WORKING PAPERS

N° 17-803

May 2017

"Honest confidence sets in nonparametric IV regression and other ill-posed models"

Andrii Babii

Toulouse School of Economics

Honest confidence sets in nonparametric IV regression and other ill-posed models

Andrii Babii*

November 15, 2016†

This paper provides novel methods for inference in a very general class of ill-posed models in econometrics, encompassing the nonparametric instrumental regression, different functional regressions, and the deconvolution. I focus on uniform confidence sets for the parameter of interest estimated with Tikhonov regularization, as in Darolles, Fan, Florens, and Renault (2011). I first show that it is not possible to develop inferential methods directly based on the uniform central limit theorem. To circumvent this difficulty I develop two approaches that lead to valid confidence sets. I characterize expected diameters and coverage properties uniformly over a large class of models (i.e. constructed confidence sets are honest). Finally, I illustrate that introduced confidence sets have reasonable width and coverage properties in samples commonly used in applications with Monte Carlo simulations and considering application to Engel curves.

Keywords: nonparametric instrumental regression, functional linear regression, density deconvolution, honest uniform confidence sets, non-asymptotic inference, ill-posed models, Tikhonov regularization

JEL classification: C14, C36

*Ph.D. Candidate, Toulouse School of Economics, Université Toulouse 1 Capitole, babii.andrii@gmail.com. I'm deeply indebted to my advisor Jean-Pierre Florens and to Eric Gautier and Ingrid Van Keilegom for many helpful suggestions and insightful conversations. This paper also benefited from discussions with Rohit Kumar, Elia Lapenta, Pascal Lavergne, and André Mas, Markus Reiss. I would also like to thank Christian Bontemps, Samuele Centorrino, Jasmin Fliegner, Emanuele Guerre, Vitalijs Jascisens, Jihyun Kim, Thierry Magnac, Nour Meddahi, Shruti Sinha, and other participants of TSE econometrics workshop for helpful comments and suggestions.

†First version: March 2016.

1 Introduction

This paper develops honest and uniform confidence sets for the structural function φ in a very broad class of ill-posed models treated with the Tikhonov regularization. The leading example is a nonparametric instrumental regression (NPIV) studied in Newey and Powell (2003), Florens (2003), Darolles et al. (2011), Hall and Horowitz (2005), and Blundell, Chen, and Kristensen (2007), and described as

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0, \tag{1}$$

where $Z \in \mathbf{R}^p$ is a vector of explanatory variables, $W \in \mathbf{R}^q$ is a vector of instruments (both continuously distributed), and φ is the function of interest. The NPIV model is ill-posed in the sense that the map from the distribution of the data to the function φ is not continuous. To have consistent estimator certain amount of regularization is needed to smooth out discontinuities.

In empirical studies the function φ represents a structural economic relation, such as Engel curve, cost function, or demand curve. To infer the magnitude of economic effects, it is not sufficient just to estimate this function. The magnitude of economic effects can only be inferred from confidence sets. This paper is the first to provide inferential methods for Tikhonov-regularized estimators. I focus on the *uniform inference*, which amounts to constructing a set containing the *entire* function φ with a high probability. Uniform inference allows to asses global features of the estimated function and to quantify the range of possible economic effects compatibly with the data. Global features may include the evidence for non-linearities, the amount of endogeneity bias comparing to the local polynomial estimator, monotonicity, concavity/convexity, or other shape properties. In contrast, pointwise confidence intervals only contain the value $\varphi(z_0)$ at some particular point z_0 with high probability and do not provide valid inference for the *entire* function φ. Another important feature of confidence sets constructed in this paper is *honesty* in the sense of Li (1989). It means that their coverage properties are valid uniformly over a large class of specified models[1]. Honesty is desirable since the underlying model is never known and coverage properties of dishonest sets may vary from one model to another[2].

Building uniform confidence sets for a function φ requires to approximate the distribution of the supremum of the variance of the estimator. I show that for a broad class of ill-posed models treated with Tikhonov regularization the variance of the estimator does not converge weakly in the space of continuous functions. As a result, *it is not possible* to build the uniform confidence set relying on the uniform central limit theorem. This calls for alternative approaches to inference.

[1]See Eq. (6) in Section 3 for a formal definition.

[2]Honest and dishonest sets are also called sets uniformly consistent in levels and pointwise consistent in levels.

In order to circumvent this problem I relax the requirement of asymptotically exact inference and focus on inference that only ensures asymptotic coverage with probability at least $1-\gamma$, for $\gamma \in (0,1)$. This relaxed coverage requirement is frequently used in econometrics, for instance, for weakly identified models in Andrews and Mikusheva (2016), moment inequalities models in Andrews and Soares (2010), or partial identification approach to NPIV model in Santos (2012).

To construct confidence sets with good coverage properties I rely on two different approaches. The first approach is to focus on a suitably normalized stochastic upper bound on the supremum of the variance of the estimator. This upper bound holds under quite general assumptions. For inference I obtain Gaussian approximation to this upper bound and use a suitable quantile of this Gaussian approximation to construct a confidence set for the function φ. This approach is flexible and valid for a large class of models.

The second inferential method developed in this paper relies on estimates of tail probabilities with a suitable concentration inequality. To the best of my knowledge this approach to inference for functions is new and this paper is the first to introduce it in econometrics. A simple illustration of this approach is to build the uniform confidence band for the empirical distribution function using the Dworetzky-Kiefer-Wolfowitz inequality. Given the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$ based on the i.i.d. sample $(X_i)_{i=1}^n$, the Dworetzky-Kiefer-Wolfowitz tells us that for any *finite sample size* n the probability of large deviations of F_n from the true distribution function F in the supremum norm declines at the exponential rate

$$\Pr\left(\|F_n - F\|_\infty > x\right) \leq 2e^{-2nx^2}, \qquad x > 0.$$

Setting $x = \sqrt{\frac{\log(2/\gamma)}{2n}}$ and $\gamma \in (0,1)$, the inequality becomes

$$\Pr\left(\|F_n - F\|_\infty \leq \sqrt{\frac{\log(2/\gamma)}{2n}}\right) \geq 1-\gamma.$$

This allows us to build the uniform confidence band for $F(x)$ with guaranteed coverage probability $1-\gamma$ taking $F_n(x) \pm \sqrt{\frac{\log(2/\gamma)}{2n}}$. In this simple example there is no coverage error (the coverage is at least $1-\gamma$ for any finite sample size n) and the diameter of the set shrinks at the rate $1/\sqrt{n}$. This inferential approach should be contrasted with the alternative asymptotic approach based on the Donsker central limit theorem, which also leads to confidence sets with diameters shrinking at the rate $1/\sqrt{n}$, but having the coverage level $1-\gamma-o(1)$. The coverage error disappears only as sample size goes to infinity. The empirical distribution function is an unbiased estimator, while the majority of estimators of functions are biased. As a result, usually confidence sets for such estimators, including ill-posed models considered in this paper, have coverage errors.

Concentration inequalities similar to the Dworetzky-Kiefer-Wolfowitz inequality for more complex statistics than $\|F_n - F\|_\infty$ are called Talagrand's concentration inequalities, see for instance Boucheron, Lugosi, and Massart (2013). In particular, I exploit the data-driven Talagrand-type concentration inequality for the supremum of the variance of the estimator to approximate quantiles of the unknown distribution. This approach does not rely on the existence of the supremum of the Gaussian process approximating the supremum of the variance of the estimator. As a result, it is valid for a wide class of data-generating processes and is especially useful in settings where all other approaches to inference fail. Lastly, they are significantly easier to implement and faster to compute than confidence sets based on the Gaussian approximation.

Whenever possible, I characterize convergence rates for coverage errors and expected diameters. It is especially important to know both rates, since consistency alone may not be very informative for inference on function φ. For instance, it may happen that the coverage error decreases slower than the confidence set shrinks, implying that larger sample sizes will be needed to achieve good coverage. I show that coverage errors of confidence sets based on the Gaussian approximation are driven by the bias of the estimator and not by the noise coming from the estimation of the operator.

Constructed confidence sets have excellent statistical properties. Their expected diameters and coverage errors may decrease at polynomial rates. Convergence rates for coverage errors are new, and have not been previously discussed neither for the NPIV model, nor for other ill-posed models considered in this paper.

Though, the Tikhonov-regularized NPIV is the leading example, the inferential methods developed in this paper are valid in other ill-posed models. This includes functional regression models and the density deconvolution model. To the best of my knowledge, despite the existing extensive literature on L_2 results for the functional regression models, no uniform convergence rates or uniform inferential methods are currently available. I provide more detailed comparison to the existing literature below.

Related literature. As was mentioned, this paper is the first to develop honest uniform inferential methods in a general and unifying framework, encompassing different ill-posed models treated with Tikhonov regularization. Previously, uniform confidence sets were studied only for sieve-type estimators of the NPIV model and the density deconvolution model. Moreover, neither honesty, nor rates for coverage errors were previously discussed. Horowitz and Lee (2012) develop uniform confidence bands for sieve NPIV estimator by first constructing pointwise confidence intervals at a finite grid of points and then, letting the number of grid points to grow at a certain speed to achieve uniform coverage. Chen and Christensen (2015) develop inferential methods for the sieve NPIV estimator without relying on discrete approximations. They focus on uniform inference for a very broad collection of linear and nonlinear functionals using Yurinskii's coupling and obtain uniform confidence bands in the special case of point evaluation functional, see also Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Tao (2014) for this approach. Chen and Christensen (2015) also propose the sieve score bootstrap procedure to construct uniform confidence bands. Lounici and Nickl (2011) obtain uniform confidence sets for the wavelet deconvolution estimator relying on the Bousquet's version of the Talagrand's concentration inequality.

Kato and Sasaki (2016) consider more general case, where the density of the noise is not known and is estimated from auxiliary sample. This paper studies the estimator based on Fourier inversion and builds on coupling inequalities developed in Chernozhukov, Chetverikov, Kato, et al. (2014) and Chernozhukov, Chetverikov, and Kato (2015).

Tikhonov-regularized estimators offer an appealing alternative to sieves. For instance, I show that confidence sets developed in this paper may enjoy polynomial convergence rates of coverage errors in mildly ill-posed and some severely ill-posed cases. At the same time, Tikhonov-regularized estimators do not require to specify sieve bases and change smoothly with respect to tuning-parameters. It is also known that even if the function φ is not identified, the Tikhonov-regularized estimator is still well-defined. In this case it converges to the best approximation to the true function φ in the orthogonal complement to the null space of the operator, see Florens, Johannes, and Van Bellegem (2011) and Babii and Florens (2016). This is especially important in the light of the fact that the identification condition in this class of models is not testable, see Canay, Santos, and Shaikh (2012). The behavior of sieve NPIV estimator in the non-identified case and coverage properties of confidence sets for sieve estimators is an open question[3]. Lastly, the computation of sieve estimators involves inversion of a random matrix, which may often be singular in applications and needs to be additionally regularized, e.g. with Tikhonov regularization. On the other hand, the Tikhonov-regularized estimator will always be well-defined.

It is also worth mentioning that to build a valid uniform confidence set, it is necessary to know uniform convergence rates of the corresponding estimator. Gagliardini and Scaillet (2012) obtain uniform convergence rates for the Tikhonov-regularized minimum distance estimator using

[3] It is known though that the penalized sieve estimator with strictly convex penalty function is consistent for some unique element in the identified set, Chen and Pouzo (2012).

Sobolev embeddings. Chen and Christensen (2015) derive minimax-optimal uniform convergence rates and show that sieve NPIV estimator can attain these rates with appropriately selected bases. Mbakop (2015) obtains convergence rates in probability for the sup-norm loss function in the Tikhonov-regularized NPIV model. Convergence rates obtained in this paper are of order of powers of $\log n/n$. To obtain good coverage properties of confidence sets, I encountered the necessity of stronger result on convergence rates of the risk of the estimator (i.e. the expected value of the sup-norm). I show that under the mild restriction, uniform convergence rates of the Tikhonov regularized estimator are of order of powers of $1/n$, which sharpens the result of Mbakop (2015). These rates are valid for mildly ill-posed and some severely ill-posed cases[4].

Lastly, some pointwise inferential results are available for spectral cut-off estimators when the ill-posed operator is known and is not estimated from the data, see Carrasco and Florens (2011), Gautier and Kitamura (2013), and Florens, Horowitz, and Keilegom (2016). In this paper I consider more realistic setting of the nonparametric IV and functional regressions, when the operator is not known and is estimated from the data. At the same time Chen and Pouzo (2015) provide pointwise inference and bootstrap confidence bands for conditional moment restriction models treated with sieve approach and nesting the NPIV model as a special case. Meanwhile Cardot, Mas, and Sarda (2007) provide results on the asymptotic normality of linear functionals, which solves the prediction problem for functional regression models. Lastly, Carrasco, Florens, and Renault (2013) study asymptotic normality under the L_2 norm for a fixed value of the regularization parameter α_n.

The paper is organized as follows. I introduce notation in the remaining part of this section. Section 2 introduces motivating examples for which I provide inferential methods. In Section 3 I describe the problem of constructing honest uniform confidence sets and introduce two inferential approaches. Under a general set of assumptions, that are verified later on in each particular application, I establish convergence rates for coverage errors and diameters of constructed sets, uniform over a general set of models. Section 4 introduces specific smoothness and regularity conditions, while Section 5 applies results of Section 3 to the NPIV model, different functional regression models, and density deconvolution. In Section 6 I show how to implement confidence sets in practice for the NPIV estimator and explore their finite-sample properties with Monte Carlo experiments. Section 7 considers the empirical application to Engel curves and Section 8

[4]It is known that minimax-optimal uniform convergence rates in the severely ill-posed case are of order of powers of $1/\log n$, see Chen and Christensen (2015). However, in the more restricted smoothness class, when the estimated function can be described by the finite number of generalized Fourier coefficients (with respect to the SVD basis), or more generally is analytic, it is possible to achieve faster uniform convergence rates of order of powers of $1/n$. This information is especially helpful since economic relations are very often believed to be represented by smooth functions.

concludes.

Notation. Let $L_2[a,b]^p$ denote the space of functions on some compact set $[a,b]^p \subset \mathbf{R}^p$, square integrable with respect to the Lebesgue measure λ. For $\varphi \in L_2[a,b]^p$, let $\|.\|$ denote the usual L_2 norm derived from the inner product $\langle .,. \rangle$. Let $C[a,b]^p$ denote the space of continuous functions endowed with supremum norm $\|.\|_\infty$. For some positive real number β, let $C^\beta_M[a,b]^p$ denote the class of β-Hölder functions on $(a,b)^p$ with $0 < M < \infty$

$$C^\beta_M[a,b]^p = \left\{ \varphi \in C[a,b]^p : \max_{|k| \le \lfloor \beta \rfloor} \|f^{(k)}\|_\infty \le M, \max_{|k| = \lfloor \beta \rfloor} \sup_{z \neq z'} \frac{\left|\varphi^{(k)}(z) - \varphi^{(k)}(z')\right|}{\|z - z'\|^{\beta - \lfloor \beta \rfloor}} \le M \right\},$$

where $k = (k_1, \dots, k_p) \in \mathbf{N}^p$ is a multi-index, $|k| = \sum_{j=1}^p k_j$, $\varphi^{(k)}(z) = \frac{\partial^{|k|}\varphi(z)}{\partial z_1^{k_1} \dots \partial z_p^{k_p}}$, and $\lfloor \beta \rfloor$ is the largest integer strictly smaller than β. Let $\mathcal{L}_2$, $\mathcal{L}_{2,\infty}$ and $\mathcal{L}_\infty$ be spaces of bounded linear operators from L_2 to L_2, from L_2 to C, and from C to C respectively. Sets on which functions are defined should be clear from the context. Spaces $\mathcal{L}_2, \mathcal{L}_{2,\infty}$, and $\mathcal{L}_\infty$ are endowed with standard operator norms, denoted by $\|K\| = \sup_{\|\varphi\| \le 1} \|K\varphi\|$, $\|K\|_{2,\infty} = \sup_{\|\varphi\| \le 1} \|K\varphi\|_\infty$, and $\|K\|_\infty = \sup_{\|\varphi\|_\infty \le 1} \|K\varphi\|_\infty$, respectively. For $K \in \mathcal{L}_2$, let K^* denote its Hilbert adjoint operator. Let $\mathcal{R}(T)$ and $\mathcal{D}(T)$ be respectively the range and the domain of the operator T. Lastly, for two real numbers a and b, I denote $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

2 Motivating examples

2.1 Nonparametric IV

The NPIV model, described in Eq. (1) is

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

The model leads to the following ill-posed equation

$$\mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w]. \tag{2}$$

Florens (2003) and Darolles et al. (2011) use Tikhonov regularization to build the estimator which solves the regularized empirical counterpart to this equation. In this case both the conditional mean-function $r(w) = \mathbb{E}[Y|W = w]$ and the conditional expectation operator T are estimated non-parametrically. In this paper I slightly modify this approach, multiplying Eq. (2) by the density function f_W

$$r(w) := \mathbb{E}[Y|W = w]f_W(w) = \int_{[a,b]^p} \varphi(z)f_{ZW}(z,w)\mathrm{d}z =: (T\varphi)(w).$$

The advantage of working with this equation is that it does not require handling random denominators in the kernel estimation of conditional mean functions, which simplifies our development of uniform asymptotic results.

For simplicity, I use product kernel estimator with equal bandwidth parameters $h_n \to 0$ as $n \to \infty$ for all coordinates to estimate the joint density f_{ZW}. Then, the estimators of different components of the model are

$$\begin{aligned}
\hat{r}(w) &= \frac{1}{nh_n^q}\sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right) \\
\hat{f}_{ZW}(z,w) &= \frac{1}{nh_n^{p+q}}\sum_{i=1}^n K_z\left(h_n^{-1}(Z_i - z)\right)K_w\left(h_n^{-1}(W_i - w)\right) \\
(\hat{T}\varphi)(w) &= \int_{[a,b]^p} \varphi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z.
\end{aligned}$$

The adjoint operator $\hat{T}^*$ is defined as a solution to $\langle \hat{T}\varphi, \psi\rangle = \langle \varphi, \hat{T}^*\psi\rangle$ and can be computed, applying Fubini's theorem to change the order of integration

$$(\hat{T}^*\psi)(z) = \int_{[a,b]^q} \psi(w)\hat{f}_{ZW}(z,w)\mathrm{d}w.$$

This operator is used in the expression of the Tikhonov-regularized estimator, see the following Section 3.

2.2 Functional regressions

In the functional linear regression, the real dependent variable Y is explained by the continuous-time stochastic process $Z(t), t \in [0,1]^p$. The distinctive feature of this class of models is that it allows to handle high-dimensional data without relying on the *sparsity* assumption, which may be restrictive in this setting. There is vast literature on estimation and testing for the functional regression models in statistics, see Hall, Horowitz, et al. (2007), Fan, James, Radchenko, et al. (2015), Comte, Johannes, et al. (2012), Cardot, Ferraty, Mas, and Sarda (2003), and some growing literature in econometrics, see Florens and Van Bellegem (2015), Benatia, Carrasco, and Florens (2015), Babii (2016). However all these papers focus on the estimation in the L_2 norm and to the best of my knowledge, there are no currently available uniform inferential methods for functional regression models.

The functional IV regression model is described by

$$Y = \int_{[0,1]^p} \varphi(t)Z(t)\mathrm{d}t + U, \qquad \mathbb{E}[UW(s)] = 0, \qquad \forall s \in [0,1]^q.$$

The slope parameter φ measures the strength of the impact of the process Z at different points $t \in [0,1]^p$. If $W = Z$, we obtain the classical functional linear regression model without endogeneity, see, for instance, Hall et al. (2007), while in the IV case of Florens and Van Bellegem (2015), W is some functional instrumental variable, uncorrelated with the error term.

The uncorelatedness assumption leads to the ill-posed equation

$$r(s) := \mathbb{E}[YW(s)] = \int_{[0,1]^p} \varphi(t)\mathbb{E}[Z(t)W(s)]\mathrm{d}t =: (T\varphi)(s).$$

The slope parameter φ is identified when the covariance operator is 1-1, which generalizes the non-singularity condition for the covariance matrix in the finite-dimensional linear regression model.

A variation of this model is studied in Babii (2016), where the identification is achieved with real-valued instrumental variable through the conditional moment restriction $\mathbb{E}[U|W] = 0$. The identifying restriction in this case is the linear completeness condition. Conditional mean-independence leads to the following ill-posed equation

$$\mathbb{E}[Y|W] = \int_{[0,1]^p} \varphi(t)\mathbb{E}[Z(t)|W]\mathrm{d}t. \tag{3}$$

Consider now a function Ψ such that the map $\psi(W) \mapsto \mathbb{E}[\psi(W)\Psi(s,W)]$ is 1-1, see Stinchcombe and White (1998). Applying this map to Eq. (3) leads to another ill-posed equation

$$r(s) := \mathbb{E}[Y\Psi(s,W)] = \int_{[0,1]^p} \varphi(t)\mathbb{E}[Z(t)\Psi(s,W)] =: (T\varphi)(s).$$

In what follows we denote by $W(s)$ either the regressor $Z(s)$, some functional IV $W(s)$, or the function $\Psi(s,W)$ of some real IV W. This notation allows to encompass three functional regression models: the model of Hall et al. (2007), the model of Florens and Van Bellegem (2015), and the model studied in Babii (2016).

Unlike the NPIV, which requires non-parametric estimation of the joint density function, all components of functional regression models can be estimated at the parametric rate using sample analogs to population moments

$$\hat{r}(s) = \frac{1}{n}\sum_{i=1}^{n} Y_i W_i(s), \qquad \hat{k}(t,s) = \frac{1}{n}\sum_{i=1}^{n} Z_i(t) W_i(s).$$

Operator $\hat{T}$ can be estimated as

$$(\hat{T}\varphi)(s) = \int_{[0,1]^p} \varphi(t)\hat{k}(t,s)\mathrm{d}t$$

The adjoint operator $\hat{T}^*$ can be obtained using the Fubini's theorem

$$(\hat{T}^*\varphi)(t) = \int_{[0,1]^q} \psi(s)\hat{k}(t,s)\mathrm{d}s.$$

2.3 Density deconvolution

Often economic data are not not measured precisely. Density deconvolution allows to estimate the density of unobserved data from the data measured with errors. Density deconvolution is encountered in a variety of econometric applications, e.g. to the earning dynamics in Bonhomme and Robin (2010), to panel data in Evdokimov (2010), or to instrumental regression in Adusumilli, Otsu, et al. (2015). In the simplest example of this model, we have some noisy scalar observations Y, of the latent variable Z, contaminated by measurement errors U

$$Y = Z + U, \qquad Z \perp\!\!\!\perp U.$$

Distributions of both Z and U are assumed to be absolutely continuous with respect to the Lebesgue measure with corresponding densities φ and f. The density of measurement errors f is assumed to be known. The goal is to recover the density of the latent variable Z from observing contaminated i.i.d. sample $(Y_i)_{i=1}^n$.

Independence and additivity of the noise imply that the density function r of Y satisfies the following convolution equation

$$r(y) = \int \varphi(z)f(y-z)\mathrm{d}z =: (T\varphi)(y), \tag{4}$$

where the operator $T : L_2 \to L_2$ in Eq. (4) is compact, whenever the density f is compactly supported. Carrasco and Florens (2011) replace Lebesgue measure by measures which yield compactness of T and study Tikhonov regularization, obtaining L_2 convergence rates and pointwise asymptotic normality. For simplicity of presentation and tractability of our results, I assume that all densities are continuous, bounded and compactly supported, with support contained inside the interval $[a, b] \subset \mathbf{R}$. This assumption is not very restrictive, since most of economic variables, e.g. reported earning, are bounded.

The adjoint operator to T can be computed by Fubini's theorem

$$(T^*\psi)(z) = \int \psi(y)f(y-z)\mathrm{d}y.$$

Notice hat operators T and T^* are known in this problem because f is known. Since r is a density function, we have $(T^*r)(z) = \mathbb{E}[f(Y-z)]$. The only component of the model that needs to be estimated to obtain Tikhonov-regularized estimator is

$$(\widehat{T^*r})(z) = \frac{1}{n}\sum_{i=1}^{n} f(Y_i - z).$$

3 Honest confidence sets

In this section we first discuss two stochastic processes that will drive the distribution of Tikhonov-regularized estimators and show that they do not converge weakly in the space of continuous functions under the uniform topology. We describe the problem of building honest uniform confidence sets and introduce two core approaches of this paper: the Gaussian approximation to upper bounds and the concentration inequality approach. We introduce two sets of general assumptions that allow us to characterize coverage errors and diameters of constructed confidence sets and state main results of the paper.

3.1 Impossibility of weak convergence

We first discuss the impossibility of using the uniform central limit theorem to obtain the distribution of the estimator. To fix notation through the rest of the paper, let Pr be a probability measure to any of ill-posed model introduced in the previous section. The model is described by the function equation $r = T\varphi$, where $\varphi : [a, b]^p \to \mathbf{R}$ is a functional parameter of interest. We aim to construct a random set C_n such that it contains the function φ with probability close to $1 - \gamma$ for $\gamma \in (0, 1)$ and that its expected diameter shrinks as the sample size increases at a certain rate. We focus on confidence sets for Tikhonov-regularized estimator defined as follows

$$\hat{\varphi}_{\alpha_n} = (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{r},$$

where $\hat{T}, \hat{T}^*$, and $\hat{r}$ are appropriate estimators[5] and α_n is some positive sequence converging to zero as $n \to \infty$. This estimator belongs to the general family of spectral regularization schemes, see Carrasco, Florens, and Renault (2007).

It will be seen in the following section that each of the three motivating examples, the leading stochastic component of the Tikhonov-regularized estimator is one of the following two stochastic processes driven by the i.i.d. centered random functions $(X_{1,i})_{i=1}^n$ and $(X_{2,i})_{i=1}^n$

$$\nu_{1,n} = \frac{1}{n}\sum_{i=1}^{n}(\alpha_n I + T^*T)^{-1}T^*X_{1,i}, \quad \text{or} \quad \nu_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\alpha_n I + T^*T)^{-1}X_{2,i}, \tag{5}$$

We assume that T is some integral operator with continuous kernel function, so that T, T^*, and T^*T map to the space of continuous functions. Thus, we can also think of the operator T as acting between $(C, \|.\|_\infty)$ spaces. This will be useful to study uniform confidence sets. On the other hand, considering T as an operator between L_2 spaces will allow to define the adjoint operator T^* and to use effectively the spectral theory and functional calculus accessible in Hilbert spaces. By Lemma 1 in the Appendix C, the operator $(\alpha_n I + T^*T)$ is invertible between spaces of continuous functions. Therefore, both processes have trajectories in the space $(C, \|.\|_\infty)$.

In order to build uniform confidence sets, we need to approximate the distribution of the supremum of these two processes. The simplest possible route to achieve this, would be to establish weak convergence of suitably normalized processes in Eq. (5) to some Gaussian processes and then to rely on quantiles of corresponding Gaussian suprema to build uniform confidence sets. Unfortunately, both processes do not converge weakly as random elements in $(C, \|.\|_\infty)$ space, as illustrated in the following proposition.

Proposition 1. *Suppose that the inverse of the operator T^*T defining stochastic processes in Eq. (5) is unbounded. Then there does not exist a normalizing sequence r_n such that $r_n\nu_{1,n}$ or $r_n\nu_{2,n}$ would converge weakly in $(C, \|.\|_\infty)$ to a non-degenerate random process.*

Proposition 1 tells us that it is not possible to rely on the asymptotic approximation with conventional central limit theorem. As a result, there does not exists a trivial way to build a uniform confidence set for the structural function φ with asymptotically exact confidence level. Given the difficulty of inference with exact coverage requirement, in this paper we relax this requirement.

[5] If some of operators are known, which is the case in the density deconvolution model, estimators are replaced by known quantities.

3.2 Honest confidence sets

We describe the problem of building the honest confidence sets. We focus on uniform confidence sets, honest to some class of models $\mathcal{F}$[6]. For a given level $\gamma \in (0,1)$, the *honest* $1-\gamma$ *uniform confidence set*, denoted $C_n = \{C_n(z) = [C_l(z), C_u(z)]\,,\ z \in [a,b]^p\}$, should satisfy

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi(z) \in C_n(z),\ \forall z \in [a,b]^p\right) \geq 1-\gamma - O(r_n), \tag{6}$$

for some sequence $r_n \to 0$. Honesty is necessary to produce confidence sets of practical value. It ensures that for a sufficiently large sample size n, not depending on the model (φ,T), the coverage of the set will be close to $1-\gamma$. In contrast, a *dishonest* set requires a weaker condition

$$\inf_{(\varphi,T)\in\mathcal{F}} \liminf_{n\to\infty} \Pr\left(\varphi(z) \in C_n(z),\ \forall z \in [a,b]^p\right) \geq 1-\gamma,$$

so that the sample size n needed for coverage close to $1-\gamma$ will depend on the unknown function[7] φ and the unknown operator T.

The most interesting choice of confidence set is such that its expected diameter under the supremum norm shrinks at the rate $\bar{r}_n$, which is the rate at which we estimate the function φ

$$\mathbb{E}\left|C_n\right|_\infty \equiv \mathbb{E}\left\|C_u - C_l\right\|_\infty = O(\bar{r}_n).$$

Uniform confidence sets, unlike the one based on other metrics, such as the L_2 distance, have appealing visualization and are easy to implement numerically. For example, if φ is the function on the real line, the confidence set becomes a band on the plane, which contains the whole graph of the function with probability close to $1-\gamma$.

3.3 Gaussian approximation and concentration

In this section we introduce two main approaches of this paper. Consider the following two processes

$$\hat{\nu}_{1,n} = \frac{1}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* X_{1,i}, \quad \text{or} \quad \hat{\nu}_{2,n} = \frac{1}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1} X_{2,i}. \tag{7}$$

The first idea suggested in this paper is to focus on some upper bound on suitably normalized

[6] The class of models will be introduced in Section 4.

[7] See also (Tsybakov, 2009, p.16-19) for the discussion why fixed (φ,T) results do not lead to a consistent notion of optimality in non-parametric problems and why it is necessary to consider results uniform over some smoothness class $\mathcal{F}$.

suprema $\|\hat{\nu}_{1,n}\|_\infty$ or $\|\hat{\nu}_{2,n}\|_\infty$, whose distribution can be easily approximated with the distribution of certain norms of Gaussian processes. It follows from the spectral theory, that for a bounded linear operator T and continuous function f on $[0, \|T\|^2]$, we have $f(T^*T)T^* = T^*f(TT^*)$, see Engl, Hanke, and Neubauer (1996). Using this fact, factorizing the operator norm, and noticing that $\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\| \leq \alpha_n^{-1}$, for any $n \in \mathbf{N}$ we have

$$\frac{\alpha_n n^{1/2}}{\|\hat{T}^*\|_{2,\infty}}\|\hat{\nu}_{1,n}\|_\infty = \frac{\alpha_n}{\|\hat{T}^*\|_{2,\infty}}\left\|\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{1,i}\right\|_\infty \leq \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{1,i}\right\|. \tag{8}$$

The latter distribution can be approximated invoking the CLT in the Hilbert space and continuous mapping theorem

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{1,i}\right\| \xrightarrow{d} \|\mathbb{G}_1\|,$$

where $\mathbb{G}_1$ is a zero-mean Gaussian process with covariance function $(w, w') \mapsto \mathbb{E}[X_{1,i}(w)X_{1,i}(w')]$ and $w, w' \in [a, b]^q$.

Similarly, by Lemma 1 in the Appendix C

$$\frac{\alpha_n^{3/2} n^{1/2}}{\|\hat{T}^*\|_{2,\infty}/2 + \alpha_n^{1/2}}\|\hat{\nu}_{2,n}\|_\infty \leq \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right\|_\infty, \tag{9}$$

and we can approximate the distribution in the right-hand side with uniform CLT

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right\|_\infty \xrightarrow{d} \|\mathbb{G}_2\|_\infty,$$

where $\mathbb{G}_2$ is a zero-mean Gaussian process with covariance function $(z, z') \mapsto \mathbb{E}[X_{2,i}(z)X_{2,i}(z')]$ and $z, z' \in [a, b]^p$. To obtain explicit convergence rates for coverage error, instead of relying on uniform central limit theorems, we will use suitable Berry-Esseen results. The distributions of both $\|\mathbb{G}_1\|$ and $\|\mathbb{G}_2\|_\infty$ are not pivotal, but it is possible to estimate the covariance structure and to simulate processes in order to obtain needed quantiles. Alternatively, we can rely on estimates of tail probabilities based Gaussian concentration or to rely on the empirical or exchangeable bootstrap, see (Van Der Vaart and Wellner, 2000, Chapter 3.6). Consistency of the multiplier bootstrap with Rademacher multiples follows from Lemma 3 and standard arguments for the multiplier bootstrap consistency for empirical processes. More detailed study of bootstrap properties is beyond the scope of this paper and is left for future research.

As an alternative to Gaussian approximation we study confidence sets based on non-asymptotic estimates of tail probabilities of the distribution of $\|\nu_{1,n}\|_\infty$ and $\|\nu_{2,n}\|_\infty$. To that end we will rely on the data-driven Talagrand-type concentration inequality, to achieve the required coverage level.

To describe confidence sets constructed with both approaches, we shall introduce additional notation. We allow for the process $X_{1,i}$ to change with the sample size n. In particular, we assume that for some i.i.d. sequence $(X'_{1,i})$ of zero-mean processes in $C[a,b]^q$ not depending on n and some known sequence of strictly positive real numbers u_n, we have $X_{1,i} = u_n X'_{1,i}$. On the other hand, the process $X_{2,i}$ is assumed not to change with n. Since stochastic processes $X'_{1,i}$ and $X_{2,i}$ may be not readily available, we denote by $\hat{X}'_{1,i}$ and $\hat{X}_{2,i}$ their respective consistent estimators and by $\hat{F}_1 = \max_{1\leq i\leq n} \|\hat{X}'_{1,i}\|$ and $\hat{F}_2 = \max_{1\leq i\leq n} \|\hat{X}_{2,i}\|_\infty$ the corresponding estimators of their envelops. Finally, for some i.i.d. sequence of Rademacher random variables $(\varepsilon_i)_{i=1}^n$, independent from the data, we denote by $\left\|\hat{\nu}^\varepsilon_{1,n}\right\|_\infty$ and $\left\|\hat{\nu}^\varepsilon_{2,n}\right\|_\infty$ the estimates of the suprema of symmetrized processes

$$\left\|\hat{\nu}^\varepsilon_{1,n}\right\|_\infty = \left\|\frac{1}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\varepsilon_i \hat{X}_{1,i}\right\|_\infty, \qquad \left\|\hat{\nu}^\varepsilon_{2,n}\right\|_\infty = \left\|\frac{1}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\varepsilon_i \hat{X}_{2,i}\right\|_\infty.$$

Confidence sets of level $1-\gamma$ with $\gamma\in(0,1)$ are described as follows

$$C^{j,s}_{1-\gamma,n}(z) = \left[\hat{\varphi}_{\alpha_n}(z) - q^{j,s}_{1-\gamma,n}, \hat{\varphi}_{\alpha_n}(z) + q^{j,s}_{1-\gamma,n}\right], \qquad z\in[a,b]^p, \quad j\in\{1,2\}, \quad s\in\{\mathrm{g},\mathrm{ci}\}, \tag{10}$$

where s $\in\{\mathrm{g},\mathrm{ci}\}$ denotes sets based on the Gaussian approximation to the upper bound or sets based on the concentration inequality, $j\in\{1,2\}$ denotes sets for each of the two processes in Eq. (5), and

$$\begin{aligned} q^{1,\mathrm{g}}_{1-\gamma,n} &= \frac{c^{1/2}_{1,1-\gamma}\|\hat{T}^*\|_{2,\infty} + c}{\alpha_n n^{1/2}}, & q^{1,\mathrm{ci}}_{1-\gamma,n} &= 2\left\|\hat{\nu}^\varepsilon_{1,n}\right\|_\infty + \frac{3\|\hat{T}^*\|_{2,\infty}\hat{F}_1\sqrt{2\log(2/\gamma)} + c}{\alpha_n n^{1/2}} u_n, \\ q^{2,\mathrm{g}}_{1-\gamma,n} &= \frac{c_{2,1-\gamma}\left(\|\hat{T}^*\|_{2,\infty}/2 + \alpha_n^{1/2}\right) + c}{\alpha_n^{3/2} n^{1/2}}, & q^{2,\mathrm{ci}}_{1-\gamma,n} &= 2\left\|\hat{\nu}^\varepsilon_{2,n}\right\|_\infty + \frac{3\left(\|\hat{T}^*\|_{2,\infty}/2 + \alpha_n^{1/2}\right)\hat{F}_2\sqrt{2\log(2/\gamma)} + c}{\alpha_n^{3/2} n^{1/2}}. \end{aligned}$$

Here $c_{1,1-\gamma}$ and $c_{2,1-\gamma}$ are $1-\gamma$ quantiles of norms of Gaussian processes $\|\mathbb{G}_1\|^2$ and $\|\mathbb{G}_2\|_\infty$ respectively and c is some positive constant.

3.4 Coverage properties and diameters

In what follows, we consider two sets of regularity conditions for the main results of the paper. These results will not be stated in the fullest possible generality, but rather at the level sufficient to cover all examples of interest. The first set of regularity conditions imposes mild restrictions on the data.

Assumption 1. *Suppose that for all models in some class $\mathcal{F}$ we have (i) $(X_{1,i})_{i=1}^n$ and $(X_{2,i})_{i=1}^n$ are sequences of i.i.d. centered random functions in $C[a,b]^q$ or $C[a,b]^p$ respectively, where $X_{2,i}$ does not depend on n; (ii) $X_{1,i} = u_n X'_{1,i}$, where $u_n^{-1} = O(1)$; (iii) trajectories of $X_{2,i}$ are in $C^s_M[a,b]^p, s \in (0,1]$ a.s.; (iv) $X_{1,i}$ does not depend on n and $\mathbb{E}\|X_{1,i}\|^3 \leq C_1 < \infty$, $\mathbb{E}\|X_{2,i}\|_\infty^3 \leq C_2 < \infty$; (v) T is 1-1 integral operator with continuous kernel function.*

Assumption (i) could be relaxed to weakly dependent environments where the functional Berry-Esseen and the data-driven concentration inequalities hold. Assumption (ii) is needed to accommodate the NPIV model, in which case $u_n = h_n^{-q}$. Assumption (iii) reduces to the mild smoothness restriction on certain densities in the NPIV and deconvolution models, or on stochastic processes in functional regression models. We shall note that the existence of the third moment in assumption (iv) is needed for functional Berry-Esseen theorems in the Gaussian approximation approach. It could be relaxed to the finiteness of the second moment at costs of not having explicit rates for coverage errors. On the other hand, for the construction with concentration inequality, we need the stronger assumption (iv) of finite envelops. We should also stress that if assumption (v) fails and T is not injective, the Tikhonov-regularized estimator $\hat{\varphi}_{\alpha_n}$ is still well-defined and converges to the best approximation to the function φ, see Florens et al. (2011) and Babii and Florens (2016). In this case, the resulting confidence set will be valid for this best approximation.

The second set of assumptions will be satisfied whenever the first set of assumptions holds. We will illustrate this in each particular application in Section 5.

Assumption 2. *There exist some sequences $r_{j,n} \to 0$ as $n \to \infty$ for $j = 1,2$ such that uniformly over some class of models $\mathcal{F}$, we have (i) $\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty \leq \|\hat{\nu}_{j,n}\|_\infty + \eta_{j,n}$, where processes $\hat{\nu}_{j,n}, j = 1,2$ are as in Eq. (7) and $\mathbb{E}\eta_{j,n} = O(r_{j,n})$; (ii) $\mathbb{E}\|\hat{T}^* - T^*\|_{2,\infty}^4 = o\left(\alpha_n^2\right)$; (iii) $\mathbb{E}\max_{1\leq i\leq n}\|\hat{X}'_{1,i} - X'_{1,i}\|^2 = o\left(1\right)$ and $\mathbb{E}\max_{1\leq i\leq n}\|\hat{X}_{2,i} - X_{2,i}\|_\infty^2 = o(1)$.*

Assumption 2 (i) tells us that we can decompose the sup-norm loss function for the estimator into the leading variance component $\|\hat{\nu}_{j,n}\|_\infty$ and the remainder $\eta_{j,n}$ that converges to zero at the rate $r_{j,n}$. Assumption 2 (ii) imposes certain requirement on the rate at which the regularization parameter α_n can converge to zero and, in particular, it ensures that the noise coming from the estimation of the operator T^* is sufficiently small for inferential purposes. As regards Assumption 2 (iii), it ensures that envelops for processes $X'_{1,i}$ and $X_{2,i}$ can be well-estimated.

Given the regularity conditions discussed above, we can state the first main result of the paper. It describes convergence rates for coverage error of constructed confidence sets based on the data-driven Talagrand-type concentration inequality (see also Appendix A for discussion of this inequality).

Theorem 1. *Suppose that Assumptions 1 (i), (v), and Assumptions 2 (i)-(iii) are satisfied, then*

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{2,\mathrm{ci}}_{1-\gamma,n}\right) \geq 1-\gamma - O(\rho_{2,n}r_{2,n}) - o(1)$$

with $\rho_{2,n} = \alpha_n^{3/2} n^{1/2}$. If additionally Assumption 1 (ii) is satisfied, the same inequality holds for $j=1$ with $\rho_{1,n} = \alpha_n n^{1/2} u_n^{-1}$.

Now we state the second main result of the paper, which is based on the Gaussian approximation. This result relies on suitable Berry-Esseen type theorems in Hilbert and Banach spaces.

Theorem 2. *Suppose that Assumptions 1 (i), (iii), (iv), (v), and Assumption 2 (i) are satisfied. Then for*

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{j,\mathrm{g}}_{1-\gamma,n}\right) \geq 1-\gamma - O\left(\rho_{j,n}r_{j,n} + \varepsilon_{j,n}\right), \qquad j=1,2,$$

where $\varepsilon_{1,n} = n^{-1/2}$, $\varepsilon_{2,n} = n^{-1/6}$, and $\rho_{j,n}, j=1,2$ are as in Theorem 1.

For the second set, we have $n^{-1/6}$ instead of $n^{-1/2}$ in the coverage error due to the fact that approximation of the distribution of $\|\hat{\nu}_{2,n}\|_\infty$ reduces to the approximation of the distribution of $\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right\|_\infty$. This is a non-smooth functional of the random functions taking values in the Banach space and Berry-Esseen estimates are not as sharp in this setting as in the case of the Hilbert space, see e.g. Bentkus, Götze, Paulauskas, and Račkauskas (2000).

For both sets we need to undersmooth in order to make the coverage error to decrease to zero as sample size increases. The next result describes the convergence rate of the diameter of constructed confidence sets.

Corollary 1. *Suppose that assumptions of Theorem 1 are satisfied, then we have the following rates for the expected diameters of confidence sets:*

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left|C^{1,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\frac{u_n}{\alpha_n n^{1/2}} + r_{1,n}\right), \qquad \sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left|C^{2,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}} + r_{2,n}\right).$$

Suppose now that assumptions of Theorem 2 are satisfied, then

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left|C^{1,\mathrm{g}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}\right), \qquad \sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left|C^{2,\mathrm{g}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right).$$

The bias-variance trade-off for the risk of the estimator $\hat{\varphi}_{\alpha_n}$ reduces to the trade-off between rates at which coverage errors decrease and the diameters of confidence sets shrink. For instance,

$r_{j,n}$ can be roughly interpreted as the order of the bias. It will be seen in the following sections that $r_{j,n}$ contains the regularization bias which is of order $\alpha_n^{\beta\wedge 1}$, and we need to set $\alpha_n \to 0$ to reduce this bias. On the other hand, the diameter of the confidence set is of the same order as the variance of the estimator and depends on α_n^{-1}. As a result, the reduction of the bias will inevitably lead to the increase of the variance and vice versa. The optimal choice of the regularization parameter α_n that balances the variance and the bias will in turn balance rates of diameters and coverage errors.

4 Ill-posedness and regularization bias

In this section we discuss regularity condition on the structural function φ and the operator T, needed to study honest and uniform confidence sets for Tikhonov-regularized estimators.

Assumption 3. *The model belongs to the source class*

$$\mathcal{F} \equiv \mathcal{F}(t,\beta,C) \equiv \left\{(\varphi,T) \in C_M^t[a,b]^p \times \mathcal{L}_{2,\infty} : \ \varphi = (T^*T)^\beta T^*\psi, \ \|T^*\|_{2,\infty} \vee \|\psi\| \vee \|T\|^{-1} \leq C\right\},$$

where $\beta, C > 0$.

Source conditions are at the heart of spectral regularization theory, and describe effectively the regularity of the problem by restricting how ill-posed the operator T is, comparing to the smoothness of the parameter of interest φ (see also Chen and Reiss (2011) for its comparability with other assumptions used in the literature). The present source condition is different from the one used to characterize L_2 rates, where we would require $\varphi = (T^*T)^\beta\psi$ only, see Carrasco et al. (2007).

We show in the next section that under this source condition, the uniform convergence rates of Tikhonov-regularized estimators are polynomial. In the severely ill-posed case, when singular values of the operator T decay to zero exponentially fast, the source condition in Assumption 3 requires that the function φ can be described by the finite number of generalized Fourier coefficients (with respect to the SVD basis of T), or more generally is analytic (also called supersmooth). Since many functions encountered in empirical work are smooth with generalized Fourier coefficients decaying rapidly, the fact that we can still achieve polynomial uniform convergence rates is a very good news for applications.

Under Assumption 3, the order of regularization bias can be established. Notice that in the Proposition that follows, the bound is uniform over the source set, and this will be needed to establish honesty of our confidence sets.

Proposition 2. *Suppose that Assumption 3 is satisfied, then*

$$\sup_{(\varphi,T)\in\mathcal{F}} \left\|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\right\|_\infty \leq R\alpha_n^{\beta\wedge 1},$$

where $R = C^2\left[\beta^\beta(1-\beta)^{1-\beta}\mathbb{1}_{0<\beta<1} + C^{2(\beta-1)}\mathbb{1}_{\beta\geq 1}\right]$.

The regularization bias is of order $O\left(\alpha_n^{\beta\wedge 1}\right)$. It is well known that for the simple Tikhonov regularization, it is not possible to characterize faster convergence rate for the bias in case of very regular problems with $\beta > 1$. This minor drawback can be fixed with iterated or extrapolated Tikhonov regularization approach.

5 Applications

In this section, we discuss how high-level conditions of Theorems 1 and 2 map to specific econometric models. We provide inferential results for three models considered in Section 2.

5.1 Nonparametric IV

The following assumptions are sufficient to characterize confidence sets for the NPIV model.

Assumption 4. *(i)* $(Y_i, Z_i, W_i)_{i=1}^n$ *is an i.i.d. sample of* (Y, Z, W) *such that* $|U| \leq F_0 < \infty$ *a.s.; (ii) the density* $f_{ZW}(z,w)$ *is bounded away from* 0 *and there exists some real number* $s > 0$ *such that* $f_{ZW} \in C_L^s[a,b]^{p+q}$*; (iii)* K_z *and* K_w *are continuous product kernel functions of order* $\lfloor t\rfloor \vee \lfloor s\rfloor$ *such that* $\int \|u\|^t|K_z(u)|\mathrm{d}u < \infty$ *and* $\int \|u\|^s|K_w(u)|\mathrm{d}u < \infty$ *with individual kernels of bounded p-variation for some* $p \geq 1$ *and supported on the rectangle which is a subset of* $[a,b]$*; (iv) the integral operator* $T : L_2[a,b]^p \to L_2[a,b]^q$ *is* 1-1.

Assumption (i) rules out distributions with unbounded supports, such as the normal distribution, but it allows, for example, for truncated normal. This should not be very restrictive since the vast majority of economic variables are bounded due to scarcity and economic constraints. Assumption (iii), in particular, allows to use boundary-corrected kernels, see Hall and Horowitz (2005) and Darolles et al. (2011)[8] for further discussion. Lastly, Assumption (v) is a completeness condition, Newey and Powell (2003). In case of its failure, the estimator is still well-defined and converge to the best approximation to function φ in the orthogonal complement of the null space of the operator T, see Babii and Florens (2016).

[8] Otherwise, to avoid problems at end-points, we will make restriction to the interior of $[a,b]^{p+q}$ and all results should be read as uniform over the interior of this set.

Theorem 3. *Suppose that Assumptions 3 and 4 are satisfied and the sequences of bandwidth parameters $h_n \to 0$ is such that $\log h_n^{-1}/(nh_n^{p+q}) = O(1)$ as $n \to \infty$. Then for $j = 1, 2$ we have*

$$\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty \leq \|\hat{\nu}_{j,n}\|_\infty + \eta_{j,n},$$

where $\hat{\nu}_{j,n}, j = 1, 2$ are as in Eq. (7) with $X_{1,i}(w) = U_i h_n^{-q} K\left(h_n^{-1}(W_i - w)\right)$, $X_{2,i}(z) = f_{Z,W}(z, W_i)U_i$, $\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\eta_{j,n} = O\left(r_{j,n}\right)$, and

$$r_{2,n} = \frac{h_n^s + h_n^t}{\alpha_n^{3/2}} + \alpha_n^{\beta\wedge 1} + \frac{1}{\alpha_n^{3/2}}\left(\frac{1}{\sqrt{nh_n^q}} + h_n^{t-q/2}\right)\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right) + \frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right),$$

$$r_{1,n} = \frac{h_n^t}{\alpha_n h_n^{q/2}} + \alpha_n^{\beta\wedge 1} + \frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right).$$

Moreover,

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\|\hat{\nu}_{1,n}\|_\infty = O\left(\frac{1}{\alpha_n n^{1/2} h_n^{q/2}}\right), \qquad \sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\|\hat{\nu}_{2,n}\|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right).$$

Rates $r_{j,n}$ have three components. The first two are the bias of the nonparametric kernel density estimators and the regularization bias respectively. The last component comes from the fact that we need to estimate the operator T. It is necessary to select tuning parameters, so that the variance is balanced with respect to the two bias terms. It is also necessary to ensure that third component, which represents the noise coming from the estimation of the operator T converges to zero faster than the leading variance and the bias terms.

The following corollary achieves both requirements by imposing minimal conditions on the smoothness of the joint density and the structural function φ. Following Darolles et al. (2011), we interpret these smoothness requirements as the "strong instrument" condition[9]. If $p = q = 1$ and the second-order derivatives of φ and f_{ZW} are sufficiently smooth in the Hölder sense, then this assumptions is not binding.

Corollary 2. *Suppose that assumptions of Theorem 3 hold. Assume additionally that components of the model are sufficiently smooth and tuning parameters converge to zero at certain speeds, in the sense that (i) for $j = 1$, we have $t > q/2 + 2p$, $s > 0.75(t - q/2)$, $h_n \sim n^{-\frac{1}{2t}}$, and $\alpha_n \sim n^{-\frac{t-q/2}{2t(\beta\wedge 1+1)}}$; (ii) for $j = 2$, we have $s \wedge t > 5(p+q)/4$, $h_n \sim n^{-\frac{1}{2(t\wedge s)}}$, and $\alpha_n \sim n^{-\frac{1}{2(\beta\wedge 1)+3}}$. Then we uniform convergence rates of the risk of the NPIV estimator are $O\left(n^{-\frac{(t-q/2)(\beta\wedge 1)}{2t(\beta\wedge 1+1)}}\right)$ for*

[9] Notice that this analytical characterization of strong instrument is different from the usual correlation requirement for the linear IV model.

$j = 1$ *and* $O\left(n^{-\frac{\beta\wedge 1}{2(\beta\wedge 1)+3}}\right)$ *for* $j = 2$.

Despite the fact that we have to estimate some parts of the model non-parametrically, rates in the second decomposition with $\nu_{2,n}$ process are free from dimension of the IV, once appropriate smoothness conditions are satisfied. This is similar to the L_2 theory developed in Darolles et al. (2011). On the other hand, convergence rates in the first decomposition depend on the dimension of the IV. Even though this "curse of dimensionality" may result in worser performance from the asymptotic point of view, frequently, in the empirical research, we only have one or two instruments and the first decomposition may lead to better performance of confidence sets in finite samples. Obtained polynomial uniform convergence rates are new and improve upon earlier results by Mbakop (2015) who obtain rates of order of powers of $\log n/n$ that depend on the dimension of both the regressors and the instrument.

The interesting feature of Tikhonov regularization is that convergence rates are of order of powers of $1/n$ in mildly and some severely ill-posed settings.

In the next result we verify conditions of Theorem 1 and 2 for the NPIV estimator show that our confidence sets provide a valid coverage. For confidence sets based on the concentration inequality, we show that the bias of the estimator will have an impact on the coverage error. On the other hand, for the confidence set based on the Gaussian approximation we characterize completely convergence rates of coverage errors, which will be driven by the bias of the estimator and the accuracy of Gaussian approximation.

Theorem 4. *Suppose that assumptions of Corollary 2 are satisfied, then*

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{j,\mathrm{ci}}_{1-\gamma,n}\right) \geq 1-\gamma - O\left(\rho_{j,n} r_{j,n}\right) - o(1), \qquad j = 1,2$$

and

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{2,\mathrm{g}}_{1-\gamma,n}\right) \geq 1-\gamma - O\left(\alpha_n^{3/2} n^{1/2} r_{2,n} + n^{-1/6}\right),$$

where $r_{j,n}, j = 1,2$ *are as in Theorem 3.*

Notice, that Assumptions of Corollary 2 do not ensure that coverage errors of confidence sets decrease to zero. To achieve this requirement, it is necessary to undersmooth, increasing the speed at which tuning parameters converge to zero. In the next corollary we assume that tuning parameters converge to zero, so that the rate at which the diameter of the confidence set shrink and the rate at which the coverage error decreases are balanced.

Corollary 3. *Suppose that Assumptions of Theorem 3 hold. Assume additionally that components of the model are sufficiently smooth and tuning parameters converge to zero at certain speeds* *in the sense that (i) for $j=1$, we have $\beta<1$, $t>q/2+p\frac{1+\beta\wedge 1}{1-\beta\wedge 1}$ and $s>(t-q/2)\frac{\beta\wedge 1+1/2}{\beta\wedge 1+1}$, $h_n\sim n^{-\frac{\beta\wedge 1+1}{(\beta\wedge 1)(t+q/2)+2t}}$, and $\alpha_n\sim n^{-\frac{t-q/2}{(\beta\wedge 1)(t+q/2)+2t}}$; (ii) for $j=2$, we have $s\wedge t>2.5(p+q)$, $h_n\sim n^{-\frac{2(\beta\wedge 1)+3}{2(s\wedge t)(\beta\wedge 1+3)}}$, and $\alpha_n\sim n^{-\frac{1}{\beta\wedge 1+3}}$. Then we have the following convergence rates for expected diameters*

$$\sup_{(\varphi,T)\in\mathcal{F}}\mathbb{E}\left|C^{1,\mathrm{ci}}_{1-\gamma,n}\right|_\infty=O\left(n^{-\frac{(\beta\wedge 1)(t-q/2)}{2((\beta\wedge 1)(t+q/2)+2t)}}\right)\quad\text{and}\quad\sup_{(\varphi,T)\in\mathcal{F}}\mathbb{E}\left|C^{2,\mathrm{ci}}_{1-\gamma,n}\right|_\infty=O\left(n^{-\frac{\beta\wedge 1}{2(\beta\wedge 1+3)}}\right).$$

Moroever, we have the same rate for $\mathbb{E}\left|C^{2,\mathrm{g}}_{1-\gamma,n}\right|_\infty$ and the coverage error of the set $C^{2,\mathrm{g}}_{1-\gamma,n}$ as for $\mathbb{E}\left|C^{2,\mathrm{ci}}_{1-\gamma,n}\right|_\infty$.

5.2 Functional regressions

The following set of assumptions provides sufficient conditions for functional regression models.

Assumption 5. *(i) (Y,Z,W) is a random vector in $(\mathbf{R}\times C[0,1]^p\times C[0,1]^q)$ a.s., and $(Y_i,Z_i,W_i)_{i=1}^n$ is an i.i.d. sample from (Y,Z,W); (ii) $(t,s)\mapsto Z(t)W(s)$ has trajectories in $C^\rho_M[0,1]^{p+q}$ a.s. for some $\rho\in(0,1]$; (iii) $\|UW\|\leq F_1<\infty$ a.s.; (iv) $\mathbb{E}\|UW\|^3\leq C_1<\infty$; (v) the integral operator $T:L_2[0,1]^p\to L_2[0,1]^q$ is injective.*

Confidence sets for functional regression models will be based on the risk decomposition with $\nu_{1,n}$ process as can be seen from the following theorem.

Theorem 5. *Suppose that Assumptions 3 and 5 are satisfied with $t=0$, then*

$$\|\hat{\varphi}_{\alpha_n}-\varphi\|_\infty\leq\|\hat{\nu}_{1,n}\|_\infty+\eta_{1,n}$$

with $X_{1,i}=U_iW_i$, $\mathbb{E}\eta_{1,n}=O\left(\alpha_n^{\beta\wedge 1}+\frac{\alpha_n^{1/2}}{\alpha_n n^{1/2}}\right)$. Moreover,

$$\sup_{(\varphi,T)\in\mathcal{F}}\mathbb{E}\left\|\hat{\varphi}_{\alpha_n}-\varphi\right\|_\infty=O\left(\frac{1}{\alpha_n n^{1/2}}+\alpha_n^{\beta\wedge 1}\right).$$

It follows that when $\alpha_n\to 0$ and $\alpha_n n^{1/2}\to\infty$, the term $\frac{\alpha_n^{1/2}}{\alpha_n n^{1/2}}$ converges to zero much faster than both the leading variance term, $\|\nu_{1,n}\|_\infty$, and the bias term, $\alpha_n^{\beta\wedge 1}$. The optimal balancing between the two is achieved when the regularization parameter is $\alpha_n\sim n^{-\frac{1}{2(\beta\wedge 1+1)}}$. In this case, the uniform convergence rate of the risk in the functional linear regression model is

$$\sup_{(\varphi,T)\in\mathcal{F}}\mathbb{E}\|\hat{\varphi}_{\alpha_n}-\varphi\|_\infty=O\left(n^{-\frac{\beta\wedge 1}{2(\beta\wedge 1+1)}}\right).$$

The next result justifies the validity of our confidence sets for functional regression models.

Theorem 6. *Suppose that Assumptions 3, 5 are satisfied and the sequence of regularization parameter is such that $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ as $n \to \infty$. Then for any $\gamma \in (0,1)$*

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{1,\mathrm{ci}}_{1-\gamma,n}\right) \geq 1-\gamma - O\left(\alpha_n^{\beta\wedge 1+1} n^{1/2}\right) - o(1)$$

and

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{1,\mathrm{g}}_{1-\gamma,n}\right) \geq 1-\gamma - O\left(\alpha_n^{\beta\wedge 1+1} n^{1/2} + \alpha_n^{1/2}\right).$$

Once again, the choice of tuning parameters optimal for the risk, does not ensure that coverage errors decrease to zero. To reduce the impact of the bias on the coverage properties we need to undersmooth by setting tuning parameter α_n to go to zero faster than optimal for the risk. This will result in confidence sets with diameter shrinking at the slower rate as can be seen from the following application of Corollary 1.

Corollary 4. *Suppose that assumptions of Theorem 6 are satisfied, then*

$$\mathbb{E}\left|C^{1,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}} + \alpha_n^{\beta\wedge 1}\right) \quad \text{and} \quad \mathbb{E}\left|C^{1,\mathrm{g}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}\right).$$

If we wish to balance the rate at which coverage errors decrease and the confidence set shrinks, we should set $\alpha_n \sim n^{-\frac{1}{\beta\wedge 1+2}}$, in which case both will be of order $O\left(n^{-\frac{\beta\wedge 1}{2(\beta\wedge 1+2)}}\right)$.

5.3 Density deconvolution

We introduce the following assumptions.

Assumption 6. *(i) $(Y_i)_{i=1}^n$ is an i.i.d. sample of Y; (ii) f, φ, r are continuous and compactly supported on some subsets of $[a,b]$ with $a < b < \infty$; (iii) $f \in C^s_M[a,b]$, for some $s \in (0,1]$; (iv) the integral operator $T : L_2[a,b] \to L_2[a,b]$ is 1-1;*

The next result gives uniform risk decomposition of the density deconvolution estimator. Confidence sets for functional regression models will be based on the variance process of type $\nu_{1,n}$.

Theorem 7. *Suppose that Assumption 3 is satisfied with $t = 0$, then*

$$\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty \leq \|\nu_{2,n}\|_\infty + R\alpha_n^{\beta\wedge 1}$$

with $X_{2,i}(z) = f(Y_i - z)$ and R as in Proposition 2. Suppose additionally that Assumption 6 *holds, then*

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi}_{\alpha_n}-\varphi\right\|_\infty = O\left(\frac{1}{\alpha_n^{3/2}n^{1/2}}+\alpha_n^{\beta\wedge 1}\right).$$

To balance the variance and the bias, we should select $\alpha_n \sim n^{-\frac{1}{2(\beta\wedge 1)+3}}$. In this case, the uniform convergence rate of the deconvolution estimator is

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi}_{\alpha_n}-\varphi\right\|_\infty = O\left(n^{-\frac{\beta\wedge 1}{2(\beta\wedge 1)+3}}\right).$$

Since operators T and T^* are known, we can consider confidence sets as in Eq. (10) with estimators replaced by the respective known quantities

$$\begin{aligned} q_{1-\gamma,n}^{2,\mathrm{ci}} &= 2\left\|\nu_{2,n}^{\varepsilon}\right\|_\infty + \frac{3\|f\|_\infty\left(\|T^*\|_{2,\infty}/2+\alpha_n^{1/2}\right)\sqrt{2\log(2/\gamma)}}{\alpha_n^{3/2}n^{1/2}}. \\ q_{1-\gamma,n}^{2,\mathrm{g}} &= c_{2,1-\gamma}\frac{\|T^*\|_{2,\infty}/2+\alpha_n^{1/2}}{\alpha_n^{3/2}n^{1/2}}, \end{aligned}$$

where $c_{2,1-\gamma}$ is $1-\gamma$ quantile of centered Gaussian processes $\|\mathbb{G}_2\|_\infty$ with covariance function $(s,t)\mapsto \mathbb{E}[f(Y-s)f(Y-t)]$, $s,t\in[a,b]$.

The next result gives coverage errors for both sets.

Theorem 8. *Suppose that Assumptions 3 and 6 are satisfied. Then if $\alpha_n^{3/2+\beta\wedge 1}n^{1/2}\to 0$ for any $\gamma\in(0,1)$*

$$\begin{aligned} \inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi\in C_{1-\gamma,n}^{2,\mathrm{ci}}\right) &\geq 1-\gamma-o(1), \\ \inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi\in C_{1-\gamma,n}^{2,\mathrm{g}}\right) &\geq 1-\gamma-O\left(n^{-1/6}\right)-o(1). \end{aligned}$$

The coverage error is effectively driven by the bias for sets built with concentration inequality. On the other hand, for sets built with Gaussian approximation we have additional $n^{-1/6}$ coming from the Berry-Esseen estimate. The following result can be obtained from a trivial application of Corollary 1.

Corollary 5. *Suppose that assumptions of Theorem 8 are satisfied, then expected diameters $\mathbb{E}\left|C_{1-\gamma,n}^{2,\mathrm{g}}\right|_\infty$ and $\mathbb{E}\left|C_{1-\gamma,n}^{2,\mathrm{ci}}\right|_\infty$ are of $O\left(\frac{1}{\alpha_n^{3/2}n^{1/2}}\right)$.*

6 Monte Carlo experiments in the NPIV model

This section reports results of Monte Carlo experiments for our confidence sets. We focus on the NPIV estimator. Samples of size $n \in \{1000, 5000\}$ are generated as follows

$$Y = \varphi(Z) + U, \qquad \varphi(z) = e^{-z^2/0.8},$$
$$\begin{pmatrix} Z \\ W \\ U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \rho\sigma_z\sigma_w & \sigma_{zu} \\ \rho\sigma_z\sigma_w & \sigma_w^2 & 0 \\ \sigma_{zu} & 0 & \sigma_u^2 \end{pmatrix} \right),$$

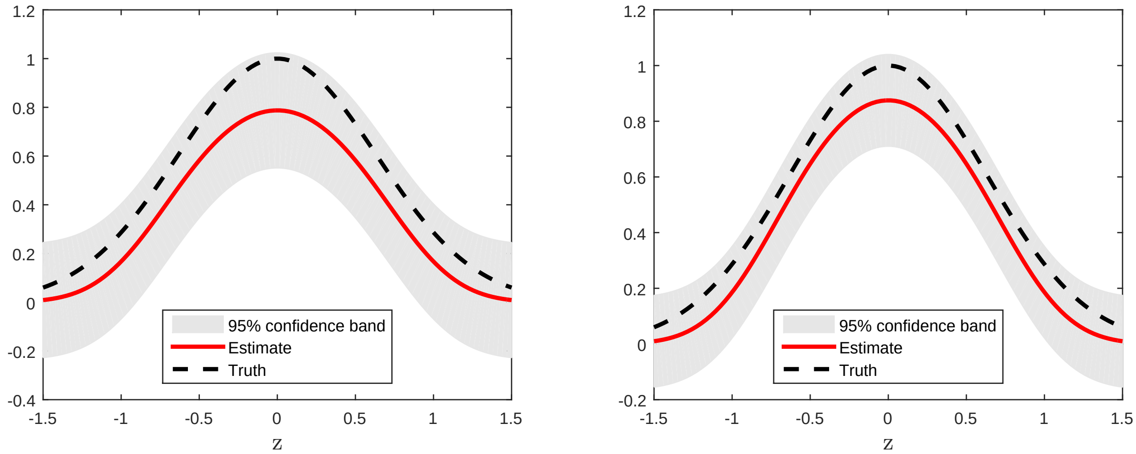
where $\sigma_z = \sigma_w = 0.3$, $\sigma_u = \sqrt{0.03}, \sigma_{zu} = 0.04$, and $\rho = 0.3$. To be consistent with the theory, we keep only observations inside sufficiently large compact set. Since the joint density of (Z, W) is bivariate normal, this choice of functional forms resembles up to a constant the convolution of Gaussian densities. Therefore, with appropriate modifications, we can easily adapt simulation design not only to functional regression, but also to the deconvolution model. Given that the structure behind all three models would be the same and given that the NPIV is the most complex among ill-posed model, we report results of MC experiments only for it. We also remark that the normal density is analytic, leading to singular values decaying exponentially fast. Therefore, this design corresponds to the most difficult severelly ill-posed setting. There were 5000 replications of each experiment. The estimate of the density function f_{ZW} is obtained using kernels. For simplicity of implementation, we do not optimize the performance with higher-order boundary kernels and simply take the product of second-order Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbb{1}_{\{|x|\leq 1\}}$.

The estimator is discretized using simple Riemann sum on the grid of 100 equidistant points. In our setting, this number of grid points ensures that the numerical errors is negligible comparing to the statistical noise in our setting. Higher number of grid points or numerical cubatures can give better approximation if needed. The discretized estimator has closed-form expressions

$$\hat{\boldsymbol{\varphi}} = \left(\alpha_n \mathbf{I} + \mathbf{K}^\top \mathbf{K}\right)^{-1} \mathbf{K}^\top \mathbf{r},$$

where $\mathbf{I}$ is the $T \times T$ identity matrix, $\mathbf{r} = \left(\frac{1}{n}\sum_{i=1}^n Y_i h_n^{-1} K\left(h_n^{-1}(W_i - w)\right)\right)_{1\leq j\leq T}$, $\mathbf{K} = \boldsymbol{f}\Delta$ with $\boldsymbol{f} = (\hat{f}_{ZW}(z_k, w_j))_{1\leq j,k\leq T}$, and Δ is the grid step. By Lemma 4, the operator norm $\|\hat{T}^*\|_{2,\infty}$ can be computed using the mixed norm of the kernel function of T.

In Figure 1 we plot estimates with 95% confidence band based on Gaussian approximation, averaged over 5000 Monte Carlo experiments. Figure 2 represents the same plot for confidence



(a) $n = 1000, \alpha_n = 0.14, h_n = 1$

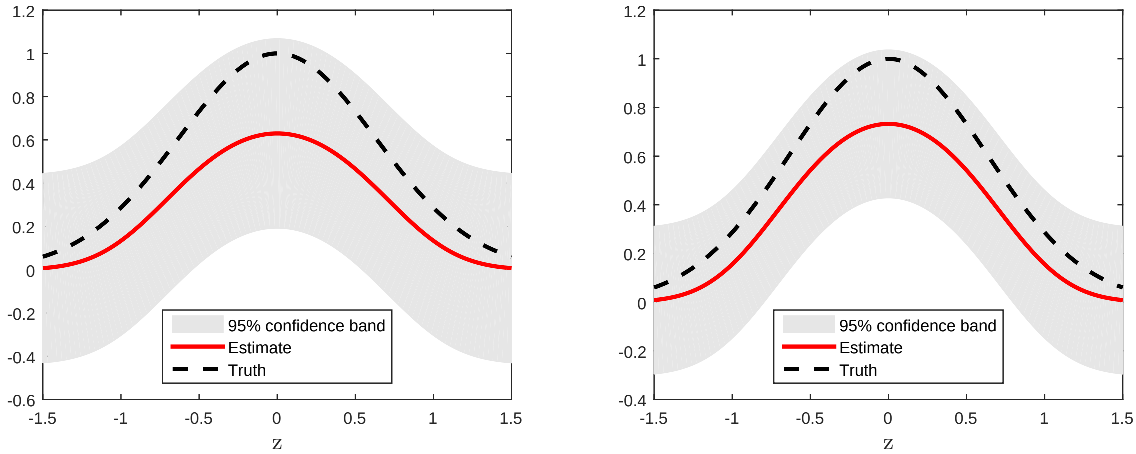
(b) $n = 5000, \alpha_n = 0.1, h_n = 1$

Figure 1: Estimates and confidence bands based on Gaussian approximation, average over 5000 experiments.

bands based on concentration inequality. In our Monte Carlo experiments, Gaussian approximation leads to better performance. Resulting confidence bands are narrower and the estimator is centered closer to the population value of the parameter of interest.

As was discussed in the Section 3, for confidence sets, the bias-variance trade-off for the risk of the estimator reduces to the trade-off between coverage errors and the diameter of confidence sets. In both figures we calibrated tuning parameters, so that the empirical coverage probability is as close to the nominal 95% as possible. For larger values of tuning parameters, the band becomes too narrow and the bias starts to dominate, reducing uniform coverage. On the other side, even thought smaller values of tuning parameters reduce bias and the estimator becomes closer to its population value, they also increase the size of the variance and lead to wider bands. The optimal choice of tuning parameters balances the two.

The question of the data-driven choice of tuning parameters is extremely important in applications. Recently Centorrino (2014) developed a cross-validation approach to the choice of the regularization parameter, optimal with respect to the L_2 loss for the NPIV model of Darolles et al. (2011). Rate-adaptive tuning parameter selection procedures in the sieve approach are discussed in Horowitz (2014) for the L_2 loss and in Chen and Christensen (2015) for the supremum loss. In this paper we study a modification of the model of Darolles et al. (2011) and we need the method to be optimal with respect to the sup-norm risk. Methods existing in the literature could be used, but their theoretical optimality for confidence sets should be studied separately. It also an open question, both for the sieve and the Tikhonov-regularized estimators, whether under some restrictions on the class of models it is possible to obtain adaptive confidence sets.



(a) $n = 1000, \alpha_n = 0.24, h_n = 1$ **(b)** $n = 5000, \alpha_n = 0.17, h_n = 0.6$

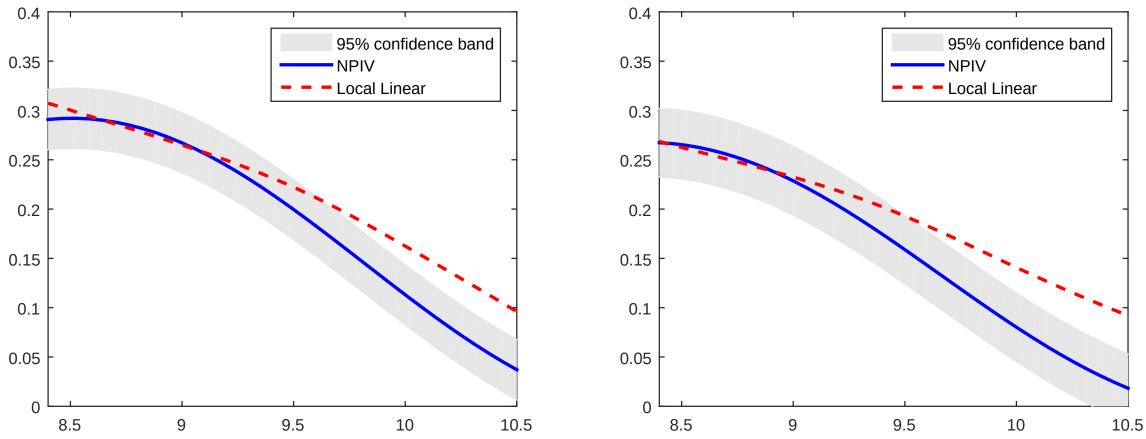
Figure 2: Estimates and confidence bands based on concentration inequality, average over 5000 experiments.

In the DGP used in MC experiments, Gaussian approximation demonstrates better performance leading to more narrow bands and smaller bias. For example, Figure 1 (b) gives an excellent confidence band with reasonably small amount of the bias. It is clear that we need to tolerate some amount of the regularization bias in small samples to have confidence bands of reasonable width and with good coverage properties. The amount of the bias can be reduced once the sample size becomes sufficiently large. It is well-known that undersmoothing and bias correction can decrease the bias at the cost of wider confidence sets, see also Florens et al. (2016) for a discussion of bias-corrected pointwise confidence intervals.

7 Confidence sets for Engel curves

In this section we estimate Engel curves using the NPIV approach and construct confidence sets using Gaussian approximation. Engel curves describe how the demand for commodity changes while the household's budget increases. Estimation of Engel curves is fundamental for the analysis of consumer behavior and has implications in different fields of empirical research. Interesting applications include the measurement of welfare losses associated with tax distortions in Banks, Blundell, and Lewbel (1997), estimation of growth and inflation in Nakamura, Steinsson, and Liu (2014), or estimation of income inequality across countries in Almås (2012).

Previously Blundell et al. (2007) estimated the shape-invariant system of Engel curves on the UK data with sieve approach. For simplicity, we focus on the non-parametric specification of the Engel curve. Partially linear specification can be easily estimated in two steps. First, we estimate



(a) With children

(b) Without children

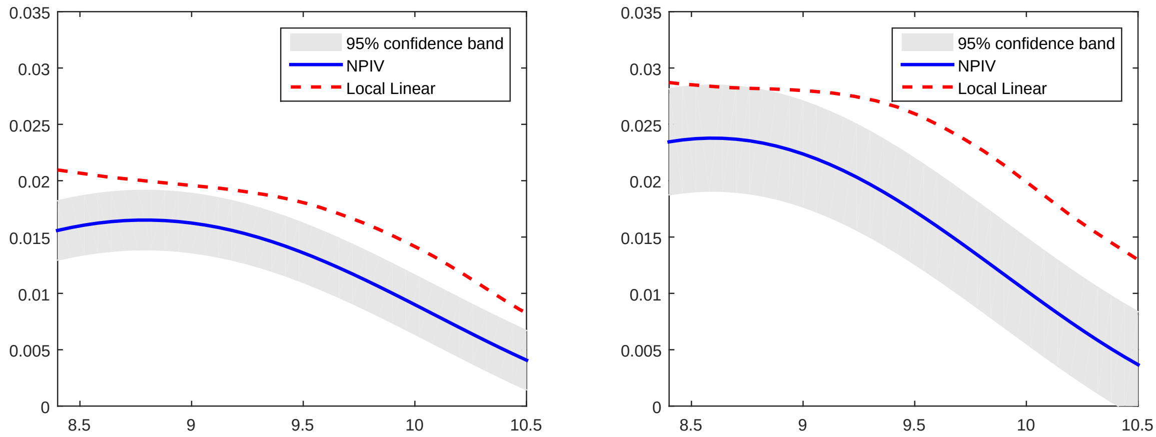
Figure 3: Engel curves for food: $\alpha_n = 0.02, h_n = 4$.

the parametric component with differencing, see Yatchew (2003). At the second stage we get rid of this parametric component and estimate the non-parametric model with IV approach.

Our dataset is drawn from the 2015 Consumer Expenditure Survey data in US and we estimate Engel curves for food and alcohol. In our subsample, we have married couples with positive income during the past 12 month, including households with and without children. The dependent variable is the share of expenditures on the particular commodity in total non-durable expenditures. The log of total expenditures on non-durable goods is used as an independent variable. As in Blundell et al. (2007) we instrument the log of total expenditures with income before tax.

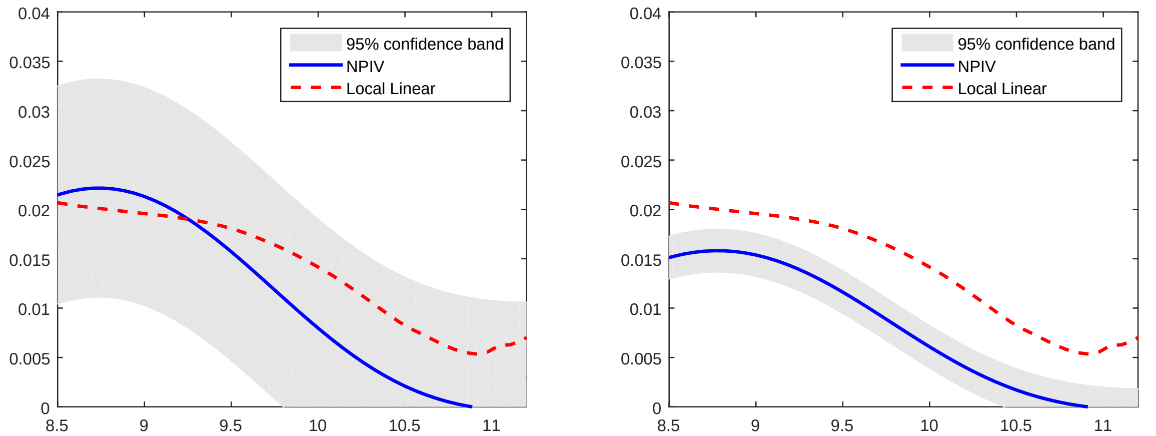
NPIV estimates with 95% confidence bands are plotted in Fig. 3 and 4. For comparison, we also include the local linear estimator, which does not correct for the endogeneity bias. We find that the Engel curve obtained with NPIV estimator is steeper and that simple local polynomial estimates are often outside of our confidence sets. For the local linear estimator we use the cross-validation to obtain the bandwidth parameter. For the NPIV model, we select tuning parameters empirically, comparing the plot for different values of tuning parameters and selecting the one at the middle of the two extreme points, after which the estimator becomes extremely wiggly or flat[10]. To illustrate the impact of tuning parameters on estimates and confidence sets, in Fig. 5 we plot several different choices of the regularization and the bandwidth parameters.

[10]This empirical selection of tuning parameters is sometimes advocated as alternative to other approaches, see, Cameron and Trivedi (2005). Notice that this methods is only appropriate for kernel or Tikhonov-regularized estimators.

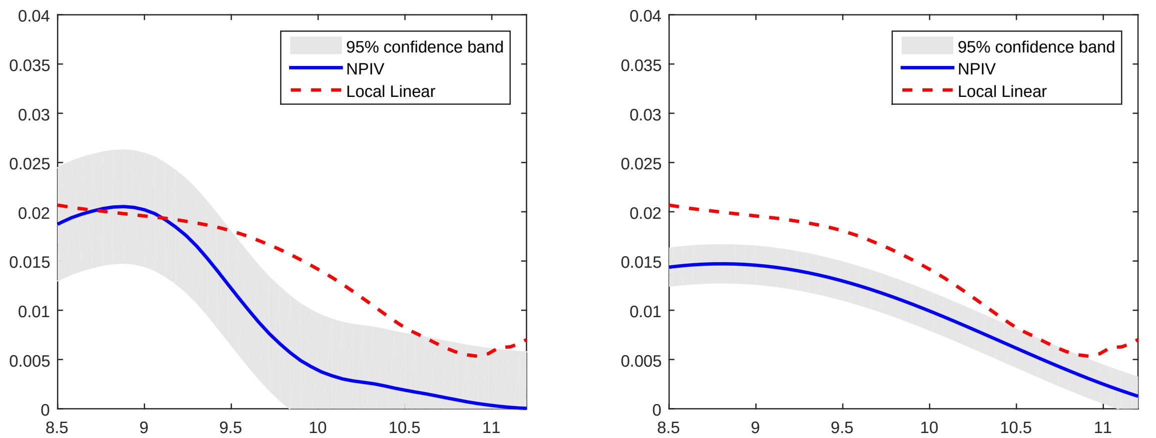


(a) With children **(b)** Without children

Figure 4: Engel curves for alcohol: $\alpha_n = 0.05, h_n = 4$.



(a) $\alpha_n = 0.02, h_n = 3$ **(b)** $\alpha_n = 0.08, h_n = 3$



(c) $\alpha_n = 0.05, h_n = 1$ **(d)** $\alpha_n = 0.05, h_n = 5$

Figure 5: Engel curves for alcohol (with children). Sensitivity to tuning parameters.

8 Conclusions

This paper studies uniform inferential methods in ill-posed models treated with Tikhonov regularization. Building uniform confidence sets in this setting is a difficult problem and requires to approximate the distribution of the supremum of a complex empirical process. I show that it is not possible to establish the weak convergence of this process in the space of continuous function under the uniform topology for a very general class of ill-posed models known in econometrics and statistics. Nonetheless, I demonstrate that it is possible to obtain uniform confidence sets with relaxed coverage requirement.

In this paper I develop two alternative approaches to uniform inference that lead to honest confidence sets. Honest confidence sets are of practical interest for several reasons. They allow to ensure that there exists a certain sample size after which the coverage level will be not significantly smaller than the nominal coverage level, regardless of how complicated the estimated function is withing a given smoothness class. Moreover, as it is widely recognized in non-parametric statistics, results for a fixed model can lead to inconsistent notions of optimality. Honest confidence sets, on the other hand, have uniform validity.

The first approach developed in this paper relies on the Gaussian approximation to a certain upper bound of the supremum of the statistics of interest. The second approach is based on the assessment of tail probabilities with a data-driven concentration inequalities. I obtain explicit convergence rate for coverage errors for confidence sets obtained with Gaussian approximation. Obtained rates may be polynomial for mildly ill-posed and some severely ill-posed problems. These rates are new and have not been previously discussed in the literature.

Confidence sets based on the concentration inequality are easier to compute. On the other hand, the Gaussian approximation seems to be less conservative in Monte Carlo experiments and leads to narrower bands with the same coverage level. As a matter of practical advice, it seems reasonable to use confidence sets based on the concentration inequality whenever the sample size is sufficiently large to produce small and informative sets and to rely on Gaussian approximation otherwise.

References

Karun Adusumilli, Taisuke Otsu, et al. Nonparametric instrumental regression with errors in variables. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, 2015.

Ingvild Almås. International income inequality: Measuring ppp bias by estimating engel curves for food. *The American Economic Review*, 102(2):1093–1117, 2012.

Donald WK Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

Isaiah Andrews and Anna Mikusheva. A geometric approach to nonlinear econometric models. *Econometrica*, 84(3):1249–1264, 2016.

Andrii Babii. Identification and estimation in the functional linear instrumental regression. *Working paper*, 2016.

Andrii Babii and Jean-Pierre Florens. Is completeness necessary? penalized estimation in non-identified models. *Working paper*, 2016.

James Banks, Richard Blundell, and Arthur Lewbel. Quadratic engel curves and consumer demand. *Review of Economics and statistics*, 79(4):527–539, 1997.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366, 2015.

David Benatia, Marine Carrasco, and Jean-Pierre Florens. Functional linear regression with functional response. *Working paper*, 2015.

V Bentkus, F Götze, V Paulauskas, and A Račkauskas. The accuracy of gaussian approximation in banach spaces. In *Limit theorems of probability theory*, pages 25–111. Springer, 2000.

R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.

Stéphane Bonhomme and Jean-Marc Robin. Generalized non-parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies*, 77(2):491–533, 2010.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford, 2013.

A Colin Cameron and Pravin K Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

Ivan Canay, Andres Santos, and Azeem Shaikh. On the testability of identification in some nonparametric models with endogeneity. Technical report, cemmap working paper, 2012.

Hervé Cardot, Frédéric Ferraty, André Mas, and Pascal Sarda. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, 30(1):241–255, 2003.

Hervé Cardot, André Mas, and Pascal Sarda. Clt in functional linear regression models. *Probability Theory and Related Fields*, 138(3-4):325–361, 2007.

M. Carrasco, J.P. Florens, and E. Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6: 5633–5751, 2007.

Marine Carrasco and Jean-Pierre Florens. A spectral method for deconvolving a density. *Econometric Theory*, 27(03):546–581, 2011.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Asymptotic normal inference in linear inverse problems. *J. S. Racine, A. Ullah and L. Su, eds, 'Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics'.*, 2013.

S Centorrino. Data driven selection of the regularization parameter in additive nonparametric instrumental regressions. Technical report, Working paper, Stony Brook, 2014.

Xiaohong Chen and Timothy Christensen. Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. 2015.

Xiaohong Chen and Demian Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 2012.

Xiaohong Chen and Demian Pouzo. Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, 2015.

Xiaohong Chen and Markus Reiss. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(03):497–521, 2011.

Victor Chernozhukov, Denis Chetverikov, Kengo Kato, et al. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597, 2014.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Coupling inequalities for suprema of non-centered empirical and bootstrap processes. *arXiv preprint arXiv:1502.00352*, 2015.

Fabienne Comte, Jan Johannes, et al. Adaptive functional linear regression. *The Annals of Statistics*, 40(6):2765–2797, 2012.

S. Darolles, Y. Fan, J.P. Florens, and E. Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Kluwer Academic Pub, 1996.

Kirill Evdokimov. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Department of Economics, Princeton University*, 2010.

Yingying Fan, Gareth M James, Peter Radchenko, et al. Functional additive regression. *The Annals of Statistics*, 43(5):2296–2325, 2015.

Jean-Pierre Florens. Inverse problems and structural econometrics. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, page 284. Cambridge University Press, 2003.

Jean-Pierre Florens and Sébastien Van Bellegem. Instrumental variable estimation in functional linear models. *Journal of Econometrics*, 186(2):465–476, 2015.

Jean-Pierre Florens, Jan Johannes, and Sébastien Van Bellegem. Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(03):472–496, 2011.

Jean-Pierre Florens, Joel Horowitz, and Ingrid Keilegom. Bias-corrected confidence intervals in a class of linear inverse problems. *Working paper*, 2016.

Patrick Gagliardini and Olivier Scaillet. Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167(1):61–75, 2012.

Eric Gautier and Yuichi Kitamura. Nonparametric estimation in random coefficients binary choice models. *Econometrica*, 81(2):581–607, 2013.

Evarist Giné and Richard Nickl. Mathematical foundations of infinite-dimensional statistical models. *Cambridge Series in Statistical and Probabilistic Mathematics*, 2015.

Peter Hall and Joel L Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.

Peter Hall, Joel L Horowitz, et al. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91, 2007.

Joel L Horowitz. Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. *Journal of Econometrics*, 180(2):158–173, 2014.

Joel L Horowitz and Sokbae Lee. Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics*, 168(2):175–188, 2012.

Kengo Kato and Yuya Sasaki. Uniform confidence bands in deconvolution with unknown error distribution. *arXiv preprint arXiv:1608.02251*, 2016.

Vladimir Koltchinskii. Rademacher penalties and structural risk minimization. *Information Theory, IEEE Transactions on*, 47(5):1902–1914, 2001.

Vladimir Koltchinskii et al. Local rademacher complexities and oracle inequalities in risk minimization. *The Annals of Statistics*, 34(6):2593–2656, 2006.

Ker-Chau Li. Honest confidence regions for nonparametric regression. *The Annals of Statistics*, pages 1001–1008, 1989.

Karim Lounici and Richard Nickl. Global uniform risk bounds for wavelet deconvolution estimators. *The Annals of Statistics*, 39(1):201–231, 2011.

Eric Mbakop. Tikhonov regularization for nonparametric iv regression in sup-norm. *Working paper*, 2015.

M Thamban Nair. *Linear operator equations: approximation and regularization.* World Scientific, 2009.

Emi Nakamura, Jón Steinsson, and Miao Liu. Are chinese growth and inflation too smooth? evidence from engel curves. Technical report, National Bureau of Economic Research, 2014.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Vygantas Paulauskas and Alfredas Rackauskas. *Approximation Theory in the Central Limit Theorem: Exact Results in Banach Spaces.* Springer Science & Business Media, 1989.

MP Rajan. Convergence analysis of a regularized approximation for solving fredholm integral equations of the first kind. *Journal of mathematical analysis and applications*, 279(2):522–530, 2003.

Sidney I Resnick. *Extreme values, regular variation and point processes.* Springer, 1987.

A. Santos. Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80(1):213–275, 2012.

Maxwell B Stinchcombe and Halbert White. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(03):295–325, 1998.

Jing Tao. Inference for point and partially identified semi-nonparametric conditional moment models. 2014.

Alexandre Tsybakov. *Introduction to nonparametric estimation*, volume 11. Springer, 2009.

Aad W Van Der Vaart and Jon A Wellner. *Weak Convergence*. Springer, 2000.

Adonis Yatchew. *Semiparametric regression for the applied econometrician*. Cambridge University Press, 2003.

VV Yurinskii. On the accuracy of normal approximation of the probability of hitting a ball. *Theory of Probability & Its Applications*, 27(2):280–289, 1982.

Appendix A: Concentration of the supremum of empirical processes

Let $(X_i)_{i\in\mathbf{N}}$ be an i.i.d. sequence of random variables taking values in some measure space $(S, \mathcal{S})$ and let $\mathcal{G}$ be a countable class of real functions defined on S. Consider the empirical process

$$\nu_n(g) \equiv \frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}g(X_i), \qquad g \in \mathcal{G}$$

and denote the symmetrized version of this process by

$$\nu_n^{\varepsilon}(g) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g(X_i),$$

where $(\varepsilon_i)_{i\in\mathbf{N}}$ is an i.i.d. sequence of Rademacher random variables, independent from $(X_i)_{i\in\mathbf{N}}$. We use $\|.\|_{\mathcal{G}}$ to denote the supremum norm over the class of functions $\mathcal{G}$.

Proposition 3. *Let $(X_i)_{i=1}^n$ be a sequence of i.i.d. random variables in some measure space $(S, \mathcal{S})$, $(\varepsilon_i)_{i=1}^n$ be a sequences of i.i.d. Radomacher random variables independent from $(X_i)_{i=1}^n$, and $\mathcal{G}$ be a countable class of functions. Then*

$$2^{-1}\mathbb{E}\,\|\nu_n^{\varepsilon}\|_{\mathcal{G}} - 2^{-1}n^{-1/2}\|Pg\|_{\mathcal{G}} \leq \mathbb{E}\|\nu_n\|_{\mathcal{G}} \leq 2\mathbb{E}\,\|\nu_n^{\varepsilon}\|_{\mathcal{G}}\,. \tag{11}$$

The second inequality in Eq. (11), is a symmetrization inequality, while the first is the desymmetrization inequality. These inequalities have important applications in the empirical process theory, see Van Der Vaart and Wellner (2000). We refer to (Koltchinskii et al., 2006, p.7) and references therein for a proof of this modification of the symmetrization inequality.

Our construction of confidence sets in this paper relies on the Talagrand-type concentration inequality. More precisely, it can be derived from the concentration inequality for functions of bounded difference, known also as McDiarmid's inequality, see excellent treatment of concentration inequalities in Boucheron et al. (2013). Talagrand's concentration inequalities only describe exponential decline of tails probabilities for the deviation of the supremum of empirical process $\|\nu_n\|_{\mathcal{G}}$ from its expected value $\mathbb{E}\|\nu_n\|_{\mathcal{G}}$. This expected value is not known in practice. However, it can be estimated using the multiplier bootstrap based on Rademacher random variables. The symmetrization inequality allows then to compare the expected value of this estimate $\|\nu_n^{\varepsilon}\|_{\mathcal{G}}$ to the original expected value, which leads to the data-driven concentration inequality. This insightful idea comes from the machine learning literature, see e.g. Koltchinskii (2001).

The next proposition states the precise version of the data-driven Talagrand-type concentration inequality used in the present paper.

Proposition 4. *Suppose that $(X_i)_{i=1}^n$, $(\varepsilon_i)_{i=1}^n$, and $\mathcal{G}$ as in Proposition 3. Suppose also that the absolute value of all $g \in \mathcal{G}$ is uniformly bounded by some constant $F < \infty$. Then for all $n \in \mathbf{N}$*

$$\Pr\left(\|\nu_n\|_{\mathcal{G}} > 2\left\|\nu_n^{\varepsilon}\right\|_{\mathcal{G}} + 3F\sqrt{\frac{2x}{n}} \right) \leq 2e^{-x}.$$

See (Giné and Nickl, 2015, Theorem 3.4.5.) for the proof of this result.

Appendix B: Proofs of main results

Impossibility of weak convergence

Proof of Proposition 1. The proof is inspired by Cardot et al. (2007) who show impossibility of weak convergence in L_2 in the special case of functional linear regression model without endogeneity. For simplicity we focus on the process $\nu_{1,n}$, since the proof for the process $\nu_{2,n}$ is similar. Recall that for some normalizing sequence r_n, the weak convergence of $r_n\nu_{1n}$ in L_2 to some random element requires $\langle r_n\nu_{1n}, \delta\rangle$ to converge weakly in $\mathbf{R}$ for all $\delta \in L_2$, (Van Der Vaart and Wellner, 2000, Theorem 1.8.4). Since $(T^*T)^{-1}$ is unbounded with $\mathcal{D}[(T^*T)^{-1}] \subset L_2[a,b]^p$, for $\delta \in \mathcal{D}[(T^*T)^{-1}]$, we need to set $r_n = n^{1/2}$, since

$$\langle r_n\nu_{1n}, \delta\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\langle T^*X_i, (\alpha_n I + T^*T)^{-1}\delta\right\rangle \xrightarrow{d} N\left(0, \mathbb{E}\left\langle T^*X_1, (T^*T)^{-1}\delta\right\rangle^2\right).$$

On the other hand, if $\delta \notin \mathcal{D}[(T^*T)^{-1}]$, $\|(\alpha_n I + T^*T)^{-1}\delta\|^2 \to \infty$, making $\mathbb{E}\langle T^*X_1, (\alpha_n I + T^*T)^{-1}\delta\rangle^2 \to \infty$, and so $\langle n^{1/2}\nu_{1n}, \delta\rangle$ can't converge in distribution. This shows that it is not possible to converge weakly in L_2. Since bounded and continuous functionals on L_2 are bounded and continuous on $(C, \|.\|_\infty)$ for finite measure spaces, it follows from the definition of weak convergence that $r_n\nu_{1n}$ does not converge weakly in $(C, \|.\|_\infty)$ for any choice of normalizing sequence r_n. □

Proofs of main results

Proof of Theorem 1. Using Assumption 2 (i) and triangle inequality, we obtain, for $j \in \{1, 2\}$

$$\begin{aligned}
\Pr\left(\varphi \in C^{j,\mathrm{ci}}_{1-\gamma,n}\right) &= \Pr\left(\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty \leq q^{j,\mathrm{ci}}_{1-\gamma,n}\right) \\
&\geq \Pr\left(\|\hat{\nu}_{j,n}\|_\infty + \eta_{j,n} \leq q^{j,\mathrm{ci}}_{1-\gamma,n}\right) \\
&\geq \Pr\left(\|\nu_{j,n}\|_\infty + \|\hat{\nu}_{j,n} - \nu_{j,n}\|_\infty + \eta_{j,n} \leq q^{j,\mathrm{ci}}_{1-\gamma,n}\right).
\end{aligned} \tag{12}$$

Now $\|\nu_{j,n}\|_\infty, j = 1, 2$ are suprema of empirical processes indexed by the following classes of functions changing with the sample size n

$$\begin{aligned}
\mathcal{G}_{1,n} &= \left\{g : x \in C[a,b]^q \to \left[(\alpha_n I + T^*T)^{-1}T^*x\right](z),\ z \in [a,b]^p\right\} \\
\mathcal{G}_{2,n} &= \left\{g : x \in C[a,b]^p \to \left[(\alpha_n I + T^*T)^{-1}x\right](z),\ z \in [a,b]^p\right\}.
\end{aligned}$$

Notice that for any $g_1 \in \mathcal{G}_{1,n}$ and $g_2 \in \mathcal{G}_{2,n}$, under Assumptions 1 (ii) and (iv) by Lemma 1, we have the following envelops

$$\|g_1(X_{1,i})\|_\infty \leq \|T^*\|_{2,\infty}\|(\alpha_n I + TT^*)^{-1}\|\|X_{1,i}\| \leq \frac{\|T^*\|_{2,\infty}}{\alpha_n}u_n F_1 \quad a.s.$$
$$\|g_2(X_{2,i})\|_\infty \leq \|(\alpha_n I + T^*T)^{-1}\|_\infty \|X_{2,i}\|_\infty \leq \frac{\|T^*\|_{2,\infty}/2 + \alpha_n^{1/2}}{\alpha_n^{3/2}}F_2 \quad a.s.,$$

where F_1 and F_2 are the smallest envelops for $\|X'_{1,i}\|$ and $\|X_{2,i}\|_\infty$ respectively.

It follows from Eq. (12), reverse triangle inequality, inequality $\Pr(X+Y \leq x+y) \geq \Pr(X \leq x) - \Pr(Y > y)$, Markov's inequality, and[11] Proposition 4

$$\begin{aligned}
\Pr\left(\varphi \in C^{1,\mathrm{ci}}_{1-\gamma,n}\right) &\geq \Pr\left(\|\nu_{1,n}\|_\infty + \|\hat{\nu}_{1,n} - \nu_{1,n}\|_\infty + \eta_{1,n} \leq 2\left\|\hat{\nu}^\varepsilon_{1,n}\right\|_\infty + 3\frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}u_n\hat{F}_1\sqrt{\frac{2\log(2/\gamma)}{n}} + \frac{u_n}{\alpha_n n^{1/2}}\right) \\
&\geq \Pr\left(\|\nu_{1,n}\|_\infty \leq 2\left\|\nu^\varepsilon_{1,n}\right\|_\infty + 3\frac{\|T^*\|_{2,\infty}}{\alpha_n}u_n F_1\sqrt{\frac{2\log(2/\gamma)}{n}}\right) - O\left(\frac{\alpha_n n^{1/2}}{u_n}\mathbb{E}Q_{1,n}\right) \\
&\geq 1 - \gamma - O\left(\frac{\alpha_n n^{1/2}}{u_n}\mathbb{E}Q_{1,n}\right),
\end{aligned} \tag{13}$$

where

$$\mathbb{E}Q_{1,n} = \frac{u_n}{\alpha_n n^{1/2}}\mathbb{E}\left|\|\hat{T}^*\|_{2,\infty}\hat{F}_1 - \|T^*\|_{2,\infty}F_1\right| + \mathbb{E}\left\|\hat{\nu}^\varepsilon_{1,n} - \nu^\varepsilon_{1,n}\right\|_\infty + \mathbb{E}\left\|\hat{\nu}_{1,n} - \nu_{1,n}\right\|_\infty + \mathbb{E}\eta_{1,n}. \tag{14}$$

Under Assumptions 1 (i), (ii), (v) and Assumptions 2 (ii)-(iii) by Lemma 3

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left\|\hat{\nu}_{1,n} - \nu_{1,n}\right\|_\infty = o\left(\frac{u_n}{\alpha_n n^{1/2}}\right) \quad \text{and} \quad \sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\left\|\hat{\nu}^\varepsilon_{1,n} - \nu^\varepsilon_{1,n}\right\|_\infty = o\left(\frac{u_n}{\alpha_n n^{1/2}}\right). \tag{15}$$

On the other hand, $\mathbb{E}\eta_{1,n} = O(r_{1,n})$ under Assumption 2 (i).

To asses the order of the first term in Eq. (14), first observe that

$$\mathbb{E}\left|\hat{F}_1 - F_1\right|^2 \leq 2\mathbb{E}\left|\max_{1\leq i\leq n}\|X'_{1,i}\| - F_1\right|^2 + 2\mathbb{E}\max_{1\leq i\leq n}\left\|\hat{X}'_{1,i} - X'_{1,i}\right\|^2, \tag{16}$$

where the first term is $o(1)$, since under Assumptions 1 (i), we have $\max_{1\leq i\leq n}\|X'_{1,i}\| \xrightarrow{a.s.} F_1$, e.g. see (Resnick, 1987, p.8-9), while the second term is $o(1)$ under Assumption 2 (iii).

Since $\alpha_n \to 0$, Assumption 2 (ii) implies

$$\|\hat{T}^* - T^*\|^2_{2,\infty} = o(1),$$

[11] By continuity of trajectories of $\nu_{j,n}, j = 1,2$ processes, their suprema can be restricted to countable sets of rational numbers.

and so combining Eq. (13)-(15), we obtain

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\varphi \in C^{1,\mathrm{ci}}_{1-\gamma,n}\right) \geq 1-\gamma - O(\rho_{1,n}r_{1,n}) - o\left(1\right).$$

This finishes the proof for the case $j=1$.

The proof of the second statement is similar and we omit it. $\square$

Proof of Theorem 2. Under Assumption 2 (i) by Markov's inequality and computations in Eq. (8), we have uniformly in $\mathcal{F}$

$$\begin{aligned}
\Pr\left(\varphi \in C^{1,\mathrm{g}}_{1-\gamma,n}\right) &= \Pr\left(\|\hat{\varphi}-\varphi\|_\infty \leq q^{1,\mathrm{g}}_{1-\gamma,n}\right) \\
&\geq \Pr\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{1,i}\right\|^2 \leq c_{1,1-\gamma}\right) - O(\rho_{1,n}r_{1,n}) \\
&\geq \Pr\left(\|\mathbb{G}_1\|^2 \leq c_{1,1-\gamma}\right) - O\left(n^{-1/2}+\rho_{1,n}r_{1,n}\right) \\
&= 1-\gamma - O\left(n^{-1/2}+\rho_{1,n}r_{1,n}\right),
\end{aligned}$$

where the second inequality follows under Assumptions 1 (i) and (iv) by the Berry-Esseen theorem in Hilbert space of Yurinskii (1982).

Similarly,

$$\Pr\left(\varphi \in C^{2,\mathrm{g}}_{1-\gamma,n}\right) \geq \Pr\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right\|_\infty \leq c_{2,1-\gamma}\right) - O(\rho_{2,n}r_{2,n}).$$

Under Assumptions 1 (i), (iii), and (iv), the process inside the second probability is Donsker and by the Berry-Esseen theorem in Banach space, see Paulauskas and Rackauskas (1989)

$$\inf_{(\varphi,T)\in\mathcal{F}} \Pr\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right\|_\infty \leq c_{2,1-\gamma}\right) \geq 1-\gamma - O\left(n^{-1/6}\right),$$

giving the second result. $\square$

Proof of Corollary 1. Since $\mathbb{E}\left|C^{j,\mathrm{s}}_{1-\gamma,n}\right|_\infty = 2\mathbb{E}\left|q^{j,\mathrm{s}}_{1-\gamma,n}\right|$ for $j=1,2$ and $\mathrm{s}\in\{\mathrm{g},\mathrm{ci}\}$, under Assumption 2 (ii), we have

$$\mathbb{E}\left|C^{1,\mathrm{g}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}\right), \qquad \mathbb{E}\left|C^{2,\mathrm{g}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right).$$

On the other hand, under Assumption 2 (ii)

$$\mathbb{E}\left|C^{1,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\mathbb{E}\|\nu^\varepsilon_{1,n}\|_\infty + \mathbb{E}\left\|\hat{\nu}^\varepsilon_{1,n} - \nu^\varepsilon_{1,n}\right\|_\infty + \frac{u_n}{\alpha_n n^{1/2}}\left(\sqrt{\mathbb{E}\left|\hat{F}_1 - F_1\right|^2} + \sqrt{\mathbb{E}\|\hat{T}^* - T^*\|^2_{2,\infty}}\right)\right).$$

Under Assumption 1 (i), by the desymmetrization inequality in Proposition 3

$$\mathbb{E}\|\nu^\varepsilon_{1,n}\|_\infty = \mathbb{E}\|\nu_{1,n}\|_\infty + O\left(\frac{1}{n^{1/2}}\right) = O\left(\frac{u_n}{\alpha_n n^{1/2}}\right).$$

By Lemma 3, we can have the same estimate for $\mathbb{E}\|\hat{\nu}^\varepsilon_{1,n} - \nu^\varepsilon_{1,n}\|$.

Therefore, using the same reasoning as the one used to obtain estimates in Eq (16), we have

$$\mathbb{E}\left|C^{1,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\frac{u_n}{\alpha_n n^{1/2}} + r_{1,n}\right).$$

The proof that

$$\mathbb{E}\left|C^{2,\mathrm{ci}}_{1-\gamma,n}\right|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}} + r_{2,n}\right)$$

is similar and so we omit it. $\square$

Proof of Proposition 2. By Lemma 1, $(\alpha_n I + T^*T)$ is invertible operator between $(C, \|.\|_\infty)$ spaces. Using $f(T^*T)T^* = T^* f(TT^*)$ with $f(x) = (\alpha_n + x)^{-1}$, and factorizing the operator norm $\|T^* g(TT^*)\phi\|_\infty \leq \|T^*\|_{2,\infty}\|g(TT^*)\|\|\phi\|$ with $g(x) = \alpha_n(\alpha_n + x)^{-1}x^\beta$, under Assumption 3 for any $(\varphi, T) \in \mathcal{F}$

$$\begin{aligned}
\left\|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\right\|_\infty &= \left\|\left[(\alpha_n I + T^*T)^{-1}T^*T - I\right]\varphi\right\|_\infty \\
&= \left\|\alpha_n(\alpha_n I + T^*T)^{-1}\varphi\right\|_\infty \\
&= \left\|\alpha_n(\alpha_n I + T^*T)^{-1}(T^*T)^\beta T^*\psi\right\|_\infty \\
&= \left\|T^*\alpha_n(\alpha_n I + TT^*)^{-1}(TT^*)^\beta\psi\right\|_\infty \\
&\leq \|T^*\|_{2,\infty}\left\|\alpha_n(\alpha_n I + TT^*)^{-1}(TT^*)^\beta\right\|\|\psi\|
\end{aligned} \tag{17}$$

Notice that for $b \in (0,1)$, the function $\lambda \mapsto \frac{\lambda^b}{\alpha_n + \lambda}$ is strictly concave on $(0,\infty)$ admitting its maximum at $\lambda = \frac{b}{1-b}\alpha_n$. On the other hand, for $b \in [1,\infty)$, this function is strictly increasing on $[0, \|T\|^2]$, reaching its maximum at the end of this interval. Therefore, by isometry of functional calculus

$$\left\|\alpha_n(\alpha_n I + TT^*)^{-1}(TT^*)^\beta\right\| = \alpha_n \sup_{\lambda \in [0,\|T\|^2]}\left|\frac{\lambda^\beta}{\alpha_n + \lambda}\right| \leq \tilde{R}\alpha_n^{\beta\wedge 1}$$

with $\tilde{R} = \beta^\beta(1-\beta)^{1-\beta}\mathbb{1}_{0<\beta<1} + C^{2(\beta-1)}\mathbb{1}_{\beta\geq 1}$. Therefore,

$$\sup_{(\varphi,T)\in\mathcal{F}} \left\|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\right\|_\infty \leq R\alpha_n^{\beta\wedge 1}$$

with $R = C^2\tilde{R}$. □

Proofs for the NPIV model

Proposition 5. *Suppose that Assumptions 4 (i)-(iii) are satisfied and the sequence of bandwidth parameters is such that $1/(nh_n^{p+q}) = O(1), \forall n \in \mathbf{N}$, then for all $1 < r < \infty$*

$$\left(\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^r\right)^{1/r} = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right).$$

Proof. By Lemma 4,

$$\begin{aligned}
\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^r &= \mathbb{E}\left[\sup_{z\in(a,b)^p}\left(\int_{[a,b]^q}\left|\hat{f}_{ZW}(z,w) - f_{ZW}(z,w)\right|^2 \mathrm{d}w\right)^{1/2}\right]^r \\
&\leq (b-a)^{qr/2}\mathbb{E}\left\|\hat{f}_{ZW} - f_{ZW}\right\|_\infty^r.
\end{aligned}$$

By triangle inequality,

$$\left(\mathbb{E}\left\|\hat{f}_{ZW} - f_{ZW}\right\|_\infty^r\right)^{1/r} \leq \left(\mathbb{E}\left\|\hat{f}_{ZW} - \mathbb{E}\hat{f}_{ZW}\right\|_\infty^r\right)^{1/r} + \left(\left\|\mathbb{E}\hat{f}_{ZW} - f_{ZW}\right\|_\infty^r\right)^{1/r}.$$

The order of the bias follows by standard computations under the assumption $f_{ZW} \in C_L^s[a,b]^{p+q}$ and Assumption 4 (iii), see Tsybakov (2009)

$$\left\|\mathbb{E}\hat{f}_{ZW} - f_{ZW}\right\|_\infty = O\left(h_n^s\right).$$

For the variance term, we apply the moment inequality in (Giné and Nickl, 2015, Theorem 5.1.5 and Theorem 5.1.15), which gives

$$\left(\mathbb{E}\left\|\hat{f}_{ZW} - \mathbb{E}\hat{f}_{ZW}\right\|_\infty^r\right)^{1/r} = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + \sqrt{\frac{1}{nh_n^{p+q}}} + \frac{1}{nh_n^{p+q}}\right).$$

Combining all estimates, the result follows. □

Proof of the Theorem 3. We focus on $j = 1$ first. The proof is based on the following decomposition

$$\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty = \left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{r} - \varphi\right\|_\infty$$
$$\leq \|\hat{\nu}_{1,n}\|_\infty + \xi_n + B_n,$$

where

$$\xi_n = \left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\left[\frac{1}{n}\sum_{i=1}^n \varphi(Z_i)h^{-q}K\left(h_n^{-1}(W_i - w)\right)\right] - (\alpha_n I + T^*T)^{-1}T^*r\right\|_\infty,$$
$$B_n = \left\|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\right\|_\infty.$$

and the variance $\|\hat{\nu}_{1,n}\|_\infty$ is as in Eq. (7) with $X_{1,i}(w) = U_i h_n^{-q}K_w\left(h_n^{-1}(W_i - w)\right)$. Under the Assumption 3 by Proposition 2, the regularization bias is $B_n = O\left(\alpha_n^{\beta\wedge 1}\right)$, uniformly over $(\varphi, T) \in \mathcal{F}$.

Decompose the remaining term $\xi_n \leq \xi_{1,n} + \xi_{2,n}$ with

$$\xi_{1,n} = \left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\left[\frac{1}{n}\sum_{i=1}^n \frac{1}{h_n^q}K_w\left(h_n^{-1}(W_i - w)\right)\{\varphi(Z_i) - [K_{h_n} * \varphi](Z_i)\}\right]\right\|_\infty,$$
$$\xi_{2,n} = \left\|\left[(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T} - (\alpha_n I + T^*T)^{-1}T^*T\right]\varphi\right\|_\infty,$$

where $[K_{h_n} * \varphi](z_0) = \int_{[a,b]^p}\varphi(z)h_n^{-p}K_z\left(h_n^{-1}(z_0 - z)\right)\mathrm{d}z$. Under Assumption 3 and Assumptions 4 (ii)-(iii), by Lemma 2 and Proposition 5, for the second term, we have

$$\mathbb{E}\xi_{2,n} = O\left(\frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right)\right).$$

Coming back to the first term, under Assumption 4, we have

$$\xi_{1,n} \leq \|\hat{T}^*\|_{2,\infty}\left\|(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\right\|\left\|\frac{1}{n}\sum_{i=1}^n \frac{1}{h_n^q}K_w\left(h_n^{-1}(W_i - .)\right)\{\varphi(Z_i) - [K_{h_n} * \varphi](Z_i)\}\right\|.$$

Under Assumption 3, $\varphi \in C_M^t[a,b]^p$, so that using standard bias computations under Assumption 4 (iii), we have

$$\|[K_{h_n} * \varphi] - \varphi\|_\infty = O\left(h_n^t\right),$$

On the other hand, the second term is α_n^{-1}. Lastly, under Assumption 4 (i)

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n}\left|K_w\left(h_n^{-1}(W_i-.)\right)\right|\right\|^2 &\leq \mathbb{E}\left\|K_w\left(h_n^{-1}(W_i-.)\right)\right\|^2\\
&= \int\int K^2\left(h_n^{-1}(w-\tilde{w})\right)\mathrm{d}\tilde{w}f_W(w)\mathrm{d}w\\
&= O(h_n^q).
\end{aligned}$$

Therefore, by the Cauchy-Schwartz inequality

$$\mathbb{E}\xi_{1,n} = O\left(\alpha_n^{-1}h_n^{t-q/2}\right).$$

Now, we focus on $j=2$. Decompose

$$\|\hat{\varphi}_{\alpha_n}-\varphi\|_\infty \leq \|\hat{\nu}_{2,n}\|_\infty + \zeta_n + B_n,$$

where

$$\begin{aligned}
\zeta_n &= \left\|(\alpha_n I+\hat{T}^*\hat{T})^{-1}\left(\hat{T}^*\hat{r}-\frac{1}{n}\sum_{i=1}^{n}U_i f_{ZW}(.,W_i)\right)-(\alpha_n I+T^*T)^{-1}T^*r\right\|_\infty\\
B_n &= \|(\alpha_n I+T^*T)T^*r-\varphi\|_\infty.
\end{aligned}$$

Similarly, to $j=1$ case, the bias is of order $O\left(\alpha_n^{\beta\wedge 1}\right)$. Likewise, decompose $\zeta_n \leq \zeta_{1,n}+\xi_{2,n}$, where $\xi_{2,n}$ is the same as in the $j=1$ case, and

$$\begin{aligned}
\zeta_{1,n} &= \left\|(\alpha_n I+\hat{T}^*\hat{T})^{-1}\left(\hat{T}^*\hat{r}-\hat{T}^*\hat{T}\varphi-\frac{1}{n}\sum_{i=1}^{n}U_i f_{ZW}(.,W_i)\right)\right\|_\infty\\
&\leq \left\|(\alpha_n I+\hat{T}^*\hat{T})^{-1}\right\|_\infty\left\|\hat{T}^*(\hat{r}-\hat{T}\varphi)-\frac{1}{n}\sum_{i=1}^{n}U_i f_{ZW}(.,W_i)\right\|_\infty\\
&=: I_n\times II_n.
\end{aligned}$$

By Lemma 1, $\mathbb{E}I_n = O\left(\alpha_n^{-3/2}\right)$. Decompose the second term as

$$\begin{aligned}
II_n &\leq \left\|T^*(\hat{r}-\hat{T}\varphi)-\frac{1}{n}\sum_{i=1}^{n}U_i f_{ZW}(.,W_i)\right\|_\infty + \left\|(\hat{T}^*-T^*)(\hat{r}-\hat{T}\varphi)\right\|_\infty\\
&\leq III_n + IV_n + V_n,
\end{aligned}$$

where

$$\begin{aligned}
III_n &= \left\| \frac{1}{n}\sum_{i=1}^n U_i \left\{[f_{ZW} * K_{h_n}](.,W_i) - f_{ZW}(.,W_i)\right\} \right\|_\infty \\
IV_n &= \left\| \frac{1}{n}\sum_{i=1}^n \left\{\varphi(Z_i) - [\varphi * K_{h_n}](Z_i)\right\} [f_{ZW} * K_{h_n}](.,W_i) \right\|_\infty \\
V_n &= \left\| (\hat{T}^* - T^*)(\hat{r} - \hat{T}\varphi) \right\|_\infty,
\end{aligned}$$

and $[f_{ZW} * K_{h_n}](z,W_i) = \int_{[a,b]^q} f_{ZW}(z,w) h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right) \mathrm{d}w$. Using standard bias computations, under Assumption 4 (i)-(iii), we have a.s. $III_n = O(h_n^s)$ and $IV_n = O(h_n^t)$. Lastly,

$$\mathbb{E}V_n = \mathbb{E}\left\| (\hat{T}^* - T^*)(\hat{r} - \hat{T}\varphi) \right\|_\infty \leq \sqrt{\mathbb{E}\left\| \hat{T}^* - T^* \right\|_{2,\infty}^2 \mathbb{E}\left\| \hat{r} - \hat{T}\varphi \right\|^2}.$$

The order of the first term follows by Lemma 5, while

$$\begin{aligned}
\mathbb{E}\left\|\hat{r} - \hat{T}\varphi\right\|^2 &\leq 2\mathbb{E}\left\| \frac{1}{n}\sum_{i=1}^n U_i \frac{1}{h_n^q} K_w\left(h_n^{-1}(W_i - .)\right) \right\|^2 + 2\mathbb{E}\left\| \frac{1}{n}\sum_{i=1}^n \left\{\varphi(Z_i) - [K_h * \varphi](Z_i)\right\} \frac{1}{h_n^q} K_w\left(h_n^{-1}(W_i - .)\right) \right\|^2 \\
&= O\left(\frac{1}{nh_n^q} + h_n^{2t-q}\right).
\end{aligned}$$

Therefore, collecting all estimates, we obtain

$$\mathbb{E}\zeta_n = O\left(\frac{1}{\alpha_n^{3/2}} \left(h_n^s + h_n^t + \left(\frac{1}{\sqrt{nh_n^q}} + h_n^{t-q/2} \right) \left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s \right) \right) + \frac{1}{\alpha_n^{1/2}} \left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s \right) \right).$$

It remains to asses the order of variance terms

$$\begin{aligned}
\mathbb{E}\left\|\hat{\nu}_{1,n}\right\|_\infty &\leq \mathbb{E}\|\hat{T}^*\|_{2,\infty} \left\| (\alpha_n I + \hat{T}^*\hat{T})^{-1} \right\| \left\| \frac{1}{n}\sum_{i=1}^n U_i h_n^{-q} K_w\left(h_n^{-1}(W_i - .)\right) \right\| \\
&= O\left(\frac{1}{\alpha_n} \sqrt{\mathbb{E}\left\| \frac{1}{n}\sum_{i=1}^n U_i h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right) \right\|^2} \right),
\end{aligned}$$

where

$$\mathbb{E}\left\| \frac{1}{n}\sum_{i=1}^n U_i h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right) \right\|^2 = \frac{1}{nh_n^{2q}} \mathbb{E}\left\|U_i K_w\left(h_n^{-1}(W_i - w)\right)\right\|^2 = O\left(\frac{1}{nh_n^q}\right).$$

In the similar way, applying Lemma 1, we obtain

$$\mathbb{E}\|\hat{\nu}_{2,n}\|_\infty = O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right).$$

□

Proof of Corollary 2. Notice that in the $j = 1$ case, it is necessary to set $h_n \sim n^{-\frac{1}{2t}}$ to balance the variance with $h_n^t/\alpha_n h^{q/2}$ term. In this case the order of the variance will be $O\left((\alpha_n n^{1/2-q/4t})^{-1}\right)$ and we need $t > q/2$ to make sure that it goes to zero. To balance this variance with remaining bias term $\alpha_n^{\beta\wedge 1}$, it is necessary to set $\alpha_n \sim n^{-\frac{2t-q}{4t(\beta\wedge 1+1)}}$. It remains to make sure that $\frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right) = o\left(\frac{1}{\alpha_n n^{1/2} h_n^{q/2}}\right)$ for the choice of tuning parameters specified above to ensure that this term converges to zero faster than the bias and the variance terms, or in other words, we need $\alpha_n h_n^{-p} \to 0$ and $\alpha_n n h_n^{2s+q} \to 0$. The first requirement reduces to $t > q/2 + p(\beta \wedge 1 + 1)$, while the second requirement holds whenever $s > 0.75(t - q/2)$.

Now we turn to the case $j = 2$. First, it is necessary to set $h_n \sim n^{-\frac{1}{2(s\wedge t)}}$ to balance bias coming from non-parametric estimation with the variance term. Second, we need to set $\alpha_n \sim n^{-\frac{1}{2(\beta\wedge 1)+3}}$ to balance the regularization bias with the variance term. To ensure that $\frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right) = o\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right)$, we need $\alpha_n h_n^{-\frac{p+q}{2}} \to 0$ and $\alpha_n n^{1/2} h_n^s \to 0$. The first condition reduces to $s \wedge t > \frac{p+q}{4}(2(\beta \wedge 1) + 3)$, while the second requirement is satisfied for any $s, t > 0$. It remains to ensure that $\alpha_n n^{-1} h_n^{-q} \to 0$ and $\alpha_n h_n^{2t-q} \to 0$. The former holds whenever $s \wedge t > \frac{5}{12} q$, while for the latter it is sufficient to have $t > q/2$. Obviously, $s \wedge t > \frac{5}{4}(p + q)$ is the minimal binding condition for any $\beta > 0$. Lastly, we show that $\frac{1}{\alpha_n^{3/2}}\left(\frac{1}{\sqrt{nh_n^q}} + h_n^{t-q/2}\right)\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right) = o\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right)$, which reduces to verifying that $n^{-1/2} h_n^{-p/2-q} \to 0$, $h_n^{t-p/2-q} \to 0$, $h_n^{s-q/2} \to 0$, and $n^{1/2} h_n^{t+s-q/2} \to 0$. Under stated rates for tuning parameters, these four conditions in turn are satisfied whenever $s \wedge t > p/2 + q$, $t > p/2 + q$, $s > q/2$, and $s \wedge t > q/2$ respectively, so that we need $s \wedge t > p/2 + q$. Therefore, in light of the previous estimates, the condition $s \wedge t > \frac{5}{4}(p + q)$ is sufficient to ensure that the variance and the bias component dominate in the risk decomposition. □

Proof of Theorem 4. We check that under Assumptions 3 and 4, all relevant conditions of Theorems 1 and 2 are satisfied. First, under Assumptions 4, we obtain Assumptions 1 with $u_n = h_n^{-q}$ and $X'_{1,i}(w) = U_i K_w\left(h_n^{-1}(W_i - w)\right)$. In particular, boundedness and compact support of the density function f_{ZW} and the kernel function K ensure both Assumptions 1 (iv) and (v). Theorem 3 gives the risk decomposition and so Assumption 2 (i) is satisfied. Assumption 2 (ii) follows, since by Proposition 5

$$\frac{1}{\alpha_n^{1/2}}\left(\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^4\right)^{1/4} = \frac{1}{\alpha_n^{1/2}}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s\right),$$

which goes to zero under the smoothness conditions and orders on tuning parameters imposed in Corollary 2. It remains to verify Assumption 2 (iii). To that end, notice that the first condition follows by the uniform convergence of the estimator, since

$$\left\|\hat{X}'_{1,i} - X'_{1,i}\right\| \leq \|K\|_\infty \|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty .$$

For the second condition we also need uniform consistency of the kernel density estimator, since

$$\left\|\hat{X}_{2,i} - X_{2,i}\right\|_\infty \leq \|\hat{\varphi} - \varphi\|_\infty \|f_{ZW}\|_\infty + F_0 \left\|\hat{f}_{ZW} - f_{ZW}\right\|_\infty + \|\hat{\varphi} - \varphi\|_\infty \left\|\hat{f}_{ZW} - f_{ZW}\right\|_\infty .$$

□

Proof of Corollary 3. For $j = 1$, $\rho_{j,n} = n^{1/2}\alpha_n h_n^{q/2}$ and to balance the variance and the bias we need $\alpha_n^{-1} n^{-1/2} h_n^{-q/2} \sim n^{1/2} h_n^t \sim n^{1/2} \alpha_n^{\beta\wedge 1+1} h_n^{q/2}$. This is achieved when $h_n \sim n^{-\frac{\beta\wedge 1+1}{(\beta\wedge 1)(t+q/2)+2t}}$ and $\alpha_n \sim n^{-\frac{t-q/2}{(\beta\wedge 1)(t+q/2)+2t}}$. For this choice of tuning parameters, it remains to verify that $\alpha_n^{-\beta\wedge 1-1/2}\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} \to 0$ and $\alpha_n^{-\beta\wedge 1-1/2} h_n^s \to 0$ to ensure that the noise coming from the estimation of the operator has smaller effect on the coverage error than the bias. The first requirement reduces to $t > q/2 + p\frac{1+\beta\wedge 1}{1-\beta\wedge 1}$ while the second requirement is satisfied whenever $s > (t - q/2)\frac{\beta\wedge 1+1/2}{\beta\wedge 1+1}$.

Now we turn to the case $j = 2$. Here to balance the variance and the bias we need $\alpha_n^{-3/2} n^{-1/2} \sim \alpha_n^{3/2+\beta\wedge 1} n^{1/2} \sim n^{1/2}(h_n^s + h_n^t)$. This requirement is achieved when $\alpha_n \sim n^{-\frac{1}{\beta\wedge 1+3}}$ and $h_n \sim n^{-\frac{2(\beta\wedge 1)+3}{2(s\wedge t)(\beta\wedge 1+3)}}$. It remains to show that additional noise has negligible impact on coverage errors, comparing to the bias. To that end, we need to verify that $h_n^{-s\wedge t-q/2-(p+q)/2} n^{-1} \to 0$, $h_n^{-s\wedge t+t-q/2-(p+q)/2} n^{-1/2} \to 0$, $h_n^{-s\wedge t-q/2+s} n^{-1/2} \to 0$, $h_n^{-s\wedge t+t-q/2+s} \to 0$, $\alpha_n h_n^{-s\wedge t-(p+q)/2} n^{-1/2} \to 0$, and $\alpha_n h_n^{-s\wedge t+s} \to 0$. These requirements are satisfied when $s\wedge t > (p+2q)\frac{2(\beta\wedge 1)+3}{6}$, $s\wedge t > (p+2q)\frac{2(\beta\wedge 1)+3}{2(\beta\wedge 1+3)}$, $s\wedge t > q\frac{2(\beta\wedge 1)+3}{2(\beta\wedge 1+3)}$, $s\wedge t > q/2$, and $s\wedge t > (p+q)\frac{2(\beta\wedge 1)+3}{2(2-\beta\wedge 1)}$ respectively (the last requirement is trivially satisfied). All requirements hold whenever $s\wedge t > 2.5(p+q)$.

A trivial application of the Corollary 1 describes convergence rates for expected diameters. □

Proofs for functional regressions

Proof of Theorem 5. Continuity of trajectories of processes Z and W ensures several useful properties. First, by the Dominated convergence theorem, the covariance function $\mathbb{E}[Z(t)W(s)]$ is continuous on $[0,1]^{p+q}$. This in turn implies that covariance operators T and T^* are in $\mathcal{L}_{2,\infty}$. Second, it also ensures that $\hat{T}, \hat{T}^* \in \mathcal{L}_{2,\infty}$ a.s.

Consider the following decomposition

$$\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty \leq \|\hat{\nu}_{1,n}\|_\infty + \xi_n + B_n,$$

where

$$\xi_n = \left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T}\varphi - (\alpha_n I + T^*T)^{-1}T^*T\varphi\right\|_\infty$$

$$\|\hat{\nu}_{1,n}\|_\infty = \left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*(\hat{r} - \hat{T}\varphi)\right\|_\infty$$

$$B_n = \left\|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\right\|_\infty$$

The order of the bias term B_n follows under the source condition (Assumption 3) by Proposition 2.

Notice that

$$\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^2 \leq \mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_iW_i - \mathbb{E}[Z_iW_i]\right\|_\infty^2$$

Using Hoffman-Jørgensen's inequality, e.g. (Giné and Nickl, 2015, Theorem 3.1.16)

$$\left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_iW_i - \mathbb{E}[Z_iW_i]\right\|_\infty^2\right)^{1/2} = O\left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_iW_i - \mathbb{E}[Z_iW_i]\right\|_\infty + n^{-1}\left(\mathbb{E}\left[\max_{1\leq i\leq n}\|Z_iW_i\|_\infty^2\right]\right)^{1/2}\right).$$

Under Assumption 5 (ii), the second term is $O(n^{-1})$, while the first term is the expected value of the supremum of the empirical process. Since the process $(t,s)\mapsto Z(t)W(s)$ has Hölder smooth trajectories, by (Giné and Nickl, 2015, Theorem 3.5.13 and 4.3.36)

$$\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_iW_i - \mathbb{E}[Z_iW_i]\right\|_\infty = O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore, by Lemma 2

$$\sup_{(\varphi,T)\in\mathcal{F}} \mathbb{E}\xi_n = O\left(\frac{1}{\alpha_n^{1/2}n^{1/2}}\right),$$

which established the first claim of the theorem.

Now we consider the variance term

$$\begin{aligned}
\mathbb{E}\|\hat{\nu}_{1,n}\|_\infty &= \mathbb{E}\left\|\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1}(\hat{r} - \hat{T}\varphi)\right\|_\infty \\
&\leq \mathbb{E}\|\hat{T}^*\|_{2,\infty}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\right\|\left\|\hat{r} - \hat{T}\varphi\right\| \\
&\leq \frac{\|T^*\|_{2,\infty}}{\alpha_n}\sqrt{\mathbb{E}\left\|\hat{r} - \hat{T}\varphi\right\|^2} + \frac{1}{\alpha_n}\sqrt{\mathbb{E}\|\hat{T}^* - T^*\|_{2,\infty}^2\mathbb{E}\left\|\hat{r} - \hat{T}\varphi\right\|^2},
\end{aligned}$$

where we applied the Cauchy-Schwartz inequality to obtain the last line.

Under Assumption 5 (i)

$$\mathbb{E}\left\|\hat{r} - \hat{T}\varphi\right\|^2 = O\left(\frac{1}{n}\right). \tag{18}$$

Therefore, taking previous estimates into account, we obtain the desired order of the variance and the second claim of theorem follows. $\square$

Proof of Theorem 6. We apply Theorem 1 and 2. Notice that we can set $u_n = 1$ so that $X_{1,i} = X'_{1,i}$ and that under Assumptions 5 (i) and (iii)-(v), the Assumption 1 is satisfied. Theorem 5 ensures Assumption 2 (i). To verify Assumption 2 (ii), notice that

$$\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^4 \leq \mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i]\right\|_\infty^4.$$

Under Assumption 5 (ii), applying the Hoffman-Jørgensen's inequality, see (Giné and Nickl, 2015, Theorem 3.1.16) and (Giné and Nickl, 2015, Theorem 3.5.13 and 4.3.36)

$$\left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i]\right\|_\infty^4\right)^{1/4} = O\left(n^{-1/2}\right).$$

This gives

$$\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^4 = O\left(n^{-2}\right),$$

which is $o(\alpha_n^2)$, since $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ as $n \to \infty$.

Also notice that by Cauchy-Schwartz inequality under Assumption 5 (ii)

$$\mathbb{E}\max_{1\leq i\leq n}\left\|\hat{X}_{1,i} - X_{1,i}\right\| = \mathbb{E}\left\|\langle\hat{\varphi}_{\alpha_n} - \varphi, Z_i\rangle W_i\right\| \leq \mathbb{E}\left\|\hat{\varphi} - \varphi\right\|_\infty,$$

which goes to zero whenever $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ by Theorem 5. This verifies Assumption 2 (iii) and finishes the proof. $\square$

Proofs for the deconvolution model

Proof of Theorem 7. The result follows from the following variance-bias decomposition

$$\begin{aligned}\|\hat{\varphi}_{\alpha_n} - \varphi\|_\infty &\leq \left\|(\alpha_n I + T^*T)^{-1}(\widehat{T^*r} - T^*r)\right\|_\infty + \|(\alpha_n I + T^*T)^{-1}T^*r - \varphi\|_\infty \\ &=: \|\nu_{2,n}\|_\infty + B_n.\end{aligned}$$

By Proposition 2, the bias can be bounded as $B_n \leq R\alpha_n^{\beta\wedge 1}$ proving the first result. For the second result, we bound the variance as

$$\|\nu_{2,n}\|_\infty \leq \left\|(\alpha_n I + T^*T)^{-1}\right\|_\infty \left\|\frac{1}{n}\sum_{i=1}^n f(Y_i - .) - \mathbb{E}f(Y - .)\right\|_\infty,$$

where the second term is the supremum of empirical process indexed by the class of functions

$$\tilde{\mathcal{G}}_n = \{k : [a, b] \to \mathbf{R},\ y \mapsto f(y - z), z \in [a, b]\} .$$

By Assumption 6 (iii), $\tilde{\mathcal{G}}_n$ is class of s-Hölder functions. Then by (Giné and Nickl, 2015, Corollary 3.5.8)

$$\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n} f(Y_i - .) - \mathbb{E}f(Y - .)\right\|_\infty = O\left(\frac{1}{\sqrt{n}}\right).$$

Applying Lemma 1 to $\left\|(\alpha_n I + T^*T)^{-1}\right\|_\infty$, the result follows. □

Proof of Theorem 8. We verify Assumptions 1 and 2. Recall that in the risk the composition given in Theorem 7, the variance process is driven by $X_{2,i}(z) = f(Y_i - z)$, which is bounded by $F_2 = \|f\|_\infty$. Therefore, under Assumptions 6, Assumption 1 (i) and (iii)-(vi) are satisfied. Assumption 2 (i) follows from Theorem 7 with $r_{2,n} = \alpha_n^{\beta\wedge 1}$, while (ii) and (iii) are trivially satisfied due to the fact that the operator T^* and the process $X_{2,i}$ are known. □

Appendix C: Additional results

We need the following inequality, known in the numerical ill-posed inverse problems literature, e.g. see (Rajan, 2003, Proposition 2.1), (Nair, 2009, Chapter 5, problem (8)). For completeness we provide a short proof here.

Lemma 1. *Suppose that $T : L_2[a,b]^p \to L_2[a,b]^q$ is an integral operator with continuous kernel function. Then for any $\alpha > 0$ $(\alpha I + T^*T)$ is invertible as an operator from $\mathcal{R}(T^*T) \subset (C, \|.\|_\infty)$ to $(C, \|.\|_\infty)$ space and*

$$\left\|(\alpha I + T^*T)^{-1}\right\|_\infty \leq \frac{\|T^*\|_{2,\infty}/2 + \alpha^{1/2}}{\alpha^{3/2}}.$$

Proof. Continuity of the kernel function ensures that T^*T maps to $C[a,b]^p$. Since for any $\alpha > 0$, $\alpha I + T^*T$ is invertible as an operator on the L_2 space, (Nair, 2009, Lemma 4.1), and $(C, \|.\|_\infty) \subset L_2$, it is also invertible as an operator between $(C, \|.\|_\infty)$.

Notice that for any $\varphi \in C[a,b]^p$

$$\left[(\alpha_n I + T^*T)^{-1}T^*T - I\right]\varphi = -\alpha_n(\alpha_n I + T^*T)^{-1}\varphi.$$

Using this identity

$$\left\|(\alpha_n I + T^*T)^{-1}\right\|_\infty = \sup_{\|\varphi\|_\infty = 1}\left\|(\alpha_n I + T^*T)^{-1}\varphi\right\|_\infty \leq \frac{\|(\alpha_n I + T^*T)^{-1}T^*T\|_\infty + 1}{\alpha_n}. \tag{19}$$

Factoring the norm as $\|T^*\psi\|_\infty \leq \|T^*\|_{2,\infty}\|\psi\|$, and using the isometry of functional calculus

$$\left\|(\alpha_n I + T^*T)^{-1}T^*T\right\|_\infty = \left\|T^*(\alpha_n I + TT^*)^{-1}T\right\|_\infty \leq \|T^*\|_{2,\infty}\sup_{\lambda\in[0,\|T\|^2]}\left|\frac{\lambda^{1/2}}{\alpha_n + \lambda}\right| = \frac{\|T^*\|_{2,\infty}}{2\alpha_n^{1/2}}. \tag{20}$$

Combining this with Eq. (19) gives the result. □

Lemma 2. *Suppose that $T : L_2[a,b]^p \to L_2[a,b]^q$ is an integral operator in $\mathcal{L}_2$ with continuous kernel function. Suppose further that Assumption 3 is satisfied, then if the risk $\mathbb{E}\|\hat{T}^* - T^*\|^2_{2,\infty}$ is bounded uniformly over $\mathcal{F}$, for all $n \in \mathbf{N}$ and all $(\varphi, T) \in \mathcal{F}$, we have*

$$\mathbb{E}\left\|\left[(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T} - (\alpha_n I + T^*T)^{-1}T^*T\right]\varphi\right\|_\infty = O\left(\sqrt{\alpha_n^{-1}\mathbb{E}\|\hat{T}^* - T^*\|^2_{2,\infty}}\right).$$

Proof. $\|\hat{T}\| \leq \|\hat{k}\| < \infty$ a.s. ensure that $\hat{T}\hat{T}^*$ is a self-adjoint operator in $\mathcal{L}_2$ a.s., so that the spectral theory for bounded self-adjoint operators in Hilbert spaces (in particular functional calculus) applies to it similarly as to TT^* and T^*T.

Decompose

$$
\begin{aligned}
\left[(\alpha_n I+\hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T}-(\alpha_n I+T^*T)^{-1}T^*T\right]\varphi &= \left[\alpha_n(\alpha_n I+\hat{T}^*\hat{T})^{-1}-\alpha_n(\alpha_n I+T^*T)^{-1}\right]\varphi \\
&= (\alpha_n I+\hat{T}^*\hat{T})^{-1}\left[\hat{T}^*\hat{T}-T^*T\right]\alpha_n(\alpha_n I+T^*T)^{-1}\varphi \\
&\leq (\alpha_n I+\hat{T}^*\hat{T})^{-1}\hat{T}^*(\hat{T}-T)\alpha_n(\alpha_n I+T^*T)^{-1}\varphi \\
&+ (\alpha_n I+\hat{T}^*\hat{T})^{-1}(\hat{T}^*-T^*)\alpha_n T(\alpha_n I+T^*T)^{-1}\varphi \\
&\equiv I_n+II_n.
\end{aligned}
$$

Both terms involve the error from the estimation of the operators and regularization bias. For the first term, we factor the operator norm $\|\hat{T}^*\psi\|_\infty \leq \|\hat{T}^*\|_{2,\infty}\|\psi\|$

$$
\|I_n\|_\infty \leq \|\hat{T}^*\|_{2,\infty}\left\|(\alpha_n I+\hat{T}\hat{T}^*)^{-1}\right\|\|\hat{T}-T\|\left\|\alpha_n(\alpha_n I+T^*T)^{-1}\varphi\right\|
$$

By isometry of functional calculus

$$
\left\|(\alpha_n I+\hat{T}^*\hat{T})^{-1}\right\| = \sup_{\lambda\in[0,\|\hat{T}\|^2]}\left|\frac{1}{\alpha_n+\lambda}\right| \leq \frac{1}{\alpha_n}, \qquad \text{a.s.}
$$

Under Assumption 3, the last term is

$$
\left\|\alpha_n(\alpha_n I+T^*T)^{-1}(T^*T)^\beta T^*\psi\right\| \leq \alpha_n \sup_{\lambda\in[0,\|T\|^2]}\left|\frac{\lambda^{\beta+1/2}}{\alpha_n+\lambda}\right| C = O\left(\alpha_n^{(\beta+1/2)\wedge 1}\right).
$$

Combining these findings with Cauchy-Schwartz inequality and using the bound

$$
\|\hat{T}-T\| = \|\hat{T}^*-T^*\| \leq (b-a)^{p/2}\|\hat{T}^*-T^*\|_{2,\infty},
$$

we have uniformly in $(\varphi,T)\in\mathcal{F}$

$$
\mathbb{E}\|I_n\|_\infty = O\left(\frac{\alpha_n^{(\beta+1/2)\wedge 1}}{\alpha_n}\sqrt{\mathbb{E}\|\hat{T}^*-T^*\|^2_{2,\infty}}\right).
$$

For the second term, we factor the operator norm in the following way

$$
\|II_n\|_\infty \leq \mathbb{E}\left\|(\alpha_n I+\hat{T}^*\hat{T})^{-1}\right\|_\infty\left\|\hat{T}^*-T^*\right\|_{2,\infty}\left\|\alpha_n T(\alpha_n I+T^*T)^{-1}\varphi\right\|
$$

By Lemma 1 and triangle inequality

$$\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\right\|_\infty \leq \frac{\|\hat{T}^* - T^*\|_{2,\infty}/2 + \|T^*\|_{2,\infty}/2 + \alpha_n^{1/2}}{\alpha_n^{3/2}}$$

Under Assumption 3,

$$\left\|\alpha_n T(\alpha_n I + T^*T)^{-1}(T^*T)^\beta T^*\psi\right\| \leq \alpha_n \sup_{\lambda\in[0,\|T\|^2]}\left|\frac{\lambda^{\beta+1}}{\alpha_n+\lambda}\right| C = O(\alpha_n).$$

Combining all above findings, we have uniformly in (φ, T)

$$\mathbb{E}\|II_n\|_\infty = O\left(\frac{\mathbb{E}\|\hat{T}^* - T^*\|_{2,\infty}^2 + \mathbb{E}\|\hat{T}^* - T^*\|_{2,\infty}}{\alpha_n^{1/2}}\right),$$

and the conclusion follows from collecting all estimates and using Jensen's inequality. $\square$

Lemma 3. *Suppose that $T : L_2[a,b]^p \to L_2[a,b]^q$ is an integral operator in $\mathcal{L}_2$ with continuous kernel function for which we have continuous estimator. Consider the variance processes as in Eq. (5) and Eq. (7), where $X_{1,i}$ and $X_{2,i}$, $i = 1,\dots,n$ are i.i.d. centered processes with continuous trajectories such that $\mathbb{E}\|X_{1,i}'\|^2 < \infty$ and $\mathbb{E}\|X_{2,i}\|^2 < \infty$ uniformly over the class of models $\mathcal{F}$. Suppose additionally that $\mathbb{E}\|\hat{T}^* - T^*\|_{2,\infty}^4 = o\left(\alpha_n^2\right)$, then for any $(\varphi, T) \in \mathcal{F}$, we have*

$$\mathbb{E}\left\|\hat{\nu}_{1,n} - \nu_{1,n}\right\|_\infty = o\left(\frac{u_n}{\alpha_n n^{1/2}}\right), \qquad \mathbb{E}\left\|\hat{\nu}_{2,n} - \nu_{2,n}\right\|_\infty = o\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\right).$$

Moreover, for symmetrized processes, for any $(\varphi, T) \in \mathcal{F}$, we have

$$\mathbb{E}\left\|\hat{\nu}_{1,n}^\varepsilon - \nu_{1,n}^\varepsilon\right\|_\infty = o\left(\frac{u_n}{\alpha_n n^{1/2}}\left(1 + \sqrt{\mathbb{E}\|\hat{X}_{1,i}' - X_{1,i}'\|^2}\right)\right)$$

and

$$\mathbb{E}\left\|\hat{\nu}_{2,n}^\varepsilon - \nu_{2,n}^\varepsilon\right\|_\infty = o\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\left(1 + \sqrt{\mathbb{E}\|\hat{X}_{2,i} - X_{2,i}\|^2}\right)\right).$$

Proof. Similarly as in the proof of Lemma 2, decompose

$$\begin{aligned}
\mathbb{E}\left\|\hat{\nu}_{2,n} - \nu_{2,n}\right\|_\infty &= \mathbb{E}\left\|\left[(\alpha_n I + \hat{T}^*\hat{T})^{-1} - (\alpha_n I + T^*T)^{-1}\right]\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\|_\infty \\
&= \mathbb{E}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}(T^*T - \hat{T}^*\hat{T})(\alpha_n I + T^*T)^{-1}\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\|_\infty \\
&\leq R_{1,n} + R_{2,n},
\end{aligned}$$

where we can show using similar computations that

$$
\begin{aligned}
R_{1,n} &= \mathbb{E}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}(\hat{T}^* - T^*)T(\alpha_n I + TT^*)^{-1}\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\|_\infty \\
&\leq \mathbb{E}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\right\|_\infty \left\|\hat{T}^* - T^*\right\|_{2,\infty} \left\|T(\alpha_n I + T^*T)^{-1}\right\| \left\|\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\| \\
&= O\left(\frac{1}{\alpha_n^2 n^{1/2}}\left(\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^4\right)^{1/4}\left(\mathbb{E}\|X_{2,i}\|^2\right)^{1/2}\right) \\
R_{2,n} &= \mathbb{E}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*(\hat{T} - T)(\alpha_n I + T^*T)^{-1}\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\| \\
&\leq \mathbb{E}\|\hat{T}^*\|_{2,\infty}\left\|(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\right\|\left\|\hat{T} - T\right\|\left\|(\alpha_n I + T^*T)^{-1}\right\|\left\|\frac{1}{n}\sum_{i=1}^n X_{2,i}\right\| \\
&= O\left(\frac{1}{\alpha_n^2 n^{1/2}}\left(\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^4\right)^{1/4}\left(\mathbb{E}\|X_{2,i}\|^2\right)^{1/2}\right)
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\mathbb{E}\left\|\hat{\nu}_{1,n} - \nu_{1,n}\right\|_\infty &= \mathbb{E}\left\|\left[\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha_n I + TT^*)^{-1}\right]\frac{1}{n}\sum_{i=1}^n X_{1,i}\right\|_\infty \\
&\leq S_{1,n} + S_{2,n},
\end{aligned}
$$

and we can show that

$$
\begin{aligned}
S_{1,n} &= \mathbb{E}\left\|T^*\left[(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - (\alpha_n I + TT^*)^{-1}\right]\frac{1}{n}\sum_{i=1}^n X_{1,i}\right\|_\infty \\
&= O\left(\frac{1}{\alpha_n^{3/2} n^{1/2}}\left(\mathbb{E}\left\|\hat{T}^* - T^*\right\|^4\right)^{1/4}\left(\mathbb{E}\|X_{1,i}\|^2\right)^{1/2}\right). \\
S_{2,n} &= \mathbb{E}\left\|(\hat{T}^* - T^*)(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\frac{1}{n}\sum_{i=1}^n X_{1,i}\right\|_\infty \\
&= O\left(\frac{1}{\alpha_n n^{1/2}}\sqrt{\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^2 \mathbb{E}\|X_{1,i}\|^2}\right).
\end{aligned}
$$

For the first symmetrized processes, using the fact that $(\varepsilon_i)_{i=1}^n$ are i.i.d. Rademacher random variables, independent from $(X_{1,i}, \hat{X}_{1,i})_{i=1}^n$, we obtain for $j = 1, 2$

$$
\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \hat{X}_{1,i}\right\|^2 = \frac{\mathbb{E}\|\hat{X}_{1,i}\|^2}{n} \leq \frac{2\mathbb{E}\|X_{1,i}\|^2}{n} + \frac{2\mathbb{E}\|\hat{X}_{1,i} - X_{1,i}\|^2}{n}
$$

and the result follows from previous computations. Similarly, we obtain result for the second process. □

Lemma 4. *Suppose that $T^* \in \mathcal{L}_{2,\infty}$ is integral operator with continuous kernel function k, then*

$$\|T^*\|_{2,\infty} = \sup_{z\in[a,b]^p} \left(\int_{[a,b]^q} |k(z,w)|^2 \mathrm{d}w \right)^{1/2} \equiv \|k\|_{2,\infty},$$

where $\|k\|_{2,\infty}$ is a mixed norm on the iterated space $L_\infty[a,b]^p(L_2[a,b]^q)$.

Proof. By Cauchy-Schwartz inequality, we have $\|T^*\|_{2,\infty} \leq \|k\|_{2,\infty}$. On the other side, continuity of k implies that $\exists z_0 \in [a,b]^p$ such that $\|k\|_{2,\infty} = \left(\int_{[a,b]^q} |k(z_0,w)|^2 \mathrm{d}w\right)^{1/2}$. Take $\psi(w) = \frac{k(z_0,w)}{\|k(z_0,.)\|}$. Then $\|\psi\| = 1$, and

$$\|T^*\|_{2,\infty} \geq \|T^*\psi\|_\infty \geq |(T^*\psi)(z_0)| = \|k(z_0,.)\| = \|k\|_{2,\infty}.$$

$\square$