Econometric Theory, **36**, 2020, 658–706.
doi:10.1017/S0266466619000380

HONEST CONFIDENCE SETS IN NONPARAMETRIC IV REGRESSION AND OTHER ILL-POSED MODELS*

ANDRII BABII
University of North Carolina at Chapel Hill

This article develops inferential methods for a very general class of ill-posed models in econometrics encompassing the nonparametric instrumental variable regression, various functional regressions, and the density deconvolution. We focus on uniform confidence sets for the parameter of interest estimated with Tikhonov regularization, as in Darolles et al. (2011, *Econometrica* 79, 1541–1565). Since it is impossible to have inferential methods based on the central limit theorem, we develop two alternative approaches relying on the concentration inequality and bootstrap approximations. We show that expected diameters and coverage properties of resulting sets have uniform validity over a large class of models, that is, constructed confidence sets are honest. Monte Carlo experiments illustrate that introduced confidence sets have reasonable width and coverage properties. Using U.S. data, we provide uniform confidence sets for Engel curves for various commodities.

1. INTRODUCTION

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

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

where $Z \in \mathbf{R}^p$ is a vector of explanatory variables, $W \in \mathbf{R}^q$ is a vector of instruments, and φ is the function of interest. The NPIV model is ill-posed in the sense that the

*First draft: March 2016. This article is a revised first chapter of my Ph.D. thesis. I'm grateful to the Co-Editor and valuable referees for constructive criticism and suggestions of how to improve the article. I'm deeply indebted to my advisor Jean-Pierre Florens and other members of my Ph.D. committee: Eric Gautier, Ingrid Van Keilegom, and Timothy Christensen for helpful suggestions and insightful conversations. This article also benefited from discussions with Christian Bontemps, Samuele Centorrino, Jasmin Fliegner, Emanuele Guerre, Vitalijs Jascisens, Jihyun Kim, Rohit Kumar, Elia Lapenta, Pascal Lavergne, Thierry Magnac, André Mas, Nour Meddahi, Markus Reiss, and Shruti Sinha. Address correspondence to Andrii Babii, University of North Carolina at Chapel Hill, Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305, USA; e-mail: babii.andrii@gmail.com

 © Cambridge University Press 2020

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

map from the distribution of the data to the function φ is not continuous. As a result, we need to introduce some amount of regularization that would smooth out discontinuities and yield a consistent estimator.

In empirical studies, the function φ represents a structural economic relation, such as an Engel curve, cost function, or demand curve. It is not enough to estimate this function to infer possible economic effects. We can only infer the range of possible economic effects from confidence sets. This article is the first to provide inferential methods for Tikhonov-regularized estimators. I focus on *uniform inference*, which amounts to constructing a set containing the *entire* function φ with a high probability. Uniform inference allows one to assess global features of the estimated function and to quantify the range of possible economic effects compatible with the data. Global features may include the evidence for nonlinearities, the amount of endogeneity bias compared with 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 a valid inference for the *entire* function φ. Another feature of confidence sets constructed in this article is *honesty* in the sense of Li (1989). Honesty means that coverage properties have uniform validity 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.

Building uniform confidence sets for a function requires approximating the distribution of the supremum of a certain stochastic process. We show that for a broad class of ill-posed models treated with Tikhonov regularization, the stochastic process driving the distribution of the estimator does not converge weakly in the space of functions. As a result, *it is not possible* to build the uniform confidence set by relying on the uniform central limit theorem which calls for alternative approaches to inference.

To construct confidence sets with good coverage properties, we rely on two different approaches. The first inferential method developed in this article relies on estimates of tail probabilities with a suitable concentration inequality. A simple illustration of this approach is to build a uniform confidence band for the empirical distribution function using the Dvoretzky–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 Dvoretzky–Kiefer–Wolfowitz tells us that for any *finite sample size* n the probability of F_n deviating from the true distribution function F in the supremum norm declines at the exponential rate

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

[1] See equation (2) in Section 3 for a formal definition.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

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 a uniform confidence band for $F(x)$ with a 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 is different from 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, confidence sets for such estimators, including the ill-posed models considered in this article, usually have coverage errors.

In this article, we rely on a concentration inequality valid for more complex statistics than $\|F_n - F\|_\infty$, see Boucheron et al. (2013). In particular, we exploit a data-driven 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 broad class of data-generating processes and is especially useful in settings where all other approaches to inference fail.

For confidence sets based on the concentration inequality, we characterize nonasymptotic rates of coverage errors and expected diameters explicitly. It is especially important to know both rates, since consistency alone may not be very informative for inference. For instance, the coverage error may decrease at a rate slower than the shrinking of the confidence set, requiring larger sample sizes to achieve good coverage. We show that the bias of the estimator, not the noise coming from the estimation of the operator, drives the coverage errors of our confidence sets. To the best of our knowledge, convergence rates for coverage errors have not been derived for the NPIV model or other ill-posed models considered in this article.

As an alternative, we also study a more traditional bootstrap inference that relies on a nontrivial application of Gaussian and bootstrap coupling inequalities developed in a seminal series of papers, Chernozhukov et al. (2014, 2016). To the best of our knowledge, our application of these coupling results to the nonparametric IV model and Tikhonov estimators is new.

Although the Tikhonov-regularized NPIV is the leading example, the inferential methods developed in this article are valid for other ill-posed models, including functional regression models and the density deconvolution model. To the best of our knowledge, despite the extensive existing literature on L_2 results for the

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

functional regression models, no uniform convergence rates or uniform inferential methods are currently available. Moreover, Tikhonov regularization plays a prominent role in the statistical learning theory. The inferential results obtained in this article can also be applied in that setting.

Contribution and related literature. This article is the first to develop honest uniform inferential methods in a general and unifying framework, encompassing different ill-posed models treated with Tikhonov regularization. This article is also the first to provide uniform inference for several ill-posed inverse problems based on the concentration inequality and to provide the nonasymptotic analysis, including convergence rates of coverage errors. Uniform confidence sets are available only for sieve-type estimators of the NPIV model and the density deconvolution model. The literature is also silent about coverage errors of confidence sets. Moreover, it is not known what sorts of restrictions are needed to construct honest confidence sets for ill-posed inverse problems. In this article, we show that the class of models should involve both the class of functions and the class of operators, which is different from classical direct nonparametric problems where we need only to restrict the function. 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 grow at a certain speed to achieve uniform coverage. Chen and Christensen (2018) develop inferential methods for the sieve NPIV estimator without relying on discrete approximations. They focus on uniform inference for a collection of linear and nonlinear functionals using Yurinskii's coupling and obtain uniform confidence bands in the particular case of point evaluation functionals, see also Belloni et al. (2015) for the conditional mean function and Tao (2014) for general conditional moment restriction models. Lounici and Nickl (2011) obtain uniform confidence sets for the wavelet deconvolution estimator by relying on Bousquet's version of Talagrand's concentration inequality.

Kato and Sasaki (2018) consider a more general case, where the density of the noise is estimated from an auxiliary sample. This article studies the estimator based on Fourier inversion and builds on the coupling inequalities developed in Chernozhukov et al. (2014, 2016).

We show that the confidence sets developed in this article may enjoy polynomial convergence rates of coverage errors in mildly ill-posed, and some severely ill-posed, cases. An appealing feature of Tikhonov's and more general spectral regularizations is that the estimator is always well-defined in finite samples and converges to the best approximation to the structural parameter under various identification failures, see Florens et al. (2011) and Babii and Florens (2018).

The theoretical validity of a uniform confidence set relies on uniform convergence rates of the corresponding estimator. Gagliardini and Scaillet (2012b) obtain uniform convergence rates for the Tikhonov-regularized minimum distance estimator relying on the Sobolev embedding. Chen and Christensen (2018) show that the sieve nonparametric IV estimator can attain optimal asymptotic uniform

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

convergence rates under a somewhat different set of assumptions. Babii and Florens (2018) derive *nonasymptotic* L_2 and uniform risk bounds for more general spectrally regularized estimators not tied to a particular estimator of the conditional mean function and show that it is possible to attain polynomial convergence rates in some severely ill-posed cases.[2]

Some pointwise inferential results are available for spectral cutoff estimators when the ill-posed operator is known and is not estimated from the data, see Carrasco and Florens (2011), Gagliardini and Scaillet (2012a), Gautier and Kitamura (2013), and Florens et al. (2017). At the same time, Chen and Pouzo (2015) provide pointwise inference and bootstrap confidence bands for conditional moment restriction models treated with a sieve approach and nesting the NPIV model as a special case. Cardot et al. (2007) provide results on the asymptotic normality of linear functionals, which solves the prediction problem for functional regression models. Lastly, Carrasco et al. (2014) study asymptotic normality of inner products for Tikhonov estimators.

The structure of the article is as follows. We introduce the notation in the remaining part of this section. Section 3 describes the problem of constructing honest uniform confidence sets and introduces two inferential approaches in the special case of the nonparametric IV model. Under a general set of assumptions that are verified later on in each particular application, we establish convergence rates for coverage errors and diameters of constructed sets, uniform over a general set of models. Section 4 extends these results to different functional regression models, and to the density deconvolution. In Section 5, we show how to implement confidence sets in practice for the NPIV estimator and explore their finite-sample properties with Monte Carlo experiments. Section 6 considers the empirical application to Engel curves, and Section 7 concludes.

Notation. Let $L_2[0,1]^p$ denote the space of functions on some compact set $[0,1]^p \subset \mathbf{R}^p$, square integrable with respect to the Lebesgue measure λ. For $\varphi \in L_2[0,1]^p$, 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 ${C_M}^\beta[a,b]^p$ denote the class of β-Hölder functions on $(a,b)^p$ with $0 < M < \infty$

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

[2]The minimax-optimal uniform convergence rates in the severely ill-posed case are logarithmically slow, see Chen and Christensen (2018). However, in the more restricted smoothness class, when the estimated function can be described by the finite number of generalized Fourier coefficients, or more generally if its smoothness matches the ill-posedness of the operator, one can achieve faster polynomial uniform convergence rates.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

where $k = (k_1, \ldots, k_p) \in \mathbf{N}^p$ is a multi-index, $|k| = \sum_{j=1}^p k_j$, $\varphi^{(k)}(z) = \frac{\partial^{|k|}\varphi(z)}{\partial z_1^{k_1} \ldots \partial z_p^{k_p}}$, and $\lfloor \beta \rfloor$ is the largest integer strictly smaller than β. For $A \subset \mathbf{R}$, let $BV(A)$ be a set of functions of bounded variation on A. 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. 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 $\|K\|_\infty = \sup_{\|\varphi\|_\infty \le 1} \|K\varphi\|_\infty$. For $K \in \mathcal{L}_2$, let K^* denote its Hilbert adjoint operator. Let $\mathcal{R}(T)$ and $\mathcal{D}(T)$ be the range and the domain of the operator T. For two real numbers a and b, I denote $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

2. HONEST UNIFORM CONFIDENCE SETS

We focus on uniform confidence sets, honest to some class of models $\mathcal{P}$. The class of $\mathcal{P}$ consists of probability distributions of the data satisfying certain restrictions. For a given level $\gamma \in (0, 1)$, the *honest* $1 - \gamma$ *uniform confidence set*, denoted $C_{n,1-\gamma} = \left\{ C_{n,1-\gamma}(z) = [C_l(z), C_u(z)],\ z \in [0,1]^p \right\}$, satisfies the following coverage

$$\inf_{P \in \mathcal{P}} \Pr\left(\varphi(z) \in C_{n,1-\gamma}(z),\ \forall z \in [0,1]^p \right) \ge 1 - \gamma - O(\delta_n), \tag{2}$$

for some sequence $\delta_n \to 0$. Honesty is desirable for theoretical and practical reasons. On the theoretical side, it is well-known that focusing on a fixed model leads to inconsistent concepts of optimality for nonparametric models; see, for example, Tsybakov (2009), pp. 16–19. Honesty also ensures that for a sufficiently large sample size n, independent from the distribution of the data, the coverage level of the set will be close to $1 - \gamma$, no matter what model in $\mathcal{P}$ nature gives us. Honesty is desirable in practice since the underlying probability distribution is unknown. In contrast, a *dishonest* set requires a weaker condition

$$\inf_{P \in \mathcal{P}} \liminf_{n \to \infty} \Pr\left(\varphi(z) \in C_{n,1-\gamma}(z),\ \forall z \in [0,1]^p \right) \ge 1 - \gamma$$

and the sample size n needed to achieve the coverage close to $1 - \gamma$ will depend on the distribution of the data.

The second requirement for the confidence set is that its expected diameter under the supremum norm, denoted $|C_{n,1-\gamma}|_\infty = \|C_u - C_l\|_\infty$, shrinks at some rate $\rho_n \to 0$

$$\sup_{P \in \mathcal{P}} \mathbb{E}\left| C_{n,1-\gamma} \right|_\infty = O(\rho_n).$$

We will also consider a variation of this requirement when the diameter is bounded in probability only.

Nonparametric estimation usually involves a bias-variance trade-off for the risk of the estimator. This trade-off translates into a trade-off between the rate δ_n at which coverage errors tend to zero and the rate ρ_n at which the expected diameter

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

tends to zero. Roughly speaking, we show that the coverage error is driven by the bias of the estimator, while the variance determines its diameter. It is vital to know both rates and requiring only the limiting coverage $1-\gamma$ in equation (2) can be misleading, as the coverage error of the set may tend to zero arbitrarily slow. In the following sections, we show that the confidence sets based on the concentration inequality allow us to characterize coverage errors explicitly.

It is also worth mentioning that among confidence sets based on different $L_p, p \in [1,\infty]$ norms, only uniform confidence sets 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 φ with high probability.

3. NONPARAMETRIC IV

3.1. The Model

The nonparametric IV model

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

leads to the functional equation

$$\mathbb{E}[Y|W=w] = \mathbb{E}[\varphi(Z)|W=w], \tag{3}$$

which is an example of an ill-posed inverse problem. Even if we knew the distribution of the data (Y,Z,W), obtaining φ requires inverting the conditional expectation operator, which is typically not continuous. Florens (2003) and Darolles et al. (2011) introduce Tikhonov regularization, see Tikhonov (1963), to smooth out discontinuities of inversion.

Multiplying equation (3) by the marginal density function f_W

$$r(w) \triangleq \mathbb{E}[Y|W=w] f_W(w) = \int_{[0,1]^p} \varphi(z) f_{ZW}(z,w)\mathrm{d}z \triangleq (T\varphi)(w).$$

Following Darolles et al. (2011), we focus on the kernel estimator, but other choices are possible. For simplicity, we use the product kernel and equal bandwidth parameters $h_n \to 0$ as $n \to \infty$ for all coordinates

$$\hat{r}(w) = \frac{1}{nh_n^q} \sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right)$$

$$\hat{f}_{ZW}(z,w) = \frac{1}{nh_n^{p+q}} \sum_{i=1}^n K_z\left(h_n^{-1}(Z_i - z)\right) K_w\left(h_n^{-1}(W_i - w)\right)$$

$$(\hat{T}\varphi)(w) = \int_{[0,1]^p} \varphi(z) \hat{f}_{ZW}(z,w)\mathrm{d}z.$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The Tikhonov-regularized estimator solves the penalized least-squares problem

$$\hat{\varphi} = \arg\min_{\phi \in L_2} \left\| \hat{T}\phi - \hat{r} \right\|^2 + \alpha_n \|\phi\|^2,$$

which admits a closed-form solution

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

where $\hat{T}^*$ is the adjoint operator of $\hat{T}$. The adjoint operator is a solution to $\langle \hat{T}\varphi, \psi \rangle = \langle \varphi, \hat{T}^* \psi \rangle$ and is computed by Fubini's theorem

$$(\hat{T}^* \psi)(z) = \int_{[0,1]^q} \psi(w) \hat{f}_{ZW}(z, w) \mathrm{d}w.$$

3.2. Ill-Posedness and Regularization Bias

We first introduce the class of functions and operators for which we wish to obtain the honest coverage.

Assumption 3.1. The structural function φ and the 1–1 operator $T : C[0,1]^p \to C[0,1]^q$ belong to the following class

$$\mathcal{F} = \mathcal{F}_{\beta,t,M,C} = \big\{ (\varphi, T) \in C_M^t[0,1]^p \times \mathcal{L}_{2,\infty} :$$

$$\varphi = (T^*T)^\beta T^* \psi,\ \kappa(\varphi, \psi, T, T^*) \le C \big\},$$

where $\kappa(\varphi, \psi, T, T^*) = \|\varphi\|_\infty \vee \|T^*\|_{2,\infty} \vee \|\psi\| \vee \|T\|^{-1}$ and $0 < \beta, t, M, C < \infty$.

Source conditions are at the heart of spectral regularization theory, and describe the regularity of the problem by restricting how ill-posed the operator T is, compared with 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, our confidence sets enjoy polynomial convergence rates of coverage errors. In the severely ill-posed case, when singular values of the operator T decay to zero exponentially fast, the source condition in Assumption 3.1 requires that the function φ is well-approximated by a small number of generalized Fourier coefficients with respect to the singular values decomposition (SVD) basis of T. It is a remarkable fact that we can still achieve polynomial convergence rates in this case, since many functions encountered in practice have rapidly declining generalized Fourier coefficients.

Assumption 3.1 allows us to control the regularization bias. Note 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.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Proposition 3.1. *Suppose that Assumption 3.1 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 $R = C^2\left[\beta^\beta(1-\beta)^{1-\beta}\mathbb{1}_{0<\beta<1} + C^{2(\beta-1)}\mathbb{1}_{\beta\geq 1}\right]$.

In terms of the bias, the regularization takes advantage of the smoothness up to $\beta = 1$. This effect is somewhat reminiscent of the saturation of convergence of the kernel density estimator and can be avoided using the iterated or the extrapolated Tikhonov regularization similarly to using higher or infinite order kernels for the kernel density estimator, see Darolles et al. (2011) and Carrasco et al. (2007) for more discussion.

3.3. Honest Confidence Sets

Put

$$\|\hat{v}_n^\eta\|_\infty = \left\|\frac{1}{n\sqrt{h_n^q}}\sum_{i=1}^n \eta_i(\alpha_n I + T^*T)^{-1}T^*\hat{X}_{ni}\right\|_\infty,$$

where $\hat{X}_{ni}(w) = \frac{1}{\sqrt{h_n^q}}\hat{U}_i K_w\left(h_n^{-1}(W_i - w)\right)$, $\hat{U}_i = Y_i - \hat{\varphi}(Z_i)$, and η_i are i.i.d. Rademacher random variables. Let $c^*_{1-\gamma}$ be $1-\gamma$ quantile of

$$\|\hat{V}_n^\varepsilon\|_\infty = \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i\hat{X}_{ni}\right\|_\infty$$

conditionally on the data $\mathcal{D} = (Y_i, Z_i, W_i)_{i=1}^n$ and ε_i are i.i.d. N(0,1).

The class of models $\mathcal{P}$ is restricted by the following assumptions.

Assumption 3.2. (i) $(Y_i, Z_i, W_i)_{i=1}^n$ is an i.i.d. sample of (Y, Z, W); (ii) $f_{ZW} \in C^s([0,1]^{p+q})$, $s > 0$ and there exists $\underline{f}, \bar{f}$ such that $0 < \underline{f} \leq f_{ZW} \leq \bar{f} < \infty$; (iii) K_z and K_w are products of the kernel function $k : \mathbf{R} \to \mathbf{R}$ with $k \in C(\mathbf{R}) \cap BV(\mathbf{R})$ having order $\lfloor s\rfloor \vee \lfloor t\rfloor$ and such that $k \in L_r(\mathbf{R})$, $r = 1, 2, 3, 4$, and $\int |u|^{s\vee t}|k(u)|du < \infty$; (iv) $\mathbb{E}|Y|^4 < \infty$, $\mathbb{E}[|U|^2|W] \geq \underline{\sigma}_2 > 0$ a.s., and $|U| \leq F$; (v) the integral operator $T : C[0,1]^p \to C[0,1]^q$ is 1–1; and (vi) h_n and α_n tend to 0 polynomially fast and (a) $h_n^t n\log^2 n \to 0$, $\alpha_n^{2(\beta\wedge 1+1)}h_n^q n\log^2 n \to 0$, and $\alpha_n h_n^{2s+q} n\log^2 n \to 0$; (b) $\alpha_n \log^3 n/h_n^p \to 0$, $h_n^{s-q/2}\log n/\alpha_n \to 0$, and $\alpha_n^{\beta\wedge 1}\log n/h_n^{q/2} \to 0$; and (c) $\alpha_n^2 n h_n^{2q}/\log n \to \infty$.

Most of these assumptions impose mild smoothness, boundedness, or moment conditions which are plausible in empirical applications. Conditions imposed on the kernel functions are extremely mild and are verified by all kernel functions commonly used in practice. Similar assumptions are usually made in other settings for uniform nonparametric estimation and inference. Assumption 3.2 imposes mild regularity conditions on the data generating process: (i) standard sampling

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

assumption and may be relaxed to time series data under some weak dependence conditions, for example, see Babii (2017); (ii) may rule out common elements between Z and W, see discussion in Section 5 how to deal with common elements; (iii) allows for boundary-corrected kernels, see Darolles et al. (2011)[3] for further discussion. Assumption (v) is a completeness condition, which may be questionable in empirical applications. Note, however, that in many cases, the completeness condition can be relaxed, and it is possible to have valid inferences for various deviations from the completeness condition, see Babii and Florens (2018) for more details.

Assumption 3.2(vi) involves several conditions on tuning parameters and deserves special attention. It imposes implicitly that different components of the model have sufficient regularity with respect to the dimensions p and q. To illustrate that these conditions are not contradictory, suppose that $h_n \sim n^{-c_1}$ and $\alpha_n \sim n^{-c_2}$ for some $c_1, c_2 > 0$ and that $\beta \geq 1$. Then Assumption 3.2 reduces to the following 6 conditions: (1) $c_2 > (p \vee q/2)c_1$, (2) $4c_2 + qc_1 > 1$, (3) $2c_2 + 2qc_1 < 1$, (4) $c_1 > 1/t$, (5) $c_1(s - q/2) > c_2$, and (6) $c_2 + c_1(2s + q) > 1$. Conditions (1)–(3) define the set of feasible value for c_1 and c_2 depending on the dimension parameters p and q. Then (4), (5), and (6) simply require that smoothness parameters t and, respectively, s are sufficiently large. For instance, when $p = q = 1$, we can take, $h_n \sim n^{-1/5}$ and $\alpha_n \sim n^{-1/4}$, and need $t > 5$ and $s > 7/4$.

The bootstrap and the concentration inequality-based confidence sets are described as

$$\begin{aligned} C^*_{n,1-\gamma} &= \left\{ [\hat{\varphi}(z) - q^*_n, \hat{\varphi}(z) + q^*_n] : \, z \in [0,1] \right\}, \\ C_{n,1-\gamma} &= \left\{ [\hat{\varphi}(z) - \hat{q}_n, \hat{\varphi}(z) + \hat{q}_n] : \, z \in [0,1] \right\} \end{aligned} \tag{4}$$

with

$$q^*_n = \frac{c^*_{1-\gamma} \|\hat{T}^*\|_{2,\infty} + \log^{-1} n}{\alpha_n \sqrt{n h_n^q}},$$

$$\hat{q}_n = 2 \left\| \hat{v}^\eta_n \right\|_\infty + \frac{3 \|\hat{T}^*\|_{2,\infty} F \|k\|^q \sqrt{2 \log(1/\gamma)} + \log^{-1} n}{\alpha_n \sqrt{n h_n^q}},$$

where the operator norm can be computed using Lemma A.4.4 as

$$\|\hat{T}^*\|_{2,\infty} = \left(\sup_z \int |\hat{f}_{ZW}(z,w)|^2 \mathrm{d}w \right)^{1/2}.$$

The next result provides theoretical justification for our honest uniform confidence sets.

[3]Otherwise, to avoid problems at end-points, we will make a restriction to the interior of $[0, 1]^{p+q}$, and all results should be read as uniform over the interior of this set.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

THEOREM 3.1. *Suppose that $\mathcal{P}$ consists of distributions satisfying Assumptions 3.1 and 3.2. Then,*

$$\inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq 1-\gamma-O(\delta_n)$$

with

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

Moreover, if Assumption 3.2(v) and (vi) is satisfied

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C^*_{n,1-\gamma}\right) \geq 1-\gamma.$$

The proof of this result can be found in the Appendix and is a consequence of more general Theorems A.1 and A.3. In particular, Theorem A.3 derives coverage errors for concentration inequality-based confidence sets, while Theorem A.1 establishes the asymptotic validity of bootstrap confidence sets for a general class of ill-posed inverse models. Theorem A.1 is based on the Gaussian and bootstrap coupling inequalities as well as the anticoncentration inequality of Chernozhukov et al. (2014, 2016). Note that concentration-based confidence sets allow us to characterize explicitly the coverage error, which has not been known for confidence sets in the nonparametric IV model. Inspection of the proof of Theorem A.1 reveals that bootstrap confidence sets have additional coverage errors due to the Gaussian and bootstrap approximations.

To illustrate that conditions needed for $\delta_n \to 0$ are not contradictory, suppose that $h_n \sim n^{-c_1}$ and $\alpha_n \sim n^{-c_2}$ for some $c_1, c_2 > 0$ and that $\beta \geq 1$. Then we need (1) $c_2 > pc_1$, (2) $4c_2 + qc_1 > 1$, (3) $2c_2 + qc_1 < 1$, (4) $c_1 > 1/t$, (5) $c_1 s > c_2$, and (6) $c_2 + c_1(2s+q) > 1$. Note that the set of feasible values of c_1, c_2 is larger than that allowed by Assumption 3.2(vi).

The following theorem describes the expected size of our confidence sets.

THEOREM 3.2. *Suppose that assumptions of Theorem 3.1 are satisfied. Then*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{1}{\alpha_n\sqrt{nh_n^q}}\right) \quad \textit{and} \quad \left|C^*_{n,1-\gamma}\right|_\infty = O_p\left(\frac{1}{\alpha_n h_n^q n^{1/2}}\right),$$

uniformly over $P \in \mathcal{P}$.

Note that we need $\alpha_n^2 nh_n^q \to \infty$ to ensure that the concentration-based confidence set shrinks and a slightly stronger condition $\alpha_n^2 nh_n^{2q} \to \infty$ to ensure that the bootstrap confidence set shrinks, which is also needed to ensure that the coverage error tends to zero.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

4. FUNCTIONAL REGRESSIONS AND DENSITY DECONVOLUTION

4.1. 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 handling high-dimensional data without relying on the *sparsity* assumption, which may be restrictive in this setting. There is vast literature on functional regression models in statistics, see Cardot et al. (2003), Hall and Horowitz (2007), and references therein. There is also some growing literature in econometrics, see Florens and Van Bellegem (2015), Benatia et al. (2017), and Babii (2017). 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 as

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

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 and Horowitz (2007), while in the IV case of Florens and Van Bellegem (2015), W is some functional instrumental variable, uncorrelated with the error term.

The moment restriction leads to the ill-posed equation

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

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

A variation of this model is studied in Babii (2017), where the identification is achieved with real-valued instrumental variables through the conditional moment restriction $\mathbb{E}[U|W] = 0$. The identifying restriction is the linear completeness condition. Conditional mean-independence leads to the following ill-posed equation

$$\mathbb{E}[Y|W] = \int_{[0,1]^p} \varphi(t) \mathbb{E}[Z(t)|W] \mathrm{d}t, \tag{5}$$

which for appropriate families of functions $\Psi(s, .)$ can equivalently be written as

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

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 encompassing all variations of the functional regression discussed above.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Unlike the NPIV, which requires a nonparametric estimation of the joint density function, all components of functional regression models are estimated at the parametric rate (in the L_2 norm) 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 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 Fubini's theorem

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

Put

$$\|\hat{v}_n^{\eta}\|_{\infty} = \left\| \frac{1}{n}\sum_{i=1}^{n} \eta_i (\alpha_n I + T^*T)^{-1} T^* \hat{X}_{ni} \right\|_{\infty},$$

where $\hat{X}_{ni}(s) = \hat{U}_i W_i(s)$, $\hat{U}_i = Y_i - \langle Z_i, \varphi\rangle$, and η_i are i.i.d. Rademacher random variables.

Let $c^*{}_{1-\gamma}$ be $1-\gamma$ quantile of

$$\|\hat{V}_n^{\varepsilon}\|_{\infty} = \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varepsilon_i \hat{X}_{ni} \right\|_{\infty}$$

and $\varepsilon_i \sim_{i.i.d.} N(0,1)$.

The following assumption restricts the class of models $\mathcal{P}$.

Assumption 4.1. (i) $(Y_i, Z_i, W_i)_{i=1}^n$ is an i.i.d. sample of (Y, Z, W); (ii) Z and W have continuous trajectories, $\|Z\|_{\infty} \le C$ and $\|UW\|_{\infty} \le F$; (iii) $W \in C_M^s[0,1]^q$ and $ZW \in C_M^s[0,1]^{p+q}$, for some $s > (p+q)/2$; (iv) $\mathbb{E}|UW(s)| \ge \underline{\sigma} > 0$ for all $s \in [0,1]^q$; (v) the integral operator $T : C[0,1]^p \to C[0,1]^q$ is 1–1; and (vi) $\alpha_n \to 0$ polynomially fast and $\alpha_n n/\log^2 n \to \infty$ and $\alpha_n^{2(\beta\wedge 1)+2} n \log^2 n \to 0$.

Note that the functional linear regressions admit a simpler characterization and involve even less stringent assumptions than the nonparametric IV model.

The bootstrap and the concentration inequality-based confidence sets are

$$C^*_{n,1-\gamma} = \left\{ [\hat{\varphi}(t) - q_n^*, \hat{\varphi}(t) + q_n^*] : t \in [0,1] \right\}$$

and

$$C_{n,1-\gamma} = \left\{ [\hat{\varphi}(t) - \hat{q}_n, \hat{\varphi}(t) + \hat{q}_n] : t \in [0,1] \right\}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

with

$$q_n^* = \frac{c_{1-\gamma}^* \|\hat{T}^*\|_{2,\infty} + \log^{-1} n}{\alpha_n \sqrt{n}},$$

$$\hat{q}_n = 2\left\|\hat{v}_n^\eta\right\|_\infty + \frac{3\|\hat{T}^*\|_{2,\infty} F\sqrt{2\log(1/\gamma)} + \log^{-1} n}{\alpha_n \sqrt{n}},$$

where the operator norm can be computed using Lemma A.4.4 as

$$\|\hat{T}^*\|_{2,\infty} = \left(\sup_t \int |\hat{k}(t,s)|^2 \mathrm{d}s\right)^{1/2}.$$

The next result provides theoretical justification for confidence sets based on the concentration inequality and bootstrap confidence sets.

THEOREM 4.1. *Suppose that Assumptions 3.1 and 4.1 are satisfied. Then for any $\gamma \in (0, 1)$,*

$$\inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq 1 - \gamma - O\left(\left(\alpha_n^{\beta\wedge 1+1} n^{1/2} + \frac{1}{\sqrt{\alpha_n n}}\right)\log n\right).$$

Moreover, if Assumption 4.2(vi) is satisfied

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_{n,1-\gamma}^*\right) \geq 1 - \gamma.$$

Note that for concentration inequality-based confidence sets, the coverage error is effectively driven by the bias induced by the regularization and kernel smoothing. For bootstrap confidence sets, we have an additional coverage error due to the Gaussian and bootstrap approximations and need additional condition $nh_n^2/\log^7 n \to \infty$. The following result describes the expected size of various confidence sets.

THEOREM 4.2. *Suppose that assumptions of Theorem 4.3 are satisfied, then*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}\right) \qquad \text{and} \qquad \left|C_{n,1-\gamma}^*\right|_\infty = O_p\left(\frac{1}{\alpha_n n^{1/2}}\right),$$

uniformly over $P \in \mathcal{P}$.

To ensure that confidence sets shrink, we need $\alpha_n^2 n \to \infty$. In both cases, sets shrink at the same speed.

4.2. Density Deconvolution

Often economic data are not measured precisely. Density deconvolution allows estimating the density of unobserved data from the data measured with errors. Density deconvolution is encountered in a variety of econometric applications, for example, to the earning dynamics in Bonhomme and Robin (2010), to panel data

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

in Evdokimov (2010), or to the instrumental regression in Adusumilli and Otsu (2018). In the simplest example of this model, see Carrasco and Florens (2011), 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 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

$$r(y) = \int \varphi(z) f(y-z) \mathrm{d}z =: (T\varphi)(y), \tag{6}$$

where the operator $T : L_2 \to L_2$. For simplicity of presentation and tractability of our results, let us assume that all densities are continuous, bounded, and compactly supported, with support contained inside the interval $[0, 1] \subset \mathbf{R}$. This assumption is not very restrictive, since most of the economic variables, for example, reported earning or costs are bounded.

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

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

We estimate the density of Y from the sample $(Y_i)_{i=1}^n$ using the kernel density estimator

$$\hat{r}(y) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{Y_i - y}{h_n}\right).$$

Put

$$\|\hat{v}_n^\eta\|_\infty = \left\| \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \eta_i (\alpha_n I + T^*T)^{-1} T^* \hat{X}_{ni} \right\|_\infty,$$

where $\hat{X}_{ni}(y) = \frac{1}{\sqrt{h_n}} K\left(\frac{Y_i - y}{h_n}\right) - \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n K\left(\frac{Y_i - y}{h_n}\right)$ and η_i are i.i.d. Rademacher random variables.

The class of models $\mathcal{P}$ is restricted by the following assumptions.

Assumption 4.2. (i) $(Y_i)_{i=1}^n$ is an i.i.d. sample of Y; (ii) f, φ, r are continuous, compactly supported on some subsets of [0, 1], and the density r is uniformly bounded away from zero and infinity by some finite constant; (iii) $r \in C_M^s[0, 1]$, for some $s > 0$; (iv) K is a symmetric continuous square-integrable kernel function of bounded variation of order $\lfloor s \rfloor$ satisfying $\int |u|^s |K(u)| \mathrm{d}u < \infty$; (v) the integral

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

operator $T : C[0,1]^p \to C[0,1]^q$ is 1–1; and (vi) $h_n \to 0$ and $\log h_n^{-1} = O(\log n)$, $nh_n^2/\log^7 n \to \infty$, $nh_n^{2s+1}\log^2 n \to 0$, and $\alpha_n^{2(\beta\wedge 1)+2} nh_n \log^2 n \to 0$.

Note that for this model, operators T and T^* are known, which simplifies the analysis. Let $c_{1-\gamma}{}^*$ be $1-\gamma$ quantile of

$$\|\hat{V}_n^{\varepsilon}\|_\infty = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \hat{X}_{ni} \right\|_\infty$$

and $\varepsilon_i \sim_{i.i.d.} N(0,1)$.

The bootstrap and the concentration inequality-based confidence sets are described as

$$C^*_{n,1-\gamma} = \left\{ [\hat{\varphi}(z) - q_n^*, \hat{\varphi}(z) + q_n^*] : \ z \in [0,1] \right\}$$

and

$$C_{n,1-\gamma} = \left\{ [\hat{\varphi}(z) - \hat{q}_n, \hat{\varphi}(z) + \hat{q}_n] : \ z \in [0,1] \right\}$$

with

$$q_n^* = \frac{c^*_{1-\gamma}\|T^*\|_{2,\infty} + \log^{-1} n}{\alpha_n \sqrt{nh_n}},$$

$$\hat{q}_n = 2\left\|\hat{v}_n^{\eta}\right\|_\infty + \frac{6\|T^*\|_{2,\infty}\|K\|\sqrt{2\log(1/\gamma)} + \log^{-1} n}{\alpha_n\sqrt{nh_n}},$$

where the operator norm can be computed using Lemma A.4.4 as

$$\|T^*\|_{2,\infty} = \left(\sup_z \int |f(y-z)|^2 \mathrm{d}y \right)^{1/2}.$$

The next result provides theoretical justification for confidence sets based on the concentration inequality and bootstrap confidence sets.

THEOREM 4.3. *Suppose that Assumptions 3.1 and 4.2 are satisfied. Then for any $\gamma \in (0,1)$,*

$$\inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq 1 - \gamma - O\left(\left(h_n^s + \alpha_n^{\beta\wedge 1+1}\right)\sqrt{nh_n}\log n + \frac{\log n}{n^{1/2}} \right).$$

Moreover, if Assumption 4.2(vi) and (vii) is satisfied

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C^*_{n,1-\gamma}\right) \geq 1 - \gamma.$$

Note that for concentration inequality-based confidence sets the coverage error is effectively driven by the bias induced by the regularization and kernel smoothing. For bootstrap confidence sets, we have an additional coverage error due to the Gaussian and bootstrap approximations and need additional condition

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$nh_n^2/\log^7 n \to \infty$. The following result describes the expected size of various confidence sets.

THEOREM 4.4. *Suppose that assumptions of Theorem 4.3 are satisfied, then*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P \left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{1}{\alpha_n\sqrt{nh_n}}\right) \qquad \text{and} \qquad \left|C^*_{n,1-\gamma}\right|_\infty = O_p\left(\frac{1}{\alpha_n h_n n^{1/2}}\right),$$

uniformly over $P \in \mathcal{P}$.

To ensure that a concentration inequality-based confidence set shrinks, we need $\alpha_n^2 h_n n \to \infty$ while for the bootstrap confidence set we need the slightly stronger requirement $\alpha_n^2 h_n^2 n \to \infty$.

5. IMPLEMENTATION AND MONTE CARLO EXPERIMENTS

This section reports results of Monte Carlo experiments for our confidence sets. We focus on the NPIV estimator. Samples of size $n \in \{1,000, 5,000\}$ 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 a sufficiently large compact set. The strength of the instrument as measured by the correlation coefficient ρ is calibrated to our empirical application.

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 the simulation design not only to functional regression but also to the deconvolution model. Since the underlying integral equation would be the same in all cases, to economize on space, we do not report Monte Carlo (MC) experiments for other models. We also remark that the normal density leads to rapidly declining singular values. This design corresponds to the most difficult severely ill-posed setting. We consider 5, 000 replications of each experiment.

We consider confidence sets described in equation (4), where the operator norm $\|\hat{T}^*\|_{2,\infty}$ is computed using the mixed norm of the kernel function of $\hat{T}^*$ by Lemma A.4.4. The estimate of the density function f_{ZW} is obtained using kernel smoothing. For simplicity of implementation, we do not optimize the performance with higher-order boundary kernels and take the product of second-order Epanechnikov kernels.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The estimator is discretized using a simple Riemann sum on the grid of 100 equidistant points. This suffices to ensure that the numerical errors are negligible compared with the statistical noise in our setting. A higher number of grid points or cubature can give better approximation if needed. The discretized estimator has closed-form expression, and its computation boils down to solving a system of linear equations

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

where $\mathbf{I}$ is the $T \times T$ identity matrix, $\mathbf{r} = \left(\frac{1}{n}\sum_{i=1}^n Y_i h_n^{-1} K\left(h_n^{-1}(W_i - w)\right)\right)_{1 \le j \le T}$, $\mathbf{K} = \boldsymbol{f}\Delta$ with $\boldsymbol{f} = (\hat{f}_{ZW}(z_k, w_j))_{1 \le j,k \le T}$, and Δ is the grid step. Therefore, the estimator is straightforward to implement and fast to compute.

Data-driven choice of tuning parameters. As discussed in 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. 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 though 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.

Existing data-driven rules for Tikhonov regularization aim to obtain an accurate estimator, see Fève and Florens (2010) or Centorrino (2016), and do not take into account the trade-off between the coverage and the diameter. Note that the risk-optimal estimation conflicts with optimal inference. In particular, our theory requires undersmoothing in order to obtain valid inferences. We suggest the following procedure inspired by Bissantz et al. (2007):

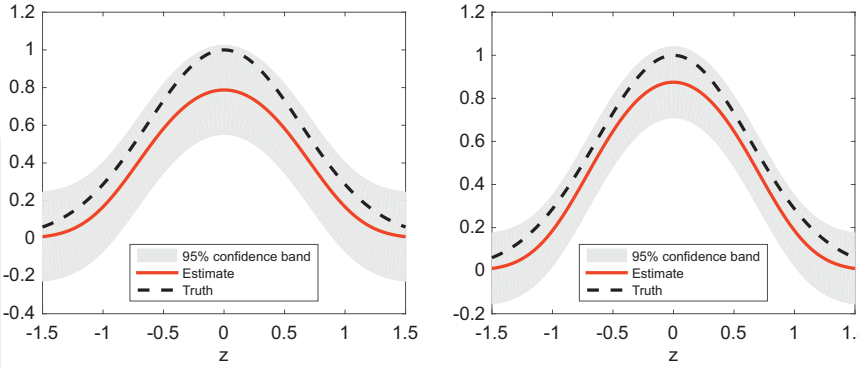
1. Use some over-smoothing data-driven rule to select pilot tuning parameters. We consider the cross-validation studied in Centorrino (2016), which consists of solving the leave-one-out empirical counterpart to $\|\hat{r} - \hat{T}\hat{\varphi}_{\alpha,h}\|$

$$(\hat{\alpha}_J, \hat{h}_J) = \underset{\alpha, h>0}{\arg\min} \frac{1}{n}\sum_{i=1}^n \times \left\{ \frac{1}{n-1}\sum_{j \neq i} (Y_j - [K_h * \hat{\varphi}_{\alpha,h}](Z_j)) \frac{1}{h} K\left(h^{-1}(W_j - W_i)\right) \right\}^2,$$

   where $[K_h * g](x) = \frac{1}{h}\int g(z) K(h^{-1}(x - z))\mathrm{d}x$ denotes the convolution operator. To ensure that tuning parameters are over-smoothing, we scale pilot tuning parameters $(\kappa_a \hat{\alpha}_J, \kappa_h \hat{h}_J)$ with some $\kappa_a, \kappa_h \ge 1$.
2. Compute $\hat{\varphi}_{h_j,\alpha_j}$ on a grid of points $(j\kappa_a \alpha_J/J, j\kappa_h h_J/J)_{j=1}^J$. Choose the largest (α_j, h_j) such that

$$\|\hat{\varphi}_{\alpha_j,h_j} - \hat{\varphi}_{\alpha_{j-1},h_{j-1}}\|_\infty > \tau \|\hat{\varphi}_{\alpha_J,h_J} - \hat{\varphi}_{\alpha_{J-1},h_{J-1}}\|_\infty$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press



(a) n= 1000, $\hat{\gamma}$= 0.87 (b) n= 5000, $\hat{\gamma}$= 0.97

Fig. 1- Colour online, B/W in print

FIGURE 1. Bootstrap uniform confidence bands, averaged over 5, 000 experiments. $\hat{\gamma}$ is the MC coverage probability.

Like Bissantz et al. (2007), we find that $J = 20$ and $\tau = 2$ work well in practice. We also find that $(\kappa_a, \kappa_h) = (3.5, 1.5)$ is suitable for concentration-based confidence sets and $(\kappa_a, \kappa_h) = (1, 1.5)$ works well for bootstrap confidence sets. Note that Bissantz et al. (2007) have a single tuning parameter, and they find that the factor 1.5 is suitable, which is what we find for the bandwidth parameter.

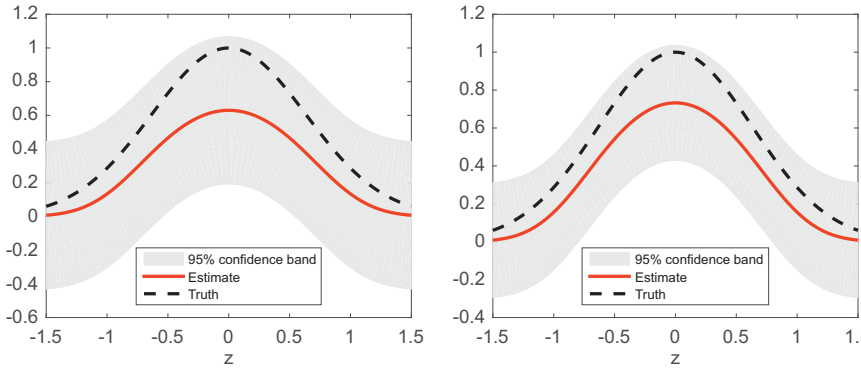
Simulation results. In Figure 1, we plot estimates with a 95% confidence band based on bootstrap approximation, averaged over 5, 000 Monte Carlo experiments. Figure 2 represents the same plot for confidence bands based on concentration inequality. We also report MC coverage probabilities $\hat{\gamma}$. In our Monte Carlo experiments, bootstrap approximation leads to slightly better performance. Resulting confidence bands are narrower, and the estimator is centered more closely to the population value of the parameter of interest. Nonetheless, confidence sets based on the concentration inequality are also informative about global shape properties of the estimated function. For example, Figure 1b gives an excellent confidence band with a reasonably small amount of bias.

Exogenous covariates. In the presence of additional exogenous covariates, one can consider several avenues. As an example, suppose that one is interested in estimating the demand function, so $Z = (Z_1, Z_2)$ with Z_1 denoting the endogenous price and Z_2 denoting exogenous covariates, such as income and other control variables. One way to accommodate the exogenous covariates is to estimate the semiparametric partially linear specification allowing for nonlinearities in prices, see Yatchew and No (2001) and Florens and Van Bellegem (2015).

Alternatively, one may want to keep the fully nonparametric specification. In this case, the model $Y = \varphi(Z_1, Z_2) + U, \mathbb{E}[U|Z_2, W] = 0$ leads to the following family of ill-posed integral equations

$$\mathbb{E}[Y|Z_2 = z_2, W = w] f_{Z_2,W}(z_2, w) = \int \varphi(z_1, z_2) f_{Z_1,Z_2,W}(z_1, z_2, w) \mathrm{d}z_1.$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press



(a) $n = 1000$, $\hat{\gamma} = 0.89$ (b) $n = 5000$, $\hat{\gamma} = 0.94$

FIGURE 2. Concentration-based uniform confidence bands, average over 5, 000 experiments. $\hat{\gamma}$ is the MC coverage probability.

For any fixed value of z_2, we are back to our setting, and our uniform inferential methods would allow us to construct uniform confidence sets for $z_1 \mapsto \varphi(z_1, z_2)$ for any fixed value of z_2. Uniform inference for the demand function for different income levels might be sufficient to reveal the amount of nonlinearities and/or endogeneity. Constructing the joint uniform confidence set for the bivariate function φ and the theoretical analysis of such sets is beyond the scope of this article.

6. ENGEL CURVES IN THE UNITED STATES

In this section, we estimate Engel curves using the NPIV approach and construct uniform confidence sets using the bootstrap. Engel curves describe how the demand for a 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 et al. (1997), estimation of growth and inflation in Nakamura et al. (2016), or estimation of income inequality across countries in Almås (2012).

Previously, Blundell et al. (2007) estimated a shape-invariant system of Engel curves for food, alcohol, and fuel on UK data with a sieve approach. For simplicity, we focus on the nonparametric specification of the Engel curve.

Our dataset is drawn from the 2015 U.S. Consumer Expenditure Survey data, and we estimate Engel curves for a set of goods, including food, tobacco, alcohol, gas and oil, and health. In our subsample, we have married couples with positive income during the past 12 months, including households with and without children. The dependent variable is the share of expenditures on the particular commodity. The log of total expenditures on nondurable goods is used as an independent variable. As in Blundell et al. (2007), we instrument the log of total expenditures

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

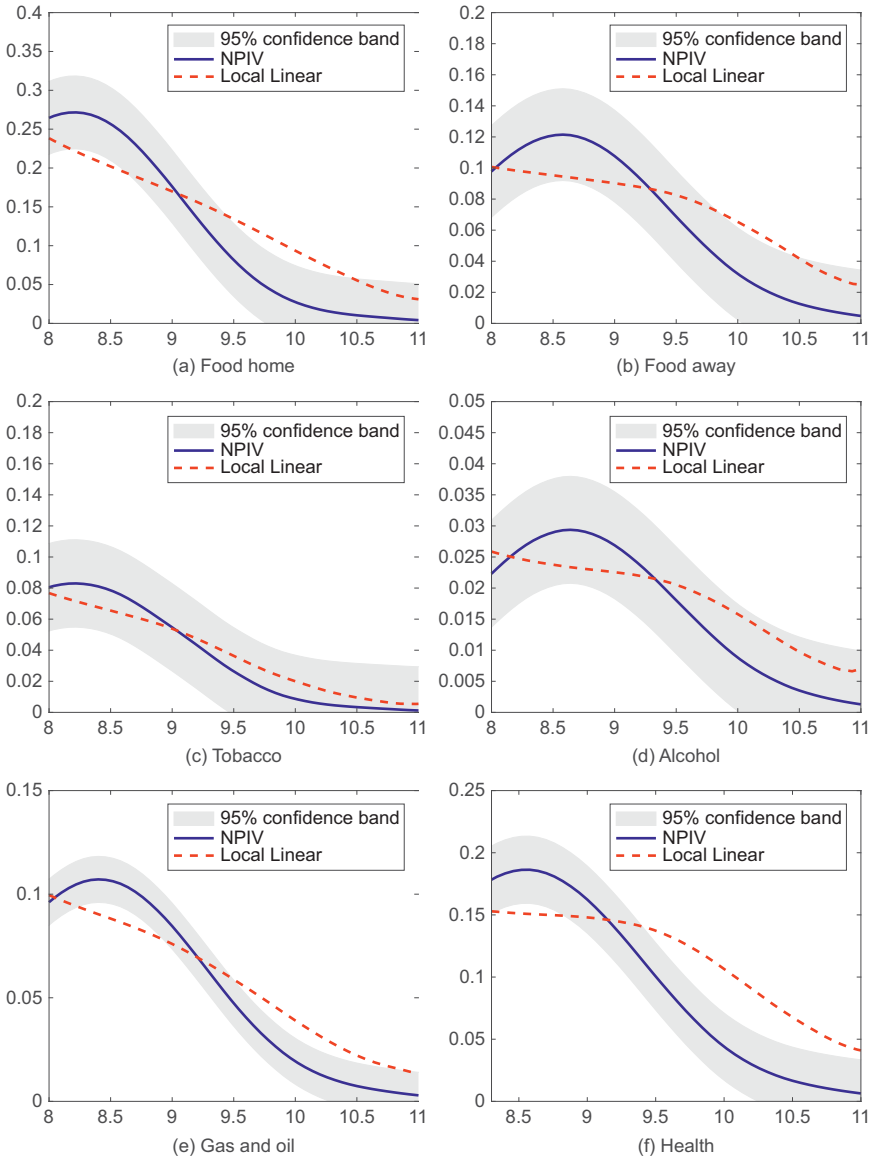


FIGURE 3. Engel curves estimated with NPIV (solid blue line) with 95% bootstrap confidence band and local polynomial estimator (dashed red line). Horizontal axis: total expenditures. Vertical axis: budget share.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

with income before tax. However, in contrast to Blundell et al. (2007), we also estimate Engel curves for tobacco and health care.

We plot NPIV estimates with a 95% uniform confidence bands in Figure 3. We also plot the local linear estimator, which does not correct for the endogeneity bias. Estimates in Figure 3 illustrate that there is significant endogeneity bias in the Engel curves. Indeed, the local polynomial estimator is often outside the confidence band. For instance, the local polynomial estimator largely overestimates the shape of the Engel curve for higher budgets. In most of the cases, the estimated Engel curves exhibit more curvature than suggested by the local polynomial estimator. Although it may look like Engel curves slightly increase at the beginning, this observation is not statistically significant and most likely can be attributed to boundary effects.

7. CONCLUSIONS

This article 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 approximating the distribution of the supremum of a complex empirical process. We show that it is not possible to establish functional convergence of this process for a very general class of ill-posed models known in econometrics and statistics. Nonetheless, we demonstrate that it is possible to obtain uniform inference relying on alternative methods.

We develop two approaches to uniform inference that lead to honest confidence sets. Honest confidence sets are of practical interest for several reasons. They ensure the existence of a certain sample size after which the coverage level will be not significantly smaller than the nominal coverage level, regardless of how complex the estimated function is within a given smoothness class. Moreover, it is widely recognized in nonparametric statistics, results for a fixed model can lead to inconsistent notions of optimality. Honest confidence sets, on the other hand, have the uniform validity.

The first approach developed in this article relies on data-driven concentration inequality and allows for nonasymptotic characterization of coverage errors and expected diameters. These rates are new and have not been previously discussed in the literature. We also develop the asymptotic bootstrap approach to inference relying on the coupling inequalities of Chernozhukov et al. (2016) and show that the resulting confidence sets have desirable theoretical properties.

Both methods demonstrate good performance in Monte Carlo experiments, and it seems that bootstrap confidence sets are slightly more narrow for a comparable coverage level. On the other hand, concentration-based confidence sets are computationally less demanding and still reveal lots of information about the global shape properties of functional parameters. Lastly, unlike bootstrap and/or Gaussian approximation-based inference, concentration inequalities do not impose any restrictions on indexing classes of empirical processes. Therefore, this approach can be easily adapted to other complex nonparametric estimators.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

REFERENCES

Adusumilli, K. & T. Otsu (2018) Nonparametric instrumental regression with errors in variables. *Econometric Theory* 34, 1256–1280.

Almås, I. (2012) International income inequality: Measuring PPP bias by estimating Engel curves for food. *The American Economic Review* 102, 1093–1117.

Babii, A. (2017) Identification and estimation in the functional linear instrumental regression. UNC Working Paper.

Babii, A. & J.-P. Florens (2018) Is completeness necessary? Estimation and inference in non-identified models. UNC Working Paper.

Banks, J., R. Blundell, & A. Lewbel (1997) Quadratic Engel curves and consumer demand. *Review of Economics and Statistics* 79, 527–539.

Belloni, A., V. Chernozhukov, D. Chetverikov, & K. Kato (2015) Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics* 186, 345–366.

Benatia, D., M. Carrasco, & J.-P. Florens (2017) Functional linear regression with functional response. *Journal of Econometrics* 201 269–291.

Bissantz, N., L. Dümbgen, H. Holzmann, & A. Munk (2007) Non-parametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 483–506.

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

Bonhomme, S. & J.-M. Robin (2010) Generalized non-parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies* 77, 491–533.

Boucheron, S., G. Lugosi, & P. Massart (2013) *Concentration inequalities: A nonasymptotic theory of independence.* OUP, Oxford, UK.

Cardot, H., A. Mas, & P. Sarda (2007) CLT in functional linear regression models. *Probability Theory and Related Fields* 138, 325–361.

Cardot, H., F. Ferraty, A. Mas, & P. Sarda (2003) Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* 30, 241–255.

Carrasco, M. & J.-P. Florens (2011) A spectral method for deconvolving a density. *Econometric Theory* 27, 546–581.

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

Carrasco, M, J.-P. Florens, & E. Renault (2014) Asymptotic normal inference in linear inverse problems. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press.

Centorrino, S. (2016) Data driven selection of the regularization parameter in additive nonparametric instrumental regressions. Working Paper, Stony Brook University.

Chen, X. & D. Pouzo (2015) Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* 83, 1013–1079.

Chen, X. & M. Reiss (2011) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27, 497–521.

Chen, X. & T.M. Christensen (2018) Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics* 9, 39–84.

Chernozhukov, V., D. Chetverikov, & K. Kato (2014). Gaussian approximation of suprema of empirical processes. *Annals of Statistics* 42, 1564–1597.

Chernozhukov, V., D. Chetverikov, & K. Kato (2016) Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and Their Applications* 126, 3632–3651.

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

Dudley, R.M (2014) *Uniform Central Limit Theorems,* vol. 142, Cambridge University Press.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Dudley, R.M (2016) V.N. Sudakov's work on expected suprema of Gaussian processes. *High Dimensional Probability VII*. pp. 37–43. Springer.

Evdokimov, K. (2010) Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Princeton University Working Paper.

Fève, F. & J.-P. Florens (2010) The practice of non-parametric estimation by solving inverse problems: The example of transformation models. *The Econometrics Journal* 13, S1–S27.

Florens, J.-P. (2003) Inverse problems and structural econometrics. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* 2, 46–85.

Florens, J.-P. & S. Van Bellegem (2015) Instrumental variable estimation in functional linear models. *Journal of Econometrics* 186, 465–476.

Florens, J.-P., J. Johannes, & S. Van Bellegem (2011) Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory* 27, 472–496.

Florens, J.-P., J.L. Horowitz, & I. Van Keilegom (2017) Bias-corrected confidence intervals in a class of linear inverse problems. *Annals of Economics and Statistics* 128, 203–228.

Gagliardini, P. & O. Scaillet (2012a) Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica* 80, 1533–1562.

Gagliardini, P. & O. Scaillet (2012b) Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics* 167, 61–75.

Gautier, E. & Y. Kitamura (2013) Nonparametric estimation in random coefficients binary choice models. *Econometrica* 81, 581–607.

Giné, E. & A. Guillou (2002) Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 38, 907–921.

Giné, E. & R. Nickl (2016) Mathematical foundations of infinite-dimensional statistical models. *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.

Hall, P. & J.L. Horowitz (2005) Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 33, 2904–2929.

Hall, P. & J.L. Horowitz (2007) Methodology and convergence rates for functional linear regression. *Annals of Statistics* 35, 70–91.

Horowitz, J.L. & S. Lee (2012) Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics* 168, 175–188.

Kato, K. & Y. Sasaki (2018) Uniform confidence bands in deconvolution with unknown error distribution. *Journal of Econometrics* 207, 129–161.

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

Koltchinskii, V. (2006) Local Rademacher complexities and oracle inequalities in risk minimization. *Annals of Statistics* 34, 2593–2656.

Koltchinskii, V. (2011) *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008,* vol. 2033, Springer.

Li, K.-C. (1989) Honest confidence regions for nonparametric regression. *Annals of Statistics* 17, 1001–1008.

Lounici, K. & R. Nickl (2011) Global uniform risk bounds for wavelet deconvolution estimators. *Annals of Statistics* 39, 201–231.

Nair, M.T. (2009) *Linear Operator Equations: Approximation and Regularization.* World Scientific.

Nakamura, E., J. Steinsson, & M. Liu (2016) Are Chinese growth and inflation too smooth? Evidence from Engel curves. *American Economic Journal: Macroeconomics* 8, 113–44.

Newey, W.K. & J.L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.

Ruymgaart, F.H (1998) A note on weak convergence of density estimators in Hilbert spaces. *Statistics* 30, 331–343.

Tao, Jing (2014) Inference for point and partially identified semi-nonparametric conditional moment models. Working Paper, University of Wisconsin-Madison.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Tikhonov, A. N. (1963) *On the Solution of Ill-Posed Problems and the Method of Regularization*, vol. 151, 501–504, Russian Academy of Sciences.
Tsybakov, A. (2009) *Introduction to Nonparametric Estimation.* Springer.
Van Der Vaart, A.W. & J.A. Wellner (2000) *Weak Convergence.* Springer.
Yatchew, A. & J.A. No (2001) Household gasoline demand in Canada. *Econometrica* 69, 1697–1709.

APPENDIX

A.1: CONCENTRATION OF THE SUPREMUM OF EMPIRICAL PROCESSES

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

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

and denote the symmetrized process by

$$\nu_n^{\eta}(h) = \frac{1}{n}\sum_{i=1}^{n} \eta_i h(X_i),$$

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

Proposition A.1.1 *The following two inequalities hold*

$$2^{-1}\mathbb{E}\left\|\nu_n^{\eta}\right\|_{\mathcal{H}} - 2^{-1}n^{-1/2}\|Ph\|_{\mathcal{H}} \leq \mathbb{E}\|\nu_n\|_{\mathcal{H}} \leq 2\mathbb{E}\left\|\nu_n^{\eta}\right\|_{\mathcal{H}}. \tag{A.1}$$

The second inequality in equation (A.1), is a symmetrization inequality, while the first is the desymmetrization inequality. These inequalities allow us to establish uniform limit theorems and play an essential role in the empirical process theory, see Van Der Vaart and Wellner (2000). We refer to (Koltchinskii, 2006, p. 7) and references therein for a proof of this modification of the symmetrization inequality.

Our construction of confidence sets relies on the concentration inequality for functions of bounded difference, also known as McDiarmid's inequality, see the excellent treatment of concentration inequalities in Boucheron et al. (2013). Such concentration inequalities tell us that the probability that the supremum of the empirical process $\|\nu_n\|_{\mathcal{H}}$ deviates from its expected value $\mathbb{E}\|\nu_n\|_{\mathcal{H}}$ declines exponentially fast. Although this expected value is unknown in practice, it can be estimated using the multiplier bootstrap with Rademacher multipliers. The symmetrization inequality allows us then to compare the expected value of this estimate $\|\nu_n^{\eta}\|_{\mathcal{H}}$ to the original expected value, which leads to the data-driven concentration inequality. This insightful idea comes from the statistical learning literature (see, e.g., Koltchinskii, 2001).

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The next proposition states the precise version of the data-driven concentration inequality used in this article.

Proposition A.1.2 *Let* $(X_i)_{i=1}^n$, $(\eta_i)_{i=1}^n$, *and* $\mathcal{H}$ *be defined as before. Suppose also that the absolute value of all* $h \in \mathcal{H}$ *is uniformly bounded by some* H. *Then for all* $x > 0$ *and* $n \in \boldsymbol{N}$

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

See Koltchinskii (2011), Thm. 4.6, for the proof of this result.

A.2: GENERAL ILL-POSED INVERSE PROBLEMS

A.2.1: 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 article, let P be a probability measure to any of the ill-posed models introduced in the previous section. The model is described by the function equation $r = T\varphi$, where $\varphi : [0, 1]^p \to \mathbf{R}$ is an infinite dimensional parameter of interest. We aim to construct a random set $C_{n,\,1-\gamma}$ containing the function φ with probability of at least $1 - \gamma$ for $\gamma \in (0, 1)$ and such that its diameter shrinks as the sample size increases at a specific rate. We focus on confidence sets for the Tikhonov-regularized estimator defined as follows

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

where $\hat{T}$, $\hat{T}^*$, and $\hat{r}$ are appropriate estimators[4] 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 et al. (2007).

For models considered in this article, the dominating stochastic component of the Tikhonov-regularized estimator is driven by a sequence of i.i.d. centered random functions

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n}(\alpha_n I + T^*T)^{-1}T^*X_{ni}. \tag{A.2}$$

Assume that T is an 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. By Lemma A.4.1 in the Appendix C, the operator $(\alpha_n I + T^*T)$ is invertible between spaces of continuous functions. Thus, trajectories of the process in equation (A.2) belong to the space $(C, \|.\|_\infty)$.

To build uniform confidence sets, we need to approximate the distribution of the supremum of this process. The most straightforward route to achieve this would be to establish its weak convergence to some Gaussian process and then to rely on quantiles of Gaussian

[4] If some operators are known, which is the case in the density deconvolution model, we replace estimators by known quantities.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

suprema to build uniform confidence sets. Unfortunately, the process in equation (A.2) cannot converge weakly as a random element in $(C, \|.\|_\infty)$ space. Indeed, we introduced regularization to smooth out discontinuities of the operator inversion. However, in the limit, as the regularization parameter tends to zero, we get back a discontinuous/unbounded operator. This discontinuity in the limit is a precise reason why the functional convergence fails.

More precisely, we can see that depending on the direction $\delta \in L_2$, the speed of convergence of inner products $\langle \hat{\varphi} - \varphi, \delta \rangle$ differs, see Carrasco et al. (2007), which makes it impossible to converge weakly in L_2. In Proposition A.2.1, we show formally that the weak convergence in $(C, \|.\|_\infty)$ is impossible generalizing the result of Cardot et al. (2007) who focus on the weak convergence in L_2 in the special case of the functional linear regression without instrumental variables. Note that our result holds for the arbitrary ill-posed inverse model, including the nonparametric IV, functional IV regressions, and the density deconvolution. Note also that our result holds regardless of what type of the estimator is used in practice, be it kernel or series estimators in the nonparametric IV model. In particular, it is worth comparing it to the impossibility of weak convergence of the kernel density estimator, where inner products do converge at the same speed, but the only compatible limiting process is a white noise process, as in Ruymgaart (1998).

Proposition A.2.1. *Suppose that the inverse of the operator T^*T in equation (A.2) is unbounded. Then there does not exist a normalizing sequence r_n such that $r_n v_n$ would converge weakly in $(C, \|.\|_\infty)$ to a nondegenerate random process.*

Proof of Proposition A.2.1 Recall that for some normalizing sequence r_n, the weak convergence of $r_n v_n$ in L_2 to some random element requires $\langle r_n v_n, \delta \rangle$ to converge weakly in $\mathbf{R}$ for all $\delta \in L_2$, Van Der Vaart and Wellner (2000), Thm. 1.8.4. Since $(T^*T)^{-1}$ is unbounded with $\mathcal{D}[(T^*T)^{-1}] \subset L_2[0,1]^p$, for $\delta \in \mathcal{D}[(T^*T)^{-1}]$, we need to set $r_n = n^{1/2}$, since

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

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 $\langle n^{1/2} v_n, \delta \rangle$ cannot 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 $r_n v_n$ does not converge weakly in $(C, \|.\|_\infty)$ for any choice of the normalizing sequence r_n. □

Proposition A.2.1 tells us that it is not possible to rely on the asymptotic approximation with the conventional central limit theorem. As a result, we need to rely on alternative approaches to inference.

A.2.2: Bootstrap Confidence Sets

Consider the empirical process

$$V_n(g) = \sqrt{n}(P_n - P)g, \qquad g \in \mathcal{G}_n = \left\{g_w : w \in [0,1]^q\right\}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

indexed by some pointwise measurable pre-Gaussian class of functions $\mathcal{G}_n \subset L_2(P)$, where $L_2(P)$ is the set of square-integrable functions with respect to P and P_n is the empirical measure. We will always assume that the data are i.i.d. In our setting, the empirical process V_n can also be considered as a stochastic process

$$V_n(g_w) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{ni}(w)$$

indexed by $w \in [0, 1]^q$. Let

$$\|V_n\|_\infty \triangleq \|V_n\|_{\mathcal{G}_n} \triangleq \sup_{g \in \mathcal{G}_n} |V_n(g)|$$

be the supremum of the empirical process and let

$$\|\mathbb{G}_n\|_\infty \triangleq \|\mathbb{G}_n\|_{\mathcal{G}_n} \triangleq \sup_{g \in \mathcal{G}_n} |\mathbb{G}_n(g)|$$

be the supremum of the tight centered Gaussian process with the same covariance structure as V_n, that is, for all $g_1, g_2 \in \mathcal{G}_n$

$$\begin{aligned}\mathbb{E}[\mathbb{G}_n(g_1)\mathbb{G}_n(g_2)] &= \mathbb{E}[V_n(g_1)V_n(g_2)] \\ &= P(g_1 - Pg_1)(g_2 - Pg_2).\end{aligned}$$

We shall note that we will work only with processes having continuous trajectories so that all suprema can be restricted to countable sets of rational numbers, and whence are measurable.

Note that by the linearity

$$\sup_{g \in \mathcal{G}_n} |V_n(g)| = \sup_{g \in \mathcal{G}_n \cup -\mathcal{G}_n} V_n(g),$$

so we can always drop the absolute value from the supremum extending the supremum to the enlarged class of functions. A similar observation is valid for the supremum of the Gaussian process Dudley (2014), Thm. 2.1. We will typically work with suprema of absolute value and will use the above facts repeatedly.

Let G_n be the envelope of the class $\mathcal{G}_n$, which we assume to be square-integrable and let $N(\mathcal{G}_n, \|.\|_{Q,2}, \epsilon)$ be the ϵ-covering of $(\mathcal{G}_n, \|.\|_{Q,2})$, with $\|g\|_{Q,q} = (Q|g|^q)^{1/q}$ for some probability measure Q. We assume that $\mathcal{G}_n$ is a VC-type class of functions.

Assumption A.2.1. Suppose that the class $\mathcal{G}_n$ is such that

$$\sup_Q N(\mathcal{G}_n, \|.\|_{Q,2}, \epsilon \|G_n\|_{Q,2}) \leq \left(\frac{A}{\epsilon}\right)^v, \qquad 0 < \epsilon \leq 1$$

for some universal constants $A \geq e$ and $v \geq 1$, where the supremum is taken over all probability measures Q with $0 < \|G_n\|_{Q,2} < \infty$.

For an i.i.d. sequence of Gaussian random variables $(\varepsilon_i)_{i=1}^n$, consider the supremum of the bootstrapped process

$$\left\|V_n^\varepsilon\right\|_\infty = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i X_{ni} \right\|_\infty .$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The process X_{ni} is typically unobserved. Let $\hat{X}_{ni}$ be its empirical counterpart estimated from the data and let $c_{1-\gamma}^{*}$ be $1-\gamma$ conditional on the data $\mathcal{D}_n$ quantile of the supremum of the bootstrapped process

$$\|\hat{V}_n^{\varepsilon}\|_{\infty} = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \hat{X}_{ni} \right\|_{\infty},$$

where $(\varepsilon_i)_{i=1}^n$ are i.i.d. $N(0, 1)$ random variables independent from the sample.

Assumption. (i) There exist constants $0 < \underline{\sigma} < \bar{\sigma}$ such that $\underline{\sigma}^2 \leq Var(X_{ni}(w)) \leq \bar{\sigma}^2$ for all $w \in [0, 1]^q$; (ii) $\sup_{g \in \mathcal{G}_n} \|g\|_{P,k}^r \leq \sigma^2 b_n^{r-2} < \infty, r = 2, 3, 4$ and $\|G_n\|_{P,4} \leq b_n < \infty$ for some $b_n = O(n^c)$.

Assumptions A.2.3 and A.2.2 consist of several technical conditions needed for the Gaussian and bootstrap approximation of Chernozhukov et al. (2014).

Assumption A.2.3. Let $\mathcal{P}$ be the class of models consisting of all probability distributions corresponding to the ill-posed model satisfying Assumptions 3.1, A.2.1, and A.2.2. Suppose additionally that $\mathcal{P}$ is such that

$$\text{(C1)} \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P \left\| \hat{T}^* - T^* \right\|_{2,\infty}^2 = O(\delta_{1n}^2),$$

$$\text{(C2)} \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P \left\| \hat{V}_n^{\varepsilon} - V_n^{\varepsilon} \right\|_{\infty} = O(\delta_{2n}),$$

$$\text{(C3)} \quad \hat{r} - \hat{T}\varphi = \frac{u_n}{n} \sum_{i=1}^{n} X_{ni} + R_{1n},$$

where R_{1n} is such that $\sup_{P \in \mathcal{P}} \mathbb{E}_P \|R_{1n}\|^2 = O(\delta_{3n}^2)$ for some sequences δ_{1n}, δ_{2n}, $\delta_{3n} \to 0$.

Assumption (C3) is asymptotic linearization of residuals in the model. In some cases, this expansion is exact and R_{1n} is trivially zero. (C1) controls the estimation error in the operator, while (C2) controls the error in the bootstrap process.

We focus on bootstrap confidence sets described as

$$C_{n,1-\gamma}^{*} = \left\{ \left[\hat{\varphi}(z) - q_n^{*}, \hat{\varphi}(z) + q_n^{*} \right] : z \in [0, 1]^p \right\} \tag{A.3}$$

with $q_n^{*} = \frac{c_{1-\gamma}^{*} \|\hat{T}^*\|_{2,\infty} + \log^{-1} n}{\alpha_n n^{1/2}} u_n$.

In the following theorem, we show that our bootstrap confidence sets are honest to $\mathcal{P}$ and characterize coverage errors explicitly. This result will be specialized to each particular model in the following sections.

THEOREM A.1. *Suppose that Assumptions 3.1, A.2.1, A.2.2, and A.2.3 are satisfied and* $b_n^4 \log^7 n / n \to 0$. *Then*

$$\inf_{P \in \mathcal{P}} \Pr(\varphi \in C_{n,1-\gamma}^{*}) \geq 1 - \gamma - O_p(\delta_n) - o_p(1)$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

with

$$\delta_n = \left(\alpha_n^{-1/2}\delta_{1n} + \alpha_n^{-1}\delta_{3n} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n + \delta_{2n}\log n.$$

Proof. Using Assumption A.2.3(C3), decompose

$$\hat{\varphi} - \varphi = \hat{v}_n + (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* R_{1n} + R_{2n} + R_{3n}$$

with

$$\hat{v}_n = (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* \frac{u_n}{n}\sum_{i=1}^n X_{ni}$$

$$R_{2n} = \left[(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T} - (\alpha_n I + T^*T)^{-1}T^*T\right]\varphi.$$

$$R_{3n} = (\alpha_n I + T^*T)^{-1}T^*T\varphi - \varphi.$$

By Lemma A.4.2 there exists a constant $C < \infty$ independent from $P \in \mathcal{P}$ such that

$$\|R_{2n}\|_\infty \le \frac{C}{\alpha_n^{1/2}}\left(\|\hat{T}^* - T^*\|_{2,\infty} + \|\hat{T}^* - T^*\|_{2,\infty}^2\right),$$

whence under Assumption A.2.3(C1)

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\|R_{2n}\|_\infty = O\left(\frac{\delta_{1n} + \delta_{1n}^2}{\alpha_n^{1/2}}\right).$$

Next, under Assumption 3.1 by Proposition 3.1 there exists a constant R independent from $P \in \mathcal{P}$ such that

$$\|R_{3n}\|_\infty \le R\alpha_n^{\beta\wedge 1}.$$

Lastly,

$$\begin{aligned}\left\|(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* R_{1n}\right\|_\infty &= \left\|\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1} R_{1n}\right\|_\infty \\ &\le \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\|.\end{aligned}$$

Then under Assumption A.2.3 by Markov's inequality

$$\begin{aligned}&\Pr\left(\varphi \in C^*_{n,1-\gamma}\right) \\ &= \Pr\left(\|\hat{\varphi} - \varphi\|_\infty \le q_n^*\right) \\ &\ge \Pr\left(\|\hat{v}_n\|_\infty + \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty \le q_n^*\right) \\ &\ge \Pr\left(\|V_n\|_\infty \le c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1}\right) \\ &\quad - \Pr\left(\frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty > \frac{u_n}{2\alpha_n n^{1/2}\log n}\right)\end{aligned}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$$\geq \Pr\left(\|V_n\|_\infty \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1}\right)$$
$$- \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty\right] u_n^{-1}\alpha_n n^{1/2}\log n$$
$$\geq \Pr\left(\|V_n\|_\infty \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1}\right)$$
$$- O\left(\left(\alpha_n^{-1}\delta_{3n} + \alpha_n^{-1/2}\delta_{1n} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n\right).$$

Under Assumptions A.2.1 and A.2.2 by the coupling inequality in Chernozhukov et al. (2016), Thm. 2.1, for any $P \in \mathcal{P}$ one can approximate the supremum of the empirical process V_n by the supremum of the Gaussian process $\mathbb{G}_n$ with the same covariance structure

$$\Pr\left(\|V_n\|_\infty \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1}\right)$$
$$\geq \Pr\left(\|\mathbb{G}_n\|_{\mathcal{G}_n} \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1} - C_1\left(\frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/3}\log n}{n^{1/6}}\right)\right)$$
$$- \Pr\left(\left|\|V_n\|_\infty - \|\mathbb{G}_n\|_{\mathcal{G}_n}\right| > C_1\left(\frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/3}\log n}{n^{1/6}}\right)\right)$$
$$\geq \Pr\left(\|\mathbb{G}_n\|_{\mathcal{G}_n} \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1} - C_1\left(\frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/3}\log n}{n^{1/6}}\right)\right) - o(1),$$

where C_1 is some absolute constants and the $o(1)$ term does not depend on $P \in \mathcal{P}$. Next, for some sequence of polynomial order δ_{4n} (to be specified below), put

$$\delta_n^I = \delta_{4n} + \frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/3}\log n}{n^{1/6}}$$
$$\delta_n^{II} = \left(\alpha_n^{-1}\delta_{3n} + \alpha_n^{-1/2}\delta_{1n} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n.$$

Then

$$\Pr\left(\varphi \in C^*_{n,1-\gamma}\right) \geq \Pr\left(\|\mathbb{G}_n\|_{\mathcal{G}_n} \leq c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1} + \delta_{4n}\right)$$
$$- \sup_{x\in\mathbf{R}} \Pr\left(\left|\|\mathbb{G}_n\|_{\mathcal{G}_n} - x\right| \leq \delta_n^I\right) - O\left(\delta_n^{II}\right) - o(1).$$

Under Assumption A.2.2(i) by the anticoncentration inequality in Chernozhukov et al. (2014), Lem. A.1, for every $\epsilon > 0$

$$\sup_{x\in\mathbf{R}} \Pr\left(\left|\|\mathbb{G}_n\|_{\mathcal{G}_n} - x\right| \leq \epsilon\right) \leq C_2\epsilon\left(\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{G}_n} + \sqrt{1 \vee \log(\underline{\sigma}/\epsilon)}\right),$$

where the constant C_2 depends only on $\underline{\sigma}$ and $\bar{\sigma}$.

Moreover, under Assumption A.2.2 by the Dudley–Sudakov's entropy bound[5] , see, for example, Dudley (2016)

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{G}_n} \leq 24\int_0^{\bar{\sigma}} \sqrt{\log N(\mathcal{G}_n, \|.\|_{P,2}, \epsilon)}\mathrm{d}\epsilon = O\left(\log^{1/2} n\right),$$

[5]Note that $N(\mathcal{G}_n, \|.\|_{P,2}, \epsilon) = 1$ for all $\epsilon \geq \mathrm{diam}(\mathcal{G}_n)/2$ and $\mathrm{diam}(\mathcal{G}_n) \leq 2\sup_{g\in\mathcal{G}_n}\|g\|_{P,2} \leq 2\bar{\sigma}$.

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

where the last equality follows under Assumptions A.2.1 and A.2.2(ii).

Next under Assumptions A.2.1 and A.2.2 by the coupling inequality in Chernozhukov et al. (2016), Thm. 2.2, we approximate the supremum of the Gaussian process by the supremum of the multiplier bootstrap process. To that end for some constant C_3 independent from $P \in \mathcal{P}$

$$\begin{aligned}
&\Pr\left(\|\mathbb{G}_n\|_\infty \le c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1} + C_3\left(\frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/2}\log n}{n^{1/4}}\right)\right)\\
&\ge \Pr\left(\|V_n^\varepsilon\|_\infty \le c^*_{1-\gamma} + (2\|\hat{T}^*\|_{2,\infty}\log n)^{-1}|\mathcal{D}_n\right) - o_p(1)\\
&\ge \Pr\left(\|\hat{V}_n^\varepsilon\|_\infty \le c^*_{1-\gamma}|\mathcal{D}_n\right) - 2\|\hat{T}^*\|_{2,\infty}\mathbb{E}\left[\left\|\hat{V}_n^\varepsilon - V_n^\varepsilon\right\|_\infty |\mathcal{D}_n\right]\log n - o_p(1),
\end{aligned}$$

where we use Markov's inequality, $\mathcal{D}_n$ denotes the data, and the $o_p(1)$ term does not depend on $P \in \mathcal{P}$. Setting $\delta_{4n} = C_3\left(\frac{b_n\log^{5/4} n}{n^{1/4}} + \frac{b_n^{1/2}\log n}{n^{1/4}}\right)$ under Assumption A.2.3(C2), we obtain

$$\Pr\left(\varphi \in C^*_{n,1-\gamma}\right) \ge 1 - \gamma - O_p\left(\delta_n^I\log^{1/2} n + \delta_n^{II} + \delta_{2n}\log n\right) - o_p(1),$$

whence the result follows under the assumption $b_n^4\log^7 n/n \to 0$. □

To characterize expected diameters of bootstrap confidence sets we need additional assumptions.

Assumption A.2.4. Suppose that (i) conditionally on the data $\|\hat{V}_n^\varepsilon\|_\infty$ is a supremum over the class of functions satisfying Assumption A.2.1;

$$(ii)\ \sup_{P\in\mathcal{P}}\max_{1\le i\le n}\mathbb{E}_P\|\hat{X}_{ni} - X_{ni}\|_\infty^2 = O(1); \qquad (iii)\ \sup_{P\in\mathcal{P}}\max_{1\le i\le n}\mathbb{E}_P\|X_{ni}\|_\infty^2 = O(u_n^2),$$

and $u_n \to \infty$.

THEOREM A.2. *Suppose that Assumption A.2.4 and assumptions of Theorem A.1 are satisfied. If $\delta_{1n} = O(1)$, then*

$$\left|C^*_{n,1-\gamma}\right| = O_p\left(\frac{u_n^2}{\alpha_n n^{1/2}}\right)$$

uniformly over $P \in \mathcal{P}$.

Proof of Theorem A.2 First, note that

$$\begin{aligned}
\left|C^*_{n,1-\gamma}\right|_\infty &= 2\left|q_n^*\right|\\
&= \frac{2u_n}{\alpha_n n^{1/2}}\left\{\left[c^*_{1-\gamma}\|\hat{T}^*\|_{2,\infty}\right] + \log^{-1} n\right\}.
\end{aligned}$$

Recall that $c^*_{1-\gamma}$ is $1-\gamma$ conditional on the data $\mathcal{D}_n$ quantile of the supremum of the Gaussian process $\|\hat{V}_n^\varepsilon\|_\infty$. By the Gaussian concentration inequality, for example, see

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Boucheron et al. (2013), Thm. 5.8, conditionally on the data

$$c^*_{1-\gamma} \leq \mathbb{E}_P\left[\|\hat{V}^{\varepsilon}_n\|_\infty | \mathcal{D}_n\right] + \sqrt{\frac{2}{n}\sum_{i=1}^n \|\hat{X}_{ni}\|^2_\infty \log(1/\gamma)}.$$

By Dudley–Sudakov's inequality, for example, see Dudley (2016), and the change of variables under Assumption A.2.1

$$\mathbb{E}_P\left[\|\hat{V}^{\varepsilon}_n\|_\infty | \mathcal{D}_n\right] \leq 24\sqrt{\frac{1}{n}\sum_{i=1}^n \|\hat{X}_{ni}\|^2_\infty}\int_0^1 \sqrt{v \log \frac{A}{\varepsilon}}\mathrm{d}\varepsilon.$$

The result follows under Assumption A.2.1(C1). □

A.2.3: Concentration-Based Confidence Sets

The confidence sets based on the concentration inequality are described as

$$C_{n,1-\gamma} = \left\{[\hat{\varphi}(z) - \hat{q}_n, \hat{\varphi}(z) + \hat{q}_n] : z \in [0,1]^p\right\},$$

where $\hat{q}_n = 2\left\|\hat{v}^{\eta}_n\right\|_\infty + \frac{3\|\hat{T}^*\|_{2,\infty} G\sqrt{2\log(1/\gamma)} + \log^{-1} n}{\alpha_n n^{1/2}} u_n$ and

$$\left\|\hat{v}^{\eta}_n\right\|_\infty = \left\|\frac{u_n}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\eta_i \hat{X}_{ni}\right\|_\infty$$

denotes the estimate of the supremum of the symmetrized process. The operator norm $\|\hat{T}^*\|_{2,\infty}$ can be computed by Lemma A.4.4, while G is selected so that $\|X_{ni}\| \leq G$.

For concentration inequality, we need a somewhat different version of Assumption A.2.3.

Assumption. Let $\mathcal{P}$ be the class of models consisting of all probability distributions corresponding to the ill-posed model satisfying Assumption 3.1. Suppose additionally that $\mathcal{P}$ is such that

$$\text{(C1)} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P \left\|\hat{T}^* - T^*\right\|^2_{2,\infty} = O(\delta^2_{1n}),$$

$$\text{(C2)} \quad \sup_{P\in\mathcal{P}} \max_{1\leq i\leq n} \mathbb{E}_P \left\|\hat{X}_{ni} - X_{ni}\right\|^2 = O(\delta^2_{2n}),$$

$$\text{(C3)} \quad \hat{r} - \hat{T}\varphi = \frac{u_n}{n}\sum_{i=1}^n X_{ni} + R_{1n},$$

where R_{1n} is a remainder term such that $\sup_{P\in\mathcal{P}} \mathbb{E}_P \|R_{1n}\|^2 = O(\delta^2_{3n})$ and $\delta_{1n}, \delta_{2n}, \delta_{3n} \to 0$ are some sequences.

THEOREM A.3. *Suppose that Assumption A.2.5 is satisfied and $\|X_{ni}\| \leq G$ for some universal constant G for all $P \in \mathcal{P}$. Then*

$$\inf_{P\in\mathcal{P}} Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq 1 - \gamma - O(\delta_n)$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

with $\delta_n = \left(\frac{\delta_{3n}}{\alpha_n} + \frac{\delta_{1n}}{\alpha_n^{1/2}} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n + \delta_{2n}\log n.$

Proof of Theorem A.3 Under Assumption A.2.5(C3), decompose

$$\hat{\varphi} - \varphi = \hat{v}_n + (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* R_{1n} + R_{2n} + R_{3n}$$

with

$$\hat{v}_n = (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^* \frac{u_n}{n}\sum_{i=1}^n X_{ni}$$
$$R_{2n} = \left[(\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T} - (\alpha_n I + T^*T)^{-1}T^*T\right]\varphi.$$
$$R_{3n} = (\alpha_n I + T^*T)^{-1}T^*T\varphi - \varphi.$$

Recall that $\hat{q}_n = 2\|v_n^\eta\|_\infty + \frac{3\|\hat{T}^*\|_{2,\infty}G\sqrt{2\log(1/\gamma)}+\log^{-1}n}{\alpha_n n^{1/2}}u_n$ and we also denote

$$v_n = (\alpha_n I + T^*T)^{-1}T^* \frac{u_n}{n}\sum_{i=1}^n X_{ni},$$
$$v_n^\eta = (\alpha_n I + T^*T)^{-1}T^* \frac{u_n}{n}\sum_{i=1}^n \eta_i X_{ni},$$

where ε_i are i.i.d. Rademacher random variables. Then by the triangle and Markov's inequalities

$$\begin{aligned}
&\Pr\left(\varphi \in C_{n,1-\gamma}\right)\\
&= \Pr\left(\|\hat{\varphi} - \varphi\|_\infty \le \hat{q}_n\right)\\
&\ge \Pr\left(\|\hat{v}_n\|_\infty + \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty \le \hat{q}_n\right)\\
&\ge \Pr\left(\|v_n\|_\infty + \left\|\hat{v}_n - v_n\right\|_\infty + \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty \le \hat{q}_n\right)\\
&\ge \Pr\left(\|v_n\|_\infty \le 2\|v_n^\eta\|_\infty + \frac{3\|T^*\|_{2,\infty}G\sqrt{2\log(1/\gamma)}}{\alpha_n n^{1/2}}u_n\right)\\
&\quad - \Pr\left(\left\|\hat{v}_n - v_n\right\|_\infty + \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty > \frac{u_n}{2\alpha_n n^{1/2}\log n}\right)\\
&\quad - \Pr\left(\left\|\hat{v}_n^\eta - v_n^\eta\right\|_\infty + \frac{3G\sqrt{2\log(1/\gamma)}}{\alpha_n n^{1/2}}u_n\|\hat{T}^* - T^*\|_{2,\infty} > \frac{u_n}{2\alpha_n n^{1/2}\log n}\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$$\geq \Pr\left(\|\nu_n\|_\infty \leq 2\|\nu_n^\eta\|_\infty + \frac{3\|T^*\|_{2,\infty} G\sqrt{2\log(1/\gamma)}}{\alpha_n n^{1/2}} u_n\right)$$
$$- 2\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\left\|\hat{\nu}_n - \nu_n\right\|_\infty + \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\|R_{1n}\| + \|R_{2n}\|_\infty + \|R_{3n}\|_\infty\right] u_n^{-1}\alpha_n n^{1/2}\log n$$
$$- 2\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\left\|\hat{\nu}_n^\eta - \nu_n^\eta\right\|_\infty u_n^{-1}\alpha_n n^{1/2}\log n + 3G\sqrt{2\log(1/\gamma)}\|\hat{T}^* - T^*\|_{2,\infty}\log n\right]. \tag{A.4}$$

Under Assumption A.2.5(C1) and (C2) by Lemma A.4.3

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left\|\hat{\nu}_n - \nu_n\right\|_\infty = O\left(\frac{u_n\delta_{1n}}{\alpha_n^{3/2}n^{1/2}}\right),$$

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left\|\hat{\nu}_n^\eta - \nu_n^\eta\right\|_\infty = O\left(\frac{u_n\delta_{1n}}{\alpha_n^{3/2}n^{1/2}} + \frac{u_n\delta_{2n}}{\alpha_n n^{1/2}}\right).$$

Therefore, another application of Assumption A.2.5 and the order of R_{2n} and R_{3n} from the proof of Theorem A.1 imply

$$\Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq \Pr\left(\|\nu_n\|_\infty \leq 2\|\nu_n^\eta\|_\infty + \frac{3\|T^*\|_{2,\infty} G\sqrt{2\log(1/\gamma)}}{\alpha_n n^{1/2}} u_n\right)$$
$$- O\left(\left(\frac{\delta_{3n}}{\alpha_n} + \frac{\delta_{1n}}{\alpha_n^{1/2}} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n + \delta_{2n}\log n\right).$$

Next, $\|\nu_n\|_\infty$ is the supremum of empirical processes indexed by the class of functions

$$\mathcal{H}_n = \left\{h : x \in C[0,1]^q \to u_n\left[(\alpha_n I + T^*T)^{-1}T^*x\right](z),\ z \in [0,1]^p\right\}.$$

For any $h \in \mathcal{H}_n$,

$$\|h(X_{ni})\|_\infty \leq u_n\|T^*\|_{2,\infty}\|(\alpha_n I + TT^*)^{-1}\|\|X_{ni}\| \leq \frac{\|T^*\|_{2,\infty}}{\alpha_n} G u_n \triangleq H_n.$$

So H_n is the envelope for the class $\mathcal{H}_n$. Then by Proposition A.1.2

$$\Pr(\|\nu_n\|_\infty \leq \hat{q}_n) = \Pr\left(\|\nu_n\|_\infty \leq 2\|\nu_n^\eta\|_\infty + 3H_n\sqrt{\frac{2\log(1/\gamma)}{n}}\right)$$
$$\geq 1 - \gamma,$$

whence the result. □

The next result describes the convergence rate of the expected diameter of the confidence set based on the concentration inequality.

THEOREM A.4. *Suppose that assumptions of Theorem A.3 are satisfied. If* $\delta_{1n}/\alpha_n^{1/2} = O(1)$, *then*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{u_n}{\alpha_n n^{1/2}}\right).$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Proof of Theorem A.4 Since $\mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = 2\mathbb{E}_P\left|\hat{q}_n\right|$, under assumptions of Theorem A.3, there exists some universal constant C such that

$$\mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty \le C\left\{\mathbb{E}_P\|v_n^\eta\|_\infty + \mathbb{E}_P\|\hat{v}_n^\eta - v_n^\eta\|_\infty + \frac{u_n}{\alpha_n n^{1/2}}\mathbb{E}_P\|\hat{T}^*\|_{2,\infty}\right\}.$$

By Jensen's inequality, since $\|X_{ni}\| \le C$

$$\begin{aligned}\mathbb{E}_P\|v_n^\eta\|_\infty &\le \frac{\|T^*\|_{2,\infty}}{\alpha_n}\mathbb{E}_P\left\|\frac{u_n}{n}\sum_{i=1}^n \eta_i X_{ni}\right\| \\ &\le \frac{u_n\|T^*\|_{2,\infty}}{\alpha_n n^{1/2}}\sqrt{\mathbb{E}_P\|X_{ni}\|^2} \\ &\le \frac{u_n\|T^*\|_{2,\infty}}{\alpha_n n^{1/2}}G.\end{aligned}$$

Therefore by Lemma A.4.3 under Assumptions 3.1 and A.2.5(C1) and (C2)

$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{u_n}{\alpha_n n^{1/2}} + \frac{u_n\delta_{1n}}{\alpha_n^{3/2}n^{1/2}} + \frac{u_n\delta_{2n}}{\alpha_n n^{1/2}}\right).$$

□

A.3: NPIV, FUNCTIONAL REGRESSIONS, AND DENSITY DECONVOLUTION

A.3.1: NPIV Model

Proposition A.3.1 *Suppose that Assumption 3.2(i)–(iii) is satisfied and the sequence of bandwidth parameters is such that* $1/(nh_n^{p+q}) = O(1)$*, then for all* $1 < r < \infty$

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

where the constant C *depends only on* $s, M, r, \bar{f}, \|k\|_\infty$.

Proof. By Lemma A.4.4,

$$\begin{aligned}\mathbb{E}\left\|\hat{T}^* - T^*\right\|_{2,\infty}^r &= \mathbb{E}\left[\sup_{z\in(0,1)^p}\left(\int_{[0,1]^q}\left|\hat{f}_{ZW}(z,w) - f_{ZW}(z,w)\right|^2 \mathrm{d}w\right)^{1/2}\right]^r \\ &\le \mathbb{E}\left\|\hat{f}_{ZW} - f_{ZW}\right\|_\infty^r.\end{aligned}$$

By Minkowski inequality,

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

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The order of the bias follows by standard computations under Assumption 3.2(i) and (iii), see Tsybakov (2009)

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

where the big-O term is independent from the distribution of (Z, W). For the variance term, we apply the moment inequality in Giné and Nickl (2016), see Thm. 5.1.5 and 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, we obtain the result. □

Proof of Theorem 3.1 We shall verify conditions of Theorems A.1 and A.3. We show first that in the nonparametric IV model Assumption A.2.3(C3) is verified with $u_n = h_n^{-q/2}$ and $X_{ni}(w) = U_i h_n^{-q/2} K_w(h_n^{-1}(W_i - w))$, where $[K_z * \varphi](Z) = h_n^{-p} \int \varphi(z) K_z(h_n^{-1}(Z - z))\mathrm{d}z$. Indeed, it is easy to see that

$$(\hat{r} - \hat{T}\varphi)(w) = \frac{h_n^{-q/2}}{n} \sum_{i=1}^n X_{ni}(w) + R_{1n}(w)$$

$$R_{1n}(w) = \frac{1}{nh_n^q} \sum_{i=1}^n \left\{\varphi(Z_i) - [K_z * \varphi](Z_i)\right\} K_w\left(h_n^{-1}(W_i - w)\right).$$

Under Assumption 3.1, $\varphi \in C_M^t[0,1]^p$, so that using standard bias computations, for example, see Tsybakov (2009), under Assumption 3.2(iii) there exists some constant $C < \infty$ such that

$$\left\| \varphi - [K_z * \varphi] \right\|_\infty \leq C h_n^t.$$

Then under Assumption 3.2(i) and (ii),

$$\begin{aligned}
\sup_{P \in \mathcal{P}} \mathbb{E}_P \|R_{1n}\|^2 &\leq \sup_{P \in \mathcal{P}} \mathbb{E}_P \left\| \frac{1}{n} \sum_{i=1}^n \left| h_n^{-q} K_w\left(h_n^{-1}(W_i - .)\right) \right| \right\|^2 C h_n^t \\
&\leq \sup_{P \in \mathcal{P}} \mathbb{E}_P \left\| h_n^{-q} K_w\left(h_n^{-1}(W_i - .)\right) \right\|^2 C h_n^t \\
&= O(h_n^{t-q}).
\end{aligned}$$

Therefore, Assumption A.2.3(C3) is verified with $\delta_{3n} = h_n^{(t-q)/2}$.

Consider now the class of functions

$$\mathcal{G}_n = \left\{ (u, w) \mapsto u \frac{1}{h_n^{q/2}} K_w\left(h_n^{-1}(w - v)\right) : v \in [0,1]^q \right\}.$$

Under Assumption 3.2(iii) and (iv), this class admits a square-integrable envelope $G_n(u, w) = h_n^{-q/2} F \|K\|_\infty$ with

$$\|G_n\|_{P,2} = h_n^{-q/2} \|K\|_\infty F$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

for all $P \in \mathcal{P}$. Recall that K_w is a product kernel of functions of bounded variation under Assumption 3.2(iii) and that the class of translations of functions of bounded variation is of the VC-type, so Assumption A.2.1 holds.

Next, for any $g_n \in \mathcal{G}_n$,

$$\text{Var}\ (g_n(v)) = \mathbb{E}\left|U h_n^{-q/2} K_w\left(h_n^{-1}(W-v)\right)\right|^2,$$

so that by the law of iterated expectation and change of variables under Assumption 3.2(ii) and (iv)

$$\underline{\sigma}_2 \|k\|^{2q} \underline{f} \leq \text{Var}(g_n(v)) \leq F\|k\|^{2q} \bar{f}.$$

This verifies Assumption A.2.2(i). Next, under Assumption 3.2(iv)

$$\begin{aligned}
\|G\|_{P,4} &= h_n^{-q/2}\|k\|_\infty^q \|U\|_{P,4} \\
&\leq F h_n^{-q/2}\|k\|_\infty^q \\
P|g_n|^3 &= h_n^{-3q/2}\mathbb{E}\left|U K_w\left(h_n^{-1}(W-v)\right)\right|^3 \\
&\leq F h_n^{-q/2}\|k\|_3^{3q} \\
P|g_n|^4 &= h_n^{-4q/2}\mathbb{E}\left|U K_w\left(h_n^{-1}(W-v)\right)\right|^4 \\
&\leq F h_n^{-q}\|k\|_4^{4q}.
\end{aligned}$$

Therefore, Assumption A.2.2(ii) is satisfied with $b_n \leq C h_n^{-q/2}$ and some universal constant $C < \infty$. Note that $b_n^4 \log^7 n/n \to 0$ under Assumption 3.2(vi) (c).

It remains to verify Assumption A.2.3(C1) and (C2). The former follows from the Proposition A.3.1 under Assumption 3.2(i)–(iii) with $\delta_{1n} = \sqrt{\frac{\log h_n^{-1}}{n h_n^{p+q}}} + h_n^s$. Lastly, conditionally on the data $\mathcal{D}_n$

$$\mathbb{E}_\varepsilon\left\|\hat{V}_n^\varepsilon - V_n^\varepsilon\right\|_\infty = \mathbb{E}_\varepsilon\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i(\hat{\varphi}(Z_i)-\varphi(Z_i))h_n^{-q/2}K_w\left(h_n^{-1}(W_i-.)\right)\right\|_\infty$$

is the expected value of the supremum of the Gaussian process indexed by the VC-type class. By Dudley–Sudakov's inequality, see Dudley (2016), and change of variables

$$\begin{aligned}
\mathbb{E}_\varepsilon\left\|\hat{V}_n^\varepsilon - V_n^\varepsilon\right\|_\infty &\leq 24\int_0^{\|H_n\|_{P_n,2}} \sqrt{\log N(\mathcal{H}_n, \|.\|_{P_n,2}, \epsilon)}\,\mathrm{d}\epsilon \\
&\leq 24\|H_n\|_{P_n,2}\int_0^1 \sqrt{v\,\log\left(\frac{A}{\epsilon}\right)}\,\mathrm{d}\epsilon,
\end{aligned}$$

where

$$\|H_n\|_{P_n,2}^2 = \frac{1}{n}\sum_{i=1}^n (\hat{\varphi}(Z_i)-\varphi(Z_i))^2 h_n^{-q}\|k\|_\infty^{2q}.$$

Therefore,

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left\|\hat{V}_n^\varepsilon - V_n^\varepsilon\right\|_\infty = \sup_{P\in\mathcal{P}} \mathbb{E}_P\left\|\hat{\varphi}-\varphi\right\|_\infty O\left(h_n^{-q/2}\right).$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Under Assumptions 3.1 and 3.2(i)–(iv), it follows from Babii and Florens (2018), Thm. 4.2, that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P \left\|\hat{\varphi}-\varphi\right\|_\infty = O\left(\frac{1}{\alpha_n}\left(\frac{1}{\sqrt{nh_n^q}}+h_n^s\right)+\frac{1}{\alpha_n^{1/2}}\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}}+\alpha_n^{\beta\wedge 1}\right).$$

Therefore, Assumption A.2.3(C2) is verified with

$$\delta_{2n} = O\left(\frac{1}{\alpha_n}\left(\frac{1}{n^{1/2}h_n^q}+h_n^{s-q/2}\right)+\frac{1}{\alpha_n^{1/2}}\frac{\log^{1/2} h_n^{-1}}{n^{1/2}h_n^{p/2+q}}+\alpha_n^{\beta\wedge 1}h_n^{-q/2}\right).$$

Combining all estimates, it follows from Theorem A.1 that if $\frac{\log^7 n}{nh_n^{2q}}\to 0$

$$\inf_{P\in\mathcal{P}} \Pr(\varphi\in C^*_{n,1-\gamma}) \geq 1-\gamma-O_p(\delta_n)-o_p(1)$$

with

$$\delta_n = \left(\alpha_n^{-1/2}\delta_{1n}+\alpha_n^{-1}\delta_{3n}+\alpha_n^{\beta\wedge 1}\right)u_n^{-1}\alpha_n n^{1/2}\log n+\delta_{2n}\log n.$$

The result then follows under Assumption 3.2(vi) after taking the probability limit.

For confidence sets based on the concentration inequality, we note that Assumptions A.2.5(C1) and (C3) hold with the same δ_{1n} and δ_{3n} sequences. Note also that $\|X_{ni}\|\leq F\|k\|^q \triangleq G$. It only remains to verify Assumption A.2.5(C2). To that end, note that

$$\begin{aligned}\mathbb{E}_P\left\|\hat{X}_{ni}-X_{ni}\right\|^2 &= \mathbb{E}_P\left\|\frac{1}{h_n^{q/2}}(\hat{\varphi}(Z_i)-\varphi(Z_i))K_w(h_n^{-1}(W_i-.))\right\|^2\\ &\leq \mathbb{E}_P\|\hat{\varphi}-\varphi\|_\infty^2\|K_w\|^2.\end{aligned}$$

Inspection of the proof of the uniform risk bound in Babii and Florens (2018) reveals that for all $r\geq 2$, we also have

$$\left(\sup_{P\in\mathcal{P}}\mathbb{E}_P\|\hat{\varphi}-\varphi\|_\infty^r\right)^{1/r} = O\left(\frac{1}{\alpha_n}\left(\frac{1}{\sqrt{nh_n^q}}+h_n^s\right)+\frac{1}{\alpha_n^{1/2}}\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}}+\alpha_n^{\beta\wedge 1}\right), \tag{A.5}$$

provided that

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

and

$$\sup_{P\in\mathcal{P}}\left(\|\hat{r}-\mathbb{E}\hat{r}\|^{2r}\right)^{\frac{1}{2r}} = O\left(\frac{1}{\sqrt{nh_n^q}}+h_n^s\right). \tag{A.6}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

The former statement is established in Proposition A.3.1, while for the latter we use the following inequality valid for any i.i.d. sequence $(\xi_{ni})_{i=1}^n$

$$\left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n \xi_{ni}\right\|^r\right)^{\frac{1}{r}} \leq C\left\{n^{-1/2}\left(\mathbb{E}\|\xi_{ni}\|^2\right)^{1/2} + n^{1/r-1}\left(\mathbb{E}\|\xi_{ni}\|^r\right)^{1/r}\right\}.$$

In particular, setting $\xi_{ni} = Y_i h_n^{-q} K_w(h_n^{-1}(W_i - w)) - \mathbb{E}\left[Y_i h_n^{-q} K_w(h_n^{-1}(W_i - w))\right]$, under Assumption 3.2(iv), we obtain equation (A.6) with $r = 2$. Therefore, Assumption A.2.5(C2) is verified with

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

and by Theorem A.3,

$$\inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_{n,1-\gamma}\right) \geq 1 - \gamma - O(\delta_n)$$

with

$$\delta_n = \delta_{2n}\log n + \left(\frac{\delta_{1n}}{\alpha_n^{1/2}} + \frac{\delta_{3n}}{\alpha_n} + \alpha_n^{\beta\wedge 1}\right) u_n^{-1}\alpha_n n^{1/2}\log n.$$

□

Proof of Theorem 3.2 For concentration-based confidence sets we apply Theorem A.4. In the proof of Theorem 3.1, we show that $\delta_{1n} = \sqrt{\frac{\log n}{nh_n^{p+q}}} + h_n^s$. Under Assumption 3.2(vi) $\delta_{1n}/\alpha_n^{1/2} \to 0$, whence since $u_n = h_n^{-q/2}$, by Theorem A.4

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{1}{\alpha_n\sqrt{nh_n^q}}\right).$$

Next we verify Assumption A.2.4. Note that (i) holds, since the class is described by the convolution kernel of bounded variation. For (ii), we note that

$$\mathbb{E}_P\left\|\hat{X}_{ni} - X_{ni}\right\|_\infty^2 \leq \mathbb{E}_P\|\hat{\varphi} - \varphi\|_\infty^2 \|K_w\|_\infty^2 h_n^{-q},$$

which tends to zero due to equation (A.5) and Assumption 3.2(vi). Lastly, (iii) follows since $\mathbb{E}_P\|X_{ni}\|_\infty^2 \leq F\|K_w\|_\infty^2 h_n^{-q}$. Therefore, we obtain the second statement by Theorem A.2. □

A.3.2: Functional Regressions

Proof of Theorem 4.1 We verify assumptions of Theorems A.1 and A.3. Under Assumption 4.1(ii), by Hoffman–Jørgensen's inequality, for example, see Giné and Nickl (2016), Thm. 3.1.15,

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$$\left(\mathbb{E}_P\left\|\hat{T}^*-T^*\right\|_{2,\infty}^2\right)^{1/2} \le \left(\mathbb{E}_P\left\|\frac{1}{n}\sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i]\right\|_\infty^2\right)^{1/2}$$
$$= 12\sqrt{3}\left(16\mathbb{E}_P\left\|\frac{1}{n}\sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i]\right\|_\infty + \frac{2FC}{n}\right).$$

Next, under Assumption 4.1(iii), by the bracketing moment inequality, for example, see Giné and Nickl (2016), Prop. 3.5.15 and 4.3.36,

$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\left\|\frac{1}{n}\sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i]\right\|_\infty = O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore, Assumptions A.2.3(C1) and A.2.5(C1) are satisfied with $\delta_{1n} = 1/\sqrt{n}$.

Since

$$\hat{r} - \hat{T}\varphi = \frac{1}{n}\sum_{i=1}^n U_i W_i,$$

Assumptions A.2.3(C3) and A.2.5(C3) are trivially satisfied with $X_{ni} = U_i W_i$, $u_n = 1$, and $\delta_{3n} = 0$.

Note also that by the Cauchy–Schwartz inequality

$$\mathbb{E}_P\left\|\hat{X}_{ni} - X_{ni}\right\|^2 = \mathbb{E}_P\left\|\frac{1}{n}\sum_{i=1}^n \langle\hat{\varphi}-\varphi, Z_i\rangle W_i\right\|^2$$
$$\le \mathbb{E}_P\|\hat{\varphi}-\varphi\|^2.$$

Under Assumptions 4.1(i) and (ii), it follows from Babii and Florens (2018) that

$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\|\hat{\varphi}-\varphi\|^2 = O\left(\frac{1}{\alpha_n n} + \alpha_n^{(2\beta+1)\wedge 2}\right), \tag{A.7}$$

whence Assumption A.2.5(C2) is verified with $\delta_{2n} = \frac{1}{\sqrt{\alpha_n n}} + \alpha_n^{(\beta+1/2)\wedge 1}$. Lastly, $\|X_{ni}\| \le 1 \triangleq G$. Then by Theorem A.3 for confidence sets based on the concentration inequality, we obtain

$$\inf_{P\in\mathcal{P}}\Pr\left(\varphi \in C_{n,1-\gamma}\right) \ge 1-\gamma - O\left(\left(\alpha_n^{\beta\wedge 1+1} n^{1/2} + \frac{1}{\sqrt{\alpha_n n}} + \alpha_n^{\beta/2\wedge 1}\right)\log n\right).$$

For bootstrap confidence sets, we need additionally to verify Assumptions A.2.1, A.2.2, and A.2.3(C2). The relevant class of functions is $\mathcal{G} = \left\{g_s(u,w) = uw(s) : s \in [0,1]^p\right\}$. This class is of VC-type under Assumption 4.1(iii), for example, see Giné and Nickl (2016), Prop. 3.6.12. Therefore Assumption A.2.1 is satisfied. Next under Assumption 4.1(iv)

$$\mathrm{Var}(g_s) = \mathbb{E}|UW(s)|^2$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

is uniformly bounded away from zero and infinity, which verifies Assumption A.2.2(i). Next, under Assumption 4.1(ii),

$$\|G\|_{P,4} = F$$
$$P|g_s|^3 = \mathbb{E}|UW(s)|^3 \leq F^3$$
$$P|g_s|^4 = \mathbb{E}|UW(s)|^4 \leq F^4.$$

Therefore, Assumption A.2.2(ii) is satisfied with universal constants b, σ. Therefore, $b^4 \log^7 n/n \to 0$.

Lastly, note that for some universal constant C depending only on $\|r\|_\infty$ and K

$$\mathbb{E}_P \left\| \hat{V}_n^\varepsilon - V_n^\varepsilon \right\|_\infty = \mathbb{E}_P \mathbb{E}_\varepsilon \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \langle \hat{\varphi} - \varphi, Z_i \rangle W_i \right\|_\infty \leq CF\mathbb{E}_P \|\hat{\varphi} - \varphi\|,$$

where the inequality follows by Dudley–Sudakov's bound. Therefore, Assumption A.2.3(C2) is verified with the same $\delta_{2n} = \frac{1}{\sqrt{\alpha_n n}} + \alpha_n^{\beta/2\wedge 1}$. Then, by Theorem A.1

$$\inf_{P\in\mathcal{P}} \Pr(\varphi \in C^*_{n,1-\gamma}) \geq 1 - \gamma - O_p\left(\left(\alpha_n^{\beta\wedge 1+1} n^{1/2} + \frac{1}{\sqrt{\alpha_n n}} + \alpha_n^{\beta/2\wedge 1}\right)\log n\right) - o_p(1)$$

and the result follows by taking the probability limit. □

Proof of Theorem 4.2 For concentration-based confidence sets, we apply Theorem A.4. In the proof of Theorem A.4, we show that $\delta_{1n} = n^{-1/2}$. Under Assumption 4.1(vi), $\delta_{1n}/\alpha_n^{1/2} \to 0$ and $\delta_{2n} \to 0$, whence since $u_n = 1$, by Theorem A.4

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P \left| C_{n,1-\gamma} \right|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}\right).$$

Next, we verify that Assumption A.2.4. (i) holds, under Assumption 4.1(iii). For (ii), under Assumption 4.1(ii)

$$\mathbb{E}_P \left\| \hat{X}_{ni} - X_{ni} \right\|_\infty^2 \leq CF\mathbb{E}_P \|\hat{\varphi} - \varphi\|^2,$$

which tends to zero due to equation (A.7) and Assumption 4.1(vi). Lastly, (iii) follows trivially under Assumption 4.1(ii). Therefore, we obtain the second statement by Theorem A.2. □

A.3.3: Density Deconvolution

Proof of Theorem 4.3 We verify assumptions of Theorems A.1 and A.3. Note that Assumptions A.2.3(C1) and A.2.5(C1) are trivially satisfied with $\delta_{1n} = 0$. Next,

$$\hat{r} - T\varphi = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{Y_i - y}{h_n}\right) - \frac{1}{h_n}\mathbb{E}\left[K\left(\frac{Y_i - y}{h_n}\right)\right] + R_{1n}$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

with

$$\|R_{1n}\|^2 = \int_0^1 \left| \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{Y_i - y}{h_n} \right) \right] - r(y) \right|^2 \mathrm{d}y.$$

Under Assumption 4.2(i)–(iv) by Tsybakov (2009), Prop. 1.5, we have

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \|R_{1n}\|^2 = O(h_n^{2s}).$$

Therefore, Assumptions A.2.3(C3) and A.2.5(C3) are verified with $X_{ni} = \frac{1}{h_n^{1/2}} K\left(\frac{Y_i - y}{h_n} \right) - \frac{1}{h_n^{1/2}} \mathbb{E}\left[K\left(\frac{Y_i - y}{h_n} \right) \right]$, $u_n = h_n^{-1/2}$, and $\delta_{3n} = h_n^s$.

Note also that

$$\mathbb{E}_P \left\| \hat{X}_{ni} - X_{ni} \right\|^2 = \mathbb{E}_P \left\| \frac{1}{n h_n^{1/2}} \sum_{i=1}^n K\left(\frac{Y_i - \cdot}{h_n} \right) - \frac{1}{h_n^{1/2}} \mathbb{E}\left[K\left(\frac{Y_1 - \cdot}{h_n} \right) \right] \right\|^2$$
$$\leq \frac{\|K\|^2}{n}.$$

Therefore, Assumption A.2.5(C2) is verified with $\delta_{2n} = n^{-1/2}$. Lastly, putting $[K_h * r](y) = h_n^{-1} \int K(h_n^{-1}(y-u)) r(u) \mathrm{d}u$

$$\|X_{ni}\|^2 = \left\| h_n^{-1/2} K\left(\frac{Y_i - \cdot}{h_n} \right) - h_n^{1/2} [K_h * r] \right\|^2$$
$$\leq \frac{2}{h_n} \int K^2\left(\frac{Y_i - y}{h_n} \right) \mathrm{d}y + 2 h_n \|K_h * r\|^2$$
$$\leq 4 \|K\|^2 \triangleq G^2,$$

where the last inequality follows by Young's inequality for convolutions. Then by Theorem A.3 for confidence sets based on the concentration inequality, we obtain

$$\inf_{P \in \mathcal{P}} \Pr\left(\varphi \in C_{n, 1-\gamma} \right) \geq 1 - \gamma - O\left(\left(h_n^s + \alpha_n^{\beta \wedge 1 + 1} \right) \sqrt{n h_n} \log n + \frac{\log n}{n^{1/2}} \right).$$

For bootstrap confidence sets, we also need to verify Assumptions A.2.1, A.2.2, and A.2.3(C2). The relevant class of functions is

$$\mathcal{G}_n = \left\{ \tilde{y} \mapsto \frac{1}{h_n^{1/2}} K\left(\frac{y - \tilde{y}}{h_n} \right) : \ y \in [0,1] \right\}.$$

This class admits the envelope $G(y) = h_n^{-1/2} \|K\|_\infty$ and is of VC-type under Assumption 4.2(vi), for example, see Giné and Nickl (2016), Prop. 3.6.12. Therefore Assumption A.2.1 is satisfied. For $g_n \in \mathcal{G}_n$,

$$\mathrm{Var}(g_n) = \mathbb{E}\left| \frac{1}{h_n^{1/2}} K\left(\frac{Y_1 - y}{h_n} \right) \right|^2 - \left(\mathbb{E}\left[\frac{1}{h_n^{1/2}} K\left(\frac{Y_1 - y}{h_n} \right) \right] \right)^2$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

so that for sufficiently large n

$$0 < \inf_{y\in[0,1]} r(y)\|K\|^2 - h_n\|K\|_1^2\|r\|_\infty^2 \le \mathrm{Var}(g_n) \le \|K\|^2\|r\|_\infty.$$

This verifies Assumption A.2.2(i). Next, under Assumption 4.2(iv),

$$\begin{aligned}
\|G\|_{P,4} &= h_n^{-1/2}\|K\|_\infty \\
P|g_n|^3 &= h_n^{-3/2}\mathbb{E}\left|K\left(h_n^{-1}(Y-y)\right)\right|^3 \\
&\le h_n^{-1/2}\|K\|_3^3\|r\|_\infty \\
P|g_n|^4 &= h_n^{-4/2}\mathbb{E}\left|K\left(h_n^{-1}(Y-y)\right)\right|^4 \\
&\le h_n^{-1}\|K\|_4^4\|r\|_\infty.
\end{aligned}$$

Therefore, Assumption A.2.2(ii) is satisfied with $b_n = O(h_n^{-1/2})$ and some universal constant σ. Note also that $b_n^4\log^7 n/n \to 0$ under Assumption 4.2.

Lastly, note that for some universal constant C depending only on $\|r\|_\infty$ and K

$$\begin{aligned}
\mathbb{E}_P\left\|\hat{V}_n^\varepsilon - V_n^\varepsilon\right\|_\infty &= \mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i\right| \mathbb{E}_P\left\|\frac{1}{nh_n^{1/2}}\sum_{i=1}^n K\left(\frac{Y_i - \cdot}{h_n}\right) - \frac{1}{h_n^{1/2}}\mathbb{E}\left[K\left(\frac{Y_1-\cdot}{h_n}\right)\right]\right\|_\infty \\
&\le C\sqrt{\frac{\log h_n^{-1}}{nh_n}},
\end{aligned}$$

where the last line follows by Giné and Guillou (2002), Thm. 2.1, under Assumption 4.2(i), (ii), (iv), and (vi). This verifies Assumption A.2.3(C2) with $\delta_{2n} = \sqrt{\frac{\log h_n^{-1}}{nh_n}}$. Therefore, by Theorem A.1,

$$\begin{aligned}
&\inf_{P\in\mathcal{P}} \Pr(\varphi \in C^*_{n,1-\gamma}) \ge 1 - \gamma - O_p \\
&\quad\times\left(\left(h_n^s + \alpha_n^{\beta\wedge 1+1}\right)\sqrt{nh_n}\log n + \sqrt{\frac{\log h_n^{-1}}{nh_n}}\log n\right) - o_p(1)
\end{aligned}$$

and the result follows by taking the probability limit. □

Proof of Theorem 4.4 For concentration-based confidence sets, we apply Theorem A.4. In the proof of Theorem A.4, we show that $\delta_{1n} = 0$. Therefore, by Theorem A.4 with $u_n = h_n^{-1/2}$

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left|C_{n,1-\gamma}\right|_\infty = O\left(\frac{1}{\alpha_n\sqrt{nh_n^q}}\right).$$

Next, we verify Assumption A.2.4. Note that (i) holds, since the class is described by the translation of the kernel function of bounded variation. For (ii), we note that by Hoffman–Jørgensen's inequality

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$$\mathbb{E}_P \left\| \hat{X}_{ni} - X_{ni} \right\|_\infty^2$$
$$\leq \mathbb{E}_P \left\| \frac{1}{nh_n^{1/2}} \sum_{i=1}^n K\left(\frac{Y_i - y}{h_n}\right) - \frac{1}{h_n^{1/2}} \mathbb{E}\left[K\left(\frac{Y_i - y}{h_n}\right)\right] \right\|_\infty^2$$
$$\leq 432 \left(\mathbb{E}_P \left\| \frac{1}{nh_n^{1/2}} \sum_{i=1}^n K\left(\frac{Y_i - y}{h_n}\right) - \frac{1}{h_n^{1/2}} \mathbb{E}\left[K\left(\frac{Y_i - y}{h_n}\right)\right] \right\|_\infty + \frac{2}{nh_n^{1/2}} \right)^2.$$

Therefore,

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left\| \hat{X}_{ni} - X_{ni} \right\|_\infty^2 = O\left(\frac{\log h_n^{-1}}{n} + h_n^{2s} + \frac{1}{n^2 h_n} \right),$$

which tends to zero under Assumption 4.2(vi). Lastly, (iii) follows since $\mathbb{E}_P \|X_{ni}\|_\infty^2 \leq 2\|K\|_\infty^2 h_n^{-1}$. Therefore, we obtain the second statement by Theorem A.2. □

A.4: TECHNICAL LEMMAS

We first characterize the bias of the regularized estimator uniformly over our class of models.

Proof of Proposition 3.1. By Lemma A.4.1, $(\alpha_n I + T^*T)$ is an 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 3.1 for any $(\varphi, T) \in \mathcal{F}$

$$\begin{aligned} \left\| (\alpha_n I + T^*T)^{-1} T^* r - \varphi \right\|_\infty &= \left\| \left[(\alpha_n I + T^*T)^{-1} T^*T - I \right] \varphi \right\|_\infty \\ &= \left\| \alpha_n (\alpha_n I + T^*T)^{-1} \varphi \right\|_\infty \\ &= \left\| \alpha_n (\alpha_n I + T^*T)^{-1} (T^*T)^\beta T^* \psi \right\|_\infty \\ &= \left\| T^* \alpha_n (\alpha_n I + TT^*)^{-1} (TT^*)^\beta \psi \right\|_\infty \\ &\leq \|T^*\|_{2,\infty} \left\| \alpha_n (\alpha_n I + TT^*)^{-1} (TT^*)^\beta \right\| \|\psi\| \end{aligned} \tag{A.8}$$

Note 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,

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

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

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 $R = C^2\tilde{R}$. □

We also need the following inequality known in the theory of numerical ill-posed inverse problems.

LEMMA A.4.1. *Suppose that $T : L_2[a, b]^p \to L_2[a, b]^q$ is an integral operator with a 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, see Nair (2009), Lem. 4.1, and $(C, \|.\|_{\infty}) \subset L_2$, it is also invertible as an operator between $(C, \|.\|_{\infty})$.

Note 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,

$$\left\| (\alpha_n I + T^*T)^{-1} \right\|_{\infty} = \sup_{\|\varphi\|_{\infty}=1} \left\| (\alpha_n I + T^*T)^{-1}\varphi \right\|_{\infty} \leq \frac{\|(\alpha_n I + T^*T)^{-1}T^*T\|_{\infty} + 1}{\alpha_n}. \tag{A.9}$$

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

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

Combining this with equation (A.9) gives the result. □

For the next result, put

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

LEMMA A.4.2. *Suppose that T and $\hat{T}$ are integral operators in $\mathcal{L}_2$ with continuous kernel functions mapping from $L_2[a, b]^p$ to a subset of $L_2[a, b]^q$. Then under Assumption 3.1, there exists a constant $C < \infty$ that does not depend on (φ, T) such that*

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

$$\|\xi_n\|_\infty \le \frac{C}{\alpha_n^{1/2}}\left(\|\hat{T}^* - T^*\|_{2,\infty} + \|\hat{T}^* - T^*\|_{2,\infty}^2\right).$$

Proof. Decompose

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

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

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

By isometry of functional calculus

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

Under Assumption 3.1, the last term is

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

Combining these findings with the following bound

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

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

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

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

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

By Lemma A.4.1,

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

Under Assumption 3.1,

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

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

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

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

and the conclusion follows from collecting all estimates and by triangle inequality. □

For the following lemma denote

$$\nu_n = \frac{u_n}{n}\sum_{i=1}^n (\alpha_n I + T^*T)^{-1}T^*X_{ni},$$
$$\hat{\nu}_n = \frac{u_n}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*X_{ni},$$
$$\nu_n^\eta = \frac{u_n}{n}\sum_{i=1}^n (\alpha_n I + T^*T)^{-1}T^*\eta_i X_{ni},$$
$$\hat{\nu}_n^\eta = \frac{u_n}{n}\sum_{i=1}^n (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\eta_i \hat{X}_{ni}.$$

LEMMA A.4.3. *For any $T, \hat{T} \in \mathcal{L}_2$, we have*

$$\left\|\hat{\nu}_n - \nu_n\right\|_\infty \le \left(\frac{1}{\alpha_n} + \frac{1}{2\alpha_n^{3/2}}\|T^*\|_{2,\infty}\right)\|\hat{T}^* - T^*\|_{2,\infty}\left\|\frac{u_n}{n}\sum_{i=1}^n X_{ni}\right\|,$$
$$\left\|\hat{\nu}_n^\eta - \nu_n^\eta\right\|_\infty \le \left(\frac{1}{\alpha_n} + \frac{1}{2\alpha_n^{3/2}}\|T^*\|_{2,\infty}\right)\|\hat{T}^* - T^*\|_{2,\infty}\left\|\frac{u_n}{n}\sum_{i=1}^n \eta_i X_{ni}\right\|$$
$$+ \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\left\|\frac{u_n}{n}\sum_{i=1}^n \eta_i(\hat{X}_{ni} - X_{ni})\right\|.$$

Proof. Decompose

$$\hat{\nu}_n - \nu_n$$
$$= \left[\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha_n I + TT^*)^{-1}\right]\frac{u_n}{n}\sum_{i=1}^n X_{ni}$$
$$= \left[(\hat{T}^* - T^*)(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1}(\hat{T}\hat{T}^* - TT^*)(\alpha_n I + TT^*)^{-1}\right]$$
$$\times \frac{u_n}{n}\sum_{i=1}^n X_{ni}$$
$$= \left[(\hat{T}^* - T^*)(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\hat{T}(\hat{T}^* - T^*)(\alpha_n I + TT^*)^{-1}\right]$$
$$\times \frac{u_n}{n}\sum_{i=1}^n X_{ni} - T^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1}(\hat{T} - T)T^*(\alpha_n I + TT^*)^{-1}\frac{u_n}{n}\sum_{i=1}^n X_{ni}.$$

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press

Note that

$$\|\hat{T} - T\| = \|\hat{T}^* - T^*\| \leq \|\hat{T}^* - T^*\|_{2,\infty}$$

and

$$\|(\alpha_n I + \hat{T}\hat{T}^*)^{-1}\hat{T}\| \leq \frac{1}{2\alpha_n^{1/2}}, \qquad \|T^*(\alpha_n I + TT^*)^{-1}\| \leq \frac{1}{2\alpha_n^{1/2}}.$$

Then

$$\|\hat{\nu}_n - \nu_n\|_\infty \leq \left(\frac{1}{\alpha_n} + \frac{1}{2\alpha_n^{3/2}}\|T^*\|_{2,\infty}\right)\|\hat{T}^* - T^*\|_{2,\infty}\left\|\frac{u_n}{n}\sum_{i=1}^n X_{ni}\right\|.$$

For the second statement note that

$$\begin{aligned}\hat{\nu}_n^\eta - \nu_n^\eta &= \left[\hat{T}^*(\alpha_n I + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha_n I + TT^*)^{-1}\right]\frac{u_n}{n}\sum_{i=1}^n \eta_i X_{ni} \\ &\quad + (\alpha_n I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\frac{u_n}{n}\sum_{i=1}^n \eta_i(\hat{X}_{ni} - X_{ni}) \\ &\triangleq I_n + II_n.\end{aligned}$$

Note that the bound on I_n follows trivially from above computations if we replace X_{ni} by $\eta_i X_{ni}$, while

$$\|II_n\|_\infty \leq \frac{\|\hat{T}^*\|_{2,\infty}}{\alpha_n}\left\|\frac{u_n}{n}\sum_{i=1}^n \eta_i(\hat{X}_{ni} - X_{ni})\right\|.$$

□

LEMMA A.4.4. *Suppose that $T^* \in \mathcal{L}_{2,\infty}$ is an integral operator with continuous kernel function k, then*

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

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

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

□

https://doi.org/10.1017/S0266466619000380 Published online by Cambridge University Press