W, \omega}, \theta_{q,k}(t) := \left( \sum_{m \in \mathbb{N}_0^p : |m| = k} |b_m(t)|^2 \right)^{1/2},
\]

and, for all \((\phi(t))_{t \geq 0} \) and \((\omega_m)_{m \in \mathbb{N}_0^p} \) increasing, \( \phi(0) = \omega_0 = 1, l, M > 0, t \in \mathbb{R}, m \in \mathbb{N}_0^p, k \in \mathbb{N}_0, \mathcal{I}_{w,W}(M) := \{ f : \| f \|_{L^2(w \otimes W \otimes \rho)} \leq M \} \), and

\[
\mathcal{H}^{q,\omega}_w(l, M) := \left\{ f : \sum_{k \in \mathbb{N}_0^p} \int_{\mathbb{R}} \phi^2(t) |\theta_{q,k}(t)|^2 dt \left\| \sum_{k \in \mathbb{N}_0^p} \omega_k^2 \| \theta_{q,k} \|_{L^2(\mathbb{R})}^2 \right\|_{2 \pi^d} \leq 2\pi^l \right\} \cap \mathcal{I}_{w,W}(M).
\]

We use the notation \( \mathcal{H}^{q,\omega}_w(l, M) \) when we require \( \| f \|_{L^2(w \otimes W \otimes \rho)} < \infty \) rather than \( \| f \|_{L^2(w \otimes W \otimes \rho)} \leq M \). The set \( \mathcal{I}_{w,W}(M) \) imposes the integrability discussed in the beginning of the section. The first set in the definition of \( \mathcal{H}^{q,\omega}_w(l, M) \) defines the notion of smoothness analyzed in this paper. It involves a maximum, thus two inequalities: the first for smoothness in the first variable and the second for smoothness in the other variables. The asymmetry in the treatment of the first and remaining variables is due to the fact that only the random slopes are multiplied by regressors which have limited variation and we make integrability assumptions in the first variable which are as mild as possible. The smoothness classes in the analysis of the Radon transform usually involve nonstandard weight functions well suited to the operator. In contrast, the ones in this paper are not too hard to interpret. To gain intuition on the inequalities, note that, by Plancherel and Parseval’s identities,

\[
2\pi \| f \|^2_{L^2(1 \otimes W \otimes \rho)} = \int_{\mathbb{R}} \| \mathcal{F}_{\text{1st}} [f] (t, \cdot, 2) \|^2_{L^2(W \otimes \rho)} dt = \int_{\mathbb{R}} \sum_{m \in \mathbb{N}_0^p} |b_m(t)|^2 dt = \sum_{k \in \mathbb{N}_0} \| \theta_{q,k} \|^2_{L^2(\mathbb{R})}, \tag{12}
\]

The multi-index \( m \) is a type of frequency associated to the space variable \( b \) and \( k \) denotes a level of all frequencies with same norm of the multi-index. The fact that \( m \) is discrete is easy to understand when \( W = i_{[-R,R]} \) and the function has compact support in \( b \). It is similar to Fourier series. The choice of the functions \( \langle \varphi^a m \rangle_{m \in \mathbb{N}_0^p} \) is well suited to the decomposition of the operator \( K \). The frequency \( t \) is related to the space variable \( a \). It is continuous because we make weaker assumptions on the tails of \( \alpha \).

The first inequality can be written as

\[
\int_{\mathbb{R}} \phi^2(t) \| \mathcal{F}_{\text{1st}} [f] (t, \cdot, 2) \|^2_{L^2(W \otimes \rho)} dt \leq 2\pi l^2.
\]
Thus, when \( \phi = 1 \sqrt{\cdot}^s \) with \( s > 0 \), it is the usual Sobolev smoothness in \( L^2(\mathbb{R}; L^2(W^{\otimes p})) \).

The second inequality is obtained by using weighted sums rather than (12). The weights are indexed by a frequency level \( |m|_q \). There is no weight function for \( t \). Hence, it can be viewed as a smoothness assumption in \( b \) only. It can be written: there exists a density \( \overline{\varphi} \) on \( \mathbb{R} \) such that

\[
\forall t \in \mathbb{R}, \sum_{m \in \mathbb{N}_0^q} \omega^2_{|m|_q} |b_m(t)|^2 \leq \overline{\varphi}(t)2\pi^2 t^2.
\]  

\( (13) \)

The different choices of sequences \( (\omega_k)_{k \in \mathbb{N}_0} \) that we shall consider correspond to different source conditions for fixed \( t \) expressed in terms of the operator \( J_{W,x_0t} \) from \( L^2(W^{\otimes p}) \) to \( L^2([-1,1]^p) \), for \( f \in L^2(W^{\otimes p}) \), by

\[
J_{W,x_0t}[f] := \sum_{m \in \mathbb{N}_0^q} \frac{1}{|m|_q} \langle \varphi^{W,x_0t}, f \rangle_{L^2(W^{\otimes p})} g^{W,x_0t}_m.
\]

When \( q = 1 \), the unbounded operator \( J_{W,x_0t}^{-1} \) can be viewed as a differential operator. When \( (\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0} \), with \( \sigma > 0 \), then, \( J_{1st}[f](t,\cdot) \) satisfies (13) if and only if it belongs to

\[
\left\{ g \in L^2(W^{\otimes p}) : g = (J_{W,x_0t}^*/J_{W,x_0t})^* v, \|v\|^2_{L^2(W^{\otimes p})} \leq \overline{\varphi}(t)2\pi^2 t^2 \right\}.
\]

\( (14) \)

When, for almost every \( a, \beta \mapsto f(a, \beta) \) has compact support, we show in Appendix E that it is possible to relate the smoothness defined via (14) to the Sobolev smoothness defined using Fourier series. There, smoothness corresponds to the function having bounded sum of squared \( L^2 \) norm of partial derivatives of degree \( \sigma \). When \( \omega \) are exponentials, the smoothness defined using Fourier series implies that all partial derivatives are square integrable. It corresponds to supersmooth classes (see, e.g., [11]),. As it is common in the literature (see [2, 10]), due to the different rate of decay of the singular values of \( \mathcal{F}_{x_0t} \) in Section 2.3, we consider slightly different supersmooth classes when \( W = i_{[-R,R]} \) and \( W = \cosh(\cdot / R) \). It is \( (\omega_k)_{k \in \mathbb{N}_0} = (e^{\kappa k \ln(k+1)})_{k \in \mathbb{N}_0} \) when \( W = i_{[-R,R]} \) and \( (\omega_k)_{k \in \mathbb{N}_0} = (e^{\kappa k})_{k \in \mathbb{N}_0} \), when \( W = \cosh(\cdot / R) \). This case is similar to nonparametric deconvolution where the density of the noise and the density of interest have Fourier transforms which both decay like \( e^{-\kappa |x|^r} \) when \( |x| \to \infty \) with same \( r > 1 \) but potentially different \( \kappa > 0 \) (see, e.g., [34, 44]). When \( W = i_{[-R,R]} \), we also consider the case where \( (\omega_k)_{k \in \mathbb{N}_0} = (e^{\kappa (k \ln(k+1))})_{k \in \mathbb{N}_0} \) and \( r > s > 1 \). In nonparametric deconvolution, this case corresponds to the case where the Fourier transforms of the noise and density of interest decay respectively as \( e^{-\kappa_1 |x|^r} \) and \( e^{-\kappa_2 |x|^r} \) for \( |x| \to \infty \), where \( \kappa_1, \kappa > 0 \) and \( r > s \) (see case 3 in Theorem 3.1 in [34]). The two values of \( q \) (1 or \( \infty \)) that we consider matter for the rates of convergence for supersmooth functions.
3. Lower bounds

The lower bounds on the minimax risk based on the mean integrated squared error (MISE) involve a function \( r \) (for rate) and take the form

\[
\exists \nu > 0 : \lim_{n \to \infty} \inf_{f_{\alpha, \beta}} \sup_{f_{\alpha, \beta} \in H_{u,W}^{0,\infty}(i) \cap D} \frac{\mathbb{E} \left[ \| \hat{f}_{\alpha, \beta} - f_{\alpha, \beta} \|_{L^2([0,R+1])}^2 \right]}{r(n)^2} \geq \nu. \tag{15}
\]

When we replace \( f_{\alpha, \beta} \) by \( f, \hat{f}_{\alpha, \beta} \) by \( \hat{f} \), remove \( D \) from the nonparametric class, and consider model (8), we refer to (15'). We use \( k_q := \mathbb{1}\{q = 1\} + p \mathbb{1}\{q = \infty\} \). We consider polynomial and exponential weights \((\omega_k)_{k \in \mathbb{N}_0}\) which yield respectively the smooth and supersmooth cases described in Section 2.6.

**Theorem 1** Let \( q \in \{1, \infty\} \), \( \phi \) increasing on \([0, \infty)\), \( 0 < l, s, \kappa, \sigma < \infty \), and \( w \) such that \( \int_0^\infty w(a)/a^4 < \infty \). When \( W = i_{[-R,R]} \),

\((T1.1a)\) \((\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0}, \) \( \phi \) is such that \( \lim_{t \to -\infty} \int_0^\infty \phi(t) e^{-2\tau t} dt = 0, f_X \) is known, and \( \| f_X \|_{L^\infty(X)} < \infty, S_X = \mathcal{X} \), then (15) holds with \( r(n) = (\ln(n)/\ln_2(n))^{-(2+qk/2\sqrt{\sigma})} \),

\((T1.1b)\) we consider model (8), \((\omega_k)_{k \in \mathbb{N}_0} = (\epsilon^{\kappa k \ln(k+1)})_{k \in \mathbb{N}_0}\), then (15') holds with \( r(n) = n^{-s/(2\kappa+2k_\alpha)}/\ln(n) \).

**Theorem 1** (Conclusion) Let \( W = \cosh(\cdot/R) \), we consider model (8),

\((T1.2a)\) \((\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0}, \) for all \( \sigma > 1/2 \), then (15') holds with \( r(n) = \ln(n/\ln(n))^{-(\kappa/\sigma)} \)

\((T1.2b)\) \((\omega_k)_{k \in \mathbb{N}_0} = (\epsilon^{\kappa k})_{k \in \mathbb{N}_0}, \) then (15') holds with \( r(n) = n^{-s/(2\kappa+2k_\alpha)} \).

For smooth functions when \((\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0}\), we obtain logarithmic lower bounds. This is consistent with the rates obtained in other severely ill-posed inverse problems with smooth functions of interest (see, e.g., [30, 34, 37]). Note that in (T1.1a) and (T1.2a), \( r(n) \) does not depend on the dimension \( p \) (see, e.g., [7]). The exponents in \( r(n) \) in (T1.1a) and (T1.2a) involve a maximum. This means that they hold without a maximum respectively when \( \sigma = 2+k_\sigma/2 \) and \( \sigma > 1/2 \). When \( \sigma \) is smaller than these values, the class of functions in the supremum in (15) is larger. So the lower bounds obtained for a larger \( \sigma \) are valid lower bounds. The result in (T1.2a) involves arbitrary \( \sigma \), but a sharp lower bound is for a sequence \( r(n) \) which goes slowly to 0, hence for \( \sigma \) close to 1/2 when \( \sigma \leq 1/2 \).

To match the rate of decay of the singular values, cases (T1.1b) and (T1.2b) consider supersmooth functions characterized by \((\omega_k)_{k \in \mathbb{N}_0} = (\epsilon^{\kappa k \ln(k+1)})_{k \in \mathbb{N}_0} \) when \( W = i_{[-R,R]} \) and by \((\omega_k)_{k \in \mathbb{N}_0} = (\epsilon^{k \kappa})_{k \in \mathbb{N}_0} \) when \( W = \cosh(\cdot/R) \). Such a situation corresponds to “2exp-severely ill-posed problems” (see, e.g., [7, 12, 44]), where the eigenvalues of the operator decay exponentially fast to zero and the weights are exponentials of the same form. Importantly, (T1.1b) and (T1.2b) will be supplemented by matching upper bounds showing that, for sufficiently smooth classes of functions, polynomial rates can be attained, for this severely ill-posed inverse problem. The lower bound (T1.2b) is, up to the logarithmic term, the one in the first case of Theorem 3.1 in [34] for nonparametric
deconvolution when the density of the noise and the density of interest have Fourier transforms decaying respectively like $e^{-k_q|x|}$ and $e^{-\kappa|x|}$ when $|x| \to \infty$. The links to classical nonparametric estimation bounds are not direct because the operator is a composition and sometimes does not have a SVD.

The proof of the lower bounds follow the usual reduction scheme where one obtains a lower bound on the left-hand side of (15) (or (15')) by replacing the supremum by a maximum over two well-chosen densities $f_{\alpha, \beta} \in \mathcal{H}_{w, W}^{q, \phi, \omega} (l) \cap D$ (respectively $f \in \mathcal{H}_{w, W}^{q, \phi, \omega} (l)$).

The standard approach is to choose the latter two as linear combinations of the singular functions of the operator $K$. However, there are two main complications. First, by Proposition 1, a SVD sometimes does not exist so we rely on the decomposition of $K$ as a composition of two operators. Second, to deal with the composition, we make use nonasymptotic (over $m$) bounds involving the SVD of $F_c$ with an explicit dependence on $c$. The proof of Theorem 1 thus relies on results in harmonic analysis that we prove in Appendix B for the PSWF and in [20] when $W = \cosh(\cdot/R)$.

4. Estimation

This section uses the following simplifying assumption.

**Assumption 2 (H2.1)** We have at our disposal i.i.d $(Y_i, X_i)_{i=1}^n$ and an estimator $\hat{f}_{X|X}$ based on $G_{n_0} = (X_i)_{i=-n_0+1}^0$ independent of $(Y_i, X_i)_{i=1}^n$.

(H2.2) $E$ is a set of densities on $X$ defined in (H1.3) such that, for $c_X, C_X \in (0, \infty)$, for all $f \in E$, $\|f\|_{L^\infty(X)} \leq C_X$ and $\|1/f\|_{L^\infty(X)} \leq c_X$, and, for $(v(n_0, E))_{n_0 \in \mathbb{N}} \in (0, 1)^\mathbb{N}$ which tends to 0, we have

$$\frac{1}{v(n_0, E)} \sup_{f_{X|X} \in E} \frac{\|\hat{f}_{X|X} - f_{X|X}\|_{L^\infty(X)}}{\|f_{X|X}\|_{L^\infty(X)}} = O_p(1).$$

We assume (H2.1) because the estimator involves estimators of $f_{X|X}$ in denominators. Alternative solutions exist when $p = 1$ (see, e.g., [28, 31, 33]). The availability of the preliminary sample $G_{n_0}$ in (H2.1) is not really an assumption but makes it explicit that the theory below relies on sample splitting. We do this for the simplicity of the proofs and relaxing it could be done in future work. In practice we do not use sample splitting in Section 5. These simulation results indicate that a practically oriented researcher does not need to implement sample splitting. By choosing well $X \subseteq S_X$ the assumption $\|f_{X|X}\|_{L^\infty(X)} \leq C_X$ and $\|1/f_{X|X}\|_{L^\infty(X)} \leq c_X$ in (H2.2) will be satisfied. By doing so, $\hat{f}_{X|X}$ effectively uses a subsample of the preliminary sample so $X$ should not be too small. To account for the sample splitting, we replace the expectation in the MISE by a conditional expectation. Denote by

$$\mathcal{R}_{n_0}^W (f_{\alpha, \beta}, f_{\alpha, \beta}) := \mathbb{E} \left[ \left\| \hat{f}_{\alpha, \beta} - f_{\alpha, \beta} \right\|_{L^2((1 \otimes W) \otimes \mathbb{P})}^2 \mid G_{n_0} \right].$$
Here the $L^2$-norm is a weighted one. It can be related to the unweighted norm. Indeed, it is $\mathbb{E}[\|\hat{f}_{\alpha,\beta} - f_{\alpha,\beta}\|_{L^2([-R+1])}^2|G_{an}]$ when $W = i_{[-R,R]}$ and $\text{supp}(\hat{f}_{\alpha,\beta}) \subseteq \mathbb{R} \times [-R,R]^p$, else
\[
\mathbb{E} \left[ \|\hat{f}_{\alpha,\beta} - f_{\alpha,\beta}\|_{L^2([-R+1])}^2 \bigg| G_{an} \right] \leq \|1/W\|_{L^\infty([-R,R])} R_{\alpha,\beta}^W \left( \hat{f}_{\alpha,\beta}, f_{\alpha,\beta} \right). \tag{16}
\]

### 4.1. Estimator

Based on the decomposition (6), the regularized inverse of $K$ is obtained in three steps. First, for $|t| \leq T$, we use an approximation of $F_{1st}^\alpha f_{\alpha,\beta}(t, \cdot \bigg|$ using the regularized inverse of the truncated Fourier operator $F_{ix_0}$. It involves spectral cut-off. Second, because the singular values of $F_{ix_0}$ go to 0 as $t$ goes to 0, using the SVD for estimation is problematic and we would rely on few to none coefficients. Rather, we rely on the interpolation of Section 2.5 for $t \in [-\epsilon, \epsilon]$, where $0 < \epsilon < 1 < T$. Third, we use a regularized inverse of the partial Fourier transform with respect to the first variable to recover $f_{\alpha,\beta}$.

Let us now make precise these three steps. For all $q \in \{1, \infty\}$, $0 < \epsilon < 1 < T$, $N \in \mathbb{R}^p$, $N(t) = |N(t)|$ for $\|t\| \leq T$, $N(t) = N(\epsilon)$ for $|t| \leq \epsilon$ and $N(t) = N(T)$ for $|t| > T$, a regularized inverse is obtained by:

(S.1) for all $t \neq 0$, obtain a first approximation of $F_0^\alpha f_{\alpha,\beta}(t, \cdot \bigg|$:
\[
F_0^q,N,T,0(t, \cdot) := \mathbb{I}\{\|t\| \leq T\} \sum_{|m| \leq N(t)} c_m(t) \varphi_{m,\alpha,\beta}^W, \tag{17}
\]
(S.2) for all $t \in [-\epsilon, \epsilon]$, we use the interpolation
\[
F_0^q,N,T,0(t, \cdot) := F_0^q,N,T,0(t, \cdot) \mathbb{I}\{\|t\| \geq \epsilon\} + \mathbb{I}_{\alpha,\beta} \left[ F_0^q,N,T,0(\cdot, \cdot) \right](t) \mathbb{I}\{\|t\| < \epsilon\}, \tag{18}
\]
(S.3) $F_0^q,N,T,0(\cdot, \cdot)$ is the inverse of partial Fourier transform with respect to the first variable.

Let us comment the choice of the estimator. For the regularized inverse of the truncated Fourier operator $F_{ix_0}$ in step (S.1), we choose spectral cut-off instead other regularization methods such as Tikhonov regularization or Landweber iteration (see, e.g., Section 1.2.2 in [2]). We do this because the SVD is fast to compute using numerical schemes developed recently (see Section 5). Moreover, the rates of spectral cut-off do not suffer from limitations due to qualification. Step (S.2) is the interpolation step. The interpolation is a nonstandard step, but it is essential to obtain the polynomial rates for supersmooth densities. The regularized inverse of the partial Fourier transform $F_{1st}$ is performed using the indicator $\mathbb{I}\{\|t\| \leq T\}$ in (S.3), which is a standard regularization device when inverting the Fourier transform which consists in removing high frequencies (see, e.g., [11, 15]). We could however use different smoothing kernels.
The regularized inverse depends on $R$ defined in Assumption 1 which should be known. This is not a smoothing parameter. When the $S_\beta$ is assumed to be compact then it is needed that $S_\beta \subseteq [-R,R]^p$, not $S_\beta = [-R,R]^p$. The choice of the parameter $a$ in the interpolation step (S.2) will be discussed later. The factor $x_0|t|^{p/2}$ in the definition of $K$ in (5) is used to show the continuity of $K$ in Proposition 1. Because it also appears on the right-hand side of (4), it does not enter the regularized inverse.

To deal with the statistical problem, we replace $c_m$ by

$$
\hat{c}_m := \frac{1}{n} \sum_{j=1}^{n} \frac{e^{i\beta_j} g_W(x_j)}{g_m(x_0)} \mathbb{I}\{X_j \in \mathcal{X}\},
$$

where $\hat{f}_{X|\mathcal{X}}(X_j) := f_{X|\mathcal{X}}(X_j) \vee \delta(n_0)$ and $\delta(n_0)$ is a trimming factor converging to zero. The intuition for (19) is that

$$
c_m(t) = \frac{1}{x_0} \int_{\mathcal{X}} E \left[ e^{i\alpha + it\beta^{\top} a} \frac{g_W(x_0)}{g_m(x_0)} \frac{x}{x_0} \mathbb{I}\{X \in \mathcal{X}\} \right].
$$

This yields the estimators $\hat{F}_{q,N,T,0}^{q,N,T,0}$, $\hat{F}_{q,N,T,\epsilon}^{q,N,T,\epsilon}$, and $\hat{f}_{\alpha,\beta}^{q,N,T,\epsilon}$. Because inverting the truncated Fourier operator $F_{tx_0}$ is more ill-posed near 0 (see Lemma B.4 and Theorem 7 in [20]), $F_{1}^{q,N,T,0}$ has a large variance for $t \in [-\epsilon, \epsilon]$. Hence we use interpolation (see Section 2.5).

We use $(\hat{f}_{\alpha,\beta}^{q,N,T,\epsilon})^+$ as a final estimator of $f_{\alpha,\beta}$ which has a smaller risk than $\hat{f}_{\alpha,\beta}^{q,N,T,\epsilon}$ (see [23, 45]). We use $n_{c} = n \wedge \lfloor \delta(n_0)/v(n_0, \mathcal{E}) \rfloor$ for the sample size required for an ideal estimator where $f_{X|\mathcal{X}}$ is known to achieve the rate of the plug-in estimator.

### 4.2. Upper bounds

The upper bounds below take the form

$$
\sup_{f_{\alpha,\beta} \in \mathcal{H}_{w,W}^{q,\phi,w}(l,M) \cap \mathcal{D}, f_{X|\mathcal{X}} \in \mathcal{E}} \frac{R_{n_0} \left( \hat{f}_{\alpha,\beta}^{q,N,T,\epsilon}, f_{\alpha,\beta} \right)}{r(n_{c})^2} = O_p(1).
$$

With the restriction $f_{\alpha,\beta} \in \mathcal{H}_{w,W}^{q,\phi,w}(l) \cap \mathcal{D}$, we refer to (20'). We also use $k'_{q} = p + 1 - k_q$.

#### Choice of the parameters.

In this section $N$, hence $N$, is a constant independent of $t$. $N$ is a function of $t$ only in Section 4.3. Denote, for $u > 0$, by $K_u(w) := a \mathbb{I}\{w \neq i[-2a] \} + u \mathbb{I}\{w = i[-2a] \}$. We use $T = \phi^f(\omega N)$, $a = w^f(\omega_{N}^2)$ when we do not know that $S_{\alpha} \subseteq [-a,a]$ and cannot take $w = i[-2a]$, and

imsart-bj ver. 2014/10/16 file: Estimation_RCLV_paper-revised-042021.tex date: April 7, 2021
Random coefficients with bounded regressors

1. when \( W = i_{[-R,R]} \), \( N \) is solution to

\[
2k_q \left( N + \frac{k_q}{2} \right) \ln \left( NK_2(1) \right) + \ln(\omega_N^2) + (p-1) \ln(N) = \ln(n_e) \tag{21}
\]

and \( \epsilon = 7e\pi / (R\epsilon_0 K_2(1)) \);

2. when \( W = \cosh(\cdot / R) \), \( N \) is solution to

\[
2k_q \left( N + \frac{k_q}{2} \right) \ln \left( K_2(\epsilon) \right) + \ln(\omega_N^2) + \frac{p-1}{q} \ln(N) = \ln(n_e) \tag{22}
\]

and \( \epsilon = 7e^2\pi / (2R\epsilon_0 K_2(14e^2)) \).

(21)-(22) describe the choices of \( N \) which realize the bias-variance trade-off. Under these choices of tuning parameters, theorems 2 and 3 provide the convergence rates \( r(n_e) \) in (20) or (20'). Results are presented in increasing order of restrictions on \( \phi \), then on \( (\omega_k)_{k \in \mathbb{N}_0} \), and on \( w \).

**Theorem 2** Let \( W = i_{[-R,R]} \), \( \mathbb{S}_A \subseteq [-R,R]^p \), \( q \in \{1, \infty\} \), and \( l, M, s, \sigma, \kappa, \mu, \gamma, \nu > 0 \). Consider \( \phi = 1 \lor |\cdot|^s \),

\( (T2.1) \) \( (\omega_k)_{k \in \mathbb{N}_0} = (k^s)_{k \in \mathbb{N}_0} \), and \( w = 1 \lor |\cdot|^s \), then (20) holds with \( r(n_e) = (\ln(n_e) / \ln(2n_e))^{-\sigma} \),

\( (T2.2) \) \( (\omega_k)_{k \in \mathbb{N}_0} = (e^{\kappa \ln(k+1)})_{k \in \mathbb{N}_0} \), with \( s \geq \kappa(p+1) / (2k_q(\nu I_{W \neq i_{[-a,a]}}) + 1) \), and

\( (T2.2a) \) \( w^f(e^{2\kappa|\cdot|^{1+1}}) = \nu^r \), then (20) holds with \( r(n_e) = n_e^{-\kappa/(2\kappa+2\nu+1)k_q} \ln(n_e)^{\Lambda(\nu)} \);

\( (T2.2b) \) \( \nu \) such that \( \mathbb{S}_A \subseteq [-a,a] \), \( w = i_{[-a,a]} \), then (20') holds with \( r(n_e) = n_e^{-\kappa/(2\kappa+2k_q)\ln(n_e)^{\Lambda(0)}} \),

where \( \Lambda := (2 + \cdot)^{p-1}(1 - (\kappa(p+1) / (2s(k_q(\cdot + 1) + \kappa))) / 2) \).

\( (T2.3) \) Consider \( \phi = e^{\gamma |\cdot|} \), \( (\omega_k)_{k \in \mathbb{N}_0} = (e^{\kappa(k \ln(k+1))})_{k \in \mathbb{N}_0} \), \( w \) such that \( w^f(e^{2\kappa|\cdot|^{1+1}}) = \nu^r \), and \( r > 1 \), then (20) holds with \( r(n_e) = \sqrt{\varphi(n_e)/n_e} \), where

\[
\varphi := \exp \left( -\sum_{k=1}^{k_0} (-1)^k \frac{d_k \ln(\cdot)^{(1/r-1)k+1}}{\ln(\cdot)^{p+1+(p-1)/r}} \right) \sqrt{\ln(\cdot)^{p+1+(p-1)/r}},
\]

\( d_0 = 2\kappa(1 + (k_q(1+\nu) + (2 + \nu)p-1)/(\ln(2 + 1/p)(1 + 1/p))^{r-1}) \), \( k_0 := \lfloor r/(r-1) \rfloor \), and for \( k \in \{1, \ldots, k_0\} \),

\[
d_k : = \left( \frac{k_q(1+\nu)(2\kappa)^{1-1/r} I_{k \equiv 0 (mod \ 2)} \kappa(1 + \ln(2)/\ln(1 + 1/p))^{r}}{\kappa(1 + \ln(2)/\ln(1 + 1/p))^{r}} \right)^{k} + \left( \frac{k_q(1+\nu) + (2 + \nu)p-1} {\kappa d_0^{1/r-1}} \right)^{k}.
\]

Theorem 1 shows the rate in (T2.1) is optimal when \( f_X \) is known and \( \mathbb{S}_X = \mathcal{X} \). It is the same as in \cite{39} for deconvolution with a known characteristic function of the noise on an
interval when the signal has compact support. The rates in Theorem 2 are independent of $p$ as common for severely ill-posed problems (see [13, 20]). The rates in (T2.2) and (T2.3) are polynomial and nearly parametric even if the problem is severely ill-posed. This means that we can lose little by going from a possibly misspecified parametric model to a nonparametric one. The rate in (T2.3) is similar to the rate obtained in [34] in the deconvolution problem with supersmooth density of interest and noise density, when the former is smoother than the later. The next theorem relaxes the condition $S_{\beta}$ is compact.

**Theorem 3** Let $W = \cosh(\cdot/R)$. For all $q \in \{1, \infty\}$, $l, M, s, \sigma, \kappa, \mu > 0$, $\phi = 1 + |\cdot|^\phi$, $\sigma 
$

- (T3.1) if $(\omega_k)_{k \in \N_0} = (k^\sigma)_{k \in \N_0}$, and $w = 1 |\cdot|^\phi$, then (20) holds with $\nu(n_e) = (\ln(n_e) / \ln_2(n_e))^{-\sigma}$,
- (T3.2) if $(\omega_k)_{k \in \N_0} = (e^{\kappa k})_{k \in \N_0}$, $a$ such that $S_\alpha \subseteq [-a, a]$, and $w = e^{-2 - \sigma}$, then (20) holds with $\nu(n_e) = n_e^{-\kappa/(2\kappa + 2k)} \ln(n_e)^{(p-1)\kappa/(2q\kappa + k)}$.

When $1/v(n_0, E) \geq n$ and $p = 1$, the rate in (T3.2) matches the lower bound (T1.2b) for model (8). Again, the results (T2.2), (T2.3), and (T3.2) are related to those for “2exp-severely ill-posed problems”, and [44] obtains the same rates up to logarithmic factor as in (T3.2) when $1/v(n_0, E) \geq n$.

### 4.3. Data-driven estimator

The estimator depends on two parameters and an optimal estimator can be obtained when they are chosen depending on unknowns such as $s, \sigma$ or $\kappa$, etc. To make it practical, we show that a nearly minimax estimator can be obtained in a data-driven way by a type of Goldenshluger-Lepski method (see, e.g., [25, 26, 35]).

To gain insight on the data-driven choices $\hat{N}$ and $\hat{T}$ of $N$ and $T$, let us sketch the proof when $\hat{f}_X|X = f_X|X$ (hence we simply write $R^W$). Consider $W = i_{[-R,R]}$. Let $\epsilon \in (0,1)$ and $q \in \{1, \infty\}$. We use the following upper bound on the risk: for all $f_{\alpha, \beta} \in \mathcal{H}_{\alpha, \beta}^q(I, M)$,

$$R^W(f_{\alpha, \beta}, \hat{N}, \hat{T}, f_{\alpha, \beta}) \leq C \int_{|t| \leq |l|} \mathbb{E} \left[ L_q^W(t, \hat{N}(t), \hat{T}(t)) \right] \, dt + CM^2 \bar{w}(\alpha),$$

where $\bar{w} := \mathbb{1}(w \neq i_{[-2,2]})/w$, and, for all $t \in \mathbb{R}$, $N \in \mathbb{N}_0$, and $T' \in [0, \infty)$,

$$L_q^W(t, N, T') := \left\| \left( F_{1st}[f_{\alpha, \beta}] - \hat{F}_1^{q,N,T',0}(t, \cdot, \cdot) \right) (t, \cdot, \cdot) \right\|_{L^2(W \otimes \rho)}^2.$$ 

This is proved in the appendix using the Plancherel identity and (A.31). The latter is derived from Proposition 3. It is important that the first integral in (23) is restricted to $\{t \in \mathbb{R} : \epsilon \leq |t| \}$ by the considerations in Section 2.5. The upper bound in (23) when $\hat{N}$ and $\hat{T}$ are nonrandom is the one we use to obtain theorems 2 and 3. The aim of the selection rules is to obtain an upper bound on the right-hand side of (23) with a similar quantity but with arbitrary nonrandom $\hat{N}$ and $\hat{T}$.

imsart-bj ver. 2014/10/16 file: Estimation_RCLV_paper-revised-042021.tex date: April 7, 2021
Random coefficients with bounded regressors

Let us start with intuitions on the selection rule for \( \hat{N} \). We define a maximum \( N_{\text{max},q}^W \) for the values that \( \hat{N}(t) \) can take. We set, for \( N_0 \in \{0, \ldots, N_{\text{max},q}^W \} \),

\[
B_1(t, N_0) := \max_{N_0 \leq N' \leq N_{\text{max},q}^W} \left( \left\| \hat{F}_{1q}^{q,q',0}(t, \cdot) - \hat{F}_{1q}^{q,q',0}(t, \cdot) \right\|_{L^2(W^{sp})}^2 - \Sigma(t, N') \right) + \]

where \( \Sigma(t, N') \) is a penalty term, called the “majorant” (see [26]), which expression is given in equation (26) below. To explain the role of \( B_1(t, N_0) \) and \( \Sigma(t, N') \) in the choice of \( \hat{N}(t) \), consider \(|t| \geq \epsilon \) and \( N', N_0 \in \{0, \ldots, N_{\text{max},q}^W \} \), \( N' \geq N_0 \). The majorant is chosen to compensate for the fluctuation of the squared \( L^2(W^{sp}) \) norm of the statistic \( b \mapsto \hat{F}_{1q}^{q,q',0}(t, b) - \hat{F}_{1q}^{q,q',0}(t, b) \). The term

\[
\left( \left\| \hat{F}_{1q}^{q,q',0}(t, \cdot) - \hat{F}_{1q}^{q,q',0}(t, \cdot) \right\|_{L^2(W^{sp})}^2 - \Sigma(t, N') \right) +
\]

is a proxy for \( \left\| \hat{F}_{1q}^{q,q',0}(t, \cdot) - \hat{F}_{1q}^{q,q',0}(t, \cdot) \right\|_{L^2(W^{sp})}^2 \). The maximum of the later is \( \left\| \hat{F}_{1q}^{q,q',0}(t, \cdot) - \hat{F}_{1q}^{q,q',0}(t, \cdot) \right\|_{L^2(W^{sp})}^2 \). It is the square norm of the bias due to the cut-off at \( N_0 \). The majorant is an upper bound on the weighted integral of the variance of \( \hat{F}_{1q}^{q,q',0}(t, \cdot) \). As a result, \( \hat{N} \) is selected as

\[
\forall t \in \mathbb{R} \setminus (-\epsilon, \epsilon), \quad \hat{N}(t) \in \arg\min_{0 \leq N \leq N_{\text{max},q}^W} \left( B_1(t, N) + c_1 \Sigma(t, N) \right), \quad (24)
\]

where \( c_1 = 1 + 1/(2 + \sqrt{5})^2 \).

In the same spirit, define a grid for the possible values of \( \hat{T} \),

\[
\mathcal{T}_n := \{ q^k : k = 1, \ldots, K_{\text{max}} \},
\]

where \( K_{\text{max}} := \lfloor \zeta_0 \ln(n)/\ln(2) \rfloor \) and \( \zeta_0 = 1/((1 + 4p(1 + \mathbb{I}(W = \cosh(\cdot/R)))) \). The choice of \( \hat{T} \) relies on, for all \( T \in \mathbb{R} \) and \( N \in \mathbb{N}^R \),

\[
B_2(T, N) := \max_{T' \in \mathcal{T}_n, T' \geq T} \left( \left\| \hat{F}_{1q}^{q,q',0}(t, \cdot) - \hat{F}_{1q}^{q,q',0}(t, \cdot) \right\|_{L^2(W^{sp})}^2 - \Sigma_2(T', N) \right) + \]

and on the majorant

\[
\Sigma_2(T, N) := \int_{|t| \leq T} \Sigma(t, N(t))dt.
\]

Then, \( \hat{T} \) is selected by

\[
\hat{T} \in \arg\min_{T \in \mathcal{T}_n} \left( B_2(T, \hat{N}) + \Sigma_2(T, \hat{N}) \right). \quad (25)
\]

The expression of the majorant is as follows: for all \( N \in \mathbb{N}_0^R, N_0 \in \mathbb{N}_0, \) and \( t \neq 0 \),

\[
\Sigma(t, N_0) := 8(2 + \sqrt{5})(1 + 2p_n) \frac{c_X}{n} \left( \frac{|t|}{2\pi} \right)^p \nu_q^W(x_0 t, N_0), \quad (26)
\]
\[
\nu_q^{W}(t, N_0) := (N_0 + 1)^{k_q} Q_q^{W}(N_0) \left( 1 + \frac{7 \epsilon \pi (N_0 + 1)}{R |t|} \right)^{2k_q N_0 + p}, \\
Q_q^{W}(N_0) := I\{q = \infty\} 2^p I\{W = i_{[-R,R]}\} + I\{W = \cosh(\cdot/R)\} + \frac{(N_0 + p - 1)^p - 1 I\{q = 1\}}{(p - 1)!},
\]

and \(p_n := 3 \sqrt{6(1 + \zeta_0) \ln(n)}\). The term \(\nu_q^{W}(t, N_0)\) is an upper bound on the sum of the inverse of the square of the singular values up to \(N_0\) (see Lemma B.2). We also take \(N_{\max,q}^{W} = [N_{\max,q}^{W}]\), where \(N_{\max,q}^{W}\) satisfies \(2k_q (N_{\max,q}^{W} + k_q / 2) \ln(7 \epsilon \pi N_{\max,q}^{W} / (R \epsilon_0)) = \ln(n)\).

Let us explain how the definition (25) of \(\hat{T}\) yields an upper bound on the right-hand side of (23) by a quantity where \(\hat{T}\) can be replaced by an arbitrary \(T\). By arguments in the proof of Lemma A.5 for the first inequality and (25) for the second, we have, for all \(T \in \mathcal{T}_n\),

\[
\int_{t \leq |t|} E \left[ \mathcal{L}_q^{W}(t, \hat{N}(t), \hat{T}) \right] dt \\
\leq \frac{2 + \sqrt{5}}{\sqrt{5}} \int_{t \leq |t|} E \left[ \mathcal{L}_q^{W}(t, \hat{N}(t), T) \right] dt \\
+ (2 + \sqrt{5}) \left( E \left[ B_2 \left( \hat{T}, \hat{N} \right) + \Sigma_2 \left( \hat{T}, \hat{N} \right) \right] + E \left[ B_2 \left( T, \hat{N} \right) + \Sigma_2 \left( T, \hat{N} \right) \right] \right) \\
\leq \frac{2 + \sqrt{5}}{\sqrt{5}} \int_{t \leq |t|} E \left[ \mathcal{L}_q^{W}(t, \hat{N}(t), T) \right] dt + 2(2 + \sqrt{5}) E \left[ B_2 \left( T, \hat{N} \right) + \Sigma_2 \left( T, \hat{N} \right) \right].
\]

We then rely on an upper bound on \(E \left[ B_2 \left( T, \hat{N} \right) \right] \) in the second term on the right-hand side. It involves a term proportional to the first one :

\[
E \left[ B_2 \left( T, \hat{N} \right) \right] \leq \left( 1 + \frac{2}{\sqrt{5}} \right) \int_{t \leq |t|} E \left[ \mathcal{L}_q^{W}(t, \hat{N}(t), T) \right] dt + O \left( \frac{1}{n} \right).
\]

The \(O(1/n)\) term is independent of \(T\) and \(\hat{N}\) and is obtained using a Talagrand’s inequality. Similarly, (24) allows to obtain yet another upper bound which replaces \(\hat{N}\) by an arbitrary nonrandom \(N\). We conclude because the final upper bound (A.68) has a similar form as the upper bound (A.29) appearing when we deal with nonrandom \(\hat{N}\) and \(\hat{T}\) in theorems 2 and 3.

The upper bounds below take the form

\[
\sup_{f_{a,b} \in \mathcal{H}^{\phi, \psi}_N (t, M) \cap \mathcal{D}} \frac{R_{n,0}^{W} \left( \hat{f}_{a,b}, \hat{T} \right)}{r(n)^2} \leq \frac{O_p}{r(n)^2 \epsilon(n_0) \ln(n_0)^{-p}}
\]

where the above \(O_p\) symbol means that the left-hand side doubly-indexed sequence of random variables, denoted for simplicity by \(X_{n,0}\), is such that, for all \(\epsilon > 0\), there exists
such that \( \mathbb{P}(|X_{n_0,n}| \geq M) \leq \epsilon \) for all \((n_0, n) \in \mathbb{N}_0^2\) satisfying the condition underneath the \(O_p\) symbol. The results in this section are for \(v(n_0, \mathcal{E})/\delta(n_0) \leq n^{-2}\ln(n)^{-p}\), in which case \(n_\alpha = n\). We refer to (27') when we use the restriction \(f_{\alpha, \beta} \in \mathcal{H}_{w, W}(l) \cap \mathcal{D}\).

**Theorem 4** Let \(W = i_{[-R, R]}\), \(S_\beta \subseteq [-R, R]^p\). For all \(l, M, s, \sigma > 0, q \in \{1, \infty\}, \phi = 1 \vee |\cdot|^s\), if

\[
\begin{align*}
(T4.1) & (\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0}, \ a = 1/\epsilon, w = 1 \vee |\cdot|, \text{ and } \epsilon = 7e^\pi/(Rx_0 \ln(n)), \text{ then (27)} \\text{ holds with } r(n) = (\ln(n) / \ln_2(n)^{-\sigma}); \\
(T4.2) & (\omega_k)_{k \in \mathbb{N}_0} = (e^{k\ln(1+k)})_{k \in \mathbb{N}_0}, \ a \text{ such that } S_\alpha \subseteq [-q, q], \ w = i_{[-q, q]}^\sigma, \ \epsilon = 7e^\pi/(Rx_0), \ \text{ and } s > (2p+1/2)\nu((\kappa(p+1)/(2k_\eta), \text{ then (27')} \text{ holds with } r(n) = n^{-\kappa/(2\kappa + 2k_\eta)} \ln(n)^{1/2} + \Lambda(0) \text{ and } \Lambda \text{ defined in (T2.2).}
\end{align*}
\]

The rate in (T4.2) matches, up to a logarithmic factor, the lower bound in Theorem (T1.1b) for model (8). (T4.2) relies on \((T4.1)\) for model (8). However, by (16), we obtain the same rate up to a logarithmic factor for the risk involving the weight \(\cosh(\cdot)/R\). Theorem 1 and (T5.2) with \(q = \infty\) show that \(\hat{f}_{\alpha, \beta}^{\hat{N}, \hat{T}}\) is adaptive.

**Theorem 5** Let \(W = \cosh(\cdot)/R\). For all \(l, M, s, \sigma > 0, q \in \{1, \infty\}, \phi = 1 \vee |\cdot|^s\), if

\[
\begin{align*}
(T5.1) & (\omega_k)_{k \in \mathbb{N}_0} = (k^\sigma)_{k \in \mathbb{N}_0}, \ a = 1/\epsilon, w = 1 \vee |\cdot|, \text{ and } \epsilon = 7e^\pi/(2Rx_0 \ln(n)), \text{ then (27)} \\text{ holds with } r(n) = (\ln(n) / \ln_2(n)^{-\sigma}); \\
(T5.2) & (\omega_k)_{k \in \mathbb{N}_0} = (e^{k\ln(1+k)})_{k \in \mathbb{N}_0}, \ a \text{ such that } S_\alpha \subseteq [-q, q], \ w = i_{[-q, q]}^\sigma, \ \epsilon = \pi/(4Rx_0), \ \text{ and } s > 4p+1/2, \text{ then (27')} \text{ holds with } r(n) = n^{-\kappa/(2\kappa + 2k_\eta)} \ln(n)^{1/2} + (p-1)\kappa/(2q(k_\eta + k_\eta)).
\end{align*}
\]

Theorem 5 is different for the lower and upper bounds in theorems 1 and 5. However, by (16), we obtain the same rate up to logarithmic factors for the risk involving the weight \(\cosh(\cdot)/R\). Theorem 1 and (T5.2) with \(q = \infty\) show that \(\hat{f}_{\alpha, \beta}^{\hat{N}, \hat{T}}\) is adaptive.

### 5. Simulations

Let \(p = 1, q = \infty, \) and \((\alpha, \beta)^T = \xi_1 D + \xi_2 (1 - D)\) with \(\mathbb{P}(D = 1) = \mathbb{P}(D = 0) = 0.5\). The law of \(X\) is a truncated normal based on a normal of mean 0 and variance 2.5 and truncated to \(X\) with \(x_0 = 1.5\). The laws of \(\xi_1\) and \(\xi_2\) are either: (Case 1) truncated normal based on a normal of mean 0 and variance 2.5 and truncated to \(X\) with \(x_0 = 1.5\).
normals based on normals with means $\mu_1 = \left( -\frac{2}{3} \right)$ and $\mu_2 = \left( \frac{1}{3} \right)$, same covariance $(\frac{7}{2} \frac{1}{2})$, and truncated to $[-6,6]^{p+1}$ or (Case 2) nontruncated. The estimators in [29, 31] cannot be used in this context. Case (1) could be treated with [5] in this particular case where $p = 1$. This requires choosing many parameters and [5] only provides a consistency result. Case (2) allows for unbounded errors as in usual linear regression models (thus is very different from tomography problems) and no other nonparametric method is available up to our knowledge.

Table 1 compares $\mathbb{E}[\|\hat{f}_{\alpha,\beta}^{\infty,N,T,\epsilon} - f_{\alpha,\beta}\|_{L^2([-7.5,7.5]^2)}^2]$ and $\min_{T \in T_n, N \in N_{n,H}} \mathbb{E}[\|\hat{f}_{\alpha,\beta}^{\infty,N,T,\epsilon} - f_{\alpha,\beta}\|_{L^2([-7.5,7.5]^2)}^2]$ (risk of the oracle) for cases 1 and 2. The Monte-Carlo experiment uses 1000 simulations.

<table>
<thead>
<tr>
<th>Case</th>
<th>MISE</th>
<th>$W = i_{[-7.5,7.5]}$</th>
<th>Case 1</th>
<th>$W = \cosh (/7.5)$, Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 300$</td>
<td>$n = 500$</td>
<td>$n = 1000$</td>
<td>$n = 300$</td>
<td>$n = 500$</td>
</tr>
<tr>
<td>data-driven</td>
<td>0.092</td>
<td>0.086</td>
<td>0.083</td>
<td>0.089</td>
</tr>
<tr>
<td>oracle</td>
<td>0.091</td>
<td>0.086</td>
<td>0.082</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Table 1. Risk

Table 1 shows that the data-driven method works well even for small samples. However, the rate of decay of the MISE is small which we suspect comes from the fact that the true density is a mixture.

Figure 1 (resp. Figure 2) displays summaries of the law of the estimator for $W = i_{[-7.5,7.5]}$ (resp. $W = \cosh (/7.5)$) in Case 1 (resp. Case 2) and $n = 1000$. As standard in the literature (see, e.g., [14, 17]), the multiplicative constant appearing in $\Sigma$ is in practice calibrated from a simulation study. $\hat{f}_{X|X \in X}$ is obtained with the same data to illustrate that sample splitting is unnecessary in practice and only used for the mathematical argument. For $\hat{f}_{X|X \in X}$ we use a Gaussian kernel density estimator using the R package ks and the multivariate plug-in bandwidth selector of [46]. $\epsilon$ is chosen as in (T4.1) and (T5.1) respectively for Case 1 and Case 2. The estimator requires the SVD of $\mathcal{F}_c$. By Proposition B.1, we have $g_m^{W(\cdot/R),\epsilon} = g_m^{W,\Re}$ for all $m \in \mathbb{N}_0$. When $W = i_{[-1,1]}$, the first coefficients of the decomposition on the Legendre polynomials are obtained by solving for the eigenvectors of two tridiagonal symmetric Toeplitz matrices (see Section 2.6 in [42]). When $W = \cosh$, we refer to Section 7 in [20]. We use the image of $g_m^{W,\Re}$ by the adjoint of $\mathcal{F}_c$ (see Appendix A.1) and that $\varphi_m^{W,\Re}$ has norm 1 to get the rest of the SVD. We obtain the Fourier inverse by fast Fourier transform. We use a resolution of $2^{13}$, which appears on simulations to realise a good trade-off between computational time and precision. For more details about the implementation, we refer to the vignette [21] of the package RandomCoefficients.
Random coefficients with bounded regressors

(a) True density
(b) Mean of estimates
(c) 97.5% quantile of estimates
(d) 2.5% quantile of estimates

Figure 1. Case 1, \( W = \iota_{[-7.5,7.5]} \)

(a) True density
(b) Mean of estimates
(c) 97.5% quantile of estimates
(d) 2.5% quantile of estimates

Figure 2. Case 2, \( W = \cosh (\cdot/7.5) \)

Acknowledgements

The authors acknowledge financial support from the grants ERC POEMH 337665 and ANR-17-EURE-0010. They are grateful to the seminar participants at Berkeley, Brown,
CREST, Duke, Harvard-MIT, Rice, TSE, ULB, University of Tokyo, those of 2016 SFDS, ISNPS, Recent Advances in Econometrics, and 2017 IAAE conferences for comments.

References


Random coefficients with bounded regressors


