Electronic Journal of Statistics
Vol. 0 (0000) 1
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Nonparametric regression for current status censored response

Sam Efromovich[1]

[1] *Department of Mathematical Sciences, University of UTDallas, Richardson, USA. e-mail:* efrom@utdallas.edu

Abstract: It is known that consistent nonparametric regression for a current status censored (CSC) response and a univariate predictor is possible. The paper, for the first time in the literature, presents sharp minimax theory of mean integrated squared error (MISE) convergence and methodology of adaptive estimation. Rate of the MISE convergence is classical, the sharp constant quantifies the effect of CSC, and the results hold under a mild assumption on smoothness of nuisance functions not tied to smoothness of the regression. Then the setting is extended to a multivariate predictor. Real and simulated examples are presented, as well as an illuminating comparison of theoretical results known for CSC and directly observed data.

Keywords and phrases: Adaptation, Anaerobic digestion, Curse of dimensionality, Sharp minimax, Survival analysis.

Contents

1. Introduction

Consider the problem of estimating a regression function

$$m(x) := \mathbb{E}\{T|X = x\}, \quad x \in [0,1] \tag{1.1}$$

where T is the continuous lifetime of interest (time to an event of interest) and X is the continuous random predictor supported on $[0,1]$. It is well known that if $m(x)$ is α-fold differentiable then, based on a sample of size n from (X,T), it can be estimated with optimal rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE (mean integrated squared error) convergence, see [12,68]. Further, if the density f^T of T is of interest and it is α-fold differentiable, then it can be estimated with the same optimal rate as the regression, see [9,68]. This is the famous principle of equivalence between the nonparametric regression and density estimation problems discussed in [13].

In the paper we consider a setting when the lifetime T is not observed directly. Instead, there exists a possibility to check status of the event of interest at some random moment of time Z, called the monitoring time. Then the available observation is the triplet of random variables (X, Z, Δ) where $\Delta := I(T \leqslant Z)$ is the status (indicator) of the event of interest, namely the status is equal to 1 if the event of interest already occurred at time Z and the status is zero otherwise. A sample from (X, Z, Δ) is called current status censored. Let us also introduce notation $\Delta' := 1 - \Delta$ that will be frequently used in the paper. It is known that for density estimation the CSC dramatically slows down the rate of MISE convergence, and exact rates will be presented shortly. If the above-mentioned principle of equivalence between density and regression holds for CSC, then this is a bad news for the regression. As we will see, the principle breaks down for CSC data and we will be able to estimate regression with the classical rates and even evaluate sharp constants for the MISE convergence. At the same time, we will also see that nonparametric regression is the exemption to the general rule that CSC slows down rate of convergence. This is an interesting specific of the CSC because there is no such phenomenon for right censored observations when the sampling is from $(\min(T, C), I(T \leqslant C))$ and C is the censoring lifetime.

Current status censoring (CSC), also known as “case I” interval censoring, arises in different applications ranging from biostatistics and engineering to econometrics. For instance, in a clinical study devoted to the time T from cancer surgery to cancer reoccurrence, the follow-up examination at time Z after the surgery determines whether or not the cancer is present. We do not observe T and instead observe the monitoring time Z and the indicator of cancer $\Delta = I(T \leqslant Z)$. Note that $\Delta' = 1$ means that at the time of examination the patient is cancer free. A number of covariates, including age and size of tumor, may be of interest. In rodent bioassay experiments, when the time from inducing a chemical to developing a disease is the lifetime of interest T, sacrifices are often used to detect the disease. Then the available information is the time Z of the sacrifice and the indicator of disease, as well as some characteristics of rodents and the chemical. In engineering experiments, destructive tests are used to find whether a system has failed. Another large cluster of CSC applications is in econometrics when the interest is in developing choice models for individual and household behavior (Nobel Prize in Economic Sciences in 2000). The following example sheds light on the binary choice model. The interest is in price T that an individual is ready to pay for an item whose asking price is Z, and the available observations are the asking price Z and the indicator of sale $\Delta' = I(T \geqslant Z)$. Note that in the econometrics the indicator Δ' is called the observable binary outcome. This and many other econometric examples can be found in [7,45,46,63,69].

CSC is a well-known problem in survival analysis, see a discussion in books [5,16,20,30,33,60], reviews in [22,59], and more recent papers [3,18,38,47] where further references may be found. Differentiable functionals were studied in [66], and the theory pointed upon regular and irregular convergence rates for CSC observations. As we will see shortly, similar phenomena exist for adaptive estimation of nonparametric curves under MISE criterion, and this requires developing a targeted methodology for a specific problem at hand. Let us also mention papers devoted specifically to CSC regression models. Proportional and additive hazards are popular models discussed in [19,27,39]. The accelerated failure time model is explored in [54]. The proportional odds regression is studied in [56]. There is a vast literature devoted to semiparametric models. These models are explored via sieve maximum likelihood, linear and additive transformation models, ensemble variable selection, EM algorithm, penalized log-likelihood estimation in [6,36,41,43,48,49,58,70,71]. Simultaneous estimation and variable selection with broken adaptive ridge regression are considered in [73]. Linear regression is studied in [37,50,57]. Interesting results for hazard regression can be found in [4]. An estimator of regression parameters in the accelerated failure time model by inverting a Wald-type test for testing a null proportional hazards model is proposed in [62]. Study of a semiparametric probit model and its applications can be found in [8,40]. Model with varying-coefficient partially linear proportional odds is investigated in [42]. Theory of semiparametric linear regression is developed in [21] where asymptotically normal estimate is proposed. Nonparametric regression of the status on the predictor is explored in [25] where a modified maximum rank correlation estimator is proposed. There is a relatively large literature devoted to sieve maximum likelihood regression, see a discussion and reviews in [44,75]. There is also a rich literature devoted to estimation of linear functionals like $\int_0^\infty g(t)S^T(t)dt$ and $\int_0^\infty g(t)f^T(t)dt$. The corresponding theory of efficient estimation, methodology and examples can be found in [2,23,28,32,72,74].

It is fair to notice that for CSC main theoretical results and a majority of literature are devoted to estimation of the distribution of T based on a sample of size n from (Z, Δ). Let us present some known results and compare them with results for a direct sample from T. For the direct sampling, a ν-fold differentiable survival function $S^T(t) := \mathbb{P}(T > t)$ can be estimated with a parametric rate n^{-1} regardless of its smoothness (regardless of ν), but CSC slows down the rate to $n^{-2\nu/(2\nu+1)}$ and makes the rate depended on the smoothness, [16,18,28,60,68]. Similarly, for a direct sample the optimal MISE rate of estimating an α-fold differentiable density $f^T(t)$ is $n^{-2\alpha/(2\alpha+1)}$, while CSC slows it down to $n^{-2\alpha/(2\alpha+3)}$ which is the optimal rate of estimating a trivariate density for direct observations, see a discussion in [20,68]. Further, CSC makes adaptation to unknown smoothness of a curve more complex, see [21]. Another important remark is as follows. For directly observed data, the familiar principle of equivalence between density and regression estimation implies that rates of convergence should be the same, see a discussion in [13], but we will see shortly that the principle breaks down for CSC data.

To shed light on complexity of CSC regression, Figure 1 presents a real environmental example of CSC regression data (the scattergram) that will be explained shortly. Note that we do not observe lifetimes of interest, and each observation only tells us the status of T with respect to a shown monitoring time Z. Accordingly, it is difficult to visualize an underlying regression function in a CSC scattergram, while it is feasible for a standard regression. As we will see shortly, special methods and a corresponding software are needed for analysis of CSC data.

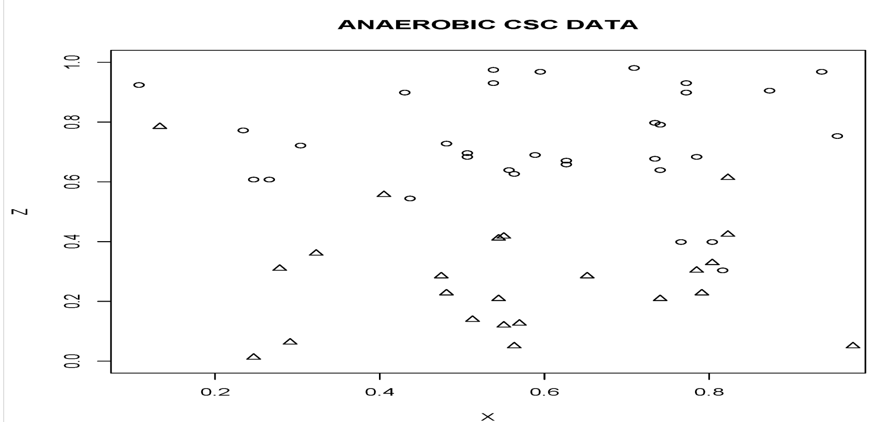

Fig 1. BIFAR data for anaerobic digestion of municipal wastewater. The circles and the triangles show pairs (X, Z) corresponding to status 1 and 0, respectively. Observations are rescaled onto the unit square.

Now let us explain the data. Anaerobic digestion of organic municipal solid waste is considered as a key element in sustainable municipal waste management due to its benefits for energy, environment, and economy. This process dramatically reduces emission of greenhouse gases, generates renewable natural gas, and produces fertilizers and soil amendments, see [1,35]. Environmental company BIFAR has been interested in a minimal time T for required anaerobic digestion of municipal sludges with different thickness X which is created by ether the gravity thickening, or gravity belt thickening, or a centrifuge. Anaerobic digestion happens in the absence of oxygen in a sealed, oxygen-free tank called an anaerobic digester, and the word anaerobic means "in the absence of oxygen". Because an anaerobic digester is sealed, the minimal time T to desired digestion cannot be directly observed. Instead, an anaerobic process may be terminated at a monitoring time Z, and then a laboratory analysis of the treated sludge will show its status $\Delta = I(T \leqslant Z)$. Results of the BIFAR's controlled experiment are shown in Figure 1, and regression analysis of the CSC data will be presented shortly in Section 5. Let us stress that the example is of interest because the CSC methodology is the only feasible method to gain access to the anaerobic process.

It is well known that, for small samples and a nonparametric estimation, constants of MISE convergence are as important as rates, see a discussion in [13]. This is why the paper develops a sharp minimax theory of regression estimation which explains how CSC affects regression estimation. Further, the presented in Section 2 lower bound is developed for oracles that know all nuisance parameters (like α) and nuisance functions (like $f^{X,Z}$), and then the aim is to propose an estimator that matches the oracle under minimal assumptions on smoothness of nuisance functions.

Now let us explain several related problems explored in the paper. Introduce the conditional survival function $S^{T|X}(t|x) := \mathbb{P}(T > t|X = x)$. Then the regression (1.1) can be written as

$$m(x) = \int_0^\infty S^{T|X}(t|x)dt, \quad x \in [0,1]. \tag{1.2}$$

Accordingly, regression is the linear functional of the conditional survival. Censoring always causes issues with estimating right tail of the distribution, and to remedy the issue it has been proposed to consider restricted linear functionals, like the restricted mean survival time (RMST) $\int_0^r S^T(t)dt$ and the conditional restricted mean survival time (CRMST)

$$\mu_r(x) := \int_0^r S^{T|X}(t|x)dt, \quad x \in [0,1], \;\; 0 < r < \infty. \tag{1.3}$$

To simplify the terminology, in what follows we are referring to the CRMST μ_r and the constant r as the *restricted regression* and the *restriction*, respectively. Estimation of more general restricted linear functionals will be also explored. The first use of the RMST for analysis of tumourless life was in [29], and the enlightening discussion of the approach can be found in [61]. In [32] estimation of RMST for right censored data with available covariates is investigated, and this paper pioneered the methodology of restricted regression. In particular, that paper presents an interesting discussion of advantages of the restricted regression with respect to other regression models and the Cox's model in particular. For now there is a relatively large literature devoted to these functionals, and reviews can be found in [23,28,72,74].

If T and Z are conditionally independent given X, and this is a standard assumption, then the following formula holds,

$$\mathbb{P}(\Delta' = 1|X = x, Z = z) = S^{T|X}(z|x). \tag{1.4}$$

This conditional probability, and equivalently the conditional survival function, is of a special interest in all the above-presented practical examples. For instance, in the cancer example this is the conditional likelihood to be free of cancer at time z after the surgery. The probability (1.4) is also of the central interest for the econometric binary choice model [7,26,45,46]. Accordingly, we will discuss nonparametric estimation of $S^{T|X}$.

Another typical complication in CSC is the missingness. In the paper we are considering a particular missing CSC (MSCS) when the sample is from $(\Delta' X, \Delta' Z, \Delta)$. Accordingly, only observations with $\Delta' = 1$ are available. For instance, in the above-mentioned econometrics example the MCSC means that only information about completed sales is available. Discussion and examples of the MCSC can be found in [16,18].

The context of the paper is as follows. Section 2 is devoted to sharp minimax lower bounds for the oracle who knows both CSC data and all nuisance parameters and functions. Then the aim is to propose a data-driven estimator that matches performance of the oracle. Section 3 presents oracle-estimators, as well as estimators for controlled CSC when the design density of (X, Z) is known. The latter is relaxed in Section 4. Analysis of real and simulated examples can be found in Section 5. Multivariate CSC regression, as well as estimation of a conditional survival function given a vector of predictors, are discussed in Section 6. Linear functionals are considered in Section 7. Conclusions and topics for future research are in Section 8. In particular, it contains an illuminating comparison between estimation of distributions and regressions for direct and CSC data. Proofs are placed in the Appendix.

Finally, let us present several notations used in the paper. Recall that $\Delta' := 1 - \Delta$. $\mathbb{P}(\cdot)$ denotes the probability, f^X denotes the density of X, $f^{X,Z}$ denotes the joint density of (X, Z), $S^{T|X}(t|x) := \mathbb{P}(T > t|X = x)$ is the conditional survival function of T given X, $S^{T|X,Z}(T|x, z) := \mathbb{P}(T > t|X = x, Z = z)$ is the conditional survival function of T given X and Z. $\mathbb{E}\{\cdot\}$ is the expectation, and we may write $E_{S^{T|X}}\{\cdot\}$ to stress that the expectation is calculated using the given $S^{T|X}$. Notations m and μ_r are used for the regression (1.2) and restricted regression (1.3), respectively. The used risk is the mean integrated squared error (MISE), for instance the MISE of a regression estimate $\tilde{m}(x)$, $x \in [0, 1]$ is $\mathbb{E}\{\int_0^1(\tilde{m}(x) - m(x))^2dx\}$. Further, $\mathbf{x} := (x_1, \dots, x_k)$ denotes a vector, $\{\varphi_0(x) := 1, \varphi_j(x) := 2^{1/2}\cos(\pi jx), j = 1, 2, \dots\}$ are elements of the cosine trigonometric orthonormal basis on $[0, 1]$, $I(\cdot)$ is the indicator, and $q_n := 3 + \lfloor \ln(n) \rfloor$ where $\lfloor x \rfloor$ is the largest integer which is smaller or equal to x. C's denote generic positive finite constants and c's denote specific constants, $o_j(1)$'s denote generic vanishing sequences as the parameter $j \to \infty$. Finally, set $\mathcal{R} := [0, 1] \times [0, \infty)$ and $\mathcal{R}_r := [0, 1] \times [0, r]$ where the positive constant r is called the restriction.

2. Sharp lower bounds for the MISE of oracle-estimators

The aim of this section is to explain what can and cannot be done, based on CSC and MCSC samples, for estimating regression (1.2) and restricted regression

(1.3). Recall that we observe a sample from (X, Z, Δ) for CSC and a sample from $(\Delta' X, \Delta' Z, \Delta)$ for MCSC, here $\Delta := I(T \leqslant Z)$ and $\Delta' := 1 - \Delta$. The approach is to consider the oracle who knows more than the statistician, and then develop a sharp lower minimax bound for the mean integrated squared error (MISE) of oracle-estimators. The idea is that the statistician cannot solve a problem better than the oracle, but may try to match performance of the oracle.

Recall that basic notations can be found at the end of the Introduction. We begin with assumptions. According to the review [30] "... the monitoring time is almost always assumed independent of the lifetime of interest." Our first assumption relaxes that independence by assuming conditional independence.

Assumption 2.1. *The predictor X is a continuous random variable supported on $[0, 1]$. The monitoring time Z and the lifetime of interest T are nonnegative continuous random variables. Given predictor X, the lifetime of interest T is conditionally independent of the monitoring time Z, that is $S^{T|X,Z}(t|x, z) = S^{T|X}(t|x)$. The pair (X, Z) may be dependent.*

The next two assumptions allow us to develop oracle's lower bounds. To explain the assumptions, we begin with several preliminary remarks. First, note that the regression and the restricted regression are linear functionals of an underlying conditional survival function,

$$m(x) = \int_0^\infty S^{T|X}(t|x)dt, \;\; \mu_r(x) = \int_0^r S^{T|X}(t|x)dt, \;\; x \in [0, 1]. \tag{2.1}$$

Second, for the oracle a minimax lower bound for estimating μ_r is always smaller than for estimating regression m because the oracle may set $S^{T|X}(t|x)$ to be known for $t > r$ (this assertion will be proved in the Appendix). Accordingly, we begin with a lower bound for restricted regression. Third, it is natural to obtain a lower bound via appropriate perturbations of $S^{T|X}(t|x)$ for $(x, t) \in \mathcal{R}_r$. Consequently, a minimax lower bound is developed for a special class of conditional survival functions. Further, as we will see shortly, a sharp minimax constant is a functional of $S^{T|X}$. This is why a local minimax approach is used when all considered conditional survival functions converge in L_∞-norm to an anchor $S_0^{T|X}$ as $n \to \infty$. Let us stress that the anchor is not an underlying conditional survival function and its primary role is to let the oracle know that all underlying conditional survival functions are near the anchor known to the oracle. Accordingly, the oracle also knows that all underlying restricted regressions μ_r are near the anchor μ_{r0} where $\mu_{r0}(x) := \int_0^r S_0^{T|X}(t|x)dt$, $x \in [0, 1]$. The anchor m_0 is defined similarly as $m_0(x) := \int_0^\infty S^{T|X}(t|x)dt$, $x \in [0, 1]$.

The above-made remarks explain the following two assumptions. Recall that $\varphi_0(x) = 1$, $\varphi_j(x) = 2^{1/2}\cos(\pi j x)$, $j = 1, 2, \ldots$ are elemenst of the cosine basis on $[0, 1]$.

Assumption 2.2. *The anchor $S_0^{T|X}(t|x)$, $(x, t) \in \mathcal{R}$ is known to the oracle. The* *anchor is continuous in x and differentiable in t on $\mathcal{R}_r$, and*

$$\min_{(x,t)\in\mathcal{R}_r} S_0^{T|X}(t|x) \geqslant c_1(r) > 0, \quad \max_{(x,t)\in\mathcal{R}_r} \frac{\partial S_0^{T|X}(t|x)}{\partial t} \leqslant -c_2(r) < 0. \tag{2.2}$$

The next assumption introduces a shrinking (toward the anchor) local Sobolev class of underlying conditional survival functions. A discussion of Sobolev classes can be found in [13,24].

Assumption 2.3. *An underlying conditional survival function $S^{T|X}$ belongs to a class*

$$\mathcal{F}_n(S_0^{T|X}, \alpha, Q, r) := \{S^{T|X}(t|x) : \int_0^r S^{T|X}(t|x)dt \in \mathcal{M}_n(\mu_{r0}, \alpha, Q, r)\}, \tag{2.3}$$

where

$$\mathcal{M}_n(\mu_{r0}, \alpha, Q, r) := \Big\{\mu_r : \ \mu_r(x) = \mu_{r0}(x) + g(x),$$
$$g \in \mathcal{S}(\alpha, Q), \ |g(x)| \leqslant 1/q_n, \ x \in [0,1]\Big\}. \tag{2.4}$$

In (2.4) $\mu_{r0}(x) := \int_0^r S_0^{T|X}(t|x)dt$ and

$$\mathcal{S}(\alpha, Q) := \Big\{g : \ g(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x),$$
$$\sum_{j=0}^{\infty}[1 + (\pi j)^{2\alpha}]\theta_j^2 \leqslant Q < \infty, \ \alpha \geqslant 1, \ x \in [0,1]\Big\} \tag{2.5}$$

is the global Sobolev class (ellipsoid).

Remark 2.1. Let us comment on the relationship between $S^{T|X}$ and the restricted regression μ_r that sheds extra light on Assumption 2.3. For the conditional survival on $\mathcal{R}_r$ we can write using the Fourier theorem,

$$S^{T|X}(t|x) = \sum_{j,i=0}^{\infty} \kappa_{ji}\varphi_j(x) r^{-1/2}\varphi_i(t/r), \quad (x,t) \in \mathcal{R}_r.$$

Here

$$\kappa_{ji} := \int_{R_r} S^{T|X}(t|x)\varphi_j(x) r^{-1/2}\varphi_i(t/r)dtdx$$

are Fourier coefficients of the conditional survival. Using $\int_0^r \varphi_i(t/r)dt = 0$ for $i \geqslant 1$, we conclude that

$$\mu_r(x) = r^{1/2}\sum_{j=0}^{\infty} \kappa_{j0}\varphi_j(x), \quad x \in [0,1].$$

Accordingly, the restricted regression is proportional to the univariate Fourier component of $S^{T|X}(t|x)$ in x.

Remark 2.2. In what follows a class $\mathcal{F}_n(S_0^{T|X}, \alpha, Q, \infty)$ is formally defined by replacing in (2.4) the restricted regression μ_r and the restricted anchor regression μ_{r0} by the regression m and the anchor regression m_0, respectively. This class will be used for analysis of estimators proposed for unbounded T.

Remark 2.3. Recall that if α is an integer, then $\mathcal{S}(\alpha, Q)$ is the global Sobolev class of α-fold differentiable functions traditionally studied in the classical non-parametric regression theory devoted to the model $T = m(X) + \sigma(X)\xi$ where ξ is independent of X standard normal variable (error), see [12,13,68]. Then it is known that, based on a direct sample of size n from (X, T), the regression function can be estimated with the classical rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE convergence. Note that a real α is also considered in the literature, see [13]. The global Sobolev class is of interest on its own and will be used in upper bounds.

Our final assumption is devoted to the joint density $f^{X,Z}(x, z)$ which is known to the oracle and to the statistician for the case of a controlled CSC study.

Assumption 2.4. *The joint density $f^{X,Z}(x, z)$ is known, continuous and positive on $\mathcal{R}_r$.*

Now we are in a position to formulate lower bounds for oracle-estimators using CSC observations.

Theorem 2.1 (Lower bounds for CSC). *(1) Suppose that Assumptions 2.1-2.4 hold and a CSC sample of size n from (X, Z, Δ) is given. Then*

$$\inf_{\tilde{\mu}_r^*} \sup_{S^{T|X} \in \mathcal{F}_n(S_0^{T|X}, \alpha, Q, r)} \Big\{ [n/d(S^{T|X}, f^{X,Z}, r)]^{2\alpha/(2\alpha+1)}$$

$$\times \mathbb{E}_{S^{T|X}} \{ \int_0^1 (\tilde{\mu}_r^*(x) - \mu_r(x))^2 dx \} \Big\} \geqslant P(\alpha, Q)(1 + o_n(1)). \quad (2.6)$$

Here the infimum is taken over all possible oracle-estimators that know data, the function class $\mathcal{F}_n(S_0^{T|X}, \alpha, Q, r)$ and the joint design density $f^{X,Z}(x, z)$,

$$P(\alpha, Q) := [\alpha/(\pi(\alpha + 1)]^{2\alpha/(2\alpha+1)}(2\alpha + 1)^{1/(2\alpha+1)} Q^{1/(2\alpha+1)}, \quad (2.7)$$

and

$$d(S^{T|X}, f^{X,Z}, r) := \int_{\mathcal{R}_r} \frac{(1 - S^{T|X}(t|x))S^{T|X}(t|x)}{f^{X,Z}(x, t)} dt dx \quad (2.8)$$

is the coefficient of difficulty for CSC.
(ii) Let us additionally assume that $d(S^{T|X}, f^{X,Z}, \infty) < \infty$ and Assumption 2.4 holds for any finite r. Then

$$\inf_{\tilde{m}*} \sup_{S^{T|X} \in \mathcal{F}_n(S_0^{T|X}, \alpha, Q, \infty)} \Big\{ [n/d(S^{T|X}, f^{X,Z}, \infty)]^{2\alpha/(2\alpha+1)}$$

$$\times\ \mathbb{E}_{S^{T|X}}\{\int_0^1 (\widetilde{m}^*(x) - m(x))^2 dx\}\Big\} \geqslant P(\alpha, Q)(1 + o_n(1)). \qquad (2.9)$$

Let us make several comments about the result. First, it will be shown in the next section that the lower bounds are sharp and attained by oracle-estimators. Second, similarly to the above-mentioned classical case of a direct sample from (X,T) and the regression model $T = m(X) + \sigma(X)\xi$, rate of the MISE convergence is the classical $n^{-2\alpha/(2\alpha+1)}$. In other words, CSC regression avoids the curse of CSC distribution estimation when the rate is slower than for direct observations. This also shows that the familiar equivalence between density and regression estimation, known for direct observations [13], breaks down for CSC. Third, complexity of a CSC regression model is captured by its coefficient of difficulty (2.8). Fourth, the coefficients of difficulty shed light on the role of restriction r which allows us to avoid improper integrals. Fifth, let us shed an extra light on the coefficient of difficulty (2.8). Consider a well known problem of estimating the RMST $\nu_r := \int_0^r S^T(t)dt$ using a sample of size n from (Z, Δ), see a comprehensive discussion in [28]. Under a mild assumption, the nonparametric maximum likelihood estimator $\widehat{\nu}_r$ is efficient and $n^{-1/2}[\widehat{\nu}_r - \nu_r] \xrightarrow{\mathcal{L}} N(0, d(S^T, f^Z, r))$ where the functional $d(\cdot,\cdot,\cdot)$ is defined in (2.8). The fact that the same functional defines asymptotic efficiency for estimation of RMSR and restricted regression is not surprising because the efficiencies are established using the same methodology of the local asymptotic normality and calculation of the Fisher information for CSC observations.

The oracle's lower bound warns us that even the oracle may not be able to propose a consistent regression estimator for unbounded lifetimes T or if the support of Z is a subset of the support of T. A similar warning, based on analysis of maximum likelihood and Fisher information, is made in [21] for linear CSC regression. Fortunately, in survival analysis an unbounded lifetime is a rare phenomenon, and in a majority of statistical applications a lifetime of interest is bounded by a known value, see the literature cited in the Introduction.

Now let us consider the MCSC sampling.

Theorem 2.2 (Lower bounds for MCSC). *Suppose that Assumptions 2.1-2.4 hold and a MCSC sample of size n from $(\Delta' X, \Delta' Z, \Delta)$ is given. Then the assertion of Theorem 2.1 holds with $d(S^{T|X}, f^{X,Z}, r)$ being replaced by*

$$d_*(S^{T|X}, f^{X,Z}, r) := \int_{\mathcal{R}_r} \frac{S^{T|X}(t|x)}{f^{X,Z}(x,t)} dt dx. \qquad (2.10)$$

Note that the MCSC sampling does not slow down the rate of MISE convergence but makes the sharp constant larger. The ratio $(d_*/d)^{2\alpha/(2\alpha+1)}$ defines the effect of missing on the accuracy of estimation under the MISE criterion. Here and in what follows we may use notations d and d_* for the functionals (2.8) and (2.10), respectively.

Remark 2.4. In the following section upper bounds for minimax MISEs will be presented. Under a minimax approach, like the one in (2.6) or (2.9), the supremum is taken over a class $\mathcal{F}$ of underlying distributions $S^{T|X}$ and the infimum over a class of estimators. Accordingly, in a lower bound it is desirable to consider a smaller $\mathcal{F}$ and in the upper bound a larger $\mathcal{F}$. In the above-presented lower bounds we are considering sequences in n of shrinking, toward a specific anchor distribution, classes $\mathcal{F}_n$ of distributions. To establish sharpness of those lower bounds, we will present oracle-estimators and estimators that attain the lower bounds for $S^{T|X} \in \mathcal{F}_n$. It is also a tradition in nonparametric curve estimation to analyze MISE over a class of estimands (in our setting over a class of regressions or restricted regressions). Traditional classes of estimands are global Sobolev classes $\mathcal{S}(\alpha, Q)$ defined in (2.5). Accordingly, in upper bounds we may simultaneously consider supremums over $S^{T|X} \in \mathcal{F}_n$ and over an estimand (m or μ_r) from $\mathcal{S}(\alpha, Q)$. As an example, we may write $\sup_{S^{T|X} \in \mathcal{F}_n,\, \mu_r \in \mathcal{S}(\alpha,Q)}$.

3. Estimation for a controlled study

A controlled study means that the joint design density $f^{X,Z}$ is known, and the case of an observational study, when the design density is unknown, will be considered in the next section. We begin with heuristic of the proposed methodology, and then consider efficient estimation for MCSC and CSC samples in turn. Recall that all general notations may be found at the end of the Introduction.

3.1. Heuristic of oracle-estimators

The aim of this subsection is threefold. First, to explain the underlying idea of used series estimation. Second, to present oracle-estimators that attain the lower bounds of Section 2. This will prove sharpness of the minimax lower bounds. Third, to explain the methodology of adaptation to unknown smoothness of regression.

We begin with a simple technical result. Recall that Cs are generic positive constants.

Lemma 3.1. *Suppose that function g belongs to the global Sobolev class $\mathcal{S}(\alpha, Q)$ defined in (2.5). Suppose that Fourier coefficients $\kappa_j := \int_0^1 g(x)\varphi_j(x)dx$ of g can be estimated by $\breve{\kappa}_j$ satisfying*

$$\mathbb{E}\{(\breve{\kappa}_j - \kappa_j)^2\} \leqslant Cn^{-1}. \tag{3.1}$$

Set $J_n^* := n^{1/(2\alpha+1)}$, and introduce the nonparametric oracle-estimator

$$\breve{g}^*(x) = \sum_{j=0}^{J_n^*} \breve{\kappa}_j \varphi_j(x). \tag{3.2}$$

Then

$$\sup_{g \in \mathcal{S}(\alpha,Q)} \mathbb{E}\{\int_0^1 (\breve{g}^*(x) - g(x))^2 dx\} \leqslant Cn^{-2\alpha/(2\alpha+1)}. \tag{3.3}$$

This assertion and its proof are simple and insightful. Note that the oracle-estimator is rate-optimal and depends only on the nuisance parameter α. Keeping in mind that in statistical practice it is often assumed that $\alpha = 1$ or $\alpha = 2$, see [13,68], we get a simple estimator. The proof is based on the Parseval identity. Write,

$$\sup_{g \in \mathcal{S}(\alpha,Q)} \mathbb{E}\{\int_0^1 (\breve{g}^*(x) - g(x))^2 dx\} = \sup_{g \in \mathcal{S}(\alpha,Q)} [\sum_{j=0}^{J_n^*} \mathbb{E}\{(\breve{\kappa}_j - \kappa_j)^2\} + \sum_{j > J_n^*} \kappa_j^2]$$

$$\leqslant C[n^{-1} J_n^* + (J_n^*)^{-2\alpha}] \leqslant C n^{-2\alpha/(2\alpha+1)}.$$

What was wished to show.

After this warming up, let us consider a more sophisticated assertion that will lead us to sharp-minimax estimation for settings considered in Section 2. Also recall Remark 2.4 about considering both local and global function classes in upper bounds.

Lemma 3.2. *(i) Suppose that a function of interest g belongs to the global Sobolev class $\mathcal{S}(\alpha, Q)$ defined in (2.5). Suppose that Fourier coefficients $\kappa_j := \int_0^1 g(x)\varphi_j(x)dx$ of g can be estimated by $\tilde{\kappa}_j$ satisfying*

$$\mathbb{E}\{\tilde{\kappa}_j\} = \kappa_j, \ \mathbb{E}\{(\tilde{\kappa}_j - \kappa_j)^2\} \leqslant d n^{-1}(1 + o_n(1) + o_j(1)), \ 0 < d < \infty. \tag{3.4}$$

Introduce the nonparametric oracle-estimator

$$\tilde{g}(x) = \sum_{j=0}^{q_n} \tilde{\kappa}_j \varphi_j(x) + \sum_{j=q_n+1}^{J_n} (1 - (j/J_n)^{\alpha}) \tilde{\kappa}_j \varphi_j(x), \tag{3.5}$$

where $J_n := q_n + 1 + \lfloor [(n/d)Q\pi^{-2\alpha}(\alpha+1)(2\alpha+1)/\alpha]^{1/(2\alpha+1)} \rfloor$. Then the following upper bound is valid for MISE of this oracle-estimator,

$$\sup_{g \in \mathcal{S}(\alpha,Q)} [n/d]^{2\alpha/(2\alpha+1)} \mathbb{E}\{\int_0^1 (\tilde{g}(x) - g(x))^2 dx\} \leqslant P(\alpha, Q)(1 + o_n(1)). \tag{3.6}$$

Here $P(\alpha, Q)$ is defined in (2.7).
(ii) Let the function of interest be $g = g_0 + g_*$ where $g_0 \in \mathcal{S}(\alpha', Q')$, $\alpha' > \alpha$, $Q' < \infty$, is the anchor function and $g_* \in \mathcal{S}(\alpha, Q)$. Suppose that (3.4) holds. Then MISE of the oracle-estimator (3.5), that does not use the anchor g_0, satisfies the following upper bound,

$$\sup_{g^* \in \mathcal{S}(\alpha,Q)} [n/d]^{2\alpha/(2\alpha+1)} \mathbb{E}\{\int_0^1 (\tilde{g}(x) - g(x))^2 dx\} \leqslant P(\alpha, Q)(1 + o_n(1)). \tag{3.7}$$

Note that the second part (ii) of the lemma does not follow from the first one because in part (ii) the estimand g no longer belongs to $\mathcal{S}(\alpha, Q)$. Part (i) sheds light on global estimation and part (ii) on local estimation centered around the anchor. Lemma 3.2 will be proved in the Appendix.

The main conclusion from Lemma 3.2 is that to construct an efficient nonparametric oracle-estimator it is sufficient to propose Fourier estimates satisfying (3.4) with d being an appropriate coefficient of difficulty. Let us show how this can be done for MCSC and CSC sampling models in turn. We are considering estimation of Fourier coefficients for restricted regression μ_r, and then comment on the case of regression.

Suppose that Assumption 2.1 holds and assume that the coefficient of difficulty $d_* := d_*(S^{T|X}, f^{X,Z}, r)$, defined in (2.10), is finite. Recall our notation $\mu_r(x) := \int_0^r S^{T|X}(t|x)dt$ for restricted regression. Using the Fourier theorem we get

$$\mu_r(x) = \int_0^r S^{T|X}(t|x)dt = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \quad x \in [0,1]. \tag{3.8}$$

Recall that $\varphi_0(x) = 1$, $\varphi_j(x) = 2^{1/2}\cos(\pi j x)$, $j = 1, 2, \ldots$ are elements of the cosine basis on $[0,1]$ and θ_j are Fourier coefficients of μ_r,

$$\theta_j := \int_0^1 \mu_r(x)\varphi_j(x)dx = \int_0^1 \Big[\int_0^r S^{T|X}(t|x)dt\Big]\varphi_j(x)dx. \tag{3.9}$$

Using Assumption 2.1 we can write

$$f^{X,Z,\Delta}(x,t,\delta) = f^{X,Z}(x,t)[1 - S^{T|X}(t|x)]^{\delta}[S^{T|X}(t|x)]^{1-\delta}. \tag{3.10}$$

Note that $f^{X,Z,\Delta}(x,t,0) = 0$ if $f^{X,Z}(x,t) = 0$ or $S^{T|X}(t|x) = 0$. This allows us to continue (3.9),

$$\theta_j = \int_0^1 \int_0^r \frac{I(f^{X,Z}(x,t) > 0) f^{X,Z,\Delta}(x,t,0)\varphi_j(x)}{f^{X,Z}(x,t)} dtdx$$

$$= \mathbb{E}\Big\{\frac{\Delta' I(Z \leqslant r)\varphi_j(X)}{f^{X,Z}(X,Z)}\Big\}. \tag{3.11}$$

The joint density $f^{X,Z}$ is known to the oracle. Accordingly, for MCSC formula (3.11) yields the sample mean Fourier estimator

$$\breve{\theta}_j := n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l I(\Delta'_l Z \leqslant r)\varphi_j(\Delta'_l X_l)}{f^{X,Z}(\Delta'_l X_l, \Delta'_l Z_l)}. \tag{3.12}$$

Further, for the sample mean Fourier estimator we get

$$\mathbb{E}\{\breve{\theta}_j\} = \theta_j, \quad \mathbb{E}\{(\breve{\theta}_j - \theta_j)^2\} = d_* n^{-1}(1 + o_j(1)). \tag{3.13}$$

Now we can invoke Lemma 2.1 and conclude that the oracle-estimator (3.2), using Fourier estimates (3.12), is rate optimal. Further, Lemma 3.2 implies that the oracle-estimator (3.5), using Fourier estimates (3.12), is efficient according to Theorem 2.2. Further, by setting $r = \infty$ in (3.12), we get a sample mean

Fourier estimate for the regression $m(x) = \mathbb{E}\{T|X = x\}$ satisfying (3.13) where the d_* is replaced by $d_*(S^{T|X}, f^{X,Z}, \infty)$.

Now we are considering the CSC sampling. So far, due to the study of MCSC, we used only a subsample with $\Delta' = 1$ to estimate restricted regression μ_r, and this is why the variance in (3.13) is larger than the coefficient of difficulty d defined in (2.8). Let us explain how remedy this issue. Consider Fourier coefficients of the conditional survival function on $\mathcal{R}_r$,

$$\beta_{ki} := \int_{\mathcal{R}_r} S^{T|X}(t|x)\varphi_k(x)r^{-1/2}\varphi_i(t/r)dtdx. \tag{3.14}$$

These Fourier coefficients allow us to introduce a special Fourier approximation of $S^{T|X}$ with a skipped subset of jth Fourier coefficients $\beta_{ji}, i = 0, 1, \ldots$,

$$S(j, n, t, x) := \sum_{k\in\{0,1,\ldots,q_n\}\setminus\{j\}} \sum_{i=0}^{q_n} \beta_{ki}\ \varphi_k(x)r^{-1/2}\varphi_i(t/r),\ \ (t, x) \in \mathcal{R}_r. \tag{3.15}$$

Note that $\int_0^1 S(j, n, t, x)\varphi_j(x)dx = 0$ and $S(j, n, t, x)$ converges to $S^{T|X}(t|x)$ as j and n increase. The oracle suggests a new unbiased Fourier coefficient estimator

$$\tilde{\theta}^*_j := n^{-1}\sum_{l=1}^{n} \frac{[\Delta'_l - S(j, n, Z_l, X_l)]\varphi_j(X_l)}{f^{X,Z}(X_l, Z_l)}. \tag{3.16}$$

The oracle's rationale is that, as it will be checked in the Proofs,

$$\mathbb{E}\{(\tilde{\theta}^*_j - \theta_j)^2\} = d(S^{T|X}, f^{X,Z}, r)n^{-1}(1 + o_j(1)). \tag{3.17}$$

Accordingly, the modified Fourier estimator $\tilde{\theta}^*_j$ can be used for sharp minimax estimation of μ_r. Of course, the conditional survival function $S^{T|X}$ is unknown, but it may be estimated with sufficient accuracy to match the oracle.

Finally, if $m_r \in \mathcal{S}(\alpha, Q)$, then how can a series estimate adapt to unknown nuisance parameters (α, Q) and d? Let us explain the heuristic of blockwise shrinkage that performs the desired sharp adaptation to the nuisance parameters.

We begin with the following classical result in point estimation. If $\bar{\theta}_j$ is unbiased estimator of parameter θ_j, then it may be beneficial to look at the shrinking estimator $\lambda_j\bar{\theta}_j$, $\lambda_j \in [0, 1]$ which minimizes the mean squared error $\mathbb{E}\{(\lambda_j\bar{\theta}_j - \theta_j)^2\}$. The oracle's solution, known as the Wiener filter, is

$$\lambda^*_j = \frac{\theta_j^2}{\theta_j^2 + \sigma_j^2 n^{-1}},\ \ \sigma_j^2 := n\mathbb{E}\{\bar{\theta}_j - \theta_j)^2\}. \tag{3.18}$$

It may be tempting to plug in appropriate estimates of θ_j^2 and σ_j^2 and replace λ^*_j by the corresponding estimate. Unfortunately, this idea is not feasible because θ_j^2 is estimable with the parametric rate n^{-1}. On the other hand, we can see from (3.5) that the smoothing weights $1 - (j/J_n)^\alpha$ are close to each other for adjacent indexes j. This leads us to a simple solution when instead of estimating individual weights $\theta_j^2/(\theta_j^2+\sigma_j^2 n^{-1})$ we estimate a single optimal weight for a block of adjacent indexes j. Namely, let $B := \{i+1,\dots,i+L\}$ be a block of length L of positive integers. Then the shrinking coefficient Λ^*, which minimizes $\mathbb{E}\{\sum_{j\in B}(\Lambda\bar{\theta}_j-\theta_j)^2\}$, is

$$\Lambda^* = \frac{L^{-1}\sum_{j\in B}\theta_j^2}{L^{-1}\sum_{j\in B}\theta_j^2 + [L^{-1}\sum_{j\in B}\sigma_j^2]n^{-1}} =: \frac{\Theta}{\Theta+\sigma^2 n^{-1}}. \tag{3.19}$$

In (3.19)

$$\Theta := L^{-1}\sum_{j\in B}\theta_j^2 \tag{3.20}$$

is the classical Sobolev functional which is the focal point of the blockwise nonparametric adaptation. The theory of estimating Sobolev functionals is well developed, and while λ_j^* may be estimated with the classical parametric rate n^{-1}, the Sobolev functional is estimable with the same rate but the constant decreases as L (the length of block) increases. This is what creates the opportunity for estimating Λ^* with sufficient accuracy for matching oracle-estimators. The corresponding theory is well developed [9,10,14,24].

The above-discussed blockwise shrinkage is the adaptation methodology used in this paper. This is the simplest and universal methodology of adaptation that matches performance of efficient oracle-estimators. Further, in Section 6 it will allow us to consider the problem of dimension reduction. Of course, there is a number of other procedures for adaptation proposed in the literature, but they are primarily concerned with rate optimal adaptation. The interested reader can find reviews in [11,13,24,68].

3.2. Efficient estimation of restricted regression for MCSC

In this subsection the lifetime of interest may be bounded or unbounded, and these two cases are considered simultaneously. Further, recall that for a bounded lifetime a restricted regression, with the restriction r equal to or larger than that bound, is the underlying regression. The available sample is MCSC meaning that we have a sample of size n from $(\Delta' X, \Delta' Z, \Delta)$. The estimand is the restricted regression $\mu_r(x) = \int_0^r S^{T|X}(t|x)dt$ where r is a finite restriction. Because the restriction is finite, we can simultaneously consider bounded and unbounded lifetimes T given Assumption 2.4. Indeed, that assumption is not tied to the support of T and only requires that the known (recall that in this section we study the controlled sampling) design density $f^{X,Z}$ is continuous and positive on $\mathcal{R}_r = [0,1]\times[0,r]$.

Introduce the Fourier coefficient estimator for MCSC observations

$$\tilde{\theta}_j := n^{-1}\sum_{l=1}^{n}\frac{\Delta' I(\Delta'_l Z_l \leqslant r)\varphi_j(\Delta'_l X_l)}{f^{X,Z}(\Delta'_l X_l, \Delta'_l Z_l)}. \tag{3.21}$$

As we will see shortly, it can be used to construct efficient restricted regression estimator. Our next step, according to the heuristic of subsection 3.1, is to define a blockwise adaptive estimator. For $j > q_n$ introduce consecutive and non-overlapping blocks B_k, $k = 1, 2, \dots$ of length $L_k := \lfloor (1 + 1/\ln(q_n))^k \rfloor$, that is $B_1 := \{q_n + 1, \dots, q_n + L_1\}$, $B_2 := \{q_n + L_1 + 1, \dots, q_n + L_1 + L_2\}$, etc. Then for each block we calculate two statistics. The first one is the U-statistic

$$\tilde{\Theta}_k := \frac{2}{L_k n(n-1)} \sum_{1 \leqslant l_1 < l_2 \leqslant n} \sum_{j \in B_k} \prod_{i=1}^{2} \frac{\Delta'_{l_i} I(\Delta'_{l_i} Z_{l_i} \leqslant r) \varphi_j(\Delta'_{l_i} X_{l_i})}{f^{X,Z}(\Delta'_{l_i} X_{l_i}, \Delta'_{l_i} Z_{l_i})}. \tag{3.22}$$

The second statistic is based on Fourier estimates $\tilde{\theta}_j$ defined in (3.21),

$$\tilde{\Theta}'_r := L_k^{-1} \sum_{j \in B_k} \tilde{\theta}_j^2. \tag{3.23}$$

Let k_n be the smallest integer such that $\sum_{k=1}^{k_n} L_k > n^{1/(2\alpha_0+1)} \ln(q_n)$ where α_0 is the smallest assumed value of parameter α. Recall that Assumption 2.3 sets $\alpha_0 = 1$, but other values also may be specified. For instance, $\alpha_0 = 2$ implies that the restricted regression is twice differentiable, and this is another traditional choice.

The proposed adaptive estimator is

$$\tilde{\mu}_r(x) := \tilde{\mu}_r(x, f^{X,Z}) := \sum_{j=0}^{q_n} \tilde{\theta}_j \varphi_j(x)$$

$$+ \sum_{k=1}^{k_n} \min(1, \tilde{\Theta}_k / \tilde{\Theta}'_k) I(\tilde{\Theta}_k > 1/[n \ln(k+3)]) \sum_{j \in B_k} \tilde{\theta}_j \varphi_j(x). \tag{3.24}$$

Note how simple the adaptive estimator is.

Theorem 3.1. *Let Assumptions 2.1 and 2.4 hold, and the anchor μ_{r0} belongs to a Sobolev class $\mathcal{S}(\alpha', Q')$ with $\alpha' > \alpha$ and $Q' < \infty$. Consider a MCSC sample of size n from $(\Delta' X, \Delta' Z, \Delta)$. Then the following upper bound holds for MISE of the adaptive estimator (3.24),*

$$\sup_{S^{T|X} \in \mathcal{F}_n(S_0^{T|X}, \alpha, Q, r),\, \mu_r \in \mathcal{S}(\alpha, Q)} \Big\{ [n/d_*]^{2\alpha/(2\alpha+1)}$$

$$\times \mathbb{E}\{ \int_0^1 (\tilde{\mu}_r(x) - \mu_r(x))^2 dx \} \Big\} \leqslant P(\alpha, Q)(1 + o_n(1)). \tag{3.25}$$

We can conclude that the lower bound of Theorem 2.2 is sharp. Further, that lower bound is also attainable for μ_r from the global Sobolev classes $\mathcal{S}(\alpha, Q)$. Accordingly, we get the same results as for the case of regressions based on direct observations from (X, T) where efficient adaptive estimators are proposed for global Sobolev classes.

3.3. *Efficient estimation of restricted regression for CSC*

In this subsection the lifetime of interest may be bounded or unbounded, and these two cases are considered simultaneously. Further, recall that if T is bounded and we set r to be equal to or larger than the upper bound for the support of T, then the restricted regression is equal to the regression.

Now let us consider a CSC sample of size n from (X, Z, Δ) and propose an efficient estimator of μ_r. Set $\mathcal{J}(j,n) := \{0, 1, \dots, q_n\}\backslash\{j\}$ if $0 \leqslant j \leqslant q_n$ and $\mathcal{J}(j,n) := \{0, 1, \dots, q_n\}$ otherwise. Introduce the cosine basis $\{\psi_0(t) := r^{-1/2}$, $\psi_j(t) = (2/r)^{1/2}\cos(\pi jt/r), j = 1, 2, \dots\}$ on $[0, r]$. Following the heuristic of subsection 3.1, introduce an estimate of $S^{T|X}$ with subtracted projection on ψ_j,

$$\widetilde{S}(j,n,z,x) := n^{-1}\sum_{l=1}^{n}\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta'_l I(Z_l \leqslant r)\psi_i(Z_l)\varphi_k(X_l)\psi_i(z)\varphi_k(x)}{f^{X,Z}(X_l,Z_l)}, \quad (3.26)$$

and the Fourier coefficient estimate

$$\widehat{\theta}_j := n^{-1}\sum_{l=1}^{n}\frac{(\Delta'_l - \widetilde{S}(j,n,Z_l,X_l))I(Z_l \leqslant r)\varphi_j(X_l)}{f^{X,Z}(X_l,Z_l)}. \quad (3.27)$$

For adaptation to unknown smoothness of μ_r, we again use blocks B_k of length L_k, introduced below line (3.21), and the sequence k_n defined below line (3.23). For each block we calculate two statistics. The first one is the U-statistic

$$\widehat{\Theta}_k := \frac{2}{L_k n(n-1)}$$
$$\times \sum_{1\leqslant l_1<l_2\leqslant n}\sum_{j\in B_k}\prod_{i=1}^{2}\frac{(\Delta'_{l_i} - \widetilde{S}(j,n,Z_{l_i},X_{l_i})))I(Z_{l_i} \leqslant r)\psi_j(X_{l_i})}{f^{X,Z}(X_{l_i},Z_{l_i})}. \quad (3.28)$$

The second statistic is based on Fourier estimates $\widehat{\theta}_j$ defined in (3.27),

$$\widehat{\Theta}'_r := L_k^{-1}\sum_{j\in B_k}\widehat{\theta}_j^2. \quad (3.29)$$

The proposed estimator is

$$\widehat{\mu}_r(x) := \widehat{\mu}_r(x, f^{X,Z}) := \sum_{j=0}^{q_n}\widehat{\theta}_j\psi_j(x)$$
$$+ \sum_{k=1}^{k_n}\min(1, \widehat{\Theta}_k/\widehat{\Theta}'_k)I(\widehat{\Theta}_k > 1/[n\ln(k+3)])\sum_{j\in B_k}\widehat{\theta}_j\psi_j(x). \quad (3.30)$$

Theorem 3.2. *Let Assumptions of Theorem 3.1 hold. Consider a CSC sample of size n from (X, Z, Δ). Then the following upper bound holds for the MISE of estimator (3.30),*

$$\sup_{S^{T|X}\in\mathcal{F}_n(S_0^{T|X},\alpha,Q,r),\ \mu_r\in\mathcal{S}(\alpha,Q)}\Big\{[n/d(S^{T|X}, f^{X,Z}, r)]^{2\alpha/(2\alpha+1)}$$

$$\times \mathbb{E}\{\int_0^1 (\hat{\mu}_r(x) - \mu_r(x))^2 dx\}\Big\} \leqslant P(\alpha, Q)(1 + o_n(1)). \quad (3.31)$$

Theorem 3.2, together with the lower bound of Theorem 2.1, allow us to conclude that not only the optimal rate $n^{-2\alpha/(2\alpha+1)}$ is preserved, but also the sharp constant is attainable. Further, the proposed estimator attains the same sharp constant over the global Sobolev classes. In short, we have the same bouquet of results for CSC as for the case of direct observations.

3.4. Estimation of regression for unbounded T

We begin with an oracle-estimator and then present an estimator. It will be convenient to use notations $D := d(S^{T|X}, f^{X,Z}, \infty)$ and $D_* := d_*(S^{T|X}, f^{X,Z}, \infty)$.

Assumption 3.1. *Conditional survival function $S^{T|X}$ and joint density $f^{X,Z}$ satisfy*

$$D_* := \int_0^1 \int_0^\infty \frac{S^{T|X}(t|x)}{f^{X,Z}(x,t)} dt dx \leqslant c_3 < \infty. \quad (3.32)$$

Remark 3.1. Assumption 3.1 yields that D is also bounded. At the same time, if D is bounded then D_* may be unbounded if $1/f^{X,Z}(x,t)$ is not integrable for small t while $(1 - S^{T|X}(t|x))/f^{X,Z}(x,t)$ is. The latter is atypical for design densities. Let is also recall that in [28], where the functional $\int_0^\tau S^T(t)dt$ is the estimand, the integral $\int_0^\tau [(1-S^T(t))S^T(t)/f^Z(t)]dt$, called the information bound, is assumed to be finite. We may conclude that Assumption 3.1 is in line with known in the CSC literature.

Set $J_n^* := q_n + 1 + \lfloor [(n/D_*)Q\pi^{-2\alpha}(\alpha+1)(2\alpha+1)/\alpha]^{1/(2\alpha+1)} \rfloor$,

$$\bar{\theta}_j := n^{-1} \sum_{l=1}^n \frac{\Delta'_l \varphi_j(\Delta'_l X_l)}{f^{X,Z}(\Delta'_l X_l, \Delta'_l Z_l)}, \quad (3.33)$$

and introduce the regression oracle-estimator motivated by Lemma 3.2,

$$\bar{m}^*(x) := \sum_{j=0}^{q_n} \bar{\theta}_j \varphi_j(x) + \sum_{j=q_n+1}^{J_n^*} (1 - (j/J_n^*)^\alpha) \bar{\theta}_j \varphi_j(x). \quad (3.34)$$

Theorem 3.3. *Suppose that Assumptions 2.1 and 3.1 hold, and the anchor $m_0 \in \mathcal{S}(\alpha', Q')$ where $\alpha' > \alpha$ and $Q' < \infty$. Then MISE of oracle-estimator (3.34) satisfies the inequality*

$$\sup_{S^{T|X} \in \mathcal{F}_n(S_0^{T|X}, \alpha, Q, \infty),\, m \in \mathcal{S}(\alpha, Q)} \Big\{ [n/D_*]^{2\alpha/(2\alpha+1)}$$

$$\times \mathbb{E}\{\int_0^1 (\bar{m}^*(x) - m(x))^2 dx\}\Big\} \leqslant P(\alpha, Q)(1 + o_n(1)). \quad (3.35)$$

Corollary 3.1. *The oracle-estimator (3.34) is asymptotically efficient for MCSC. Its MISE is also within factor* $(D_*/D)^{\frac{2\alpha}{2\alpha+1}}$ *from the lower bound (2.9) for CSC. Accordingly, the rougher the regression the smaller the factor.*

Now we are in a position to propose a blockwise adaptive estimator whose heuristic was explained in subsection 3.1. Recall that blocks B_k with length L_k, as well as the sequence k_n, were introduced below line (3.21). For each block we calculate statistics

$$\bar{\Theta}_k := \frac{2}{L_k n(n-1)} \sum_{1\leqslant l_1<l_2\leqslant n} \sum_{j\in B_k} \prod_{i=1}^{2} \frac{\Delta'_{l_i}\varphi_j(\Delta'_{l_i}X_{l_i})}{f^{X,Z}(\Delta'_{l_i}X_{l_i}, \Delta'_{l_i}Z_{l_i})}. \tag{3.36}$$

and

$$\bar{\Theta}'_r := L_k^{-1} \sum_{j\in B_k} \bar{\theta}_j^2. \tag{3.37}$$

The proposed adaptive regression estimator is

$$\bar{m}(x, f^{X,Z}) := \sum_{j=0}^{q_n} \bar{\theta}_j\varphi_j(x)$$

$$+ \sum_{k=1}^{k_n} \min(1, \bar{\Theta}_k/\bar{\Theta}'_k) I(\bar{\Theta}_k > 1/[n\ln(k+3)]) \sum_{j\in B_k} \bar{\theta}_j\varphi_j(x). \tag{3.38}$$

Theorem 3.4. *Let Assumption 2.1 hold and*

$$\int_0^\infty \frac{S^{T|X}(t|x)}{[f^{X,Z}(x,t)]^8} dt < c_4 < \infty. \tag{3.39}$$

Then the assertion of Theorem 3.3 holds for the regression estimator (3.38).

Remark 3.2. Consider a lifetime of interest T supported on $[0,\infty)$. For CSC observations, the regression estimator (3.38) is within factor $(D_*/D)^{2\alpha/(2\alpha+1)}$ from the lower bound (2.6). For MCSC observations the estimator is efficient. Further, we can conclude that the regression function can be estimated with the classical rate $n^{-2\alpha/(2\alpha+1)}$ known for the case of direct observations. Accordingly, the nonparametric regression "breaks" the CSC curse known for distribution estimation, and the CSC does not slow down the optimal rate known for direct observations.

4. Estimation for an observational study

So far we have considered the case of a known design density $f^{X,Z}$. In this section this assumption is relaxed. It will be shown that under a mild assumption a plug in methodology is feasible. Because both MCSC and CSC are considered, for

CSC we can use the available sample of size n from (X, Z) to estimate $f^{X,Z}$. For MCSC, when we observe a sample from $(\Delta' X, \Delta' Z, \Delta)$, this sample does not allow us to estimate $f^{X,Z}$ because the data is biased [16]. Accordingly, for MCSC it is assumed that an extra sample of size n from (X, Z) is available.

Assumption 4.1. *An underlying joint probability density $f^{X,Z}(x,z)$ has a continuous mixed derivative $\partial^2 f^{X,Z}(x,z)/\partial x \partial z$ on $\mathcal{R}_r$ and*

$$\int_{\mathcal{R}_r} \Big[[\partial f^{X,Z}(x,z)/\partial x]^2 + [\partial f^{X,Z}(x,z)/\partial z]^2 + [\partial^2 f^{X,Z}(x,z)/\partial x \partial z]^2\Big] dx dz$$

$$\leqslant c_5 < \infty. \tag{4.1}$$

An underlying conditional survival function $S^{T|X}(t|x)$ has a continuous mixed derivative $\partial^2 S^{T|X}(t|x)/\partial t \partial x$ on $\mathcal{R}_r$ and

$$\int_{\mathcal{R}_r} \Big[[\partial S^{T|X}(t|x)/\partial x]^2 + [\partial S^{T|X}(t|x)/\partial t]^2 + [\partial^2 S^{T|X}(t|x)/\partial t \partial x]^2\Big] dx dt$$

$$\leqslant c_6 < \infty. \tag{4.2}$$

The left sides of (4.1) and (4.2) are classical Sobolev functionals. They do not involve second-order derivatives with respect to either of the arguments. As we will see shortly, this mild differentiability allows us to match performance of the oracle who knows joint density $f^{X,Z}$.

Recall that $\{\psi_s(t), s = 1, 2, \ldots, t \in [0, r]\}$ is the cosine basis on $[0, r]$ defined above line (3.26). For CSC model introduce the joint density estimator

$$\hat{f}^{X,Z}(x,z) := \max\Big(\frac{1}{\ln(\ln(n+3))},$$

$$\frac{1}{n} \sum_{i,s=0}^{\lfloor 1+n^{1/4} \rfloor} \sum_{l=1}^{n} \varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)\Big). \tag{4.3}$$

For MCSC the same estimator is used only it is based on an extra sample from (X, Z). We use (4.3) in place of an underlying $f^{X,Z}$. Note that the estimate is separated from zero by the iterated logarithm $1/\ln(\ln(n+3))$ and hence may be used in a denominator.

Theorem 4.1. *Let Assumption 4.1 and the assumption of Theorem 3.1 hold, only now the joint density $f^{X,Z}$ is unknown and $\alpha \geqslant \alpha_0 = 2$. Introduce the plug-in estimators $\tilde{\mu}_r(x, \hat{f}^{X,Z})$ and $\hat{\mu}_r(x, \hat{f}^{X,Z})$ defined in (3.24) and (3.30), respectively. Then the assertions of Theorems 3.1 and 3.2 hold for the plug-in estimators.*

We may conclude that the data-driven estimation is possible and the lower bounds of Section 2 are sharp. Further, neither CSC not MCSC slow down rate of the regression estimation with respect to direct data observations.

Now we are in a position to complement the theory by examples.

5. Analysis of real and simulated examples

The context of this section is as follows. Subsection 5.1 begins with a visual analysis of two simulated CSC datasets when we know the underlying regression function. In the former simulation X and Z are independent , and in the latter they are dependent. For these simulations the above-presented estimates are shown and discussed. Then the more complicated simulation with dependent predictor and monitoring time is repeated 10 times and residuals of the data-driven CSC estimator are shown. This experiment sheds light on the bias and variance of the regression estimator. Results of an intensive numerical study are presented in subsection 5.2 via histograms of ratios between integrated squared errors of the proposed estimates. The study supports the proposed methodology of plugged in estimates of the design density $f^{X,Z}$. Regression analysis of the environmental CSC data, presented in the Introduction, can be found in the last subsection.

5.1. Two simulated CSC regressions

The aim of this subsection is to shed light on CSC regression via visual analysis of simulated datasets and performance of the proposed estimates. Particular simulations for two regression experiments are shown in the two columns of Figure 2. The top diagrams show by crosses the underlying direct observations of (X, T). The bottom diagrams show corresponding CSC samples from (X, Z, Δ) with triangles and circles indicating observations with $\Delta = 0$ and $\Delta = 1$, respectively. Full description of the underlying experiments and the diagrams can be found in the caption.

Let us comment on the scattergrams and the estimates shown in Figure 2. The direct data scattergrams are complicated due to the strong heteroscedasticity. Still, it is possible to visualize the underlying regression as a curve that goes through the “middle” of data. The reader may try to make a guess about the regression, and then compare the guess with the solid line (the underlying regression) and the dashed line (the data-driven nonparametric estimate). Note how the estimates fit the data. The bottom diagrams show CSC modifications of direct data. Overall, if not due to the indicators of censoring, shown by the triangles ($\Delta = 0$) and the circles ($\Delta = 1$), it is impossible to visualize the underlying regression. Indeed, in the both bottom diagrams the observations are spread over the unit square and the best bet may be a horizontal line. Now let us look at the underlying regression (the solid line) and the data paying attention to the triangles and circles. Overall, with the help of the solid line, it is possible to appreciate the special structure of CSC scattergrams and even get a “feeling” of the underlying regression. With some training in visualization of different CSC scattergrams, it is possible to get a general feeling of the shape of an underlying regression, but overall this is a complicated task. Only a special software can estimate the CSC regression because here one needs first to figure out an underlying conditional survival and then evaluate its integral.

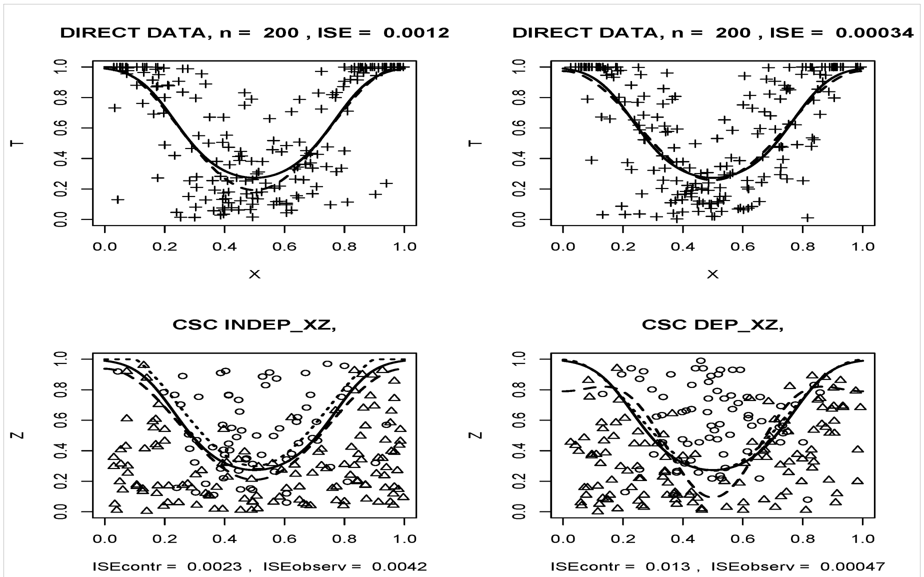

Fig 2. Two simulated direct regressions and the corresponding CSC regressions are shown in the top and bottom diagrams, respectively. Direct observations are shown by the crosses. CSC observations are shown by the triangles and the circles for $\Delta = 0$ and $\Delta = 1$, respectively. The left column presents the experiment with independent X and Z. Here the predictor X is distributed according to the density $f^X(x) = (4/5)(1 + 0.5x)I(x \in [0, 1])$, and the conditional distribution of the response T given the predictor is Beta$(1, f_2(x))$. As a result, $m(x) = 1/[1 + f_2(x)]$ for $x \in [0, 1]$. Function f_2 is the Normal corner function defined in [18], page 32. Namely, f_2 is the normal density with mean 0.5 and standard deviation 0.15. The underlying monitoring lifetime Z is independent of the predictor X and $f^Z(z) = (1 + 0.5\cos(\pi z))I(z \in [0, 1])$. In the left-bottom diagram $\sum_{l=1}^n \Delta_l = 68$. The right column presents the experiment with the same conditional distribution of T given X but with dependent X and Z. Here the design density is $f^{X,Z}(x, z) = [1 + (1/4)\cos(\pi x) + (1/2)\cos(\pi x)\cos(\pi z)]I((x, z) \in [0, 1]^2)$. In the right-bottom diagram $\sum_{l=1}^n \Delta_l = 74$. In all diagrams the solid line is the underlying regression. In the top diagrams the dashed lines are the regression estimates of [16, section 2.3]. In the bottom diagrams the dashed lines are CSC estimates for controlled regressions (the $f^{X,Z}$ is known) and the dotted lines are data-driven estimates for observational regressions when $f^{X,Z}$ is unknown. The estimates use the information $0 \leqslant m(x) \leqslant 1$ for $x \in [0, 1]$. The titles show the type of regression, sample size n, and the empirical integrated squared errors (ISE) of the used estimates. ISE, ISEcontr, and ISEobserv show corresponding ISE for the estimates based on direct data, controlled ($f^{X,Z}$ is known) CSC data and observational CSC data, respectively.

The two CSC estimates in the bottom diagrams, shown by the dashed and dotted lines, are the estimates for controlled (the design density $f^{X,Z}$ is known) and observational (the design density $f^{X,Z}$ is unknown) CSC models, respectively. The chosen simulations show us two possible outcomes. The left-bottom diagram exhibits a “reasonable” outcome when the controlled regression with

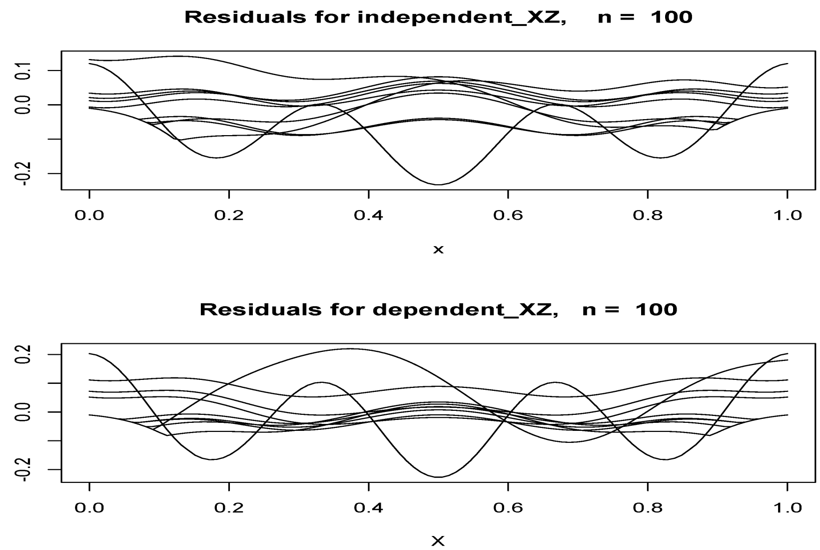

Fig 3. Residuals $m(x) - \widehat{m}(x)$ of the data-driven CSC regression estimators for ten simulations. The underlying experiments are the same as in Figure 2, the sample size $n = 100$.

known $f^{X,Z}$ yields a better regression estimate than the observational regression where $f^{X,Z}$ is estimated. Further, here both estimates are significantly lag the estimate based on direct data.

The outcome is more nuanced for the data in the right-bottom diagram where X and Z are dependent. Here the estimate for observational regression dramatically outperforms the estimate for controlled regression. Further, its performance is comparable with the estimate based on direct observations. Is this an abnormal simulation? A surprising answer will be presented in the next subsection.

Now let us look at Figure 3. It sheds additional light on performance of the proposed data-driven estimator for observational CSC data when the design density $f^{X,Z}$ is unknown. The underlying CSC models are the same as in Figure 2, only now the smaller sample size $n = 100$ is used. Ten simulations are performed, and then the residuals are shown. The residuals help us to visualize the bias and variance of the estimates. As we see, the higher frequencies are often present, and they can be easily removed from a series estimate if additional information about smoothness or number of possible modes is available, see [13].

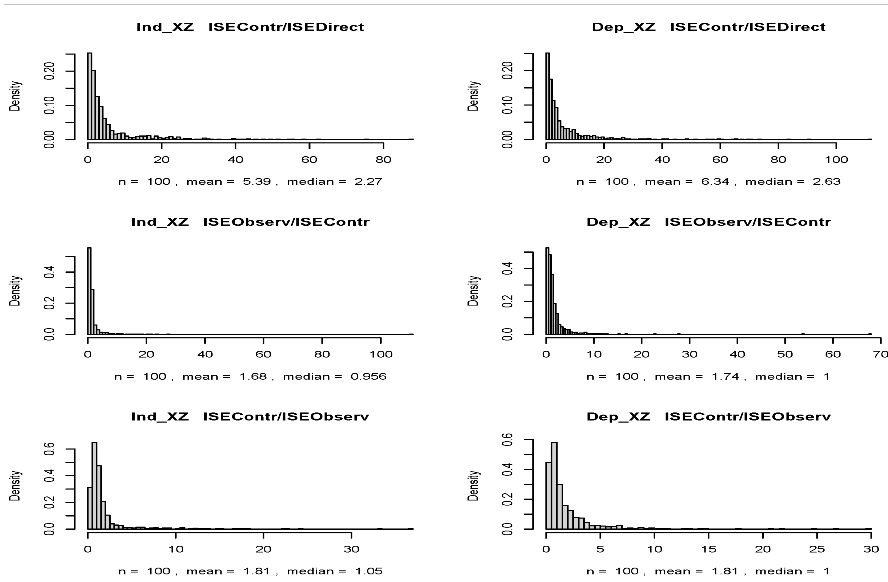

FIG 4. Histograms based on 1000 simulations, $n = 100$*. Mean and median of the ratios are shown in the subtitle.*

5.2. Numerical study

Here we return to the earlier formulated question of why a data-driven regression estimate may perform better than an estimate that uses an underlying design density $f^{X,Z}$. Further, we would like to shed light on relative performance, measured by the ISE (integrated squared error), of the three regression estimates discussed in the previous subsection. Namely, we would like to compare the estimates based on direct data, controlled CSC data when $f^{X,Z}$ is known, and observational CSC data when $f^{X,Z}$ is estimated. We already observed in Figure 2 the outcome when knowing the design density was not helpful. A reasonable explanation of that phenomenon is that a simulated data, especially a bivariate one, may be far from an “expected” data corresponding to the underlying design density. In that case a data-driven estimate, based on the estimated design bivariate density, may be better and yield a smaller ISE.

Let us check the above-made conjecture via the following intensive simulations. We use the two underlying CSC experiments of Figure 2 with independent and dependent X and Z, repeat each experiment 1000 times, for each simulation calculate the above-explained ISEDirect, ISEContr and ISEObserv for the 3 corresponding estimates, and then visualize histograms of ratios between the three ISEs for different sample sizes n in turn.

We begin with Figure 4 which shows us the histograms of ratios for the sample size $n = 100$. The titles explain the underlying ratios. Similarly to Figure 2, the

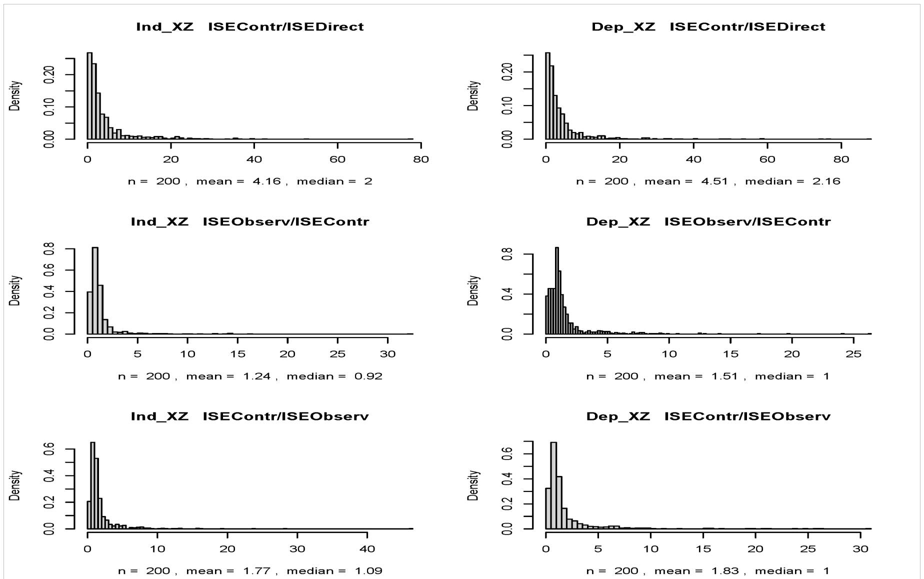

FIG 5. *Histograms based on 1000 simulations, $n = 200$. Mean and median of the ratios are shown in the subtitle.*

left column presents results for independent X and Z, the right for dependent. Let us look at diagrams in the left column. The left-top diagram exhibits ratios of ISEContr (the ISE of the CSC regression estimate using the underlying design density $f^{X,Z}$) to ISEDirect (the ISE of regression estimate based on direct observations of (X, T)). As it could be expected, the ratio can be relatively large with the mean 5.39 and median 2.27. Note that for some simulations the direct data estimate may be overwhelmingly better. Now let us look at the left-middle diagram. Here we compare ISE of the data-driven CSC regression estimate (ISEObserv) with ISE of the CSC regression estimate knowing the $f^{X,Z}$ (ISEContr). It could be expected that the ratios ISEObserv/ISEContr would be larger than 1. And indeed, the mean ratio is 1.68. Further, some ratios are very large implying that the used design density estimate is far from being perfect. At the same time, the median ratio is 0.956. This tells us that in a majority of simulations knowing the design density is not needed. This supports and sheds light on the above-made conjecture about why the data-driven estimator may outperform the estimator knowing the underlying design density. The bottom diagram adds extra evidence in support of the conjecture.

Now let us look at results for the experiment with dependent X and Z in the right column of diagrams in Figure 4. The outcomes are similar but they are clearly magnified due to the dependence between X and Z, and correspondingly due to the more complicated underlying design density. It is difficult for the CSC regression estimate to compete with the regression estimate based on direct

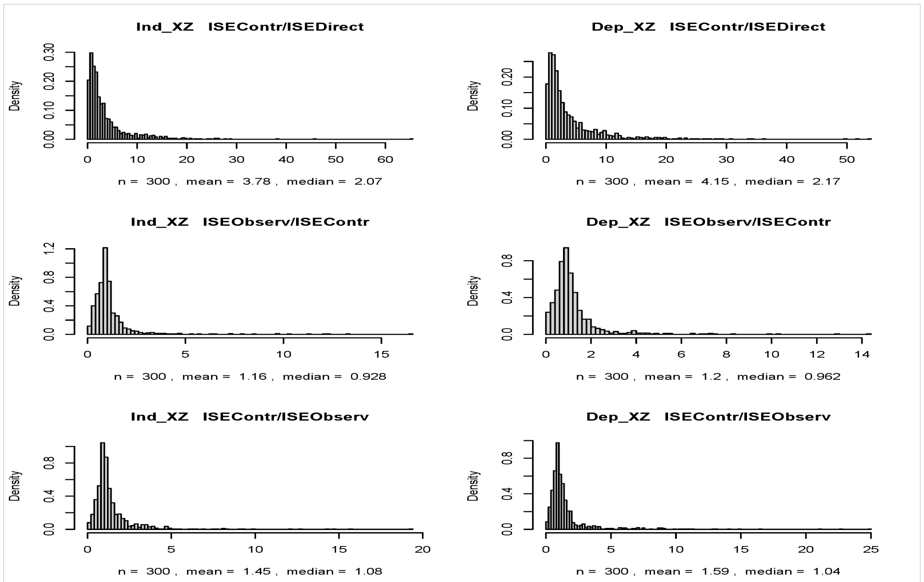

Fig 6. Histograms based on 1000 simulations, $n = 300$. Mean and median of the ratios are shown in the subtitle.

observations. Note that some ratios are above 100, but overall the mean and median are reasonable for this sample size. The right-middle diagram shows that the mean ratio is 1.74 and some ratios are about 70. At the same time, the median ratio is 1, that is for a half of simulations the data-driven CSC estimate performs exceptionally well with respect to the CSC estimate knowing the design density. The right-bottom diagram sheds an extra light on that conclusion.

What will be if we increase the sample size? Figures 5 and 6 exhibit similar histograms for sample sizes 200 and 300, respectively. Overall, the conclusions are the same and they support the above-made observation, only with larger samples we see less outliers in terms of extremely large ratios and the improvement in relative performance with respect to the oracle-estimator using direct observations of (X, T). This is a natural outcome for larger samples, and it supports the proposed methodology of CSC regression estimation.

5.3. Real example

In the Introduction a real CSC example was presented in Figure 1. The reader is advised to look at that figure one more time and try to visualize an underlying regression using the experience gained in subsection 5.1. Next let us look at the estimate (the solid line) shown in the top diagram in Figure 7. After the previous training in “reading” CSC scattergrams, the regression looks reasonable. Further, the decreasing regression reflects the underlying physics of anaerobic

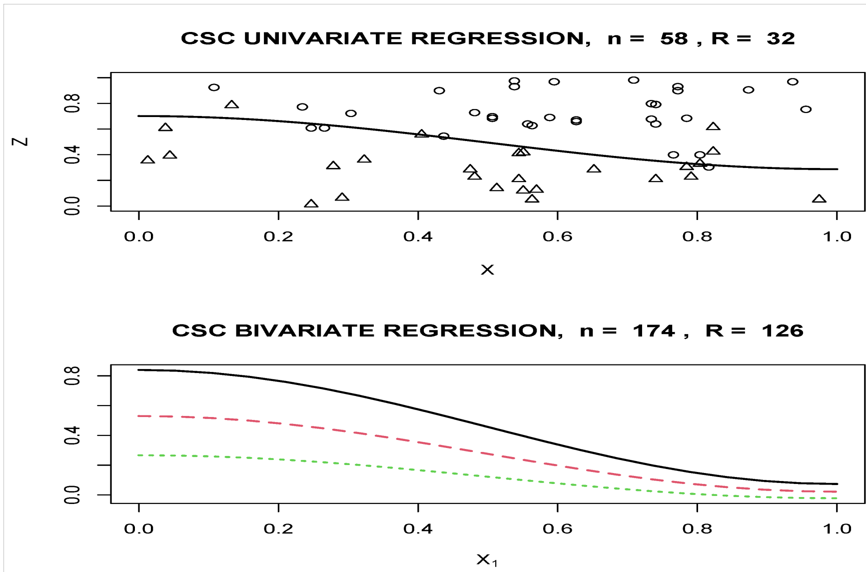

Fig 7. Regression analysis of two BIFAR anaerobic digestion datasets. The top diagram exhibits regression analysis of CSC data presented in Figure 1. The solid line shows the regression estimate, and $R := \sum_{l=1}^{n} \Delta_l$. The bottom diagram exhibits bivariate regression analysis of controlled CSC anaerobic digestion of an industrial sludge discussed in subsection 6.1. Here X_1 is thickness and X_2 is concentration of bacteria rescaled onto $[0,1]^2$. The solid, dashed and dotted lines show regression estimates $\hat{m}(x_1, x_2)$ corresponding to x_2 equal to 0.1, 0.5, and 0.8, respectively.

digestion, and the outcome was well accepted by the BIFAR. The bottom diagram will be explained in the next section.

6. Multivariate regression

We begin with a multivariate anisotropic regression and show that it is possible to match the classical result of [24]. The theory is complemented by analysis of real data. We continue with discussion of several interesting open topics in multivariate CSC regression. The final subsection is devoted to estimation of conditional survival function that will help us to learn more about nonparametric estimation for CSC data.

6.1. Rate optimal estimation for anisotropic regression

This is a natural extension of the explored univariate setting. Following [24], where the case of direct observations was considered, we begin with the classical setting of an anisotropic multivariate regression. An underlying multivariate regression function is $m(\mathbf{x}) = \mathbb{E}\{T|\mathbf{X} = \mathbf{x}\}$, $\mathbf{x} := (x_1, \ldots, x_k) \in [0,1]^k$. Its smoothness (the number of derivatives) may be different for each covariate, namely it is assumed that the regression function belongs to an anisotropic Sobolev class

$$\mathcal{S}(\alpha_1, \ldots, \alpha_k) := \Big\{m(\mathbf{x}) : \ m(\mathbf{x}) = \sum_{i_1,\ldots,i_k=0}^{\infty} \theta_{\mathbf{i}} \varphi_{\mathbf{i}}(\mathbf{x}),$$

$$\sum_{i_1,\ldots,i_k=0}^{\infty} \theta_{\mathbf{i}}^2 [1 + \sum_{s=1}^{k} i_s^{2\alpha_s}] \leqslant Q < \infty\Big\}, \tag{6.1}$$

where $\mathbf{i} = (i_1, \ldots, i_k)$, and $\varphi_{\mathbf{i}}(\mathbf{x}) = \prod_{s=1}^{k} \varphi_{i_s}(x_{i_s})$ are elements of the cosine tensor-product basis on $[0,1]^k$. Introduce an effective multivariate smoothness $\alpha_* := [\sum_{s=1}^{k} \alpha_s^{-1}]^{-1}$. The pioneering result of [24] shows that for a sample of size n from $(\mathbf{X}, T)$ the optimal rate of MISE convergence is $n^{-2\alpha_*/(2\alpha_*+1)}$, and in particular if $\alpha_1 = \ldots = \alpha_k = \alpha$ then the rate is $n^{-2\alpha/(2\alpha+k)}$. Recall that k is dimensionality of the regression, and the decreased rate defines the curse of multidimensionality.

Let us show that this rate is also achievable for a CSC sample of size n from $(\mathbf{X}, Z, \Delta)$, and hence we again can break the curse of CSC in terms of slower rates of convergence for nonparametric estimation of distributions. The following result matches [14,24].

Theorem 6.1. *Consider a controlled CSC sampling from* $(\mathbf{X}, Z, \Delta)$ *where* $\mathbf{X} = (X_1, \ldots, X_k)$ *is supported on* $[0,1]^k$. *Suppose that* $S^{T|\mathbf{X},Z}(t|\mathbf{x}, z) = S^{T|\mathbf{X}}(t|\mathbf{x})$ *and*

$$\int_0^\infty \Big[\int_{[0,1]^k} \frac{S^{T|\mathbf{X}}(z|\mathbf{x})}{f^{\mathbf{X},Z}(\mathbf{x}, z)} d\mathbf{x}\Big] dz < \infty. \tag{6.2}$$

Introduce a multivariate regression estimator

$$\widehat{m}(\mathbf{x}) := \sum_{\mathbf{i} \in \mathcal{J}} \widetilde{\theta}_{\mathbf{i}} \varphi_{\mathbf{i}}(\mathbf{x}), \tag{6.3}$$

where $\mathcal{J} := \{0, 1, \ldots, J_1\} \otimes \{0, 1, \ldots, J_2\} \otimes \ldots \otimes \{0, 1, \ldots, J_k\}$, $J_s := 1 + \lfloor n^{\frac{\alpha_s^{-1}}{2+\alpha_*^{-1}}} \rfloor$,

$$\widetilde{\theta}_{\mathbf{i}} := n^{-1} \sum_{l=1}^{n} \frac{\Delta_l' \varphi_{\mathbf{i}}(\mathbf{X}_l)}{f^{\mathbf{X},Z}(\mathbf{X}_l, Z_l)}. \tag{6.4}$$

Then the estimator is rate optimal and

$$\sup_{m \in \mathcal{S}(\alpha_1, \ldots, \alpha_k)} \mathbb{E}\Big\{\int_{[0,1]^k} (\widehat{m}(\mathbf{x}) - m(\mathbf{x}))^2 d\mathbf{x}\Big\} \leqslant C n^{-2\alpha_*/(2\alpha_*+1)}. \tag{6.5}$$

We may conclude that even for the case of anisotropic multivariate regression the CSC does not slow down rates known in the theory of regression for direct observations.

Let us complement the theory by example of bivariate regression for anaerobic digestion. Additions of microorganisms to industrial sludges for improving anaerobic digestion have been explored intensively in recent years, see a review in [52]. When the process of anaerobic digestion is improved by the microbes, it is referred to as bioaugmentation. BIFAR is interested in how thickness (recall our discussion of Figures 1 and 7) and concentration of added microbes affect time T of digestion. BIFAR conducted a controlled CSC experiment similar to the one described in the Introduction only now with a pair of predictors (X_1, X_2) where X_1 is thickness and X_2 is concentration of microbes. The bottom diagram in Figure 7 exhibits three slices of estimated bivariate regression for a small concentration (the solid line), a moderate concentration (the dashed line), and a large concentration (the dotted line). As we see, addition of microbes is more beneficial for sludges with lower thickness (implies active bioaugmentation), and the effect diminishes as thickness increases but it is still clearly present. The regression sheds light on how performance of anaerobic digestion can be regulated by enriching the microbial community and by thickening sludges.

6.2. Several topics in multivariate regression

In this section several topics, suggested by the reviewers, are discussed. They are the adaptation to smoothness of continuous covariates, dimension reduction, categorical covariates, and a special setting with necessity of adding a covariate that makes T and Z conditionally independent but which by itself is not of interest for the regression.

We begin with adaptation and dimension reduction. In general these are two different topics. The former is to match performance of the oracle who knows smoothness of the regression function. For the setting of Theorem 6.1 the adaptation means that an estimator yields the rate $n^{-2\alpha_*/(2\alpha_*+1)}$ without knowing nuisance parameters $(\alpha_1, \dots, \alpha_k)$. Dimension reduction is when an estimator matches performance of the oracle who knows that only a subset of covariates defines the regression. To make the presentation shorter, we are considering these two problems together. Namely, the estimator should match the oracle who knows the subset of covariates that define the regression and also knows the corresponding smoothness of regression in those covariates.

It is sufficient to explain the heuristic for the case of a bivariate regression, the general case is considered absolutely similarly. Introduce the tensor-product cosine basis $\{\varphi_j(x_1)\varphi_i(x_2), j, i = 0, 1, \dots\}$ on $[0, 1]^2$, and write down a bivariate regression as the Fourier series,

$$m(x_1, x_2) = \sum_{j,i=0}^{\infty} \theta_{ji}\varphi_j(x_1)\varphi_i(x_2), \quad (x_1, x_2) \in [0, 1]^2. \tag{6.6}$$

Next, we rewrite (6.6) as the sum of four terms,

$$m(x_1, x_2) = \theta_{00}\varphi_0(x_1)$$

$$+\sum_{j=1}^{\infty}\theta_{j0}\varphi_j(x_1)+\sum_{i=1}^{\infty}\theta_{0i}\varphi_i(x_2)+\sum_{j,i\geqslant 1}\theta_{ji}\varphi_j(x_1)\varphi_i(x_2). \tag{6.7}$$

Following the blockwise adaptation methodology explained in Section 3, we introduce blocks for Fourier coefficients θ_{j0}, θ_{0i}, and θ_{ji} with indexes $j, i \geqslant 1$. Blocks B_k of length L_k, defined in subsection 3.2, are used for the first and second sums in (6.7). Tensor-product blocks $B_{k_1k_2} := B_{k_1} \otimes B_{k_2}$ are used for the third sum. Let $\widetilde{\theta}_{ji}$ be the Fourier estimator (6.4). Then, following (3.19), the blockwise oracle-estimator is

$$\widetilde{m}^*(x_1,x_2)=\widetilde{\theta}_{ji}\varphi_0(x)+\sum_{k=1}^{k_n}\frac{\sum_{j\in B_k}\theta_{j0}^2}{\sum_{j\in B_k}[\theta_{j0}^2+\mathbb{E}\{(\widetilde{\theta}_{j0}-\theta_{j0})^2\}]}\widetilde{\theta}_{j0}\varphi_j(x_1)$$

$$+\sum_{k=1}^{k_n}\frac{\sum_{i\in B_k}\theta_{0i}^2}{\sum_{i\in B_k}[\theta_{0i}^2+\mathbb{E}\{(\widetilde{\theta}_{0i}-\theta_{0i})^2\}]}\widetilde{\theta}_{0i}\varphi_i(x_2)$$

$$+\sum_{k_1,k_2=1}^{k_n'}\frac{\sum_{(j,i)i\in B_{k_1k_2}}\theta_{ji}^2}{\sum_{(j,i)\in B_{k_1k_2}}[\theta_{ji}^2+\mathbb{E}\{(\widetilde{\theta}_{ji}-\theta_{ji})^2\}]}\widetilde{\theta}_{ji}\varphi_j(x_1)\varphi_i(x_2). \tag{6.8}$$

Here, similarly to subsection 3.2, sequences k_n and k_n' are chosen based on the assumption about minimal smoothness of the regression.

Now let us look at the three sums in (6.8). If the regression $m(x_1, x_2)$ depends only on x_1, then the second and third sums are equal to zero. If the regression $m(x_1, x_2)$ depends only on x_2, then the first and third sums are equal to zero. Further, it is known from [13,14,24] that the oracle's blockwise shrinkage can be mimicked by statistics with accuracy preserving the oracle's rate of the MISE convergence. Accordingly, this special blockwise shrinkage allows us to solve the adaptation and dimension reduction problems.

Next we are considering the case of a categorical covariate Y supported on $\{0, 1, \dots, K-1\}$. It is natural to incorporate the Y into a series estimator by using a discrete cosine basis on $\{0, 1, \dots, K-1\}$,

$$\zeta_0(y) := 1,\ \zeta_k(y) := 2^{1/2}\cos(\pi(2y+1)k/(2K)), k = 1,\dots,K-1. \tag{6.9}$$

The used inner product for this basis is $\langle g_1, g_2\rangle = K^{-1}\sum_{y\in\{0,1,\dots,K-1\}} g_1(y)g_2(y)$.

As an example, consider a regression $m(x, y) = \mathbb{E}\{T|X = x, Y = y\}$ where X is a continuous predictor supported on $[0, 1]$ and Y is categorical. The regression can be written as the Fourier series $m(x,y) = \sum_{j=0}^{\infty}\sum_{k=0}^{K-1}\theta_{jk}\varphi_j(x)\zeta_k(y)$, and the Fourier coefficients θ_{jk} can be estimated by unbiased sample mean estimates (compare with (6.4))

$$\breve{\theta}_{jk} := K^{-1}n^{-1}\sum_{l=1}^{n}\frac{\Delta_l'\varphi_j(X_l)\zeta_k(Y_l)}{f^{X,Y,Z}(X_l,Y_l,Z_l)}. \tag{6.10}$$

Finally, let us consider the following situation. Recall that the developed theory of regression estimation for the CSC is based on the assumption that the lifetime of interest T and the monitoring time Z are conditionally independent given covariates. Consider the case of two continuous covariates X_1 and X_2 supported on $[0,1]^2$. We are interested in regression $m(X_1) := \mathbb{E}\{T|X_1\}$ but need to consider the bivariate regression $m(X_1, X_2) := \mathbb{E}\{T|X_1, X_2\}$ because T and Z are independent only given the pair (X_1, X_2). Two possible scenarios are discussed.

First, suppose that $m(x_1, x_2) = m(x_1)$. Then we are dealing with the already discussed problem of dimension reduction. In other words, we treat the problem as a bivariate regression and use the above-described estimator that adapts to the underlying univariate dimensionality.

Second, suppose that the regression $m(x_1, x_2)$ is bivariate but we are interested in the univariate regression $m(x_1) = \mathbb{E}\{T|X_1 = x_1\}$. For simplicity, assume that the CSC is controlled and hence we know the joint density $f^{X_1,X_2,Z}$. Write,

$$m(x_1) = \int_0^\infty S^{T|X_1}(t|x_1)dt$$

$$= \int_0^\infty \int_0^1 f^{X_2|X_1}(x_2|x_1) S^{T|X_1,X_2}(t|x_1,x_2)dx_2dt. \quad (6.11)$$

Fourier coefficients of the univariate regression of interest can be written as

$$\theta_j := \int_0^1 m(x_1)\varphi_j(x_1)dx_1$$

$$= \int_0^\infty \int_{[0,1]^2} f^{X_2|X_1}(x_2|x_1) S^{T|X_1,X_2}(t|x_1,x_2)\varphi_j(x_1)dx_1dx_2dt. \quad (6.12)$$

Note that $S^{T|X_1,X_2}(t|x_1,x_2) = f^{X_1,X_2,Z,\Delta'}(x_1,x_2,t,1)/f^{X_1,X_2,Z}(x_1,x_2,t)$ and continue (6.12),

$$\theta_j = \mathbb{E}\{\frac{\Delta' f^{X_2|X_1}(X_2|X_1)\varphi_j(X_1)}{f^{X_1,X_2,Z}(X_1,X_2,Z)}\} = \mathbb{E}\{\frac{\Delta'\varphi_j(X_1)}{f^{X_1}(X_1) f^{Z|X_1,X_2}(Z|X_1,X_2)}\}.$$

This formula yields the unbiased sample mean Fourier coefficient estimator

$$\tilde{\theta}_j := n^{-1}\sum_{l=1}^n \frac{\Delta'_l \varphi_j(X_{1l})}{f^{X_1}(X_{1l}) f^{Z|X_1,X_2}(Z_l|X_{1l},X_{2l})}. \quad (6.13)$$

Finally, the methodology of Section 3 can be used to calculate a series regression estimator.

6.3. Estimation of conditional survival function

Following the above-presented setting of a multivariate regression, let us consider estimation of a conditional survival function $S^{T|\mathbf{X}}(t|\mathbf{x}) := \mathbb{P}(T > t|\mathbf{X} = \mathbf{x})$. Note that now we are estimating a $(1+k)$-dimensional function in t and $\mathbf{x}$. Recall that, according to [18], estimation of a survival function $S^T(t)$ based on a CSC sample from (Z,Δ) is ill-posed, no longer the classical rate n^{-1} is achievable, and for α-fold differentiable survival function the optimal rate of MISE convergence is $n^{-2\alpha/(2\alpha+1)}$. As we will see shortly, estimation of a conditional survival function is also ill-posed with respect to its direct data counterpart. To make presentation of the following proposition shorter, let us assume that T is supported on $[0,1]$ and that $S^{T|\mathbf{X}}(t|\mathbf{x})$, as a $(1+k)$-dimensional function in $(t,\mathbf{x})$, belongs to a Sobolev class $\mathcal{S}(\alpha_1,\dots,\alpha_{k+1})$ defined in (6.1).

Theorem 6.2. *Suppose that assumptions of Theorem 6.1 hold and $S^T(1)=0$. Introduce an estimator*

$$\widehat{S}^{T|\mathbf{X}}(t|\mathbf{x}) := \sum_{\mathbf{i}\in\mathcal{J}} \widehat{\theta}_{\mathbf{i}}\varphi_{\mathbf{i}}(t,\mathbf{x}). \tag{6.14}$$

Here $\mathbf{i} := (i_1,\dots,i_{k+1})$, $\mathcal{J} := \{0,1,\dots,J_1\}\otimes\{0,1,\dots,J_2\}\otimes\dots\otimes\{0,1,\dots,J_{k+1}\}$, $J_s := 1+\lfloor n^{\frac{\alpha_s^{-1}}{2+\alpha_*^{-1}}}\rfloor$, $\alpha_* = [\sum_{s=1}^{k+1}\alpha_s^{-1}]^{-1}$, $\varphi_{\mathbf{i}}(t,\mathbf{x}) = \varphi_1(t)\prod_{s=2}^{k+1}\varphi_{i_s}(x_{i_s-1})$, *and*

$$\widehat{\theta}_{\mathbf{i}} := n^{-1}\sum_{l=1}^{n}\frac{\Delta_l'\varphi_{\mathbf{i}}(Z_l,\mathbf{X}_l)}{f^{\mathbf{X},Z}(\mathbf{X}_l,Z_l)}. \tag{6.15}$$

Then the estimator (6.14) is rate optimal and

$$\sup_{S^{T|\mathbf{X}}\in\mathcal{S}(\alpha_1,\dots,\alpha_{k+1})} \mathbb{E}\Big\{\int_{[0,1]^{k+1}}(\widehat{S}^{T|\mathbf{X}}(t|\mathbf{x}) - S^{T|\mathbf{X}}(t|\mathbf{x}))^2 dtd\mathbf{x}\Big\}$$
$$\leqslant Cn^{-2\alpha_*/(2\alpha_*+1)}. \tag{6.16}$$

To get a feeling of the rate and compare it with a regression setting of the previous subsection, let us assume that all α_s are the same and equal to α, that is $S^{T|\mathbf{X}}(t|\mathbf{x})$ and $m(\mathbf{x})$ are α-fold differentiable with respect to each variable. Then the optimal rate for regression is $n^{-2\alpha/(2\alpha+k)}$ versus a slower $n^{-2\alpha/(2\alpha+k+1)}$ for the conditional survival function. The difference in one dimension is explained by the integral $m(\mathbf{x}) = \int_0^\infty S^{T|\mathbf{X}}(t|\mathbf{x})dt$. On the other hand, because $S^{T|\mathbf{X}}(t|\mathbf{x}) = \mathbb{E}\{I(T>t)|\mathbf{X}=\mathbf{x}\}$, for direct data conditional survival is estimated with the classical rate $n^{-2\alpha/(2\alpha+k)}$.

7. Conditional linear functionals

Recall formula (1.2) and write

$$m(x) := \mathbb{E}\{T|X=x\} = \int_0^\infty S^{T|X}(t|x)dt. \tag{7.1}$$

We also have

$$m(x) = \int_0^\infty tf^{T|X}(t|x)dt. \tag{7.2}$$

As we see, the regression can be considered as a linear functional of the conditional survival or of the conditional density. There is a rich survival analysis literature devoted to linear functionals [23,28,29,31,32,61,66,72,74]. As we will see shortly, in general it is preferable to work with linear functionals of $S^{T|X}$ because they may be more accurately estimated.

In what follows we assume that ψ is a known and continuous on $[0,\infty)$ function, and $\Psi(t) := \int_0^t \psi(u)du$. Introduce a restricted linear functional

$$M_r(x) := M_r(x,\psi,S^{T|X}) := \int_0^r \psi(t)S^{T|X}(t|x)dt. \tag{7.3}$$

Note that $\psi(t) = 1$ yields the restricted regression μ_r, while the power function $\psi(t) = kt^{k-1}$, $k = 2,3,\ldots$ is often used to evaluate higher moments [28,32,72].

To present results for estimation of M_r, it is convenient to use Assumption 2.3 where μ_r and $\mu_{r0}(x)$ are replaced by M_r and $M_{r0}(x) := \int_0^r \psi(t)S_0^{T|X}(t|x)dt$, respectively. To stress the change, the modified function class (2.3) is denoted as $\mathcal{F}_n(S_0^{T|X},\alpha,Q,r,\psi)$.

Theorem 7.1 (Lower bounds for oracle-estimators). *Suppose that Assumptions 2.1-2.4 hold. Suppose that a CSC sample of size n from (X,Z,Δ) is given. Then*

$$\inf_{\widetilde{M}_r^*} \sup_{S^{T|X}\in\mathcal{F}_n(S_0^{T|X},\alpha,Q,r,\psi)} \Big\{[n/d'(S^{T|X},f^{X,Z},r,\psi)]^{2\alpha/(2\alpha+1)}$$
$$\times \mathbb{E}_{S^{T|X}}\{\int_0^1 (\widetilde{M}_r^*(x) - M_r(x))^2dx\}\Big\} \geqslant P(\alpha,Q)(1+o_n(1)). \tag{7.4}$$

Here P is defined in (2.7) and

$$d'(S^{T|X},f^{X,Z},r,\psi) := \int_{\mathcal{R}_r} \frac{\psi^2(t)(1-S^{T|X}(t|x))S^{T|X}(t|x)}{f^{X,Z}(x,t)}dtdx. \tag{7.5}$$

Now suppose that a MCSC sample of size n from $(\Delta'X,\Delta'Z,\Delta)$ is given. Then

$$\inf_{\widetilde{M}_r^*} \sup_{S^{T|X}\in\mathcal{F}_n(S_0^{T|X},\alpha,Q,r,\psi)} \Big\{[n/d'_*(S^{T|X},f^{X,Z},r,\psi)]^{2\alpha/(2\alpha+1)}$$
$$\times \mathbb{E}_{S^{T|X}}\{\int_0^1 (\widetilde{M}_r^*(x) - M_r(x))^2dx\}\Big\} \geqslant P(\alpha,Q)(1+o_n(1)). \tag{7.6}$$

Here

$$d'_*(S^{T|X},f^{X,Z},r,\psi) := \int_{\mathcal{R}_r} \frac{\psi^2(t)S^{T|X}(t|x)}{f^{X,Z}(x,t)}dtdx. \tag{7.7}$$

In (7.4) and (7.6) the infimum is taken over all oracle-estimators that know the corresponding sample, $\mathcal{F}_n(S_0^{T|X},\alpha,Q,r,\psi)$, $f^{X,Z}$ and ψ.

Let us present an oracle-estimator that attains the lower bound (7.6), and it is within a constant factor from (7.4).

Theorem 7.2. *Let Assumptions 2.1 and 2.4 hold. Consider the oracle-estimator*

$$\overline{M}_r^*(x) := \sum_{j=0}^{q_n} \bar{\theta}_j \varphi_j(x) + \sum_{j=q_n+1}^{J_n^*} (1-(j/J_n^*)^\alpha)\bar{\theta}_j\varphi_j(x). \tag{7.8}$$

Here $J_n^* := q_n + 1 + \lfloor [(n/d'_*)Q\pi^{-2\alpha}(\alpha+1)(2\alpha+1)/\alpha]^{1/(2\alpha+1)} \rfloor$, *and*

$$\bar{\theta}_j := n^{-1} \sum_{l=1}^{n} \frac{\Delta' I(\Delta'_l Z_l \leqslant r)\psi(\Delta'_l Z_l)\varphi_j(\Delta'_l Z_l)}{f^{X,Z}(\Delta'_l X_l, \Delta'_l Z_l)}. \tag{7.9}$$

The MISE of this oracle-estimator attains the lower bound (7.6).

We can also use the blockwise shrinkage methodology of Section 3 to prove that the adaptive estimator is also efficient.

Now let us look at linear functionals of $f^{T|X}$. Write,

$$M_r(x, \Psi, f^{T|X}) = \int_0^r \Psi(t) f^{T|X}(t|x)dt$$

$$= [\int_0^r \psi(t) S^{T|X}(t|x)dt] - [\Psi(r)S^{T|X}(r|x)]$$

$$= M_r(x, \psi, S^{T|X}) - [\Psi(r)S^{T|X}(r|x)]. \tag{7.10}$$

A particular example is $M_b(x, t^k, f^{T|X}) = \mathbb{E}\{T^k|X=x\}$ where b is the endpoint of the support of T, see an interesting discussion in [64,65]. Let us look at the two terms on the right side of (7.10). Suppose that $\Psi(r) \neq 0$ to avoid triviality. The first term is the already studied linear functional of $S^{T|X}$, and we know that it can be estimated with the univariate rate defined by smoothness of the conditional survival in x. The second term, unless $S^{T|X}(r|x) = 0$ due to $r \geqslant b$, is the bivariate function in (r, x), and it is estimated with a slower bivariate rate [65].

The latter is not the only complication in estimating linear functionals of $f^{T|X}$. First, recall that adaptive nonparametric pointwise estimation triggers the Lepski's penalty for the rate of convergence [13,68]. Second, an intriguing outcome may occur if $b = b(x)$, that is when the endpoint of the support of T depends on X. In this case for some x we may have a univariate rate of estimating $M_r(x, \Psi, f^{X|T})$ and for others a bivariate rate.

In conclusion, the survival analysis uses the above-discussed linear functionals as interpretable and meaningful survival metrics [28,32,72,74]. It is reasonable to conclude that, whenever it is possible to choose between linear functionals of $S^{T|X}$ or $f^{T|X}$, it is prudent to choose the former.

8. Conclusions and further research

1. Under current status censoring (CSC), the lifetime of interest T to the event of interest is not observed. Instead, there is a possibility to know at a monitoring

time Z if the event of interest already occurred or not. Accordingly, under CSC we observe a sample of size n from pair (Z, Δ) where $\Delta := I(T \leqslant Z)$ is the status, and if a predictor X of T is available, then we observe a sample of size n from the triplet (X, Z, Δ). The CSC literature is primarily devoted to estimation of the distribution of T, and it is shown that the problem is ill-posed and all traditional rates for estimating the survival function and density of T dramatically slow down. In particular, under CSC the rate of estimating the density is the same as estimating a trivariate density based on direct observations. This is why CSC is considered as an extremely complicated problem.

2. For directly observed data there is a familiar principle of equivalence between nonparametric regression and density estimation problems. In particular, the principle implies that rates of the MISE convergence must be the same. Fortunately, the paper shows that CSC limits applicability of the equivalence principle, and CSC nonparametric regression can be estimated with the same rate as a nonparametric regression based on direct observations. This is a dramatic relief for CSC because it has a wide range of applications where T cannot be observed directly.

3. The established fact that rate of a multivariate CSC regression is the same as for the case of direct observations is of a key importance due to the familiar curse of dimensionality. At the same time, unfortunately the outcome is worse for estimating conditional survival functions that suffer slower rates than estimates based on direct observations. On the other hand, there is an interesting theoretical situation with the survival functions. Namely, consider a conditional survival function $S^{T|X}(t|x)$. This is a bivariate function, but for direct data it is estimated with a univariate rate, but CSC "corrects" that and it is estimated with a classical bivariate rate.

4. Nuisance functions and their required smoothness are a hot topic in modern nonparametric literature. It is shown, for the first time in the literature, that a very mild assumption that does not involve second derivatives, is sufficient for a data-driven estimation of regression $m(x) := \mathbb{E}\{T|X = x\}$ that may be as smooth as desired. Accordingly, it is possible to untie smoothness of nuisance functions and smoothness of the regression.

5. Presented in Theorem 1 coefficient of difficulty sheds additional light on CSC via explanation of how an underlying conditional survival function, together with nuisance functions, affect regression estimation. The latter allows us, at least theoretically, compare regressions for direct and CSC data. Namely, consider a classical normal regression model $Y = m(X) + \sigma(X)\xi$. Here a standard normal regression error ξ is independent of predictor X, $\sigma(x)$ is the scale function, and density $f^X(x)$ is continuous, positive and supported on $[0, 1]$. For direct data the coefficient of difficulty for this regression is $\int_0^1 \frac{\sigma^2(x)}{f^X(x)}dx$, and note that it does not depend on an underlying regression function, see [12,13]. This nonparametric result mimics a familiar one in the theory of point estimation. Namely, consider a sample of size n from a Normal variable with mean θ and variance σ^2. Then the sample mean is efficient estimator of θ and its variance σ^2/n attains the famous Cramer-Rao lower bound. We may say that for this

classical parametric problem the coefficient of difficulty is σ^2 (of course, it is the reciprocal of Fisher information), and similarly to the normal regression the coefficient of difficulty does not depend on the estimand. Now let us return to our CSC regression and write down coefficient of difficulty (2.8) as

$$d = \int_0^1 \frac{\int_0^r \frac{S^{T|X}(z|x)(1-S^{T|X}(z|x))}{f^{Z|X}(z|x)}dz}{f^X(x)}dx =: \int_0^1 \frac{\sigma^2_{CSC}(x)}{f^X(x)}dx. \tag{8.1}$$

The expression on the right side of (8.1) is motivated by the above-presented coefficient of difficulty for the normal regression. If one would like to think about equivalence between the two nonparametric models, then (8.1) sheds light on how CSC "creates" a scale function. Further, note how coefficient of difficulty σ^2 for the classical parametric model transfers into $\int_0^1 \frac{\sigma^2(x)}{f^X(x)}dx$ for the direct-data nonparametric regression, and then into (8.1) for CSC regression.

6. What will be if the support of Z is a subset of the support of T, say T is supported on $[0, a]$ while Z is supported on $[0, b]$ with $b < a$? Then no consistent estimation of the regression is possible. A feasible ad hoc remedy is to estimate the distribution over $[0, b]$ and then test reasonable parametric models.

7. Anaerobic digestion of organic municipal solid waste is a key element in sustainable municipal waste management due to its benefits for energy, environment, and economy. This process dramatically reduces emission of greenhouse gases, generates renewable natural gas, and produces fertilizers and soil amendments. At the same time, it is impossible to directly evaluate the minimal time of digestion but collecting CSC observations is possible. Then the nonparametric CSC regression has allowed the environmental company BIFAR to "look" at the hidden minimal time and choose optimal parameters for anaerobic digestion.

8. Let us compare what we know about rates of estimation of parameters and nonparametric functions for directly observed and CSC data. In what follows by a rate we mean an optimal rate under MSE or MISE for a sample of size n. Further, for a multivariate function it is assumed that it has the same number α of derivatives for each variable. We begin with estimation of a population mean $\mu = \mathbb{E}\{T\}$ and a survival function $S^T(t) = \mathbb{E}\{I(T > t)\}$. Note that the first estimand is a parameter and the second is a function of one variable. Nonetheless, for direct observations of T the rate of their estimation is the same n^{-1} because the estimands are expectations of observable variables T and $I(T > t)$, respectively. For CSC the situation changes, and while the population mean is still estimated with the parametric rate n^{-1}, the survival function is estimated with the rate $n^{-2\alpha/(2\alpha+1)}$ which is the same as for estimation of an α-fold differentiable univariate regression $m(x)$ based on direct observations. Relation $\mu = \int_0^\infty S^T(t)dt$ sheds light on why the parametric rate for μ is preserved, and the fact that indicator $I(T > t)$ is no longer directly observed explains "return to normality" in estimation of a univariate survival function $S^T(t)$. The interested reader may find an insightful theory of estimating functionals in [66]. After this warm-up let us turn our attention to nonparametric estimation of a k-variate regression function $m(\mathbf{x}) = \mathbb{E}\{T|\mathbf{X} = \mathbf{x}\}$ and a corresponding conditional survival function $S^{T|\mathbf{X}}(t|\mathbf{x}) = \mathbb{E}\{I(T > t)|\mathbf{X} = \mathbf{x}\}$. The regression

function is k-variate, the conditional survival function is $(k+1)$-variate, but for direct observations they are estimated with the same optimal k-dimensional rate $n^{-2\alpha/(2\alpha+k)}$. The fact that conditional survival is the conditional expectation of an observed indicator $I(T > t)$ explains the result. For CSC the outcome changes. Because the indicator $I(T > t)$ is no longer observed, for the survival function the rate slows down to the classical $(k+1)$-dimensional $n^{-2\alpha/(2\alpha+k+1)}$, and note that the rate "fits" the dimensionality of $S^{T|\mathbf{X}}$. The integral formula $m(\mathbf{x}) = \int_0^\infty S^{T|\mathbf{X}}(t|\mathbf{x})dt$ explains the faster rate $n^{-2\alpha/(2\alpha+k)}$ for CSC regression, and note that the integral effectively performs one dimension reduction with respect to the conditional survival. Let us also stress that while there exists the asymptotic equivalence between density and regression settings for direct data, see [13], there is no such equivalence for CSC data. In other words, the CSC limits the classical theory of asymptotic equivalence between density and regression problems.

9. Let us present several open problems that may be solved via further developing of the presented nonparametric methodology. (i) Regression with doubly CSC data is a practically important topic where the interest is in the length of time between two events that occur sequentially. Classical examples are the length of time between: The infection and its diagnosis; The infection of an individual and the subsequent infection of another individual; Marriage and divorce. Under the doubly CSC, lifetimes of the two consecutive events of interest are not available and instead, at a monitoring time Z, statuses of the two events are observed. An excellent overview of the doubly CSC regression can be found in [55,65], and more recent results in [67]. (ii) Another closely related problem is regression with interval-censored data when survival times are not known exactly, but are only known to have occurred between intermittent examination times, see a discussion in the book [60]. (iii) An interesting developing of the considered CSC regression is a sequential sampling with assigned risk and minimal cost of the sampling. (iv) Efficient nonparametric estimation of the hazard rate is another feasible expansion of the developed theory, see a discussion of the problem for right-censored observations in [15]. It may be expected that the problem is ill-posed, and this presents new challenges for small samples. (v) A CSC problem becomes extremely complicated when T and Z are dependent given X. A straightforward calculation shows that in this case the proposed regression estimators are biased and they estimate a function $g(x) := \int_0^\infty S^{T|X,Z}(z|x,z)dz$. Because in a CSC sample the lifetime of interest is never observed, it is impossible to use CSC data for testing the conditional independence. To tackle the dependence, one of the possible approaches is to use parametric models that define the dependence, see a discussion and interesting results in [44] where the case of a known copula model is explored. An appropriate alternative is to develop an adaptive nonparametric approach based on an extra experiment devoted to exploring the dependence. (vi) It is of interest to consider other nonparametric estimators like kernels, splines and series estimators with wavelet and other bases considered in the literature, explore different loss functions like L_∞ and MSE at a point, develop confidence bands, explore more general Sobolev classes using the methodology of [17]. (vii) An open and interesting topic is to

explore a vast class of known dimension-reduction techniques for multivariate CSC regressions. (viii) Case-control study is another interesting and related problem where sampling is done from subpopulations with $\Delta = 1$ and $\Delta = 0$, see [34]. No optimal nonparametric results have been developed so far. (ix) It is an open and interesting problem to consider a restricted regression with the restriction r being a function of the predictor. (x) Missing data is a familiar complication in survival analysis, see [16]. Theory of nonparametric regression with missing predictors and responses is well developed for direct data, and it will be of interest to test robustness of the standard approaches to the CSC.

Acknowledgements

The author would like to thank the editor, Prof. Gang Li, an Associate Editor and the reviewers for valuable comments and suggestions that broaden the scope and improved the quality of this paper.

Fundings

The research is supported in part by NSF Grant DMS-1915845 and Grants from CAS and BIFAR.

Appendix: Proofs

Recall that all basic definitions can be found at the end of the Introduction. Whenever it is not confusing, in what follows we may skip subscripts.

Proof of Theorem 2.1. It is worthwhile to begin with explanation of the proof's heuristic. Let us present it via classical example of nonparametric regression [11,12]. We begin with the nonparametric regression $Y = m(X) + \sigma\xi$ where ξ is standard normal and independent of the predictor X supported on $[0, 1]$, σ is a positive constant (the scale). A sample of size n from (X, Y) is available and it is known that the regression m belongs to the global Sobolev class $\mathcal{S}(\alpha, Q)$. The idea of obtaining a sharp minimax lower bound is as follows. First, the nonparametric problem of estimating m is replaced by considering a parametric regression $m(x) = m_0(x) + \sum_{j=J_1}^{J_2} \theta_j\varphi_j(x)$ with some specially chosen sequences J_1 and J_2. Then a Bayesian approach is used with special Gaussian distributions for the Fourier coefficients θ_j, and Fisher informations for each Fourier coefficient are calculated. The Fisher informations yield the desired coefficient of difficulty σ^2 that can be attained by estimators. Now consider a more complicated heteroscedastic regression $Y = m(X) + \sigma(X)\xi$ where $\sigma(x)$ is the scale function. We can use the above-described approach, but it yields the coefficient of difficulty $d' = [\int_0^1 \frac{f^X(x)(x)}{\sigma^2(x)} dx]^{-1}$. This coefficient of difficulty is too small. Best regression estimators can attain a larger coefficient of difficulty $d = \int_0^1 \frac{\sigma^2(x)}{f^X(x)} dx$ (the relation between d' and d follows from the Cauchy-Schwarz inequality). To get the larger coefficient of difficulty, a less favorable prior should be proposed. The idea is to divide the support $[0, 1]$ into a slowly increasing in n number of subintervals, on each subinterval approximate $m(x)$ by its own Fourier series, apply to that series the above-presented Bayes approach, and then add the lower bounds for each subinterval. This is how we get the larger coefficient of difficulty d. Finally, for the CSC regression we can employ the above-described methodology of obtaining a lower bound, but it yields the coefficient of difficulty $d' = [\int_{R_r} (f^{X,Z}(x,z)/[(1-S^{T|X}(t|x))S^{T|X}(t|x)])dtdx$ which is too small. Again, the idea is to propose a less favorable prior. We divide the rectangle $\mathcal{R}_r$ into a slowly increasing sequence of subrectangles, on each subrectangle consider its own Fourier series, calculate Fisher informations and finish with local Bayes lower bounds. Then we add together obtained lower bounds for the subrectangles and get the wished coefficient of difficulty. This is the approach used in the following proof.

We begin with introducing several new notations, and also note that we are interested in asymptotic in n. In the proof (and only in this proof) we use the sequence $s := s_n := 5+\lfloor \ln(q_n)\rfloor$, divide the interval $[0,1]$ (the support of X) into s subintervals, and then use an additive perturbation only at the inner intervals. Set

$$\mathcal{M}_s := \{\mu_r(x) : \mu_r(x) := \mu_{r0}(x) + \sum_{k=1}^{s-2} g_k(x) I(1/s \leqslant x \leqslant 1-1/s),\ g_k(x) \in \mathcal{M}_{sk}\}.$$

To define the above-mentioned function classes $\mathcal{M}_{sk}$, we need more definitions. Let $\phi(x) := \phi(n,x)$ be a sequence of flattop kernels defined on a real line such that for a given n it is zero beyond $(0,1)$, it is α-fold continuously differentiable on $(-\infty,\infty)$, $0 \leqslant \phi(x) \leqslant 1$, $\phi(x) = 1$ for $2(\ln(n))^{-2} \leqslant x \leqslant 1-2(\ln(n))^{-2}$, and $|\phi^{(\alpha)}| \leqslant C(\ln(n))^{2\alpha}$, see Section 7.1 in [13]. Set $\phi_{sk}(x) := \phi(sx-k)$ and recall that $\varphi_0(x) := 1$, $\varphi_j(x) := 2^{1/2}\cos(\pi j x)$, $j = 1,2,\ldots$ For $1 \leqslant k \leqslant s-2$ define: $\varphi_{skj}(x) := \sqrt{s}\varphi_j(sx-k)$, $g_{[k]}(x) := \sum_{j=J'(k)}^{J(k)} \nu_{skj}\varphi_{skj}(x)$, $g_{(k)}(x) := g_{[k]}(x)\phi_{sk}(x)$, $J(k) := \lceil [n(2\alpha+1)(\alpha+1)s^{-2\alpha}Q_{sk}(\alpha\pi^{2\alpha})^{-1}]^{1/(2\alpha+1)}\rceil$, $J'(k) := \lceil J(k)/\ln(n)\rceil$, $Q_{sk} := (Q-1/s)(\overline{I_s^{-1}}I_{sk})^{-1}$, $I_{sk}^{-1} := \int_0^r [S_0^{T|X}(t|k/s)[1-S_0^{T|X}(t|k/s)]/f^{X,Z}(k/s,t)]dt$, $\overline{I_s^{-1}} := \sum_{k=1}^{s-2} I_{sk}^{-1}$.

Using the above-introduced definitions we set

$$\mathcal{M}_{sk} := \Big\{g : g(x) = g_{(k)}(x)I(k/s \leqslant x \leqslant (k+1)/s),$$
$$\sum_{j=J'(k)}^{J(k)} (\pi j)^{2\alpha}\nu_{skj}^2 \leqslant s^{-2\alpha}Q_{sk},\ |g_{[k]}(x)|^2 \leqslant s^3\ln(n)J(k)n^{-1}\Big\}.$$

The class $\mathcal{M}_s$ of considered regression functions is defined, and note that any μ_r from this class may be written as

$$\mu_r(x) = \mu_{r0}(x) + \sum_{k=1}^{s-2}\sum_{j=J'(k)}^{J(k)} \nu_{skj}\varphi_{skj}(x)\phi_{sk}(x). \tag{A.1}$$

Now recall that the class $\mathcal{M}_n(\mu_{r0},\alpha,Q,r)$ of underlying μ_r is introduced in Assumption 3. Our next step is to show that for large n the considered in Theorem 1 function class $\mathcal{M}_n(\mu_{r0}, \alpha, Q, r)$ includes the above-defined class $\mathcal{M}_s$. If the latter is correct, then we will be able to replace the larger class by the smaller one in establishing the lower bound. To check this fact, we first note that using the flattop kernel implies that $\mu_r(x) - \mu_{r0}(x)$ is α-fold continuously differentiable over $[0, 1]$. Now we are exploring the derivatives. By Leibniz rule $\int_0^1 [(g_{[k]}(x)\phi_{sk}(x))^{(\alpha)}]^2 dx = \int_0^1 [\sum_{l=0}^{\alpha} \mathbf{C}_l^\alpha g_{[k]}^{(\alpha-l)}(x)\phi_{sk}^{(l)}(x)]^2 dx$ where $\mathbf{C}_l^\alpha := \alpha!/((\alpha-l)!l!)$. Note that $\max_{0\leqslant l\leqslant\alpha} \int_0^1 (\phi_{sk}^{(l)}(x))^2 dx < C(s(\ln(n))^2)^{2\alpha}$, and for $0 < l \leqslant \alpha$ we can write,

$$|g_{[k]}^{(\alpha-l)}(x)|^2 = |\sum_{j=J'(k)}^{J(k)} \nu_{skj}\varphi_{skj}^{(\alpha-l)}(x)|^2$$

$$\leqslant Cs^{2(\alpha-l)+1}(\sum_{j=J'(k)}^{J(k)} j^{2\alpha}\nu_{skj}^2)(\sum_{j=J'(k)}^{J(k)} j^{-2l}) \leqslant C\ln^2(n)/J(k).$$

Further,

$$\int_0^1 [g_{[k]}^{(\alpha)}(x)\phi_{sk}(x)]^2 dx \leqslant \int_{k/s}^{(k+1)/s} (g_{[k]}^{(\alpha)}(x))^2 dx \leqslant Q_{sk}.$$

What was wished to check.

Following [10,16] and using [31], let us introduce specific parameters

$$\tau_{skj} := [n^{-1}(1-3\rho^{-1})I_{sk}^{-1}\max(\rho^{-1}, \min(\rho, (J(k)/j)^\alpha - 1))]^{1/2},$$

where $\rho > 3$ is a constant that may be as large as desired. A direct calculation (see the above-mentioned references) shows that if we set $\nu_{skj} = \tau_{skj}$, then these particular parameters satisfy the definition of classes $\mathcal{M}_{sk}$. Namely, introduce the class of vectors (recall that functions $g_{[k]}(x)$ are used in the definition of $\mathcal{M}_{sk}$)

$$\Theta_{sk} := \Big\{\vec{\nu}_{sk} : \sum_{j=J'(k)}^{J(k)} (\pi j)^{2\alpha}\nu_{skj}^2 \leqslant s^{-2\alpha}Q_{sk},$$

$$|g_{[k]}(x)|^2 \leqslant s^3\ln(n)J(k)n^{-1}\Big\}, \;\; k = 1, \ldots, s-2, \tag{A.2}$$

where $\vec{\nu}_{sk} := \{\nu_{skJ'(k)}, \ldots, \nu_{skJ(k)}\}$, then $\vec{\tau}_{sk} := \{\tau_{skJ'(k)}, \ldots, \tau_{skJ(k)}\} \in \Theta_{sk}$.

Now recall that the oracle knows the anchor μ_{r0}, and hence we can write down an oracle-estimator as $\tilde{\mu}_r^*(x) = \mu_{r0}(x) + \tilde{g}^*(x)$, $x \in [0, 1]$ and convert the considered problem into estimation of the additive perturbation g. Using this fact, the already established $\mathcal{M}_s \subset \mathcal{M}_n(\mu_{r0}, \alpha, Q, r)$, and the Parseval identity we can write,

$$\sup_{\mu_r \in \mathcal{M}_n(\mu_{r_0}, \alpha, Q, r)} E\left\{\int_0^1 (\tilde{m}^*(x) - m(x))^2 dx\right\}$$

$$\geqslant \sup_{m \in \mathcal{M}_s} E\left\{\int_0^1 (\widetilde{m}^*(x) - m(x))^2 dx\right\}$$

$$= \sum_{k=1}^{s-2} \sup_{\vec{\nu}_{sk} \in \Theta_{sk}} E\left\{\int_{k/s}^{(k+1)/s} (\widetilde{g}^*(x) - g_{(k)}(x))^2 dx\right\}. \tag{A.3}$$

Next we are evaluating a particular integral on the right side of (A.3),

$$\int_{k/s}^{(k+1)/s} (\widetilde{g}^*(x) - g_{(k)}(x))^2 dx$$

$$\geqslant (1 - s^{-1}) \int_{k/s}^{(k+1)/s} (\widetilde{g}^*(x) - g_{[k]}(x))^2 dx$$

$$-s \int_{k/s}^{(k+1)/s} [g_{[k]}(x)(1 - \phi_{sk}(x))]^2 dx$$

$$\geqslant (1 - s^{-1}) \int_{k/s}^{(k+1)/s} (\widetilde{g}^*(x) - g_{[k]}(x))^2 dx$$

$$+ o_n(1) s (\ln(n))^{-1/2} n^{-2\alpha/(2\alpha+1)} . \tag{A.4}$$

By using this lower bound in (A.3) and with the help of the Parseval's identity we conclude that

$$\sup_{\mu_r \in \mathcal{M}_n(\mu_{r0}, \alpha, Q, r)} E\left\{\int_0^1 (\widetilde{\mu}_r^*(x) - \mu_r(x))^2 dx\right\}$$

$$\geqslant (1 - s^{-1}) \sum_{k=1}^{s-2} \sup_{\vec{\nu}_{sk} \in \Theta_{sk}} \sum_{j=J'(k)}^{J(k)} \mathbb{E}\left\{(\widetilde{\nu}_{skj}^* - \nu_{skj})^2\right\} + o(1) n^{-2\alpha/(2\alpha+1)}, \tag{A.5}$$

where $\widetilde{\nu}_{skj}^* := \int_{k/s}^{(k+1)/s} \widetilde{g}^*(x) \varphi_{skj}(x) dx$. The last term on the right side of (A.5) is in order smaller than the verified lower bound, and hence we can concentrate on the $s - 2$ sums.

To estimate a particular sum, we use a classical minimax theory technique when a minimax risk is bounded from below by a Bayesian risk. To do this, we need to introduce a prior distribution for $\vec{\nu}_{sk}$. Let us explain how this may be done. Introduce independent and zero mean normal random variables ζ_{skj} with the above-defined corresponding variances τ_{skj}^2. To use that normal distribution for creating a bona fide prior on Θ_{sk}, it is projected onto Θ_{sk}, and then we are dealing with random vector $\vec{\eta}$ such that $\mathbb{P}(\vec{\eta} \in A) := \mathbb{P}(\vec{\zeta} \in A | \vec{\zeta} \in \Theta_{sk})$, $A \in \Theta_{sk}$. If we were dealing with a traditional regression, then we could calculate Fisher information corresponding to each ν_{skj} and then follow the proof in [10]. Here, because we are dealing with CSC regression when the response is not observed directly, that path should be modified. Below two steps are proposed and they include: (1) New method of defining normal variables ζ_{skj} via a vector of other independent normal variables; (2) Calculation of Fisher information matrices for the new variables.

For $(x,t) \in \mathcal{R} := [0,1] \times [0,\infty)$ introduce a function

$$S_*^{T|X}(t|x) := S_0^{T|X}(t|x) + \sum_{k=1}^{s-2} \sum_{j=J'(k)}^{J(k)} \sum_{v=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} \kappa_{kjvi} \varphi_{skj}(x) \phi_{sk}(x) \psi_{skvi}(t), \tag{A.6}$$

where $S_0^{T|X}$ is introduced in Assumption 2.2, $\psi_{skvi}(t) := (s/r)^{1/2}\psi_i(st/r - v)$, and $\psi_i(t) = 2^{1/2}\sin(\pi i t) I(t \in [0,1])$. Denote by $S^*(x,t)$, $(x,t) \in \mathcal{R}_r$ the sum on the right side of (A.6). The function S^* is continuous in t due to the used sine bases on the subintervals of $[0,r]$, and it is also differentiable in $t \in [0,r]$ with the exception of the end points of the subintervals where S^* is equal to zero. Now note that if $|\kappa_{kjvi}| \leqslant n^{-1/3}/s^5$ (and compare this bound with τ_{skj} being of order $n^{-1/2}$), then for all inner points of the subintervals $\max_{x\in[0,1]} |\partial S^*(x,t)/\partial t| = o_n(1)$. Combining these properties, we conclude that according to Assumption 2.2 the function $S_*^{T|X}(t|x)$ is a bona fide survival function for all large n. Further, under this assumption we get a bona fide regression

$$m_*(x) = m_0(x) + \sum_{k=1}^{s-2} \sum_{j=J'(k)}^{J(k)} \Big[\sum_{v=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} b_{skvi} \kappa_{kjvi} \Big] \varphi_{skj}(x) \phi_{sk}(x). \tag{A.7}$$

Here $b_{skvi} := \int_0^r \psi_{skvi}(t)dt$, and then the Parseval identity implies that

$$\sum_{i=1}^{\infty} b_{skvi}^2 = r/s. \tag{A.8}$$

Now we can compare terms in the square brackets on the right side of (A.7) with ν_{skj} in (A.1). Recall that we used a zero-mean normal prior for ν_{skj} with standard deviation $\tau_{skj} := [n^{-1}(1-3\rho^{-1}) I_{sk}^{-1} \max(\rho^{-1}, \min(\rho, (J(k)/j)^\alpha - 1))]^{1/2}$, where $I_{sk}^{-1} := \int_0^r [S_0^{T|X}(t|k/s)(1-S_0^{T|X}(t|k/s))/f^{X,Z}(k/s,t)]dt$. Introduce independent normal random variables ζ_{skjvi} with zero mean and variance

$$[S_0^{T|X}(v/s|k/s)(1 - S_0^{T|X}(v/s|k/s))/f^{X,Z}(k/s, v/s)] \times [n^{-1}(1-3\rho^{-1}) \max(\rho^{-1}, \min(\rho, (J(k)/j)^\alpha - 1))]$$

(compare with τ_{skj}^2). Then using (A.8) we get

$$\mathbb{E}\Big\{ \Big(\sum_{v=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} b_{skvi} \zeta_{skjvi} \Big)^2 \Big\} = \tau_{skj}^2 (1 + o_n(1)). \tag{A.9}$$

We conclude that the above-defined more complicated model of generating priors for Fourier coefficients ν_{skj} of the regression function yields asymptotically the same prior as the above-described ζ_{skj} used in [10]. This finishes the first above-outlined step.

The second step is to calculate Fisher matrices $\mathcal{I}_{skj}$ for vector-parameter $\vec{\kappa}_{skj} := \{\kappa_{kjvi}, v = 1, \ldots, s-2, i = 1, \ldots, \lceil \ln(s) \rceil\}$. For two particular pairs of indexes (v_1, i_1) and (v_2, i_2) the corresponding element $\mathcal{I}_{skj}(v_1, i_1, v_2, i_2)$ of the Fisher matrix is

$$\mathcal{I}_{skj}(v_1, i_1, v_2, i_2) := \mathbb{E}\Big\{ \prod_{u=1}^{2} [\partial \ln(f^{X,Z,\Delta}(X, Z, \Delta)/\partial \kappa_{kjv_u i_u}] \Big\}. \tag{A.10}$$

To simplify notation, set $\theta_u := \kappa_{kjv_u i_u}$, $u = 1, 2$. Also recall that $\phi_{sk}(x)$ and $\psi_{skvi}(t)$ are supported on $(k/s, (k+1)/s)$ and $(rv/s, r(v+1)/s)$, respectively. This allows us to write for the considered indexes,

$$\mathcal{I}_{skj}(v_1, i_1, v_2, i_2) = \mathbb{E}\Big\{(1-\Delta) \prod_{u=1}^{2} \Big[[\partial S^{T|X}(Z|X)/\partial \theta_u]/[S^{T|X}(Z|X) \Big] \Big\}$$

$$+\mathbb{E}\Big\{\Delta \prod_{u=1}^{2} \Big[[\partial(1 - S^{T|X}(Z|X))/\partial \theta_u]/[1 - S^{T|X}(Z|X) \Big] \Big\}$$

$$= \int_0^1 \int_0^r [f^{X,Z}(x,t)(\frac{1}{S^{T|X}(t|x)} + \frac{1}{1 - S^{T|X}(t|x)})]$$

$$\times \psi_{skv_1 i_1}(t) \psi_{skv_2 i_2}(t) \varphi^2_{skj}(x) \phi^2_{sk}(x) dt dx. \tag{A.11}$$

To evaluate the right side of the last equality we use Assumptions 2.1-2.3 and relations

$$\int_{k/s}^{(k+1)/s} [\varphi_{skj}(x)\phi_{sk}(x)]^2 dx = 1 + \int_{k/s}^{(k+1)/s} \varphi^2_{skj}(x)(\phi^2_{sk}(x) - 1)dx,$$

$$|\int_{k/s}^{(k+1)/s} \varphi^2_{skj}(x)(\phi^2_{sk}(x) - 1)dx| = o_n(1)s^{-1},$$

and

$$|\int_{k/s}^{(k+1)/s} \varphi_{skj}(x)\phi_{sk}(x)dx| = |\int_{k/s}^{(k+1)/s} \varphi_{skj}(x) \times [\phi_{sk}(x) - 1]dx| = o_n(1)s^{-1}.$$

Further, we have $\int_0^r w(t)\psi_{skjv_1 i_1}(t)\psi_{skjv_2 i_2}(t)dt = 0$ for any integrable function $w(t)$ whenever $v_1 \neq v_2$ (note that in this case supports of the two basis functions are disjoint), and the Cauchy-Schwarz inequality yields $\int_0^r |\psi_{skjvi}(t)|dt \leqslant Cs^{-1/2}$. We may conclude that the above-defined Fisher matrix $\mathcal{I}_{skj}$ is a block-diagonal matrix and $\mathcal{I}_{skj} = \mathrm{diag}(A_1, \ldots, A_{s-2})$ where each block A_v is a $\lceil \ln(s) \rceil \times \lceil \ln(s) \rceil$ matrix with diagonal elements

$$A_v(i_1, i_1) = \Big[S_0^{T|X}(rv/s|k/s)$$

$$\times (1 - S_0^{T|X}(rv/s|k/s))/f^{X,Z}(k/s, rv/s)\Big]^{-1}(1 + o_n(1)), \quad \text{(A.12)}$$

where $|o_n(1)| < C/s$ for all considered parameters. Further, absolute values of all other elements in a block-matrix are bounded by Cs^{-1}. Now recall that $s := s_n \to \infty$ as $n \to \infty$ and that the inverse of a block-diagonal matrix is again a block-diagonal matrix created by corresponding inverses of the blocks. This and (A.8) allow us to conclude that the inverse Fisher matrix $\mathcal{I}_{skj}^{-1}$ satisfies for the vector-row $\vec{b}_{sk} := (b_{sk11}, \dots, b_{sk(s-2)\lceil \ln(s)\rceil})$ and its transpose $\vec{b}'_{sk}$ the relation

$$\vec{b}_{sk}\mathcal{I}_{skj}^{-1}\vec{b}'_{sk} = \int_0^r [S_0^{T|X}(t|k/s)(1 - S_0^{T|X}(t|k/s))/f^{X,Z}(k/s, t)]dt(1 + o_n(1))$$

$$= I_{sk}^{-1}(1 + o_n(1)). \quad \text{(A.13)}$$

Now we are ready to straightforwardly follow the proof of Theorem 1 in [10] and conclude that

$$\inf \sup_{\vec{\nu}_{sk}\in\Theta_{sk}} \sum_{j=J'(k)}^{J(k)} \mathbb{E}\{(\tilde{\nu}^*_{skj} - \nu_{skj})^2\} \geqslant (nI_{sk})^{-2\alpha/(2\alpha+1)}P_*(1 + o_n(1)), \quad \text{(A.14)}$$

where the infimum is over all possible oracle-estimators of $\vec{\nu}_{sk}$ considered in Theorem 1 and $P_* := (\alpha/\pi(\alpha+1))^{2\alpha/(2\alpha+1)}(2\alpha+1)^{1/2\alpha+1)}$. Using (A.14) on the right side of (A.5) we can write,

$$\inf_{\tilde{\mu}^*_r} \sup_{\mu_r\in\mathcal{M}_n(\mu_{r0},\alpha,Q,r)} \mathbb{E}\{\int_0^1 (\tilde{\mu}^*_r(x) - \mu_r(x))^2 dx\}$$

$$\geqslant [s^{-1}\sum_{k=1}^{s-2} I_{sk}^{-1}]^{2\alpha/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)}\ P_* Q^{1/(2\alpha+1)}(1 + o_n(1))$$

$$= \Big[\sum_{k=1}^{s-2}\int_{k/s}^{(k+1)/s}\int_0^r [S_0^{T|X}(t|x)(1 - S_0^{T|X}(t|x))/f^{X,Z}(x, t)]dtdx\Big]^{2\alpha/(2\alpha+1)}$$

$$\times n^{-2\alpha/(2\alpha+1)}P_* Q^{1/(2\alpha+1)}(1 + o_n(1))$$

$$= \Big(\int_0^1\int_0^r [S_0^{T|X}(t|x)(1 - S_0^{T|X}(t|x))/f^{X,Z}(x, t)]dtdx\Big)^{2\alpha/(2\alpha+1)}$$

$$\times\, n^{-2\alpha/(2\alpha+1)}P(\alpha, Q)(1 + o_n(1)). \quad \text{(A.15)}$$

This proves the lower bound (2.6).

Now note that in the proof proof of (2.6) we considered only conditional survival functions $S^{T|X}(t|x) = S_0^{T|X}(t|x)$, $(x, t) \in [0, 1] \times [r, \infty)$. Thus that proof also verifies (2.7).

The assumed in part (ii) of Theorem 2.1 inequality $d(S^{T|X}, f^{X,Z}, \infty) < \infty$ yields that $d(S^{T|X}, f^{X,Z}, r) \to d(S^{T|X}, f^{X,Z}, \infty)$ as $r \to \infty$. This and (2.6) verify (2.9). Theorem 2.1 is proved.

Proof of Theorem 2.2. The proof follows along line of the proof of Theorem 2.1 with the following two changes. The new $I_{sk}^{-1} := \int_0^\tau [S_0^{T|X}(t|k/s)/f^{X,Z}(k/s,t)]dt$ are used. Then we repeat the already made calculations in (A.10)–(A.13) only now, due to the MCSC, in place of the sum $[S^{T|X}(t|x)]^{-1} + [1 - S^{T|X}(t|x)]^{-1}$ in (A.11) we have only the first term $[S^{T|X}(t|x)]^{-1}$. This is how the new I_{sk} is calculated in (A.13). Theorem 2.2 is proved.

Proof of Lemma 3.2. Recall that C denotes a generic positive constant. Using the Parseval identity we can write for all sufficiently large n,

$$\mathbb{E}\Big\{\int_0^1 (\tilde{g}(x) - g(x))^2\Big\}$$

$$= \sum_{j=0}^{q_n} \mathbb{E}\{(\tilde{\kappa}_j - \kappa_j)^2\} + \sum_{j=q_n+1}^{J_n} \mathbb{E}\{(1-(j/J_n)^\alpha)\tilde{\kappa}_j - \kappa_j)^2\} + \sum_{j>J_n} \kappa_j^2$$

$$\leqslant Cq_n n^{-1} + \sum_{j=q_n+1}^{J_n} \mathbb{E}\{[(1-(j/J_n)^\alpha)(\tilde{\kappa}_j - \kappa_j) - (j/J_n)^\alpha \kappa_j]^2] + \sum_{j>J_n} \kappa_j^2. \quad \text{(A.16)}$$

Using (3.4) we continue (A.16),

$$\mathbb{E}\Big\{\int_0^1 (\tilde{g}(x) - g(x))^2\Big\} \leqslant Cq_n n^{-1}$$

$$+ n^{-1} d(g) \sum_{j=q_n+1}^{J_n} (1-(j/J_n)^\alpha)^2 (1+o_n(1)) + (\pi J_n)^{-2\alpha} \sum_{j>q_n} (\pi j)^{2\alpha}\kappa_j^2. \quad \text{(A.17)}$$

Now we are evaluation the two sums on the right side of (A.17). A direct calculation yields

$$\sum_{j=q_n+1}^{J_n} (1-(j/J_n)^\alpha)^2 = \frac{2\alpha^2}{(\alpha+1)(2\alpha+1)} J_n(1+o_n(1)). \quad \text{(A.18)}$$

Definition (2.5) of the global Sobolev class yields

$$\sup_{g\in\mathcal{S}(\alpha,Q)} \sum_{j>q_n} (\pi j)^{2\alpha}\kappa_j^2 \leqslant Q. \quad \text{(A.19)}$$

Using (A.18), (A.19) and the definition of J_n proves part (i) of Lemma 3.2.

Now let us verify part (ii) of Lemma 3.2. Denote by κ_{0j} and κ_{*j} Fourier coefficients of g_0 and g_*, respectively. Now note that the underlying g affects only the second sum on the right side of (A.17). Let us evaluate it for the considered $g = g_0 + g_*$. Write for some $\gamma > 0$,

$$\sum_{j>q_n} (\pi j)^{2\alpha}\kappa_j^2 \leqslant \sum_{j>q_n} (\pi j)^{2\alpha}[(1+\gamma)\kappa_{*j}^2 + (1+\gamma^{-1})\kappa_{0j}^2]$$

$$\leqslant (1+\gamma)Q + (1+\gamma^{-1}) \sum_{j>q_n} (\pi j)^{2(\alpha-\alpha')}(\pi j)^{2\alpha'}\kappa_{0j}^2$$

$$\leqslant (1+\gamma)Q + (1+\gamma^{-1})q_n^{2(\alpha-\alpha')}Q'.$$

Set $\gamma := q_n^{\alpha-\alpha'}$ and get

$$\sup_{g_* \in \mathcal{S}(\alpha,Q)} \sum_{j>q_n} (\pi j)^{2\alpha}\kappa_j^2 \leqslant Q(1+o_n(1)). \tag{A.20}$$

Using (A.20) in place of (A.19) in (A.17) verifies part (ii) of Lemma 3.2. Lemma 3.2 is proved.

Proof of Theorem 3.1. Recall that the lifetime of interest T may be unbounded, the sample is MCSC and the joint design density $f^{X,Z}$ is known. The proof is based on using Lemma 3.2.

We begin with analysis of the proposed Fourier estimator

$$\tilde{\theta}_j := n^{-1}\sum_{l=1}^{n} \frac{\Delta'_l I(\Delta'_l Z_l \leqslant r)\varphi_j(\Delta'_l X_l)}{f^{X,Z}(\Delta'_l X_l, \Delta'_l Z_l)}.$$

Using $f^{X,Z,\Delta}(x,t,0) = f^{X,Z}(x,t)S^{T|X}(t|x)$ we get

$$\mathbb{E}\{\tilde{\theta}_j\} = \mathbb{E}\Big\{\frac{\Delta' I(\Delta' Z \leqslant r)\varphi_j(\Delta' X)}{f^{X,Z}(\Delta' X, \Delta' Z)}\Big\}$$

$$= \int_0^1 [\int_0^r S^{T|X}(t|x)dt]\varphi_j(x)dx = \int_0^1 m(x)\varphi_j(x)dx = \theta_j. \tag{A.21}$$

We conclude that the Fourier estimator is unbiased. For its variance we may write,

$$\mathbb{E}\{(\tilde{\theta}_j - \theta_j)^2\} = n^{-1}[\mathbb{E}\Big\{\Big[\frac{\Delta' I(\Delta' Z \leqslant r)\varphi_j(\Delta' X)}{f^{X,Z}(\Delta' X, \Delta' Z)}\Big]^2\Big\} - \theta_j^2].$$

To evaluate the last expectation we use $\varphi_j^2(x) = 1 + \varphi_{2j}(x)$, $j \geqslant 1$ and get

$$\mathbb{E}\Big\{\Big[\frac{\Delta' I(\Delta' Z \leqslant r)\varphi_j(\Delta' X)}{f^{X,Z}(\Delta' X, \Delta' Z)}\Big]^2\Big\}$$

$$= \int_{R_r} \frac{S^{T|X}(t|x)\varphi_j^2(x)}{f^{X,Z}(x,t)}dtdx = d_*(S^{T|X}, f^{X,Z}, r) + o_j(1).$$

Further, we have $\theta_j^2 \to 0$, $j \to \infty$.

Now we are ready to evaluate MISE of the blockwise-shrinkage estimator (3.24). Ee begin with analysis of the oracle-estimator

$$\hat{m}_r^*(x) := \sum_{j=0}^{q_n} \tilde{\theta}_j\varphi_j(x) + \sum_{k=1}^{k_n} \frac{\Theta_k}{\Theta_k + d_* n^{-1}} \sum_{j \in B_k} \tilde{\theta}_j\varphi_j(x) \tag{A.22}$$

is sharp-minimax, that is its MISE attains the lower bound of Theorem 2.2. In (A.22) we used notation $\Theta_k := L_k^{-1}\sum_{j\in B_k}\theta_j^2$ for Sobolev functionals. Note that the studied adaptive estimator (3.24) mimics the oracle. To make formulae shorter, set $\Lambda_k := \Theta_k/(\Theta_k + d_* n^{-1})$ and write,

$$\sum_{j\in B_k}\mathbb{E}\{(\Lambda_k\widetilde{\theta}_j - \theta_j)^2\}$$

$$= \sum_{j\in B_k}[\Lambda_k^2\mathbb{E}\{(\widetilde{\theta}_j - \theta_j)^2\} + (1-\Lambda_k)^2\theta_j^2 - 2\theta_j\Lambda_k(1-\Lambda_k)\mathbb{E}\{(\widehat{\theta}_j - \theta_j)\}]. \quad \text{(A.23)}$$

To simplify further references on the proof, note that the Fourier estimator $\widetilde{\theta}_j$ satisfies the inequalities

$$|\mathbb{E}\{\widetilde{\theta}_j - \theta_j\}| \leqslant Cn^{-1}q_n^2, \quad \mathbb{E}\{(\widetilde{\theta}_j - \theta_j)^2\} \leqslant n^{-1}[d_* + o_i(1) + o_n(1)]. \quad \text{(A.24)}$$

These inequalities will be sufficient.

Using (A.24) and Cauchy inequality $|2\theta_j\Lambda_r(1-\Lambda_r)| \leqslant n^{-1/2}\Lambda_k^2 + n^{1/2}\theta_j^2(1-\Lambda_k)^2$, we continue (A.23)

$$\sum_{j\in B_k}\mathbb{E}\{(\Lambda_k\widetilde{\theta}_j - \theta_j)^2\}$$

$$\leqslant \Lambda_k^2 L_k n^{-1}d_* + (1-\Lambda_k)^2 L_k\Theta_k + \sum_{j\in B_k}[\Lambda_k^2 n^{-1} + (1-\Lambda_k)^2\theta_j^2](o_n(1) + o_j(1))$$

$$= [\Lambda_k^2 L_k n^{-1}d_* + (1-\Lambda_k)^2 L_k\Theta_k](1 + o_n(1) + o_j(1)).$$

Using definition of Λ_k we continue,

$$\sum_{j\in B_k}\mathbb{E}\{(\Lambda_k\widetilde{\theta}_j - \theta_j)^2\}$$

$$\leqslant L_k\Big[\frac{\Theta_k^2 n^{-1}d_*}{(\Theta_k + d_* n^{-1})^2} + \frac{d_*^2 n^{-2}\Theta_k}{(\Theta_k + d_* n^{-1})^2}\Big][1 + o_n(1) + o_k(1)]$$

$$= L_k\frac{\Theta_k n^{-1}d_*}{\Theta_k + d_* n^{-1}}[1 + o_n(1) + o_k(1)]. \quad \text{(A.25)}$$

Now we may return to the MISE of the oracle-estimator and write using the Parseval identity and (A.24),

$$\mathbb{E}\{\int_0^1(\widehat{m}_r^*(x) - m_r(x))^2dx\}$$

$$= \sum_{j=0}^{b_n}\mathbb{E}\{(\widetilde{\theta}_j - \theta_j)^2\} + \sum_{k=1}^{k_n}\mathbb{E}\{(\Lambda_k\widetilde{\theta}_j - \theta_j)^2\} + \sum_{k>k_n}L_k\Theta_k$$

$$\leqslant Cb_n n^{-1} + \sum_{k=1}^{k_n}L_k\frac{\Theta_r n^{-1}d_*}{\Theta_r + d_* n^{-1}}[1 + o_n(1) + o_k(1)] + \sum_{k>k_n}L_r\Theta_r. \quad \text{(A.26)}$$

Next, a direct calculation shows that

$$\sup_{m_r in \mathcal{S}(\alpha,Q)} \Big[\sum_{k=1}^{k_n} L_k \frac{\Theta_k n^{-1} d_*}{\Theta_k + d_* n^{-1}} [1 + o_n(1) + o_k(1)] + \sum_{k>k_n} L_k \Theta_k \Big]$$

$$\leqslant P(\alpha, Q)(d_* n^{-1})^{2\alpha/(2\alpha+1)}(1 + o_n(1)). \tag{A.27}$$

Using the proof of Lemma 3.2, the same upper bound holds for the supremum over $\mathcal{F}_n(S_0^{T|X}, \alpha, Q, r)$. This verifies that the oracle-estimator is sharp-minimax over the local and global Sobolev classes.

Note that the proposed estimator (3.24) mimics the oracle-estimator's smoothing coefficients $\Theta_k/(\Theta_k + d_* n^{-1})$ by statistic

$$\min(1, \widetilde{\Theta}_k/\widetilde{\Theta}'_k)I(\widetilde{\Theta}_k > 1/[n \ln(k + 3)).$$

This mimicking is well-known in the literature and the proof that it preserves the verified sharp-minimaxity of the oracle-estimator may be found in [12]. Theorem 3.1 is verified.

Proof of Theorem 3.2. The proposed Fourier estimator (3.27) is motivated by the oracle-estimator

$$\hat{\theta}_j^* := n^{-1} \sum_{l=1}^{n} \frac{[\Delta'_l - S(j, n, Z_l, X_l)]I(Z_l \leqslant r)\varphi_j(X_l)}{f^{X,Z}(X_l, Z_l)}, \tag{A.28}$$

where (recall our notation $\mathcal{J}(j, n) := \{0, 1, \dots, q_n\}\backslash\{j\}$)

$$S(j, n, z, x)$$

$$:= \sum_{k\in\mathcal{J}(j,n)} \sum_{i=0}^{q_n} \Big[\int_0^1 \int_0^r S^{T|X}(z|x)\psi_i(u)\varphi_k(v)dudv \Big] \psi_i(z)\varphi_k(x). \tag{A.29}$$

Note that $S(j, n, z, x)$ is a special Fourier series approximation of $S^{T|X}(z|x)$ with deleted jth Fourier component.

Recall that $\{1, \varphi_j(x), j = 1, 2, \dots\}$ is the orthonormal basis on $[0, 1]$ and write,

$$\mathbb{E}\{\hat{\theta}_j^*\} = \theta_j - \int_0^1 \int_0^r S(j, n, z, x)\varphi_j(x)dzdx = \theta_j - 0 = \theta_j. \tag{A.30}$$

We conclude that the oracle's Fourier estimator is unbiased.

Next we evaluate the variance of $\hat{\theta}_j^*$. Write,

$$n\mathbb{E}\{(\hat{\theta}_j^* - \theta_j)^2\} = \mathbb{E}\Big\{\Big[\frac{[\Delta' - S(j, n, Z, X)]I(Z \leqslant r)\varphi_j(X)}{f^{X,Z}(X, Z)}\Big]^2\Big\} - \theta_j^2. \tag{A.31}$$

Consider the expectation on the right side of (A.31). Note that $(\Delta')^2 = \Delta'$ and write,

$$\mathbb{E}\Big\{\Big[\frac{[\Delta' - S(j, n, Z, X)]I(Z \leqslant r)\varphi_j(X)}{f^{X,Z}(X, Z)}\Big]^2\Big\}$$

$$= \mathbb{E}\Big\{\frac{[\Delta' - 2\Delta' S(j,n,Z,X) + S^2(j,n,Z,X)]I(Z \leqslant r)\varphi_j^2(X)}{[f^{X,Z}(X,Z)]^2}\Big\}$$

$$= \int_0^1 \Big[\int_0^r \frac{S^{T|X}(z|x) - 2S^{T|X}(z|x)S(j,n,z,x) + S^2(j,n,z,x)}{f^{X,Z}(x,z)}dz\Big]\varphi_j^2(x)dx$$

$$= \int_0^1 \Big[\int_0^r \frac{S^{T|X}(z|x) - (S^{T|X}(z|x))^2}{f^{X,Z}(x,z)}dz\Big]\varphi_j^2(x)dx$$

$$+ \int_0^1 \Big[\int_0^r \frac{[2S^{T|X}(z|x)(S^{T|X}(z|x) - S(j,n,z,x))}{f^{X,Z}(x,z)}dz\Big]\varphi_j^2(x)dx$$

$$+ \int_0^1 \Big[\int_0^r \frac{S^2(j,n,z,x) - (S^{T|X}(z|x))^2}{f^{X,Z}(x,z)}dz\Big]\varphi_j^2(x)dx. \qquad \text{(A.32)}$$

The first integral on the right side of (A.32) can be evaluated as

$$\int_0^1 \Big[\int_0^r \frac{S^{T|X}(z|x)[1 - (S^{T|X}(z|x)]}{f^{X,Z}(x,z)}dz\Big]\varphi_j^2(x)dx = d + o_j(1). \qquad \text{(A.33)}$$

Here $d = d(S^{T|X}, f^{X,Z}, r)$. The second and third integrals, using definition (A.29), can be evaluated as $o_n(1)$. We conclude that

$$\mathbb{E}\{(\widehat{\theta}_j^* - \theta_j)^2\} = n^{-1}[d + o_j(1) + o_n(1)]. \qquad \text{(A.34)}$$

Now note that the proposed Fourier estimator $\widehat{\theta}_j$ mimics the oracle's Fourier estimator by replacing unknown function $S(j,n,z,x)$ by its estimate

$$\widetilde{S}(j,n,z,x) := n^{-1}\sum_{l=1}^{n}\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta_l'\psi_i(Z_l)\varphi_k(X_l)\psi_i(z)\varphi_k(x)}{f^{X,Z}(X_l,Z_l)}.$$

Further, we may write,

$$\widehat{\theta}_j = \widehat{\theta}_j^* + n^{-1}\sum_{l=1}^{n}\frac{S(j,n,Z_l,X_l) - \widetilde{S}(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l) =: \widehat{\theta}_j^* + \widetilde{A}_j. \qquad \text{(A.35)}$$

Now we evaluate the mean and variance of $\widetilde{A}_j$. Set $\mathcal{N}(l) := \{1,2,\dots,n\}\backslash\{l\}$ and write

$$\mathbb{E}\{\widetilde{A}_j\} = \mathbb{E}\Big\{\frac{(S(j,n,Z_1,X_1) - \widetilde{S}(j,n,Z_1,X_1))\varphi_j(X_1)}{f^{X,Z}(X_1,Z_1)}\Big\}$$

$$= -\mathbb{E}\Big\{\frac{\sum_{l\in\mathcal{N}(1)}\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta_l'\psi_i(Z_l)\varphi_k(X_l)\psi_i(Z_1)\varphi_k(X_1)}{f^{X,Z}(X_l,Z_l)} - S(j,n,Z_1,X_1)}{nf^{X,Y}(X_1,Z_1)}$$

$$\times \varphi_j(X_1)\Big\} - n^{-1}\mathbb{E}\Big\{\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta_1'\psi_i^2(Z_1)\varphi_k^2(X_1)}{[f^{X,Z}(X_1,Z_1)]^2}\varphi_j(X_1)\Big\}. \qquad \text{(A.36)}$$

The first expectation on the right side of (A.36) is zero because

$$\mathbb{E}\{\varphi_j(X_1)[f^{X,Z}(X_1, Z_1)]^{-1}\} = 0,$$

and the second is of order $n^{-1}q_n^2$. We conclude that

$$|\mathbb{E}\{\widetilde{A}_j\}| \leqslant Cn^{-1}q_n^2. \tag{A.37}$$

Next we evaluate the second moment of $\widetilde{A}_j$. Write,

$$\mathbb{E}\{\widetilde{A}_j^2\} = n^{-2}\sum_{l,u=1}^{n}\mathbb{E}\Big\{\frac{\widetilde{S}(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)$$

$$\times\frac{\widetilde{S}(j,n,Z_u,X_u) - S(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}. \tag{A.38}$$

There are two types of terms in the double sum. The first one is when $u = l$ and the second when $u \neq l$. We explore them in turn. For the case $u = l$ we note that

$$\mathbb{E}\Big\{\Big[\frac{\widetilde{S}(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)\Big]^2\Big\} \leqslant Cn^{-1}q_n^2. \tag{A.39}$$

If $u \neq l$, set $\mathcal{N}(u,l) := \{1,2,\dots,n\}\backslash\{u,l\}$ and write,

$$\widetilde{S}(j,n,z,x) = n^{-1}\sum_{s\in\mathcal{N}(u,l)}\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta'_s\psi_i(Z_s)\varphi_k(X_s)\psi_i(z)\varphi_k(x)}{f^{X,Z}(X_s,Z_s)}$$

$$+n^{-1}\sum_{s\in\{u,l\}}\sum_{k\in\mathcal{J}(j,n)}\sum_{i=0}^{q_n}\frac{\Delta'_s\psi_i(Z_s)\varphi_k(X_s)\psi_i(z)\varphi_k(x)}{f^{X,Z}(X_s,Z_s)}$$

$$=: \widetilde{S}_1(j,n,z,x) + \widetilde{S}_2(j,n,z,x). \tag{A.40}$$

Using the new notation we can write (recall that we are considering $r \neq l$)

$$\mathbb{E}\Big\{\frac{\widetilde{S}(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)$$

$$\times\frac{\widetilde{S}(j,n,Z_u,X_u) - S(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}$$

$$= \mathbb{E}\Big\{\frac{[\widetilde{S}_1(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)] + \widetilde{S}_2(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)$$

$$\times\frac{[\widetilde{S}_1(j,n,Z_u,X_u) - S(j,n,Z_u,X_u)] + \widetilde{S}_2(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}$$

$$= \mathbb{E}\Big\{\frac{\widetilde{S}_1(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)$$

$$\times \frac{\widetilde{S}_1(j,n,Z_u,X_u) - S(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}$$

$$+2\mathbb{E}\Big\{\frac{\widetilde{S}_1(j,n,Z_l,X_l) - S(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)\frac{\widetilde{S}_2(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}$$

$$+\mathbb{E}\Big\{\frac{\widetilde{S}_2(j,n,Z_l,X_l)}{f^{X,Z}(X_l,Z_l)}\varphi_j(X_l)\frac{\widetilde{S}_2(j,n,Z_u,X_u)}{f^{X,Z}(X_u,Z_u)}\varphi_j(X_u)\Big\}$$

$$=: B_1 + B_2 + B_3. \tag{A.41}$$

Term B_1 on the right side of (A.41) is zero because $\int_0^1 \varphi_j(x)dx = 0$. Using (A.39), $\mathbb{E}\{[S_2(j,n,X_l,Z_j)]^2\} \leqslant Cn^{-2}q_n^4$ and Cauchy-Schwarz inequality we get $|B_2| \leqslant Cn^{-3/2}q_n^2$. Finally, $|B_3| \leqslant Cn^{-2}q_n^4$. Combining these results in (A.41), and then using the obtained relation and (A.39) in (A.38) we conclude that

$$\mathbb{E}\{\widetilde{A}_j^2\} \leqslant Cn^{-3/2}q_n^4. \tag{A.42}$$

Now we can combine the obtained results and conclude that

$$|\mathbb{E}\{\widehat{\theta}_j\} - \theta_j| \leqslant Cn^{-1}q_n^2, \quad \mathbb{E}\{(\widehat{\theta}_j - \theta_j)^2\} = n^{-1}[d + o_j(1) + o_n(1)]. \tag{A.43}$$

Properties (A.43) of the Fourier coefficient estimator $\widehat{\theta}_j$ are the same as (A.24) used in the proof of efficiency of the blockwise shrinkage estimator, recall the proof of Theorem 3.1. This finishes the proof of Theorem 3.2.

Remark A.1 It follows from the established properties of the Fourier estimator $\widehat{\theta}_j$ that it matches performance of the oracle's Fourier estimator $\widehat{\theta}_j^*$ with sufficient accuracy for the efficient nonparametric estimation.

Proof of Theorem 3.3 Fourier estimator (3.33) is the sample mean estimator. Accordingly, we get $\mathbb{E}\{\bar{\theta}_j\} = \theta_j$ and $\mathbb{E}\{(\bar{\theta}_j - \theta_j)^2|\] = D_*(1 + o_j(1))$. These relations allow us to use Lemma 3.2, and it proves Theorem 3.3.

Proof of Theorem 3.4. The proof follow the steps of the proof of Theorem 3.1. First, it follows from Theorem 3.3 that the smoothing oracle (3.34) is efficient for MCSC and within factor $(D_*/D)^{2\alpha/(2\alpha+2)}$ from being efficient for CSC. Second, introduce the blockwise oracle-estimator

$$\bar{m}_*(x) := \sum_{j=0}^{q_n} \bar{\theta}_j\varphi_j(x) + \sum_{k=1}^{k_n} \frac{\Theta_k}{\Theta_k + d_* n^{-1}} \sum_{j\in B_k} \bar{\theta}_j\varphi_j(x). \tag{A.44}$$

Here $\Theta_k := L_k^{-1}\sum_{j\in B_k}\theta_j^2$ are Sobolev functionals. Note that the studied adaptive estimator (3.38) mimics the oracle. Directly following the proof of Theorem 3.1 we establish that the blockwise oracle estimator $\bar{m}_*$ is efficient for MCSc and that (3.38) is also efficient. Theorem 3.4 is verified.

Proof of Theorem 4.1. We are considering MCSC and CSC samplings in turn. For MCSC there is an extra sample of size n^* (we are considering a more general setting than just sample size n) from (X, Z) that is used to estimate the joint density $f^{X,Z}$. Due to the extra sample, we can separate observations used to estimate the nuisance function and the regression, and at the same time prove a number of needed technical propositions. Then we explore the CSC when the same observations are used to estimate the nuisance joint density and the regression.

For MCSC the Fourier estimator is

$$\tilde{\theta}_j := n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l I(\Delta_l Z_l \leqslant r)\varphi_j(\Delta'_l X_l)}{\hat{f}(\Delta'_l X_l, \Delta'_l Z_l)} = n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l I(Z_l \leqslant r)\varphi_j(X_l)}{\hat{f}(X_l, Z_l)}. \quad (A.45)$$

Note that we skipped some factors Δ'_l to simplify the formula. In (A.45)

$$\hat{f}(x,z) := \max(1/\ln\ln(n_* + 3), \tilde{f}(x,z)) \quad (A.46)$$

and $\tilde{f}(x,z)$ is the projection density estimate

$$\tilde{f}(x,z) := n_*^{-1} \sum_{r=1}^{n_*} \sum_{(i,s)\in\mathcal{N}} \varphi_i(X_{Er})\psi_s(Z_{Er})\varphi_i(x)\psi_s(z). \quad (A.47)$$

In (A.47) $(X_{E1}, Z_{E1}), \ldots, (X_{En_*}, Z_{En_*})$ is the extra sample from (X,Z), $\mathcal{N} := \{0, 1, \ldots, \lfloor 1 + n_*^{1/4} \rfloor\}^2$, $\{\varphi_j\}$ and $\{\psi_s\}$ are cosine bases on $[0,1]$ and $[0,r]$, respectively. The truncation from below in (A.46) allows us to use $\hat{f}$ in the denominators of (A.45).

Recall that the nuisance joint density $f^{X,Z}(x,z)$ is supported and bounded below from zero on $\mathcal{R}_r := [0,1] \times [0,r]$, and in what follows we are considering only $(x,z) \in \mathcal{R}_r$.

The following elementary relation will be useful (to simplify formulae in what follows we may write $f(x,z) := f^{X,Z}(x,z)$)

$$\frac{1}{\hat{f}(x,z)} = \frac{1}{f(x,z)} + \frac{f(x,z) - \hat{f}(x,z)}{f^2(x,z)} + \frac{(f(x,z) - \hat{f}(x,z))^2}{\hat{f}(x,z)f^2(x,z)}. \quad (A.48)$$

Using this relation we can rewrite (A.45) as

$$\begin{aligned} \tilde{\theta}_j = n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l \varphi_j(X_l)}{f(X_l, Z_l)} + n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l \varphi_j(X_l)(f(X_l, Z_l) - \hat{f}(X_l, Z_l))}{f^2(X_l, Z_l)} \\ + n^{-1} \sum_{l=1}^{n} \frac{\Delta'_l \varphi_j(X_l)(f(X_l, Z_l) - \hat{f}(X_l, Z_l))^2}{\hat{f}(X_l, Z_l) f^2(X_l, Z_l)} \\ =: \tilde{\theta}_j^* + \tilde{A}_1 + \tilde{A}_2. \quad (A.49) \end{aligned}$$

Note that $\tilde{\theta}_j^*$ is the Fourier estimator (3.21) used in Theorem 3.1, and hence we only need to explore terms $\tilde{A}_1$ and $\tilde{A}_2$. We begin with the following technical result.

Lemma A.1. *Suppose that a bivariate function* $g(x, z)$, $(x, z) \in \mathcal{R}_\tau$ *has a continuous mixed derivative* $\partial^2 f(x, z)/\partial x \partial z$. *Set*

$$\gamma_{is} := \int_0^r \int_0^1 g(x, z)\varphi_i(x)\psi_s(z)dxdz$$

for Fourier coefficients of the function. If

$$\int_0^r \int_0^1 [\partial^2 g(x, z)/\partial x \partial z]^2 dxdz < \infty, \tag{A.50}$$

then

$$\sum_{i,s=0}^{\infty} (is)^2 \gamma_{is}^2 = (r^2/\pi^4) \int_0^r \int_0^1 [\partial^2 g(x, z)/\partial x \partial z]^2 dxdz. \tag{A.51}$$

If

$$\int_0^r \int_0^1 [\partial g(x, z)/\partial x]^2 dxdz < \infty, \tag{A.52}$$

then

$$\sum_{i,s=0}^{\infty} i^2 \gamma_{is}^2 = \pi^{-2} \int_0^r \int_0^1 [\partial g(x, z)/\partial x]^2 dxdz. \tag{A.53}$$

If

$$\int_0^r \int_0^1 [\partial g(x, z)/\partial z]^2 dxdz < \infty, \tag{A.54}$$

then

$$\sum_{i,s=0}^{\infty} s^2 \gamma_{is}^2 = (r^2/\pi^2) \int_0^a \int_0^1 [\partial g(x, z)/\partial z]^2 dxdz. \tag{A.55}$$

If (A.50), (A.52) and (A.54) hold, then

$$\sum_{i,s=0}^{\infty} (1 + i^2)(1 + s^2)\gamma_{is}^2 = G_g, \tag{A.56}$$

where

$$G_g := \int_0^r \int_0^1 \Big[[g(x, z)]^2 + \pi^{-2}[\partial g(x, z)/\partial x]^2$$
$$+ (r^2/\pi^2)[\partial g(x, z)/\partial z]^2 + (r^2/\pi^4)[\partial^2 g(x, z)/\partial x \partial z]^2 \Big] dxdz, \tag{A.57}$$

and for any pair (i_0, s_0) *of nonnegative integers*

$$\sum_{i \geqslant i_0, s \geqslant s_0} |\gamma_{is}| \leqslant c_*(1 + i_0)^{-1/2}(1 + s_0)^{-1/2} \ln(3 + i_0) \ln(3 + s_0) G_g^{1/2}, \tag{A.58}$$

where c_* *is an absolute finite constant.*

Proof of Lemma A.1. We begin with (A.51). Using the Parseval identity, the sine basis $\{2^{1/2}\sin(\pi jx),\ j=1,2,\ldots\}$ on $[0,1]$, and integration by parts we can write for any $z\in[0,r]$,

$$\int_0^1 [\partial^2 g(x,z)/\partial x\partial z]^2 dx = \sum_{i=1}^\infty \Big[\int_0^1 [\partial^2 g(x,z)/\partial x\partial z] 2^{1/2}\sin(\pi ix)dx\Big]^2$$

$$= \sum_{i=1}^\infty \Big[[\partial g(x,z)/\partial z]2^{1/2}\sin(\pi ix)\Big|_{x=0}^{x=1} - (\pi i)\int_0^1 [\partial g(x,z)/\partial z]2^{1/2}\cos(\pi ix)dx\Big]^2$$

$$= \sum_{i=1}^\infty (\pi i)^2 \Big[\int_0^1 [\partial g(x,z)/\partial z]\varphi_i(x)dx\Big]^2 =: \sum_{i=1}^\infty (\pi i)^2\phi_i^2(z). \tag{A.59}$$

Using the Leibnitz theorem we note that

$$\phi_i(z) = \int_0^1 [\partial g(x,z)/\partial z]\varphi_i(x)dx = \frac{d\int_0^1 g(x,z)\varphi_i(x)dx}{dz} =: \frac{dG_i(z)}{dz}, \tag{A.60}$$

where $G_i(z) := \int_0^1 g(x,z)\varphi_i(x)dx$.

By repeating steps made in (A.59), only now using the sine basis $\{(2/r)^{1/2}\sin(\pi sz/r),\ s=1,2,\ldots\}$ on $[0,r]$, we can write for ϕ_i defined in (A.60),

$$\int_0^r \phi_i^2(z)dz = \sum_{s=1}^\infty \Big[\int_0^r \phi_i(z)(2/r)^{1/2}\sin(\pi sz/r)dz\Big]^2$$

$$= \sum_{s=1}^\infty \Big[G_i(z)(2/r)^{1/2}\sin(\pi sz/r)\Big|_{z=0}^{z=r} - (\pi s/r)\int_0^r G_i(z)(2/r)^{1/2}\cos(\pi sz/r)dz\Big]^2$$

$$= \sum_{s=1}^\infty (\pi s/r)^2\Big[\int_0^r G_i(z)r^{-1/2}\varphi_s(z/r)dz\Big]^2. \tag{A.61}$$

Recall that $G_i(z) := \int_0^1 g(x,z)\varphi_i(x)dx$, and using the Fubini theorem we conclude that

$$\int_0^r\int_0^1 [\partial^2 g(x,z)/\partial x\partial z]^2 dxdz$$

$$= \sum_{i,s=1}^\infty (\pi^2 is/r)^2\Big[\int_0^r [\int_0^1 g(x,z)\varphi_i(x)dx]r^{-1/2}\varphi_s(z/a)dz\Big]^2$$

$$= (\pi^4/r^2)\sum_{i,s=0}^\infty (is)^2\gamma_{is}^2.$$

Equality (A.51) is proved.

Now we verify (A.53). Following (A.59) we can write,

$$\int_0^1 [\partial g(x,z)/\partial x]^2 dx = \sum_{i=1}^\infty \Big[\int_0^1 [\partial g(x,z)/\partial x]2^{1/2}\sin(\pi ix)dx\Big]^2$$

$$= \sum_{i=1}^{\infty} \Big[g(x,z)2^{1/2}\sin(\pi i x)\Big|_{x=0}^{x=1} - (\pi i)\int_0^1 g(x,z)2^{1/2}\cos(\pi i x)dx \Big]^2$$

$$= \sum_{i=1}^{\infty} (\pi i)^2 \Big[\int_0^1 g(x,z)\varphi_i(x)dx \Big]^2. \tag{A.62}$$

Next, using the Parseval identity we get

$$\int_0^r \Big[\int_0^1 g(x,z)\varphi_i(x)dx \Big]^2 dz$$

$$= \sum_{s=0}^{\infty} \Big[\int_0^r \int_0^1 g(x,z)\varphi_i(x) r^{-1/2}\varphi_s(z/r)dxdz \Big]^2 = \sum_{s=0}^{\infty} \gamma_{is}^2. \tag{A.63}$$

Combining (A.62) and (A.63) we may write,

$$\int_0^r \Big[\int_0^1 [\partial g(x,z)/\partial x]^2 dx \Big] dz = \sum_{i=1}^{\infty} (\pi i)^2 \int_0^r \Big[\int_0^1 g(x,z)\varphi_i(x)dx \Big]^2 dz$$

$$= \sum_{i=1}^{\infty} (\pi i)^2 \Big[\sum_{s=0}^{\infty} \gamma_{is}^2 \Big] = \sum_{i,s=0}^{\infty} (\pi i)^2 \gamma_{is}^2.$$

Relation (A.53) is verified. Relation (A.55) is verified similarly to (A.53).

To verify (A.56), note that $(1+i^2)(1+s^2) = 1 + i^2 + s^2 + i^2 s^2$. This, together with the Parseval identity, (A.45), (A.53), and (A.55), verify (A.56).

Now we are verifying (A.52). Using the Cauchy-Schwartz inequality and (A.56) we may write,

$$\Big[\sum_{i \geqslant i_0, s \geqslant s_0} |\gamma_{is}| \Big]^2 \leqslant \Big[\sum_{i \geqslant i_0, s \geqslant s_0} (1+i^2)^{-1/2}(1+s^2)^{-1/2}[\ln(3+i)\ln(3+s)]^{-2} \Big]$$

$$\times \Big[\sum_{i \geqslant i_0, s \geqslant s_0} (1+i^2)^{1/2}(1+s^2)^{1/2}[\ln(3+i)\ln(3+s)]^2 \gamma_{is}^2 \Big] \tag{A.64}$$

Because $\sum_{i \geqslant 2} i^{-1}[\ln(i)]^{-2} < \infty$, the first factor on the right side of (A.64) is bounded from above by a finite absolute constant calculated for the case $i_0 = s_0 = 0$ (recall that To evaluate the second factor, we note that $(1+i^2)^{1/2}\ln^2(3+i_0)/(1+i_0^2)^{1/2} \geqslant \ln^2(3+i)$ for any $i \geqslant i_0 \geqslant 0$, and that $(1+i_0^2)^{-1/2} \leqslant 2^{1/2}(1+i_0)^{-1}$. This and (A.56) allow us to write for the second factor on the right side of (A.64),

$$\sum_{i \geqslant i_0, s \geqslant s_0} (1+i^2)^{1/2}(1+s^2)^{1/2}[\ln(3+i)\ln(3+s)]^2 \gamma_{is}^2$$

$$\leqslant (1+i_0^2)^{-1/2}\ln^2(3+i_0)(1+s_0^2)^{-1/2}\ln^2(3+s_0) \sum_{i \geqslant i_0, s \geqslant s_0} (1+i^2)(1+s^2)\gamma_{is}^2$$

$$\leqslant 2(1+i_0)^{-1}(1+s_0)^{-1}\ln^2(3+i_0)\ln^2(3+s_0)G_g.$$

Combining the obtained results in (A.62) we verify (A.58). Lemma A.1 is proved.

We need one more technical result about the proposed density estimate.

Lemma A.2. *Let density $f^{X,Z}(x,z)$ satisfies the assumption*

$$G_{f^{X,Z}} < \infty, \tag{A.65}$$

where $G_{f^{X,Z}}$ is defined in (A.57). Then, for some positive constants c_0 and c_1, the defined in (A.47) estimate $\widetilde{f}(x,z)$ satisfies the exponential inequality

$$\mathbb{P}\Big(\max_{(x,z)\in\mathcal{R}_r} |\widetilde{f}(x,z) - f^{X,Z}(x,z)| > c_1 n_*^{-1/8}\Big) \leqslant 2n_*^2 e^{-c_0 n_*^{1/4}}. \tag{A.66}$$

Proof of Lemma A.2. We begin with the following relation,

$$\widetilde{f}(x,z) - f^{X,Z}(x,z) = -\sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\varphi_i(x)a^{-1/2}\varphi_s(z/a)$$

$$+ \sum_{i,s\in\mathcal{N}} [n_*^{-1}\sum_{l=1}^{n_*}\varphi_i(X_{El})\psi_s(Z_{El}) - \kappa_{is}]\varphi_i(x)\psi_s(z)$$

$$=: D_1(x,z) + \widetilde{D}_2(x,z) =: D_1 + \widetilde{D}_2. \tag{A.67}$$

Here $\mathcal{N} := \{0,1,\dots,N\}^2$ with $N := \lfloor 1 + n_*^{1/4}\rfloor$. The first term D_1 is evaluated using Assumption 4.1, Lemma A.1, and the Cauchy-Schwarz inequality (below a more general than needed relation is presented for future use)

$$|D_1| = |\sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\varphi_i(x)a^{-1/2}\varphi_s(z/a)| \leqslant 2a^{-1/2}\sum_{(i,s)\notin\mathcal{N}}|\kappa_{is}|$$

$$\leqslant 2a^{-1/2}\Big[\sum_{(i,s)\notin\mathcal{N}}(1+i^2)^{-1}(1+s^2)^{-1}\sum_{(i,s)\notin\mathcal{N}}(1+i^2)(1+s^2)\kappa_{is}^2\Big]^{1/2}$$

$$\leqslant CN^{-1/2} \leqslant Cn_*^{-1/8}. \tag{A.68}$$

Recall that Cs denote generic constants.

Analysis of the second term in (A.67) is more involved. We begin with a remark that we may rewrite $\widetilde{D}_2$ as

$$\widetilde{D}_2 = n_*^{-1}\sum_{l=1}^{n_*}\Big[[\sum_{(i,s)\in\mathcal{N}}\varphi_i(X_{El})\psi_s(Z_{El})\varphi_i(x)\varphi_s(z)]$$

$$-[\sum_{(i,s)\in\mathcal{N}}\kappa_{is}\varphi_i(x)\psi_s(z)]\Big] =: n_*^{-1}\sum_{l=1}^{n_*}\Big[\widetilde{D}_{2l} - D_{2l}\Big]. \tag{A.69}$$

Note that $\mathbb{E}\{\widetilde{D}_{2l} - D_{2l}\} = 0$. To evaluate the second moment of the difference we note that Lemma 1 yields

$$\sum_{i,s=0}^{\infty} |\kappa_{is}| < C. \tag{A.70}$$

Next, using the fact that the basis functions are bounded, (A.70) and the elementary trigonometric identity

$$\varphi_i(x)\varphi_s(x) = 2^{-1/2}[\varphi_{i-s}(x) + \varphi_{i+s}(x)],$$

we can write,

$$\begin{aligned}
&\mathbb{E}\{(\widetilde{D}_{2l} - D_{2l})^2\} \\
&= \sum_{(i,s)\in\mathcal{N}} \sum_{(i',s')\in\mathcal{N}} \Big[\mathbb{E}\{(\varphi_i(X_{El})\psi_s(Z_{El}) - \kappa_{is})(\varphi_{i'}(X_{El})\psi_{s'}(Z_{El}) - \kappa_{i's'})\} \\
&\qquad \times \varphi_i(x)\psi_s(z)\varphi_{i'}(x)\varphi_{s'}(z/a)\Big] \\
&\leqslant C \sum_{(i,s)\in\mathcal{N}} \sum_{(i',s')\in\mathcal{N}} [|\kappa_{i\pm i',s\pm s'}| + |\kappa_{is}\kappa_{i's'}|] \leqslant CN^2 < Cn_*^{1/2}.
\end{aligned} \tag{A.71}$$

Here $\kappa_{i\pm i',s\pm s'}$ denotes any possible plus or minus combination of the indexes which is created by the above-mentioned trigonometric identity. Let us also note that

$$|\widetilde{D}_{2l} - D_{2l}| \leqslant CN^2 < Cn_*^{1/2}. \tag{A.72}$$

These results allow us to use the Bernstein exponential inequality [18], p.19 and conclude that

$$\mathbb{P}\Big(|\widetilde{D}_2| \geqslant u\Big) \leqslant 2\exp\Big\{-C\frac{n_* u^2}{n_*^{1/2} + tn_*^{1/2}}\Big\} \leqslant 2\exp\Big\{-C\frac{n_*^{1/2}u^2}{1+u}\Big\}. \tag{A.73}$$

Set $u := c_1 n_*^{-1/8}$ for some positive constant c_1. Also recall notations $\widetilde{D}_2 = \widetilde{D}_2(x,z)$ introduced in (A.67), that $(x,z) \in \mathcal{R}_\tau$, and then conclude with the help of (A.73) that for any k pairs $(x_i, z_i) \in [0,1] \times [0,a]$ we have

$$\sum_{i=1}^{k} \mathbb{P}\Big(|\widetilde{D}_2(x_i,z_i)| \geqslant c_1 n_*^{-1/8}\Big) \leqslant 2k\exp\{-c_0 n_*^{1/4}\}. \tag{A.74}$$

The inequality (A.72), together with (A.68), is “almost” what we want to prove, only additionally we need to check that a similar inequality holds simultaneously for all points $(x,z) \in \mathcal{R}_\tau$. To check this we use a rough inequality $|\partial^2\widetilde{D}_2(x,z)/\partial x\partial z| \leqslant CN^4 < Cn_*$ a.s. (it is indeed a rough but sufficient for our purposes inequality, and a better result may be obtained via analog of the generalized Bernstein inequality for the derivative of a trigonometric polynomial discussed in [41], p.96. Keeping in mind the maximum value of the derivative, we divide the rectangle $[0,1] \times [0,a]$ into $k := n_*^2$ identical sub-rectangles, note that if (x_i, z_i) belongs to a sub-rectangle and $|\widetilde{D}_2(x_i, z_i)| \leqslant c_1 n_*^{-1/8}$, then the mean value theorem implies that for all points from this sub-rectangle we have $|\widetilde{D}_2(x, z)| \leqslant c_0 n_8^{-1/8}$ for some finite c_0. This concludes the proof of Lemma A.2.

Remark A.2 The exponential inequality of Lemma A.2 allows us, whenever it is convenient in a proof, to replace the density estimate $\widehat{f}(x, z)$ by $\widetilde{f}(x, z)$ and vise versa. Indeed, it is known that the underlying bivariate density $f(x, z)$ is bounded below from zero on $\mathcal{R}_r$, and hence probability of the event $\widetilde{f}(x, z) \neq \widehat{f}(x, z)$ is exponentially small in n_* (and hence in n) while all the studied risks decrease as some power of n.

Following Remark 6.2, denote by $\widetilde{A}_1$ and $\widetilde{A}_2$ the expressions for $\widehat{A}_1$ and $\widehat{A}_2$, defined in (6.55), only with $\widehat{f}$ being replaced by $\widetilde{f}$, and recall that definitions of these joint density estimates are presented in (6.52) and (6.53). We begin with evaluation of second moments of $\widetilde{A}_1$ and $\widetilde{A}_2$ in turn. Recall that to simplify formulae we may write $f(x, z)$ in place of $f^{X,Z}(x, z)$. For the second moment of $\widetilde{A}_1$ we get

$$\mathbb{E}\{\widetilde{A}_1^2\} = n^{-1}\mathbb{E}\Big\{\Big[\frac{\Delta'\varphi_j(X)(f(X,Z) - \widetilde{f}(X,Z))}{f^2(X,Z)}\Big]^2\Big\}$$
$$+ n^{-2}n(n-1)\mathbb{E}\{\widetilde{V}^2\} =: A_{11} + n^{-2}n(n-1)A_{12}. \quad (A.75)$$

Here

$$\widetilde{V} := \int_0^r \int_0^1 S^{T|X}(z|x) f^{-1}(x, z)\varphi_j(x)[f(x, z) - \widetilde{f}(x, z)]dxdz$$
$$= \int_0^r \int_0^1 S^{T|X}(z|x) f^{-1}(x, z)\varphi_j(x)\Big[\sum_{(i,s)\in\mathcal{N}} (\kappa_{is} - \widetilde{\kappa}_{is})\varphi_i(x)\psi_s(z)$$
$$+ \sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\varphi_i(x)\psi_s(z)\Big]dxdz$$
$$= \sum_{(i,s)\in\mathcal{N}} (\kappa_{is} - \widetilde{\kappa}_{is})\nu_{i\pm j,s} + \sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\nu_{i\pm j,s}, \quad (A.76)$$

and we used notation

$$\nu_{i\pm j,s} := \int_0^a \int_0^1 S^{T|X}(z|x) f^{-1}(x, z)\psi_j(x)\varphi_i(x)\psi_s(z)dxdz. \quad (A.77)$$

This notation is motivated by the trigonometric identity (recall that we already used it before)

$$\varphi_i(x)\varphi_j(x) = 2^{-1/2}[\varphi_{i-j}(x) + \varphi_{i+j}(x)]. \quad (A.78)$$

Further, for the considered bivariate functions $S(z|x)$ and $f(x, z)$, with the help of Lemma A.1, we get

$$\max_{j\geqslant 0} \sum_{i=0}^{\infty} \sum_{s=0}^{\infty} |\nu_{i\pm j,s}| \leqslant C. \quad (A.79)$$

Also, recall our notations for Fourier coefficients of the density and its estimate,

$$\kappa_{is} := \int_0^r \int_0^1 f(x,z)\varphi_i(x)\varphi_s(z)dxdz, \tag{A.80}$$

and

$$\widetilde{\kappa}_{is} := \int_0^r \int_0^1 \widetilde{f}(x,z)\varphi_i(x)\varphi_s(z)dxdz$$
$$= n_*^{-1}\sum_{r=l}^{n_*}\varphi_i(X_{El})\varphi_s(Z_{El}). \tag{A.81}$$

Now we are in a position to evaluate the two terms on the right side of (A.75). For the first term we get using Lemma A.2 that

$$A_{11} \leqslant Cn_*^{-1/4}n^{-1}. \tag{A.82}$$

Evaluation of A_{12} is more involved, and we begin with the following useful proposition.

Lemma A.3. *Let Assumption 4.1 hold. Consider κ_{is} and $\widetilde{\kappa}_{is}$ defined in (A.80) and (A.81), respectively. Then*

$$\mathbb{E}\{\widetilde{\kappa}_{is}\} = \kappa_{is}, \tag{A.83}$$

$$\mathbb{E}\{(\widetilde{\kappa}_{is}-\kappa_{is})^2\} \leqslant Cn_*^{-1}, \tag{A.84}$$

and using notations $\varphi^*_{i\pm i'}(x) := \varphi_j(x)\varphi_{i'}(x)$ and $\psi^*_{i\pm i'}(x) := \psi_j(x)\psi_{i'}(x)$, we have

$$|\mathbb{E}\{(\widetilde{\kappa}_{is}-\kappa_{is})(\widetilde{\kappa}_{i's'}-\kappa_{i's'})\}|$$
$$= n_*^{-1}\Big[\int_0^r\int_0^1 f(x,z)\varphi^*_{i\pm i'}(x)\psi^*_{s\pm s'}(z)dxdz - \kappa_{is}\kappa_{i's'}\Big]$$
$$=: n_*^{-1}[\kappa_{i\pm i',s\pm s'} - \kappa_{is}\kappa_{i's'}], \tag{A.85}$$

where

$$\sum_{i,s=1}^{\infty}[|\kappa_{i\pm i',s\pm s'}| + |\kappa_{is}\kappa_{i's'}|] < C. \tag{A.86}$$

Proof of Lemma A.3. Recall that

$$\widetilde{\kappa}_{is} := n_*^{-1}\sum_{l=1}^{n_*}\varphi_i(X_{Er})\psi_s(Z_{Er}),$$

and this yields (A.83) because $\kappa_{is} = \mathbb{E}\{\varphi_i(X)\psi_s(Z)\}$. Relation (A.84) follows from (A.83) and independence of pairs (X_{El}, Z_{El}) in the extra sample. To check (A.85) we write using (A.83),

$$\mathbb{E}\{(\widetilde{\kappa}_{is}-\kappa_{is})(\widetilde{\kappa}_{i's'}-\kappa_{i's'})\}$$

$$= n_*^{-2} \sum_{l_1,l_2=1}^{n_*} \mathbb{E}\{(\varphi_i(X_{El_1})\psi_s(Z_{El_1}) - \kappa_{is})(\varphi_{i'}(X_{El_2})\varphi_{s'}(Z_{El_2}) - \kappa_{i's'})\}$$

$$= n_*^{-1}\mathbb{E}\{(\varphi_i(X)\psi_s(Z) - \kappa_{is})(\varphi_{i'}(X)\psi_{s'}(Z) - \kappa_{i's'})\}$$

$$= n_*^{-1}\Big[\int_0^a\int_0^1 f(x,z)\varphi_{i\pm i'}(x)\psi_{s\pm s'}(z)dxdz - \kappa_{is}\kappa_{i's'}\Big]$$

$$=: n_*^{-1}[\kappa_{i\pm i',s\pm s'} - \kappa_{is}\kappa_{i's'}]. \quad (A.87)$$

This relation and (A.71) verify (A.85). Lemma A.3 is proved.

Using Lemma A.3 and (A.76)-(A.81) we can evaluate A_{12} defined in (A.75). Write,

$$A_{12} = \mathbb{E}\{\widetilde{V}^2\}$$

$$= \mathbb{E}\Big\{\sum_{(i,s),(i',s')\in\mathcal{N}} (\kappa_{is} - \widetilde{\kappa}_{is})(\kappa_{i's'} - \widetilde{\kappa}_{i's'})\nu_{i\pm j,s}\nu_{i'\pm j,s'}\Big\} + \Big[\sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\nu_{i\pm j,s}\Big]^2$$

$$=: A_{121} + A_{122}. \quad (A.88)$$

For the first term, using (A.70), (A.79) and (A.87), we conclude that

$$A_{121} = n_*^{-1} \sum_{(i,s),(i',s')\in\mathcal{N}} [\kappa_{i\pm i',s\pm s'} - \kappa_{is}\kappa_{i's'}]\nu_{i\pm j,s}\nu_{i'\pm j,s'} \leqslant Cn_*^{-1}. \quad (A.89)$$

To evaluate A_{122} we recall that an underlying parameter α is at least $\alpha_0 = 2$. This yields that for all sufficiently large considered in Theorem 4.1 sample sizes n_* and n we have $j \leqslant N/2$ (we have even stronger relation $j \leqslant o_{n*}(1)N$). Then, using Lemma A.1 and the Cauchy-Schwarz inequality we get,

$$A_{122} = \Big[\sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\nu_{i\pm j,s}\Big]^2 \leqslant \sum_{(i,s)\notin\mathcal{N}} \kappa_{is}^2 \sum_{(i,s)\notin\mathcal{N}} \nu_{i\pm j,s}^2 \leqslant N^{-4}. \quad (A.90)$$

This yields that

$$A_{12} \leqslant C[n_*^{-1} + N^{-4}], \quad (A.91)$$

which in its turn, together with (A.82) and (A.75), yield the following result for $\widetilde{A}_1$,

$$\mathbb{E}\{\widetilde{A}_1^2\} \leqslant C[n_*^{-1} + N^{-4}]. \quad (A.92)$$

Now we are evaluating the second moment of $\widetilde{A}_2$. Recall that this statistic is defined in (A.49) and we can write it as a sum of two terms,

$$\widetilde{A}_2 = n^{-1}\sum_{l=1}^n \frac{\Delta'_l\varphi_j(X_l)(f(X_l,Z_l) - \hat{f}(X_l,Z_l))^2}{f^3(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^n \frac{\Delta'_l\varphi_j(X_l)(f(X_l,Z_l) - \hat{f}(X_l,Z_l))^3}{\hat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$= \widetilde{A}_{21} + \widetilde{A}_{22}. \tag{A.93}$$

Recall our earlier explanation that for the purpose of estimating moments with the desired accuracy, and according to Lemma A.2 we may replace in $\widetilde{A}_{21}$ the density estimate $\widehat{f}$ by $\widetilde{f}$. Let us denote this modified term as $\widehat{A}_{21}$ and explore the second moment of $\widehat{A}_{21}$. Using Lemma A.3 and formula

$$f(x,z) - \widetilde{f}(x,z)$$
$$= \sum_{(i,s)\in\mathcal{N}} (\kappa_{is} - \widetilde{\kappa}_{is})\varphi_i(x)\psi_s(z) + \sum_{(i,s)\notin\mathcal{N}} \kappa_{is}\varphi_i(x)\psi_s(z), \tag{A.94}$$

we may write,

$$\mathbb{E}\{\widehat{A}_{21}^2\} = n^{-1}\mathbb{E}\Big\{\Big[\frac{\Delta'\varphi_j(X)(f(X,Z)-\widetilde{f}(X,Z))^2}{f^3(X,Z)}\Big]^2\Big\}$$
$$+n^{-2}n(n-1)\mathbb{E}\Big\{\Big[\int_0^r\int_0^1 \frac{S^{T|X}(z|x)\varphi_j(x)(f(x,z)-\widetilde{f}(x,z))^2}{f^2(x,z)}dxdz\Big]^2\Big\}$$
$$\leqslant Cn^{-1}n_*^{-1/8} + \mathbb{E}\Big\{\Big[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}} (\kappa_{i_1s_1}-\widetilde{\kappa}_{i_1s_1})(\kappa_{i_2s_2}-\widetilde{\kappa}_{i_2s_2})\nu_{i_1\pm i_2\pm j,s_1\pm s_2}$$
$$+2\sum_{(i_1,s_1)\in\mathcal{N}}\sum_{(i_3,s_3)\notin\mathcal{N}} (\kappa_{i_1s_1}-\widetilde{\kappa}_{i_1s_1})\kappa_{i_3s_3}\nu_{i_1\pm i_3\pm j,s_1\pm s_3}$$
$$+\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j,s_3\pm s_4}\Big]^2\Big\}$$
$$\leqslant C\Big[n^{-1}n_*^{-1/8} + \mathbb{E}\Big\{[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}} (\kappa_{i_1s_1}-\widetilde{\kappa}_{i_1s_1})(\kappa_{i_2s_2}-\widetilde{\kappa}_{i_2s_2})\nu_{i_1\pm i_2\pm j,s_1\pm s_2}]^2\Big\}$$
$$+[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j,s_3\pm s_4}]^2$$
$$+[\sum_{(i_1,s_1)\in\mathcal{N}}\sum_{(i_3,s_3)\notin\mathcal{N}} (\kappa_{i_1s_1}-\widetilde{\kappa}_{i_1s_1})\kappa_{i_3s_3}\nu_{i_1\pm i_3\pm j,s_1\pm s_3}]^2\Big]$$
$$=: C[n^{-1}n_*^{-1/8} + A_{211} + A_{212} + A_{213}]. \tag{A.95}$$

Here

$$\nu_{i_1\pm i_2\pm j,s_1\pm s_2}$$
$$:= \int_0^r\int_0^1 \frac{S^{T|X}(z|x)\varphi_j(x)\varphi_{i_1}(x)\varphi_{i_2}(x)\psi_{s_1}(z)\psi_{s_2}(z)}{f^2(x,z)}dxdz. \tag{A.96}$$

Applying the Cauchy inequality we evaluate A_{211},

$$A_{211} \leqslant \mathbb{E}\Big\{\Big[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}} (\kappa_{i_1s_1}-\widetilde{\kappa}_{i_1s_1})^2|\nu_{i_1\pm i_2\pm j,s_1\pm s_2}|\Big]^2\Big\}. \tag{A.97}$$

Now note that $\mathbb{E}\{(\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^4\} \leqslant Cn_*^{-2}$ and relation (A.79) holds for parameters defined in (A.96), namely $\sum_{i_2,s_2=0}^{\infty} |\nu_{i_1\pm i_2\pm j, s_1\pm s_2}| \leqslant C$. Using these results we continue (A.97),

$$A_{211} \leqslant C\Big[\sum_{(i_1,s_1)\in\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^2\Big]^2 \leqslant Cn_*^{-2}N^4. \tag{A.98}$$

For evaluation of A_{212} we can write using the Cauchy inequality,

$$\Big[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j, s_3\pm s_4}\Big]^2$$

$$\leqslant \Big[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}^2 |\nu_{i_3\pm i_4\pm j, s_3\pm s_4}|\Big]^2 \leqslant CN^{-4}. \tag{A.99}$$

Evaluation of A_{213} may be converted to the previously considered A_{211} and A_{212} by using the inequality

$$2|(\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})\kappa_{i_3s_3}\nu_{i_1\pm i_3\pm j, s_1\pm s_3}|$$

$$\leqslant (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^2|\nu_{i_1\pm i_3\pm j, s_1\pm s_3}| + \kappa_{i_3s_3}^2|\nu_{i_1\pm i_3\pm j, s_1\pm s_3}|.$$

Combining the obtained results in (A.95) we establish the following upper bound,

$$\mathbb{E}\{\widehat{A}_{21}^2\} \leqslant C[n^{-1}n_*^{-1/8} + n_*^{-2}N^4 + N^{-4}]. \tag{A.100}$$

Now we are evaluating the second moment of $\widetilde{A}_{22}$ defined in (A.93). We begin with writing $\widetilde{A}_{22}$ as

$$\widetilde{A}_{22} = n^{-1}\sum_{l=1}^{n} \frac{\Delta_l'\varphi_j(X_l)(f(X_l,Z_l) - \widetilde{f}(X_l,Z_l))^3}{\widehat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n} \frac{\Delta_l'\varphi_j(X_l)[(f(X_l,Z_l) - \widehat{f}(X_l,Z_l))^3 - (f(X_l,Z_l) - \widetilde{f}(X_l,Z_l))^3]}{\widehat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$=: \widetilde{A}_{221} + \widetilde{A}_{222}. \tag{A.101}$$

Recall that $\widehat{f}$ is bounded below from zero by $1/\ln(\ln(n_* + 3))$, and according to Lemma 2 the deviation $|f(x,z) - \widetilde{f}(x,z)|$ is larger than $Cn_*^{-1/8}$ with the probability decreasing exponentially in n_*. These two facts imply that

$$\mathbb{E}\{\widetilde{A}_{221}^2\} \leqslant C[\ln(\ln(n_* + 3))]^2 n_*^{-2/8}$$

$$\times\, \mathbb{E}\Big\{n^{-1}\sum_{l=1}^{n}\Big[\frac{\Delta_l'\varphi_j(X_l)(f(X_l,Z_l) - \widetilde{f}(X_l,Z_l))^2}{f^3(X,Z)}\Big]^2\Big\}. \tag{A.102}$$

Then the expectation on the right side of (A.102) is evaluated following the lines (A.95)-(A.100). Similarly, recall that $\widehat{f}(x,z) \neq \widetilde{f}(x,z)$ only if $\widetilde{f}(x,z) <$

$1/\ln(\ln(n_*+3))$, and then according to Lemma A.2 that event occurs with the probability that decreases exponentially in n_*. This yields a rough but sufficient for our purposes inequality $\mathbb{E}\{\widetilde{A}_{222}^2\} \leqslant Cn_*^{-2}$. Combining the results we conclude that

$$\mathbb{E}\{\widetilde{A}_{22}^2\} \leqslant Cn_*^{-10/9}. \tag{A.103}$$

Using (A.100) and (A.103) in (A.93) we get

$$\mathbb{E}\{\widetilde{A}_2^2\} \leqslant C[n^{-1}n_*^{-1/8} + n_*^{-2}N^4 + N^{-4} + n_*^{-10/9}]. \tag{A.104}$$

Finally, combining (A.92) and (A.104) in (A.49) and recalling the used $N = \lfloor 1 + n_*^{1/4} \rfloor$, we get the verified inequality

$$\mathbb{E}\{(\widetilde{\theta}_j - \widetilde{\theta}_j^*)^2\} \leqslant Cn_*^{-1} \quad \text{whenever} \quad n_* > Cn. \tag{A.105}$$

We conclude that the mean squared error of data-driven Fourier estimator $\widehat{\theta}_j$ decreases with the same rate as the mean squared error of the oracle-estimator $\widetilde{\theta}_j$ that knows the underlying joint density $f^{X,Z}(x,z)$. Moreover, if $n_*/n \to \infty$ as $n \to \infty$ then

$$\mathbb{E}\{(\widetilde{\theta}_j - \theta_j)^2\} \leqslant (1 + o_n(1))\mathbb{E}\{(\widetilde{\theta}_j^* - \theta)^2\} + o_n(1)n^{-1}. \tag{A.106}$$

We have proved that for MCSC the plug-in methodology works and the Fourier coefficients can be estimated as well as by the oracle-estimators using the nuisance bivariate density $f^{X,Z}$. The presented results also give us all technical propositions needed to prove Theorem 4.1 for MCSC.

Now we are in a position to consider CSC. Recall that the studied Fourier estimator is

$$\widehat{\theta}_j := n^{-1} \sum_{l=1}^{n} \frac{[\Delta_l' - \widehat{S}(j,n,Z_l,X_l)]I(Z_l \leqslant r)\varphi_j(X_l)}{\widehat{f}(X_l,Z_l)}. \tag{A.107}$$

Here

$$\widehat{f}(x,z) := \max(1/\ln(\ln(n+3)), \widetilde{f}(x,z)), \tag{A.108}$$

$$\widetilde{f}(x,z) = n^{-1} \sum_{l=1}^{n} \sum_{(i,s)i\mathcal{N}} \varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z), \tag{A.109}$$

$\mathcal{N} = \{0,1,\dots,N\}^2$, $N := \lfloor 1 + n^{1/4} \rfloor$,

$$\widehat{S}(j,n,z,x) := n^{-1} \sum_{l=1}^{n} \sum_{k \in \mathcal{J}(j,n)} \sum_{i=0}^{q_n} \frac{\Delta_l'\psi_k(Z_l)\varphi_i(X_l)\psi_k(z)\varphi_k(x)}{\widehat{f}(X_l,Z_l)}, \tag{A.110}$$

and $\mathcal{J}(j,n) = \{0,1,\dots,q_n\}\backslash\{j\}$ if $0 \leqslant j \leqslant b_n$ and $\mathcal{J}(j,n) = \{0,1,\dots,b_n\}$ otherwise.

Our aim is to evaluate $\mathbb{E}\{(\widehat{\theta}_j - \theta_j)^2\}$. Write,

$$\widehat{\theta}_j = n^{-1} \sum_{l=1}^{n} \frac{[\Delta_l' - S(j,n,Z_l,X_l)]I(Z_l \leqslant r)\varphi_j(X_l)}{\widehat{f}(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{[S(j,n,Z_l,X_l)-\widehat{S}(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)}{\widehat{f}(X_l,Z_l)}$$

$$=n^{-1}\sum_{l=1}^{n}\frac{[\Delta_l'-S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)}{f(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{[S(j,n,Z_l,X_l)-\widehat{S}(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)}{\widehat{f}(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{[\Delta_l'-S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)(f(X_l,Z_l)-\widehat{f}(X_l,Z_l))}{f^2(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{[\Delta_l'-S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)(f(X_l,Z_l)-\widehat{f}(X_l,Z_l))^2}{\widehat{f}(X_l,Z_l)f^2(X_l,Z_l)}$$

$$=:\widehat{\theta}_j^*+\widehat{A}_j+\widehat{B}_1+\widehat{B}_2. \tag{A.111}$$

Recall that $\widehat{\theta}_j^*$ is the oracle's Fourier estimator that was used in Theorem 3.2. The term $\widehat{A}_j$ is the analogue of statistic $\widetilde{A}_j$ defined in (A.35), it is analyzed similarly to (A.36)-(A.41), and accordingly $\mathbb{E}\{(\widehat{A}_j')^2\}$ is in order smaller than n^{-1}. The terms $\widehat{B}_1$ and $\widehat{B}_2$ are analogues of statistics $\widetilde{A}_1$ and $\widetilde{A}_2$ defined in (A.49). It was established that $\mathbb{E}\{\widehat{A}_1^2\}$ and $\mathbb{E}\{\widehat{A}_2^2\}$ are of order n_*^{-1}. Here we need to establish a stronger result and show that the corresponding second moments of $\widehat{B}_1$ and $\widehat{B}_2$ decrease faster.

We are evaluating the second moments of $\widehat{B}_1$ and $\widehat{B}_2$ in turn. Using Remark A.2 we may replace $\widehat{f}(x,z)$ by $\widetilde{f}(x,z)$ and then denote the new term as $\widetilde{B}_1$. Let is rewrite $\widetilde{f}(x,z)$ as

$$\widetilde{f}(x,z)=(n-1)^{-1}\sum_{l\in\{1,\dots,n\}\setminus\{j\}}\ \sum_{(i,s)\in\mathcal{N}}\varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)$$

$$+n^{-1}\Big[-(n-1)^{-1}\sum_{l\in\{1,\dots,n\}\setminus\{j\}}\ \sum_{(i,s)\in\mathcal{N}}\varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)$$

$$+\sum_{(i,s)\in\mathcal{N}}\varphi_i(X_j)\psi_s(Z_j)\varphi_i(x)\psi_s(z)\Big]$$

$$=:\widetilde{f}(x,z,-j)+\widetilde{d}_1(x,z,j). \tag{A.112}$$

Here $\widetilde{f}(x,z,-j)$ is a density estimate based on all observations apart of (X_r,Z_j), and $\widetilde{d}_1(x,z,j)$ is a small remaining part of the density estimate which is of order n^{-1}. Similarly, we may separate two pairs (X_j,Z_j) and (X_k,Z_k) of observations and rewrite the estimate $\widetilde{f}(x,z)$ as

$$\widetilde{f}(x,z)=(n-2)^{-1}\sum_{l\in\{1,\dots,n\}\setminus\{j,k\}}\ \sum_{(i,s)\in\mathcal{N}}\varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)$$

$$+\Big[\frac{-2}{n(n-2)}\sum_{l\in\{1,\dots,n\}\setminus\{j,k\}}\ \sum_{(i,s)\in\mathcal{N}}\varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)$$

$$+n^{-1}a^{-1}\sum_{l\in\{j,k\}}\ \sum_{(i,s)\in\mathcal{N}}\varphi_i(X_l)\psi_s(Z_l)\varphi_i(x)\psi_s(z)\Big]$$

$$=: \widetilde{f}(x,z,-j,-k)+\widetilde{d}_2(x,z,j,k). \tag{A.113}$$

For the second moment of $\widetilde{B}_1$ (recall that according to our notation $\widetilde{B}_1$ is $\widehat{B}_1$ with $\widehat{f}$ being replaced by $\widetilde{f}$) we may write using the new notations,

$$\mathbb{E}\{\widetilde{B}_1^2\}$$

$$= n^{-1}\mathbb{E}\Big\{\Big[\frac{[\Delta_1'-S(j,n,Z_1,X_1)]I(Z_1\leqslant r)\varphi_j(X_1)(f(X_1,Z_1)-\widetilde{f}(X_1,Z_1))}{f^2(X_1,Z_1)}\Big]^2\Big\}$$

$$+(1-n^{-1})\mathbb{E}\Big\{\prod_{l=1}^{2}\frac{[\Delta_l'-S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)(f(X_l,Z_l)-\widetilde{f}(X_l,Z_l))}{f^2(X_l,Z_l)}\Big\}$$

$$= n^{-1}\mathbb{E}\Big\{\Big[[\Delta_1'-S(j,n,Z_1,X_1)]I(Z_1\leqslant r)\varphi_j(X_1)$$

$$\times\frac{(f(X_1,Z_1)-\widetilde{f}(X_1,Z_1,-1)-\widetilde{d}_1(X_1,Z_1,1))}{f^2(X_1,Z_1)}\Big]^2\Big\}$$

$$+n^{-1}(n-1)\mathbb{E}\Big\{\prod_{l=1}^{2}[\Delta_l'-S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)$$

$$\times\frac{(f(X_l,Z_l)-\widetilde{f}(X_l,Z_l,-1,-2)-\widetilde{d}_2(X_l,Z_l,1,2)}{f^2(X_l,Z_l)}\Big\}$$

$$=: B_{11}+B_{12}. \tag{A.114}$$

Using the Cauchy inequality and Lemma A.2 we evaluate B_{11},

$$B_{11}\leqslant 2n^{-1}$$

$$\times\mathbb{E}\Big\{\Big[\frac{(\Delta_1'-S(j,n,Z_1,X_1))I(Z_1\leqslant r)\varphi_j(X_1)(f(X_1,Z_1)-\widetilde{f}(X_1,Z_1,-1))}{f^2(X_1,Z_1)}\Big]^2\Big\}$$

$$+n^{-1}2\mathbb{E}\Big\{\Big[\frac{(\Delta_1'-S(j,n,Z_1,X_1))I(Z_l\leqslant r)\varphi_j(X_1)\widetilde{d}_1(X_1,Z_1,1))}{f^2(X_1,Z_1)}\Big]^2\Big\}$$

$$\leqslant Cn^{-9/8}. \tag{A.115}$$

Next we are considering B_{12}. Write,

$$B_{12}\leqslant\mathbb{E}\Big\{\Big[\int_0^r\int_0^1\frac{(S(z|x)-S(j,n,z,x))\varphi_j(x)(f(x,z)-\widetilde{f}(x,z,-1,-2))}{f(x,z)}\Big]^2\Big\}$$

$$+2\mathbb{E}\Big\{[\Delta_1'-S(j,n,Z_1,X_1)]I(Z_1\leqslant r)\varphi_j(X_1)$$

$$\times \frac{(f(X_1,Z_1)-\widetilde{f}(X_1,Z_1,-1,-2))\widetilde{d}_2(X_2,Z_2,1,2)}{f^2(X_1,Z_1)f^2(X_2,Z_2)}\Big\}$$

$$+ C\mathbb{E}\{|\widetilde{d}_2(X_1,Z_1,1,2)\widetilde{d}_2(X_2,Z_2,1,2)|\} =: B_{121} + B_{122} + B_{123}. \qquad (A.116)$$

Note that $|\widetilde{d}_2(x,z,1,2)| \leqslant Cn^{-1}q_n^2$ almost sure, and this together with Lemma A.2 imply

$$B_{122} + B_{123} \leqslant Cn^{-10/9}. \qquad (A.117)$$

Now we are evaluating B_{121}. This term is the analogue of $\mathbb{E}\{\widetilde{V}^2\}$ introduced in (A.75)-(A.76). In what follows we are using notation $\widetilde{\kappa}_{is} := (n-2)^{-1}\sum_{l=3}^n \varphi_i(X_l)\psi_s(Z_l)$ for Fourier estimates of $\widetilde{f}(x,z,-1,-1)$. It reflects the fact that $\widetilde{f}(x,z,-1,-2)$ is based on pairs $(X_3,Z_3),\dots,(X_n,Z_n)$ but otherwise it is our density estimator based on $n-2$ pairs of observations, check (A.113). Then we may write for the integral in B_{121},

$$\int_0^r\int_0^1 (S(z|x)-S(j,n,z,x))f^{-1}(x,z)\varphi_j(x)[f(x,z)-\widetilde{f}(x,z,-1,-2)]dxdz$$

$$= \int_0^r\int_0^1 (S(z|x)-S(j,n,z,x))f^{-1}(x,z)\varphi_j(x)\Big[\sum_{(i,s)\in\mathcal{N}}(\kappa_{is}-\widetilde{\kappa}_{is})\varphi_i(x)\psi_s(z)$$

$$+ \sum_{(i,s)\notin\mathcal{N}}\kappa_{is}\varphi_i(x)\psi_s(z)\Big]dxdz$$

$$= \sum_{(i,s)\in\mathcal{N}}(\kappa_{is}-\widetilde{\kappa}_{is})\nu_{i\pm j,s} + \sum_{(i,s)\notin\mathcal{N}}\kappa_{is}\nu_{i\pm j,s}, \qquad (A.118)$$

where here (compare with (A.77))

$$\nu_{i\pm j,s} := \int_0^r\int_0^1 (S(z|x)-S(j,n,z,x))$$

$$\times f^{-1}(x,z)\varphi_j(x)\varphi_i(x)\psi_s(z)dxdz. \qquad (A.119)$$

Recall that this notation is motivated by the trigonometric identity

$$\varphi_i(x)\varphi_j(x) = 2^{-1/2}[\varphi_{i-j}(x)+\varphi_{i+j}(x)]. \qquad (A.120)$$

Recall that in the first step of the proof the key property of ν_{is} was (see (6.85))

$$\max_{j\geqslant 0}\sum_{i=0}^\infty\sum_{s=0}^\infty |\nu_{i\pm j,s}| \leqslant C. \qquad (A.121)$$

It is also valid for (A.119) with formally setting $S(j,n,z,x) = 0$ when the corresponding functional G_g of Lemma A.1 is bounded by a constant. In our case, because the Fourier projection approximation $S(j,n,z,x)$ is subtracted from the underlying conditional survival function $S(z|x)$, in Lemma A.1 the functional G_g is defined by the bivariate function

$$g(x,z) := (S(z|x) - S(j,n,z,x))/f^{X,Z}(x,z),$$

and accordingly the functional is bounded by $o_j(1) + o_n(1)$. This and (A.58) yield

$$\max_{j\geqslant 0}\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}|\nu_{i\pm j,s}| = o_j(1) + o_n(1). \tag{A.122}$$

This relation explains why the sample size n is sufficient for estimating the underlying joint density $f^{X,Z}$ with the accuracy yielding

$$B_{121} = [o_j(1) + o_n(1)]n^{-1}. \tag{A.123}$$

This result, together with (A.115)–(A.117), yield

$$\mathbb{E}\{\widetilde{B}_1^2\} = [o_j(1) + o_n(1)]n^{-1}. \tag{A.124}$$

Now we are evaluating $\mathbb{E}\{\widehat{B}_2^2\}$ where $\widehat{B}_2$ is defined in (A.111). Write,

$$\widehat{B}_2 = n^{-1}\sum_{l=1}^{n}\frac{[\Delta_l' - S(j,n,Z_l,X_l]I(Z_l\leqslant r)\varphi_j(X_l)(f(X_l,Z_l) - \widehat{f}(X_l,Z_l))^2}{f^3(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{[\Delta_l' - S(j,n,Z_l,X_l)]I(Z_l\leqslant r)\varphi_j(X_l)(f(X_l,Z_l) - \widehat{f}(X_l,Z_l))^3}{\widehat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$= \widehat{B}_{21} + \widehat{B}_{22}. \tag{A.125}$$

Recall Remark A.1 which explains that according to Lemma A.2 we may replace in $\widehat{B}_{21}$ the density estimate $\widehat{f}$ by $\widetilde{f}$. Denote the corresponding modified term as $\widetilde{B}_{21}$ and explore it using Lemma A.3 and the equality

$$f(x,z) - \widetilde{f}(x,z) = \sum_{(i,s)\in\mathcal{N}}(\kappa_{is} - \widetilde{\kappa}_{is})\varphi_i(x)\psi_s(z)$$

$$+\sum_{(i,s)\notin\mathcal{N}}\kappa_{is}\varphi_i(x)\psi_s(z/a). \tag{A.126}$$

We may write,

$$\mathbb{E}\{\widetilde{B}_{21}^2\} = n^{-1}\mathbb{E}\Big\{\Big[\frac{[\Delta' - S(j,n,Z_l,X_l)]I(Z\leqslant r)\varphi_j(X)(f(X,Z) - \widetilde{f}(X,Z))^2}{f^3(X,Z)}\Big]^2\Big\}$$

$$+(1-n^{-1})\mathbb{E}\Big\{\Big[\int_0^r\int_0^1\frac{[S(z|x) - S(j,n,Z_l,X_l)]\varphi_j(x)(f(x,z) - \widetilde{f}(x,z))^2}{f^2(x,z)}dxdz\Big]^2\Big\}$$

$$\leqslant Cn^{-9/8} + \mathbb{E}\Big\{\Big[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}}(\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})(\kappa_{i_2s_2} - \widetilde{\kappa}_{i_2s_2})\nu_{i_1\pm i_2\pm j,s_1\pm s_2}$$

$$+2 \sum_{(i_1,s_1)\in\mathcal{N}} \sum_{(i_3,s_3)\notin\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})\kappa_{i_3s_3}\nu_{i_1\pm i_3\pm j,s_1\pm s_3}$$

$$+ \sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j,s_3\pm s_4}\Big]^2\Big\}$$

$$\leqslant C\Big[n^{-9/8} + \mathbb{E}\Big\{[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})(\kappa_{i_2s_2} - \widetilde{\kappa}_{i_2s_2})\nu_{i_1\pm i_2\pm j,s_1\pm s_2}]^2\Big\}$$

$$+[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j,s_3\pm s_4}]^2$$

$$+\mathbb{E}\Big\{[\sum_{(i_1,s_1)\in\mathcal{N}} \sum_{(i_3,s_3)\notin\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})\kappa_{i_3s_3}\nu_{i_1\pm i_3\pm j,s_1\pm s_3}]^2\Big\}\Big]$$

$$=: C[n^{-9/8} + B_{211} + B_{212} + B_{213}]. \quad (A.127)$$

In (A.127) we used notation

$$\nu_{i_1\pm i_2\pm j,s_1\pm s_2}$$

$$:= \int_{\mathcal{R}_r} \frac{[S(z|x) - S(j,n,Z_l,X_l)]\varphi_j(x)\varphi_{i_1}(x)\varphi_{i_2}(x)\psi_{s_1}(z)\varphi_{s_2}(z)}{f^2(x,z)} dxdz. \quad (A.128)$$

Now note that similarly to (A.122) we have

$$\max_{i_2,s_2} \sum_{i_1,s_1=0}^{\infty} \nu_{i_1\pm i_2\pm j,s_1\pm s_2} = o_j(1) + o_n(1). \quad (A.129)$$

Next we are evaluating terms on the right side of (A.127) in turn. Applying the Cauchy-Schwarz inequality and $\mathbb{E}\{(\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^4\} \leqslant Cn^{-2}$ we get

$$B_{211} \leqslant \mathbb{E}\Big\{\Big[\sum_{(i_1,s_1),(i_2,s_2)\in\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^2|\nu_{i_1\pm i_2\pm j,s_1\pm s_2}|\Big]^2\Big\}$$

$$= [o_j(1)+o_n(1)]\mathbb{E}\Big\{\Big[\sum_{(i_1,s_1)\in\mathcal{N}} (\kappa_{i_1s_1} - \widetilde{\kappa}_{i_1s_1})^2\Big]^2\Big\} = [o_j(1)+o_n(1)]n^{-1}. \quad (A.130)$$

Term B_{212} is evaluated with the help of the Cauchy-Schwarz inequality,

$$\Big[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa_{i_3s_3}\kappa_{i_4s_4}\nu_{i_3\pm i_4\pm j,s_3\pm s_4}\Big]^2$$

$$\leqslant \Big[\sum_{(i_3,s_3),(i_4,s_4)\notin\mathcal{N}} \kappa^2_{i_3s_3}|\nu_{i_3\pm i_4\pm j,s_3\pm s_4}|\Big]^2 = [o_j(1) + o_n(1)]n^{-1}. \quad (A.131)$$

Term B_{213} is evaluated via the Cauchy inequality and then using (A.130)-(A.131), and this yields $\mathbb{E}\{B_{213}\} = [o_j(1) + o_n(1)]n^{-1}$. Combining the obtained results in (A.127) we get

$$\mathbb{E}\{\widehat{B}^2_{21}\} = [1 + o_n(1)]\mathbb{E}\{\widetilde{B}^2_{21}\} = [o_j(1) + o_n(1)]n^{-1}. \quad (A.132)$$

Next we are considering $\widehat{B}_{22}$ on the right side of (A.125). Write,

$$\widehat{B}_{22} = n^{-1}\sum_{l=1}^{n}\frac{[\Delta'_l - S(j,n,Z_l,X_l)]I(Z_l \leqslant r)\varphi_j(X_l)(f(X_l,Z_l)-\widetilde{f}(X_l,Z_l))^3}{\widehat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$+n^{-1}\sum_{l=1}^{n}[\Delta'_l - S(j,n,Z_l,X_l)]I(Z_l \leqslant r)\varphi_j(X_l)$$

$$\times\frac{[(f(X_l,Z_l)-\widehat{f}(X_l,Z_l))^3-(f(X_l,Z_l)-\widetilde{f}(X_l,Z_l))^3]}{\widehat{f}(X_l,Z_l)f^3(X_l,Z_l)}$$

$$=: \widehat{B}_{221}+\widehat{B}_{222}. \qquad (A.133)$$

We begin with analysis of $\widehat{B}_{222}$. The estimate $\widehat{f}$ is bounded below from zero by $1/\ln(\ln(n+3))$, according to Lemma 2 the deviation $|f(x,z)-\widetilde{f}(x,z)|$ is larger than $Cn^{-1/8}$ with the probability decreasing exponentially in n, $\widehat{f}(x,z) \neq \widetilde{f}(x,z)$ only if $\widetilde{f}(x,z) < 1/\ln(\ln(n+3))$ and according to Lemma 2 the event occurs with the probability that decreases exponentially in n. This yields $\mathbb{E}\{\widehat{B}^2_{222}\} = o_n(1)n^{-1}$. Next,

$$\mathbb{E}\{\widehat{B}^2_{221}\} \leqslant Cn^{-9/8} + C[\ln(\ln(n+3))]^2 n^{-2/8}\mathbb{E}\Big\{n^{-1}\sum_{l=1}^{n}[\Delta'_l - S(j,n,Z_l,X_l)]$$

$$\times\Big[\frac{I(Z_l \leqslant r)\varphi_j(X_l)(f(X_l,Z_l)-\widetilde{f}(X_l,Z_l))^2}{f^3(X,Z)}\Big]^2\Big\}. \qquad (A.134)$$

The expectation on the right side of (A.134) is evaluated in (A.127)-(A.132), and we conclude that

$$\mathbb{E}\{B^2_{22}\} \leqslant Cn^{-9/8}. \qquad (A.135)$$

Using (A.131) and (A.135) in (A.125) we conclude that

$$\mathbb{E}\{\widehat{B}^2_2\} = [o_j(1)+o_n(1)]n^{-1}. \qquad (A.136)$$

Combining the obtained results we get

$$\mathbb{E}\{(\widetilde{\theta}_j - \widetilde{\theta}^*_j)^2\} = [o_j(1)+o_n(1)]n^{-1}. \qquad (A.137)$$

This is the result that shows feasibility of the plug-in methodology. The rest of the proof follows along lines of the proof of Theorem 3.2.

Theorem 4.1 is verified.

Proof of Theorem 6.1 Recall that for CSC triplet (X,Z,Δ) the following formula is valid,

$$f^{\mathbf{X},Z,\Delta}(\mathbf{x},z,0) = f^{\mathbf{X},Z}(\mathbf{x},z)\mathbb{P}(\Delta=0|\mathbf{X}=\mathbf{x},Z=z)$$

$$= f^{\mathbf{X},Z}(\mathbf{x},z)S^{T|\mathbf{X},Z}(z|\mathbf{x},z). \qquad (A.138)$$

Using the assumption about conditional independence of T and Z given $\mathbf{X}$, we continue (A.138) and get

$$f^{\mathbf{X},Z,\Delta}(\mathbf{x},z,0) = f^{\mathbf{X},Z}(\mathbf{x},z)S^{T|\mathbf{X}}(z|\mathbf{x}). \tag{A.139}$$

This formula allows us to evaluate the mean and variance of the proposed Fourier estimator (6.4). We begin with the mean. Write,

$$\mathbb{E}\{\widetilde{\theta}_{\mathbf{i}}\} = \mathbb{E}\Big\{\frac{\Delta'\varphi_{\mathbf{i}}(\mathbf{X})}{f^{\mathbf{X},Z}(\mathbf{X},Z)}\Big\} = \int_0^\infty \int_{[0,1]^k} S^{T|\mathbf{X}}(z|\mathbf{x})\varphi_{\mathbf{i}}(\mathbf{x})d\mathbf{x}dz. \tag{A.140}$$

Now note that

$$m(\mathbf{x}) = \mathbb{E}\{T|\mathbf{X}=\mathbf{x}\} = \int_0^\infty S^{T|\mathbf{X}}(z|\mathbf{x})dz. \tag{A.141}$$

Using this relation in (A.140) yields

$$\mathbb{E}\{\widetilde{\theta}_{\mathbf{i}}\} = \int_{[0,1]^k}\Big[\int_0^\infty S^{T|\mathbf{X}}(z|\mathbf{x})dz\Big]\varphi_{\mathbf{i}}(\mathbf{x})d\mathbf{x}$$

$$= \int_{[0,1]^k} m(\mathbf{x})\varphi_{\mathbf{i}}(\mathbf{x})d\mathbf{x} = \theta_{\mathbf{i}}. \tag{A.142}$$

We conclude that the proposed Fourier estimator is unbiased. Next we evaluate the variance,

$$\mathbb{V}(\widetilde{\theta}_{\mathbf{i}}) = \mathbb{E}\{(\widetilde{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} = n^{-1}\Big[\mathbb{E}\Big\{\Big[\frac{\Delta'\varphi_{\mathbf{i}}(\mathbf{X})}{f^{\mathbf{X},Z}(\mathbf{X},Z)}\Big]^2\Big\} - \theta_{\mathbf{i}}^2\Big]$$

$$= n^{-1}\Big[\int_0^\infty\Big[\int_{[0,1]^k}\frac{S^T(z|\mathbf{x})[\varphi_{\mathbf{i}}(\mathbf{x})]^2}{f^{\mathbf{X},Z}(\mathbf{x},z)}d\mathbf{x}\Big]dz - \theta_{\mathbf{i}}^2\Big] \leqslant Cn^{-1}. \tag{A.143}$$

In the last inequality we used the assumed inequality (6.2).

Now we can verify (6.5). First of all, note that $\varphi_{\mathbf{i}}$ are elements of the cosine tensor-product basis on $[0,1]^k$. Then the Parseval identity yields,

$$\mathbb{E}\Big\{\int_{[0,1]^k}(\widehat{m}(\mathbf{x})-m(\mathbf{x}))^2d\mathbf{x}\Big\} = \mathbb{E}\{\sum_{\mathbf{i}\in\mathcal{J}}(\widetilde{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} + \sum_{\mathbf{i}\notin\mathcal{J}}\theta_{\mathbf{i}}^2. \tag{A.144}$$

Here the first term is the integrated variance and the second term is the integrated squared bias of the proposed projection regression estimator. Using (A.143) we can bound the integrated variance,

$$\mathbb{E}\{\sum_{\mathbf{i}\in\mathcal{J}}(\widetilde{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} \leqslant Cn^{-1}\prod_{s=1}^k J_s$$

$$\leqslant Cn^{-1}n^{\alpha_*^{-1}/(2+\alpha_*^{-1})} = Cn^{-2\alpha_*/(2\alpha_*+1)}. \tag{A.145}$$

The integrated squared bias is evaluated using (6.1),

$$\sup_{m\in\mathcal{S}(\alpha_1,\ldots,\alpha_k)} \sum_{\mathbf{i}\notin\mathcal{J}} \theta_{\mathbf{i}}^2 \leqslant C \max_{s\in\{1,\ldots,k\}} J_s^{-2\alpha_s} \sum_{\mathbf{i}\notin\mathcal{J}} [1+\sum_{r=1}^{k} i_r^{2\alpha_r}]\theta_{\mathbf{i}}^2$$

$$\leqslant Cn^{-2/(2+\alpha_*^{-1})} = Cn^{-2\alpha_*/(2\alpha_*+1)}. \tag{A.146}$$

Using (A.145) and (A.146) in (A.144) verifies (6.5). Theorem 6.1 is proved.

Proof of Theorem 6.2. For the considered setting formula (A.139) holds. Using it we get

$$S^{T|\mathbf{X}}(t|\mathbf{x}) = \mathbb{E}\{\Delta'|Z=t,\mathbf{X}=\mathbf{x}\}. \tag{A.147}$$

What we see in (A.147) is the classical Bernoulli regression of Δ' on $k+1$ covariates $(Z,\mathbf{X})$. Accordingly, we cannot estimate $S^{T|X}(t|\mathbf{x})$ with a rate faster than the optimal $n^{-2\alpha_*/(2\alpha_*+1)}$ for the $(k+1)$-dimensional regression $\mathbb{E}\{\Delta'|Z=t,\mathbf{X}=\mathbf{x}\}$ based on direct observations from $((Z,\mathbf{X}),\Delta')$. This yields the lower bound for rate of the MISE. Next we show that our estimator attains this rate. We begin with analysis of the proposed Fourier estimator (6.15). For its mean we can write,

$$\mathbb{E}\{\hat{\theta}_{\mathbf{i}}\} = \mathbb{E}\Big\{\frac{\Delta'\varphi_{\mathbf{i}}(Z,\mathbf{X})}{f^{\mathbf{X},Z}(\mathbf{X},Z)}\Big\} = \int_{[0,1]^{k+1}} S^{T|\mathbf{X}}(z|\mathbf{x})\varphi_{\mathbf{i}}(z,\mathbf{x})dzd\mathbf{x} =: \theta_{\mathbf{i}}. \tag{A.148}$$

We conclude that the proposed Fourier estimator is unbiased. Next we evaluate its mean squared error,

$$\mathbb{E}\{(\hat{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} = n^{-1}\Big[\mathbb{E}\Big\{\Big[\frac{\Delta'\varphi_{\mathbf{i}}(Z,\mathbf{X})}{f^{\mathbf{X},Z}(\mathbf{X},Z)}\Big]^2\Big\} - \theta_{\mathbf{i}}^2\Big]$$

$$= n^{-1}\Big[\int_{[0,1]^{k+1}} \frac{S^{T|\mathbf{X}}(z|\mathbf{x})[\varphi_{\mathbf{i}}(z,\mathbf{x})]^2}{f^{\mathbf{X},Z}(\mathbf{x},z)}dzd\mathbf{x} - \theta_{\mathbf{i}}^2\Big] \leqslant Cn^{-1}. \tag{A.149}$$

Now we can use the Parseval identity and write,

$$\mathbb{E}\Big\{\int_{[0,1]^{k+1}} (\hat{S}^{T|\mathbf{X}}(t|\mathbf{x}) - S^{T|\mathbf{X}}(t|\mathbf{x}))^2 dzd\mathbf{x}\Big\}$$

$$= \mathbb{E}\{\sum_{\mathbf{i}\in\mathcal{J}}(\hat{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} + \sum_{\mathbf{i}\notin\mathcal{J}}\theta_{\mathbf{i}}^2. \tag{A.150}$$

Using (A.149) we get

$$\mathbb{E}\{\sum_{\mathbf{i}\in\mathcal{J}}(\tilde{\theta}_{\mathbf{i}}-\theta_{\mathbf{i}})^2\} \leqslant Cn^{-1}\prod_{s=1}^{k+1} J_s \leqslant Cn^{-2\alpha_*/(2\alpha_*+1)}, \tag{A.151}$$

and following (A.146) we get

$$\sup_{S^{T|\mathbf{X}}\in\mathcal{S}(\alpha_1,\ldots,\alpha_{k+1})} \sum_{\mathbf{i}\notin\mathcal{J}} \theta_{\mathbf{i}}^2 \leqslant Cn^{-2\alpha_*/(2\alpha_*+1)}. \tag{A.152}$$

Using the last two inequalities in (A.150) finishes the proof of Theorem 6.2.

Proof of Theorem 7.1. The assertion is established similarly to the proofs of Theorems 2.1 and 2.2. The only remark to make is that in line (A.3) the factor $[\psi(k/s)]^2$ appears in each sum. Theorem 7.1 is verified.

Proof of Theorem 7.2. Note that $\bar{\theta}_j$ is unbiased estimator of θ_j and

$$\mathbb{E}\{(\bar{\theta}_j - \theta_j)^2\} \leqslant d'_*(S^{T|X}, f^{X,Z}, r, \psi)n^{-1}(1 + o_j(1)). \tag{A.153}$$

Then the assertion follows from Lemma 3.1. Theorem 7.2 is proved.

References

[1] Ampesea, L., Sganzerla, W., Zieroa, H., Mudho, A., Martins, G. and Forster-Carneiroa, T. (2022). Research progress, trends, and updates on anaerobic digestion technology: A bibliometric analysis. *Journal of Cleaner Production*, 331: 130-154.

[2] Andersen, P., Hansen, M. and Klein, J. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis*, 10: 335-350.

[3] Becker, D., Braun, W. and White, B. (2017). Interval-censored unimodal kernel density estimation via data sharpening. *Journal of Statistical Computation and Simulation*, 87: 2023-2037.

[4] Cai, T. and Betensky, R. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* 59: 570-579.

[5] Chen, D., Sun, J. and Peace, K. (2012). *Interval-Censored Time-to-Event Data: Methods and Applications.* Boca Raton: CRC Press.

[6] Cheng G, and Wang, X. (2011) Semiparametric additive transformation model under current status data. *Electronic Journal of Statistic*, 5: 1735-1764.

[7] Cosslett, S. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrics*, 51: 765-782.

[8] Du, M., Hu, T. and Sun, J. (2019). Semiparametric probit model for informative current status data, *Statistics in Medicine.* https://doi.org/10.1002/sim.8106

[9] Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and Applications* 30: 557–568.

[10] Efromovich, S. (1989). On sequential nonparametric estimation of density. *Theory of Probability and Its Applications* 34: 228-239.

[11] Efromovich, S. (1996). On nonparametric regression for iid observations in a general setting. *Annals of Statistics* 26: 273-278.

[12] Efromovich, S. and Pinsker, M. (1996). Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica* 6: 925-942.

[13] Efromovich, S. (1999). *Nonparametric Curve Estimation.* New York: Springer.

[14] Efromovich, S. (2000). On sharp adaptive estimation of nonparametric curves. *Mathematical Methods of Statistics*, 9: 117-139.

[15] Efromovich, S. (2016). Minimax theory of nonparametric hazard rate estimation: efficiency and adaptation. *Annals of Institute of Mathematical Statistics*, 68: 25-75.

[16] Efromovich, S. (2018). *Missing and Modified Data in Nonparametric Estimation with R Examples,* Boca Raton: Chapman&Hall.

[17] Efromovich, S. (2019). On sharp nonparametric estimation of differentiable functions. *Statistics and Probability Letters* **152** 9-14.

[18] Efromovich, S. (2021). Sharp minimax distribution estimation for current status censoring with or without missing. *The Annals of Statistics*, 49: 568-589.

[19] Feng, Y. and Chen, Y. (2018). Oracle inequalities for the Lasso in the additive hazards model with interval-censored data. *Communications in Statistics - Theory and Methods*, 47: 2927-2949

[20] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics,* Cambridge Series in Statistical and Probabilistic Mathematics 38, New York: Cambridge Univ. Press.

[21] Groeneboom, P. and Hendrickx, K. (2018). Current status linear regression. *Annals of Statistics*, 46: 1415-1444.

[22] Grummer-Strawn, L. (1993). Regression analysis of current status data: An application to breast feeding. *Journal of the American Statistical Association*, 88: 758-765.

[23] Han, K. and Jung, I. (2022). Restricted mean survival time for survival analysis: a quick guide for clinical researchers. *Korean Journal of Radiology*, 23: 495-499.

[24] Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. With Discussion. *The Annals of Statistics*, 30: 325-396.

[25] Honda, T. (2004). Nonparametric regression with current status data. *Annals of the Institute of Statistical Mathematics*, 56: 49-72.

[26] Horowitz, J. and Savin, N. (2000). Binary response models: logits, probits and semiparametrics. *Journal of Economic Perspectives*, 15: 43-56.

[27] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*, 24: 540-566.

[28] Huang, J. and Wellner, J.A. (1997). Interval Censored Survival Data: A Review of Recent Progress. In: Lin, D.Y., Fleming, T.R. (eds) Proceedings of the First Seattle Symposium in Biostatistics. Lecture Notes in Statistics, vol 123. Springer, New York.

[29] Irwin, J. (1949). The standard error of an estimate of expectational life. *Journal of Hygiene*, 47: 188-189.

[30] Jewell, N. and van der Laan, M. (2004). Current status data: review, recent developments and open problems. In: *Advances in Survival Analysis*, eds. N. Balakrishnan and C.R. Rao, Amsterdam: Elsevier, 625-643.

[31] Kahane, J.-P. (1985). *Some Random Series of Functions*, Cambridge: Cambridge University Press.

[32] Karrison, T. (1987). Restricted mean life with adjustment for covariates, *Journal of the American Statistical Association*, 82: 1169-1176.

[33] Klein, J., van Houwelingen, H., Ibrahim, J. and Sheike, T. (2014). *Handbook of Survival Analysis.* Boca Raton: Chapman&Hall.

[34] Keogh, R. and Cox, D. (2014). *Case-Control Studies*. Cambridge: Cambridge Univ. Press.

[35] Kunatsa, T. and Xia, X. (2022). A review on anaerobic digestion with focus on the role of biomass codigestion, modeling and optimization on biogas production and enhancement. *Bioresource Technology*, 344: 126-163.

[36] Lam, K. and Xue, H. (2005). A semiparametric regression cure model with current status data.*Biometrika*, 92: 573-586.

[37] Li, G. and Zhang, C.-H. (1998). Linear regression with interval censored data. *Annals of Statistics*, 26: 1306-1327.

[38] Li, S., Wang, P. and Sun, J. (2017). Regression analysis of current status data in the presence of depended censoring with applications to tumorigenicity experiments. *Computational Statistics and Data Analysis*, 110: 75-86.

[39] Lin, D., Oakes, D, and Ying, Z. (1998). Additive hazard regression with current status data. *Biometrika*, 85: 289–298.

[40] Liu, H. and Qin, J. (2018). Semiparametric probit models with univariate and bivariate current status data. *Biometrics*, 74: 68-76.

[41] Lu, M. and McMahan, C. (2018). A partially linear proportional hazards model for current status data. *Biometrics*, 74: 1240-1249.

[42] Lu, S., Wu, J. and Lu, X. (2019). Efficient estimation of the varying-coefficient partially linear proportional odds model with current status data. *Metrika*, 82: 173-194.

[43] Ma, S. and Kosorok, M. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Annals of Statistics*, 33: 2256-2290.

[44] Ma, L., Hu, T. and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika*, 102: 731-738.

[45] Manski, C. and McFadden, D. (1981). *Structural Analysis of Discrete Data with Econometric Applications.* The MIT Press, Cambridge.

[46] Manski, C. (1988). Identification of binary response models. *JASA*, 83: 729-738.

[47] Malov, S. (2019). Nonparametric estimation for a current status right-censored data model. *Statistica Neerlandica*, 73: 475-495.

[48] Martinussen, T. and Scheike, T. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89: 649–658.

[49] McMahan, C., Wang, L. and Tebbs, J. (2013). Regression analysis for current status data using the EM algorithm. *Statistics in Medicine*, 32: 4452-4466.

[50] Murphy, S. A., van der Vaart, A. W. and Wellner, J. A. (1999). Current status regression. *Mathematical Methods of Statistic*, 8: 407–425.

[51] Nikolskii, S.M. (1975). *Approximation of Functions of Several Variables and Embedding Theorems* New York, Springer.

[52] Paritosh, K., Ydav, M., Chawade,A., Sahoo, D., and Parek, N. (2020). Additives as a Support Structure for Specific Biochemical Activity Boosts in Anaerobic Digestion: A Review. *Frontiers in Energy Research*

https://doi.org/10.3389/fenrg.2020.00088

[53] Qiu, J., Gu, E., Zhou, D., Lawrence, J., Bai, S. and H. Hung (2019) Estimation of conditional restricted mean survival time with counting process, *Journal of Biopharmaceutical Statistics*, 29: 800-809.

[54] Rabinowitz, D., Tsiatis, A. and Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, 82: 501-5013.

[55] Rabinowitz,D. and Jewell, N. (1996). Regression with doubly censored current status data. *Journal of Royal Statistical Society*, 58: 541–550.

[56] Rossini, A. and Tsiatis, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of American Statistical Association*, 91: 713-721.

[57] Shen, X. (2000). Linear regression with current status data. *Journal of American Statistical Association*, 95: 842–852.

[58] Shin, S., Liu, Y., Cole, S. and Fine, J. (2020). Ensemble estimation and variable selection with semiparametric regression models. *Biometrika*, 107: 433-448.

[59] Sun, J. (2005). Interval censoring. *Encyclopedia of biostatistics*, New York: Wiley Online Library.

[60] Sun, J. and Zhao, X. (2013). *The Statistical Analysis of Interval-Censored Failure Time Data.* New York: Springer.

[61] Susarla, V. and Van Ryzin, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *The Annals of Statistics*, 8: 1002-1016.

[62] Tian, L. and Cai, T. (2006). On the accelerated failure time model for current status and interval censored data. *Biometrika*, 93: 329-342.

[63] Train, K. (2009). *Discrete Choice Methods with Simulations.* Cambridge: Cambridge University Press.

[64] van der Laan, M. and Hubbard, A. (1997). Estimation with interval censored data and covariates. *Lifetime Data Analysis*, 3: 77-91.

[65] van der Laan, M., Bickel, P. and Jewell, N. (1997). Singly and doubly censored current status data: estimation, asymptotics and regression. *Scandinavian Journal of Statistics*, 24: 289-307.

[66] van der Vaart, A. (1991). On differentiable functionals. *The Annals of Statistics*, 19: 178-204.

[67] Wang, P., Tong, X. and Sun, J. (2018). A semiparametric regression cure model for doubly censored data. *Lifetime Data Analysis*, 24: 492-508.

[68] Wasserman, L. (2006). *All of Nonparametric Statistics.* New York: Springer.

[69] Windmeijer, F. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews*, 14: 101-116.

[70] Xue, H., Lam, K. and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of American Statistical Association*, 99: 346-356

[71] Xu, D., Zhao. S., Hu, T. and Sun, J. (2019). Regression analysis of informatively interval-censored failure time data with semiparametric linear transformation model. *Journal of Nonparametric Statistics*, 31: 663-679.

[72] Zhang, C., Wu, Y. and Yin, G. (2020). Restricted mean survival time for

interval-censored data. *Statistics in Medicine*, 39: 3879-3895.

[73] Zhao, H., Wu, Q., Li, G. and Sun, J. (2020). Simultaneous estimation and variable selection for interval censored data with broken adaptive ridge regression. *Journal of the American Statistical Association*, 115: 204-216.

[74] Zhong, Y. and Schaubel, D. (2022). Restricted mean survival time as a function of restricted time. *Biometrica*, 78: 192-201.

[75] Zhou, Q., Hu, T. and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112: 664-672.