Econometric Theory, **41**, 2025, 1284–1321.
doi:10.1017/S0266466625000039

IS COMPLETENESS NECESSARY? ESTIMATION IN NONIDENTIFIED LINEAR MODELS

ANDRII BABII
Economics Department, UNC Chapel Hill

JEAN-PIERRE FLORENS
Toulouse School of Economics

Modern data analysis depends increasingly on estimating models via flexible high-dimensional or nonparametric machine learning methods, where the identification of structural parameters is often challenging and untestable. In linear settings, this identification hinges on the completeness condition, which requires the nonsingularity of a high-dimensional matrix or operator and may fail for finite samples or even at the population level. Regularized estimators provide a solution by enabling consistent estimation of structural or average structural functions, sometimes even under identification failure. We show that the asymptotic distribution in these cases can be nonstandard. We develop a comprehensive theory of regularized estimators, which include methods such as high-dimensional ridge regularization, gradient descent, and principal component analysis (PCA). The results are illustrated for high-dimensional and nonparametric instrumental variable regressions and are supported through simulation experiments.

1. INTRODUCTION

Structural nonparametric and high-dimensional models are often ill-posed. Notable examples include nonparametric instrumental variable (IV) regression, various high-dimensional regressions, measurement error models, and random coefficient models. These cases commonly lead to an ill-posed functional equation

$$K\varphi = r,$$

where φ is the structural parameter of interest, r is a known function, and K is a linear operator. The classical literature on numerical ill-posed inverse problems

We are grateful to the Editor, Co-Editor, and three anonymous referees for helpful comments. We are also grateful to the participants of the Duke workshop, TSE Econometrics seminar, Triangle Econometrics Conference, 4th ISNPS Conference, 2018 NASMES Conference, and Bristol Econometric Study Group. Jean-Pierre Florens acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissement d'Avenir, grant ANR-17-EURE-0010) and grant ANR-24-CE26-3681-01. All remaining errors are ours. Address correspondence to Andrii Babii, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, e-mail: babii.andrii@gmail.com.

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

(see Engl, Hanke, and Neubauer, 1996) addresses *deterministic* problems, where K is known and where r is observed with a deterministic numerical error. However, in econometric applications, both K and r must be estimated from data, posing a statistically ill-posed inverse problem.

Identification has long been a central concept in econometrics, as demonstrated by foundational works such as Koopmans (1949), Koopmans and Reiersol (1950), and Rothenberg (1971) for parametric models. In nonparametric and high-dimensional settings, K and r are typically derived from the data-generating process, with φ uniquely identified if $K\varphi = r$ has a unique solution. Uniqueness is achieved when K is a one-to-one (or nonsingular) operator, meaning that $K\phi = 0 \implies \phi = 0$ for all ϕ in the domain of K. However, in econometric applications, K is often unknown, and the estimated operator $\hat{K}$, being finite-rank, lacks one-to-one properties in finite samples.

In cases of nonidentification, the maximum likelihood estimator can yield a flat likelihood in some parameter space regions, leading to an ambiguously defined maximum. It is thus important to characterize the asymptotic behavior of estimators in potentially nonidentified models. In ill-posed models without identification, the identified set is a linear manifold, $\phi + \mathcal{N}(K)$, where ϕ solves $K\phi = r$ and where $\mathcal{N}(K)$ is the null space of K. This identified set, which is generally unbounded, lacks informativeness on φ without additional constraints.

Given that K or $\hat{K}$ may fail to be one-to-one and often has a discontinuous generalized inverse, regularization is essential for estimating the structural parameter consistently. In this article, we focus on spectral regularization methods that modify the spectrum of $\hat{K}$. A prominent example is Tikhonov regularization (functional ridge regression) (see Tikhonov, 1963). Other methods include iterated Tikhonov, spectral cutoff (or principal component analysis, see Yao, Müller, and Wang, 2005), and Landweber iteration (or functional gradient descent).[1] We show that estimators based on spectral regularization are uniformly consistent for the best approximation of the structural parameter in the orthogonal complement of the null space of K, and we explore the distributional implications of identification failures.

In some cases, the best approximation may coincide with, or closely approximate, the structural parameter even if K is singular and has a nontrivial null space. This provides an insightful interpretation of nonparametric IV regression under identification failure, akin to the least-squares best approximation under misspecification (see Angrist and Pischke, 2008, Chap. 3).[2] Moreover, the best approximation can also be used to construct set estimators in partial identification frameworks.

[1] Functional gradient descent is also foundational in boosting estimators (see Friedman, 2001).

[2] In contrast, the parametric 2SLS estimator lacks this best approximation property (see Escanciano and Li, 2021) for an example of an IV estimator with this interpretation.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Contribution and related literature.

Our paper contributes to several research areas. First, it advances the literature on regularization in statistical inverse problems, as discussed in Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2012), Chen and Christensen (2018), Carrasco, Florens, and Renault (2007, 2014), Darolles et al. (2011), Florens, Johannes, and Van Bellegem (2011), Gagliardini and Scaillet (2012), Hall and Horowitz (2005), Babii (2020, 2022), and other studies. Canay, Santos, and Shaikh (2013) demonstrate that the completeness condition is untestable in these models (see also D'Haultfoeuille, 2011, Hu and Shiu, 2018, and Andrews, 2017 for more on this condition). Our contribution is the derivation of convergence rates in Hilbert space and L_∞ norms for estimators based on generic spectral regularization schemes, including Tikhonov, iterated Tikhonov, Landweber iteration, and spectral cutoff. These results, which do not require the completeness condition, are illustrated for nonparametric IV and high-dimensional regressions.[3]

Second, we add to the literature on identifying linear functionals of a structural function φ. For instance, Severini and Tripathi (2006, 2012) show that linear functionals of φ can be identified under conditions weaker than the completeness condition and derive the corresponding efficiency bounds when $\sqrt{n}$ estimation is possible. Santos (2011) establishes the asymptotic normality of linear functionals of nonparametric IV regression without the completeness condition, whereas Escanciano and Li (2021) provides necessary and sufficient conditions for the identification and $\sqrt{n}$-estimability of the best linear approximation to the nonparametric IV structural function. We contribute by showing that the asymptotic distribution of linear functionals of φ may include a chi-square component when the completeness condition fails and the functional is estimable at a rate slower than $\sqrt{n}$.

Third, our work connects to the partial identification literature in nonparametric IV regression (see Santos, 2012, Freyberger, 2017, and Chernozhukov, Newey, and Santos, 2023). Our findings on consistency for the best approximation in the L_∞ norm, alongside Freyberger (2017), have applications in partial identification.

The remainder of this article is structured as follows. Section 2 introduces the concept of identification in linear ill-posed models. Section 3 derives nonasymptotic risk bounds in Hilbert space and L_∞ norms for Tikhonov-regularized estimators, generalizing these results to other spectral regularization schemes in Appendix Section A.1. Section 4 illustrates the transition between Gaussian and weighted chi-square asymptotics under identification failure. Section 5 presents a Monte Carlo study, while Section 6 concludes the article. All proofs are presented in Appendix A.3. Appendix A.4 discusses partial identification, and Appendix A.4 examines Tikhonov-regularized estimators under extreme identification failures.

[3]This generalizes Florens et al. (2011), which covers only L_2 rates for Tikhonov regularization under certain high-level conditions.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Finally, Sections S.1 and S.2 of the Supplementary Material, review relevant results from generalized inverse theory and Hilbert space-valued U-statistics.

2. IDENTIFICATION

Consider the functional linear equation

$$K\varphi = r,$$

where $K : \mathcal{E} \to \mathcal{H}$ is a compact linear operator between Hilbert spaces $(\mathcal{E}, \langle .,. \rangle_{\mathcal{E}})$ and $(\mathcal{H}, \langle .,. \rangle_{\mathcal{H}})$ and where $\varphi \in \mathcal{E}$ represents the structural parameter of interest. For φ to be point identified, K must be one-to-one (or nonsingular), meaning that the null space of K, denoted by $\mathcal{N}(K) = \{\phi \in \mathcal{E} : K\phi = 0\}$, is reduced to $\{0\}$. Thus, for any $\phi \in \mathcal{E}$,

$$K\phi = 0 \implies \phi = 0.$$

When the completeness condition fails, the identified set becomes a closed linear manifold, which is denoted by $I_0 = \varphi + \mathcal{N}(K)$. Since $\mathcal{N}(K)$ forms a closed linear subspace in $\mathcal{E}$, we can decompose $\varphi = \varphi_1 + \varphi_0$, where φ_0 is the unique orthogonal projection of φ on $\mathcal{N}(K)$ and where φ_1 is the projection of φ on its orthogonal complement, denoted by $\mathcal{N}(K)^{\perp}$. We let $K^* : \mathcal{H} \to \mathcal{E}$ denote the Hilbert adjoint operator of $K : \mathcal{E} \to \mathcal{H}$, defined by the relation[4]

$$\langle \psi, K\phi \rangle = \langle K^*\psi, \phi \rangle, \qquad \forall (\psi, \phi) \in \mathcal{H} \times \mathcal{E}.$$

Since $\mathcal{N}(K)^{\perp} = \overline{\mathcal{R}(K^*)}$ (see Luenberger (1997, p. 157)), the best approximation φ_1 equals the structural parameter φ whenever $\varphi \in \overline{\mathcal{R}(K^*)}$, which is the closure of the range of the adjoint operator K. Note that $\mathcal{R}(K^*) = \mathcal{R}(K^*K)^{1/2}$, as shown in Engl et al. (1996, Prop. 2.18). Therefore, the identification of φ can be related to a regularity condition known as the source condition, where $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$ with $\beta \geq 1$ (see Carrasco et al., 2007). In other words, additional regularity of φ can compensate for the failure of the completeness condition.[5] The following two examples further illustrate this point (see also Bontemps, Florens, and Meddahi, 2025 for further applications to ecological inference).

Example 2.1 (Nonparametric IV regression). Consider

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

where $(Y, Z, W) \in \mathbf{R} \times \mathbf{R}^p \times \mathbf{R}^q$. The exclusion restriction leads to

$$r(w) \triangleq \mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w] \triangleq (K\varphi)(w),$$

[4] To reduce visual clutter, we omit subscripts and use $\langle .,. \rangle$ and $\|.\|$ to denote the inner products and norms in Hilbert spaces. For example, it is clear that the inner product on the left side is $\langle .,. \rangle_{\mathcal{H}}$, whereas the inner product on the right side is $\langle .,. \rangle_{\mathcal{E}}$.

[5] A similar observation holds for high-dimensional sparse regressions, where sparse coefficients can be identified under the restricted eigenvalue condition (see Tsybakov, Bickel, and Ritov, 2009).

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

where $K : L_2(Z) \to L_2(W)$ is a conditional expectation operator, and for a random vector X, we define $L_2(X) = \{\phi : \mathbb{E}|\phi(X)|^2 < \infty\}$. If K is compact, then according to the spectral theorem, there exists a sequence $(\sigma_j, e_j, h_j)_{j\geq 1}$, where $\sigma_j \to 0$ are singular values, $(e_j)_{j\geq 1}$ is a complete orthonormal system of $\mathcal{N}(K)^\perp = \overline{\mathcal{R}(K^*)}$, and $(h_j)_{j\geq 1}$ is the complete orthonormal system of $\mathcal{N}^\perp(K^*) = \overline{\mathcal{R}(K)}$ (see Kress (2014, Thm. 15.16)). φ_1 is the same as φ whenever it can be represented in terms of $(e_j)_{j\geq 1}$.

Example 2.2 (Nonparametric IV with a discrete instrumental variable). Following the previous example, we suppose that the instrumental variable is discrete, namely, $W \in \{w_k : k \geq 1\}$. Substituting $f_k(z) = f_{Z|W=w_k}(z)$ for every $k \geq 1$ yields

$$\mathcal{N}(K) = \{\phi \in L_2(Z) : \langle \phi, f_k \rangle = 0, \ \forall k \geq 1\},$$

where $\langle \phi, f_k \rangle = \mathbb{E}[\phi(Z) f_k(Z)]$. If $\varphi \in \operatorname{span}\{f_k : k \geq 1\}$, then $\varphi \in \mathcal{N}(K)^\perp$. φ_1 matches φ whenever it can be expressed in terms of $(f_k)_{k\geq 1}$.

3. NONASYMPTOTIC RISK BOUNDS

In this section, we derive risk bounds for the Tikhonov estimator, which solves

$$\min_{\phi} \left\| \hat{K}\phi - \hat{r} \right\|^2 + \alpha_n \|\phi\|^2,$$

where $\alpha_n > 0$ is a regularization parameter. The solution to this problem is

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

which is straightforward to verify. Note that the estimator $\hat{\varphi}$ is well defined even if K or $\hat{K}$ is not one-to-one.

3.1. Hilbert Space Risk

We first describe the relevant class of structural functions and operators:

Assumption 3.1. The structural parameter $\varphi = \varphi_1 + \varphi_0$ and the operator K belong to

$$\mathcal{F} = \mathcal{F}(\beta, C) = \left\{(\varphi, K) : \ \varphi_1 = (K^*K)^{\beta/2}\psi, \ \|\psi\|^2 \vee \|\varphi_0\| \vee \|K\| \leq C\right\}$$

for some $\beta, C > 0$, where $\|K\| = \sup_{\|\phi\|\leq 1} \|K\phi\|$ and $a \vee b = \max\{a, b\}$.

To illustrate this assumption, we let $(\sigma_j, e_j, h_j)_{j\geq 1}$ be the spectral decomposition of $K : \mathcal{E} \to \mathcal{H}$; see Kress (2014, Thm. 15.16). Then, $\varphi_1 = \sum_{j\geq 1} \langle \varphi_1, e_j \rangle e_j$, and by Parseval's identity, since $\psi = (K^*K)^{-\beta/2}\varphi_1$, we have

$$\|\psi\|^2 = \sum_{j\geq 1} \frac{|\langle \varphi_1, e_j \rangle|^2}{\sigma_j^{2\beta}}.$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Assumption 3.1 restricts the relative rates of decline of singular values $(\sigma_j)_{j\geq 1}$, which describe the ill-posedness, and Fourier coefficients $\langle \varphi_1, e_j\rangle_{j\geq 1}$, which describe the regularity of φ_1.

We estimate (r, K) with $(\hat{r}, \hat{K})$ and impose the following assumption.

Assumption 3.2. (i) $\mathbb{E}\|\hat{r} - \hat{K}\varphi_1\|^2 \leq C_1\delta_n$ and (ii) $\mathbb{E}\|\hat{K} - K\|^2 \leq C_2\rho_{1n}$, where the constants $C_1, C_2 < \infty$ do not depend on (φ, K) and $\delta_n, \rho_{1n} \to 0$ as $n \to \infty$.

Assumption 3.2 restricts the convergence rate of the residuals and the estimated operator. Note that Florens et al. (2011, Thm. 2.2), imposes a stronger high-level condition:

$$\mathbb{E}\|\hat{K} - K\|^4 \leq C\left(\mathbb{E}\|\hat{K} - K\|^2\right)^2$$

and certain assumptions on the tuning parameters that have not been verified for specific models.

The following result applies to the Hilbert space norm, $\|.\|$:

THEOREM 3.1. *Suppose that Assumptions 3.1 and 3.2 hold. Then, for all* $\beta \in (0, 2]$,

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi} - \varphi_1\right\|^2 = O\left(\frac{\delta_n + \rho_{1n}\alpha_n^{\beta\wedge 1}}{\alpha_n} + \alpha_n^{\beta}\right),$$

where $a \wedge b = \min\{a, b\}$.

Theorem 3.1 shows that the convergence rate in the Hilbert space norm is driven by the rates of

- the residuals, $O(\delta_n/\alpha_n)$;
- the estimated operator, $O\left(\rho_{1n}\alpha_n^{\beta\wedge 1}/\alpha_n\right)$;
- the regularization bias, $O\left(\alpha_n^{\beta}\right)$.

Notably, a stronger version of Assumption 3.1 is commonly assumed: $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$ with $\beta \geq 1$ (see Carrasco et al., 2007). In this case, we actually have $\varphi_0 = 0$, and the Tikhonov regularized estimator can consistently estimate the structural parameter φ, despite the failure of the completeness condition. More generally, we distinguish the following possibilities:

- the identified case: $\varphi_0 = 0$ and $\rho_{1n}\alpha_n^{\beta\wedge 1} \lesssim \delta_n$, where the rate of convergence to φ is driven by the residuals;
- the weakly identified case: $\varphi_0 = 0$ and $\delta_n \lesssim \rho_{1n}\alpha_n^{\beta\wedge 1}$, where the rate of convergence to φ is driven by the estimated operator;
- the nonidentified case: $\varphi_0 \neq 0$ and $\rho_{1n}\alpha_n^{\beta\wedge 1} \lesssim \delta_n$, where the rate of convergence to φ_1 is driven by the residuals;

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

- strongly nonidentified models: $\varphi_0 \neq 0$ and $\delta_n \lesssim \rho_{1n}\alpha_n^{\beta\wedge 1}$, where the rate of convergence to φ_1 is driven by the estimated operator.

In the identified case, the optimal choice of the regularization parameter is $\alpha_n \sim \delta_n^{1/(\beta+1)}$, which yields a convergence rate of $O\left(\delta_n^{\beta/(\beta+1)}\right)$ for $\beta \in (0,2]$. For $\beta > 2$, we demonstrate in Appendix Section A.1 that the optimal rate can be achieved with additional Tikhonov iterations or alternative regularization schemes.

Finally, it is worth mentioning that in the special case of L_2 spaces, Theorem 3.1 provides a sharper result than does Florens et al. (2011, Thm. 2.2), where the rate is always driven by ρ_{1n}.

3.2. L_∞ Risk

We consider the space of continuous functions on a compact set $D \subset \mathbf{R}^p$, denoted by $(C(D), \|.\|_\infty)$, where $\|.\|_\infty$ is the uniform norm. We suppose that this space is embedded into $\mathcal{E}$ and that $\mathcal{R}(K^*) \subset C(D)$ and that $\varphi_1 \in C(D)$. The mixed operator norm of K^* is defined as $\|K^*\|_{2,\infty} = \sup_{\|\phi\|\leq 1} \|K^*\phi\|_\infty$. The following assumption captures the approximation quality of the operator K^* by its estimate $\hat{K}^*$ in the $\|.\|_{2,\infty}$ norm.

Assumption 3.3. We suppose that $\|K^*\|_{2,\infty} \leq C_3$ and $\mathbb{E}\|\hat{K}^* - K^*\|_{2,\infty}^2 \leq C_3\rho_{2n}$, where $C_3 < \infty$ is independent of (φ, K) and $\rho_{2n} \to 0$.

The following result establishes a uniform convergence rate:

THEOREM 3.2. *Under Assumptions 3.1, 3.2, and 3.3, for all* $\beta \in (0,2]$,

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi} - \varphi_1\right\|_\infty = O\left(\frac{\delta_n^{1/2} + \rho_{1n}^{1/2}\alpha_n^{\frac{\beta+1}{2}\wedge 1} + \rho_{2n}^{1/2}\alpha_n^{1/2}}{\alpha_n} + \alpha_n^{\beta/2}\right).$$

Theorem 3.2 provides a uniform convergence rate for generic ill-posed inverse problems. Unlike Theorem 3.1, the L_∞ convergence rate is affected by the estimation error in K^*, assessed in the $\|.\|_{2,\infty}$ norm. For further details on the L_∞ convergence rates of more general regularized estimators, see Appendix Section A.1.

3.3. Applications

3.3.1. *High-dimensional regressions.* We consider the model

$$Y = \langle Z, \varphi\rangle + U, \qquad \mathbb{E}[UW] = 0,$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

where $(Y,Z,W) \in \mathbf{R} \times \mathcal{E} \times \mathcal{H}$ (see Florens and Van Bellegem, 2015).[6] The exclusion restriction leads to the functional linear equation

$$r \triangleq \mathbb{E}[YW] = \mathbb{E}[W \otimes Z]\varphi \triangleq K\varphi,$$

where $K : \mathcal{E} \to \mathcal{H}$ is a covariance operator and where $\otimes$ denotes the tensor product. An econometrician observes an i.i.d. sample $(Y_i, Z_i, W_i)_{i=1}^n$.[7] Then,

$$r = \mathbb{E}[YW], \qquad K = \mathbb{E}[W \otimes Z], \qquad K^* = \mathbb{E}[Z \otimes W]$$

are estimated as

$$\hat{r} = \frac{1}{n}\sum_{i=1}^n Y_i W_i, \qquad \hat{K} = \frac{1}{n}\sum_{i=1}^n W_i \otimes Z_i, \qquad \hat{K}^* = \frac{1}{n}\sum_{i=1}^n Z_i \otimes W_i.$$

It can be shown that

$$\mathbb{E}\left\|\hat{r} - \hat{K}\varphi_1\right\|^2 = \frac{\mathbb{E}\|(Y - \langle Z, \varphi_1\rangle)W\|^2}{n} \qquad \text{and} \qquad \mathbb{E}\left\|\hat{K} - K\right\|^2 \le \frac{\mathbb{E}\|ZW\|^2}{n}.$$

We let $\mathcal{F}(\beta, C)$ represent the class of models satisfying Assumption 3.1 and assume that $\mathbb{E}\|UW\|^2 \vee \mathbb{E}\|ZW\|^2 \le C$ for all the models in this class. Then, $\delta_n = \rho_{1n} = n^{-1}$, and by Theorem 3.1, the Hilbert space risk is of order

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi} - \varphi_1\right\|^2 = O\left(\frac{1}{\alpha_n n} + \alpha_n^\beta\right).$$

For high-dimensional regression, the model is either identified ($\varphi_0 = 0$) or nonidentified ($\varphi_0 \neq 0$). The conditions $\alpha_n \to 0$ and $\alpha_n n \to \infty$ as $n \to \infty$ are sufficient for consistency in the Hilbert space norm. The optimal choice $\alpha_n \sim n^{-1/(\beta+1)}$ yields a convergence rate of order $O(n^{-\beta/(\beta+1)})$, which is optimal when $W = Z$ (see Babii, Carrasco, and Tsafack 2024, Thm. 2.2).

For uniform convergence, we suppose that $\mathcal{E} = L_2(D)$, where $D \subset \mathbf{R}^p$ is bounded, and that the functions in $L_2(D)$ are square integrable with respect to the Lebesgue measure. To satisfy Assumption 3.3, we also assume that the models in $\mathcal{F}(\beta, C)$ are such that $\|Z\|_\infty \vee \|W\|_\infty \le C < \infty$ and that stochastic processes Z and W belong to a Hölder ball with smoothness $s > d/2$. Using the Hoffman–Jørgensen and moment inequalities (see Giné and Nickl (2016, p. 129 and p. 202)),

$$\mathbb{E}\left\|\hat{K}^* - K^*\right\|_{2,\infty}^2 = O\left(\frac{1}{n}\right).$$

[6] For example, when $W = Z$ and $\mathcal{E} = L_2$ with a counting measure, this yields a high-dimensional regression model for nonsparse data

$$Y = \sum_{j\ge 1} \varphi_j Z_j + U, \qquad \mathbb{E}[UZ_j] = 0, \qquad \forall j \ge 1.$$

When $\mathcal{E} = L_2$ with the Lebesgue measure, this model is also referred to as functional regression (see Florens and Van Bellegem, 2015 and Babii, 2022).

[7] This assumption can be relaxed to covariance stationarity and absolute summability of autocovariances (see Babii, 2022).

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Therefore, Assumption 3.3 holds with $\rho_{2n} = n^{-1}$, and by Theorem 3.2,

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi}-\varphi_1\right\|_\infty = O\left(\frac{1}{\alpha_n n^{1/2}}+\alpha_n^{\beta/2}\right).$$

The conditions $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ as $n \to \infty$ ensure the uniform consistency of $\hat{\varphi}$. The optimal choice $\alpha_n \sim n^{-1/(\beta+2)}$ leads to a convergence rate of order $O(n^{-\beta/(2\beta+4)})$. Determining the minimax-optimal convergence rate for the class $\mathcal{F}(\beta, C)$ remains an open problem.

3.3.2. *Nonparametric IV regression.* Following Example 2.1, we can reformulate the model as

$$r(w) \triangleq \mathbb{E}[Y|W=w]f_W(w) = \int \varphi(z) f_{ZW}(z,w)\mathrm{d}z \triangleq (K\varphi)(w),$$

where $K : L_2([0,1]^p) \to L_2([0,1]^q)$. To estimate r and K, we use kernel smoothing:

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

$$(\hat{K}\phi)(w) = \int \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z,$$

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

where K_w and K_z are symmetric kernel functions and where $h_n \to 0$ is a bandwidth parameter. Under mild assumptions, Proposition A.2.1 shows that $\delta_n = \frac{1}{nh_n^q} + h_n^{2s}$ and $\rho_{1n} = \frac{1}{nh_n^{p+q}} + h_n^{2s}$, where s denotes the Hölder smoothness of f_{ZW}. Thus, by Theorem 3.1, the mean-integrated squared error has the following rate:

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\left\|\hat{\varphi}-\varphi_1\right\|^2 = O\left(\frac{1}{\alpha_n}\left(\frac{1}{nh_n^q}+h_n^{2s}\right)+\frac{1}{nh_n^{p+q}}\alpha_n^{(\beta-1)\wedge 0}+\alpha_n^\beta\right),$$

where the class $\mathcal{F}(\beta, C)$ includes additional moment restrictions (see Babii, 2020). In the nonparametric IV model, all four identification cases are possible, depending on the regularity parameter β. For consistency of $\hat{\varphi}$, we require $\alpha_n nh_n^q \to \infty$, $\alpha_n^{(1-\beta)\wedge 0} nh_n^{p+q} \to \infty$, and $h_n^{2s}/\alpha_n \to 0$ as $n \to \infty, \alpha_n \to 0$, and $h_n \to 0$.

Moreover, it is known that $\rho_{2n}^{1/2} = \sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s$ (see, for instance, Babii 2020, Prop. A.3.1). Thus, by Theorem 3.2, we find

$$\sup_{(\varphi,K)\in\mathcal{F}} \mathbb{E}\|\hat{\varphi}-\varphi_1\|_\infty = O\left(\frac{1}{\alpha_n}\left(\frac{1}{\sqrt{nh_n^q}}+h_n^s\right)+\frac{1}{\alpha_n^{1/2}}\sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}}+\alpha_n^{\beta/2}\right).$$

For uniform consistency of $\hat{\varphi}$, we need $\alpha_n^2 nh_n^q \to \infty$, $\alpha_n nh_n^{p+q} \to \infty$, and $h_n^s/\alpha_n \to 0$ as $n \to \infty$, $\alpha_n \to 0$, and $h_n \to 0$.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

It remains an open question whether these convergence rates for the NPIV model are minimax-optimal for the class $\mathcal{F}(\beta, C)$. Note that this class does not require differentiability or specific Hölder smoothness of φ_1 (see Chen and Christensen, 2018), which established minimax-optimal convergence rates for the NPIV model over the Hölder ball.

4. CONTINUOUS LINEAR FUNCTIONALS

In many economic applications, the primary focus is on a continuous linear functional of a structural function φ. This section demonstrates that the asymptotic distribution of such functionals can become nonstandard under identification failure. To simplify the exposition, we concentrate on the high-dimensional regression example in Section 3.3.1; further results are provided in Appendix Section A.4.

According to the Riesz representation theorem, any continuous linear functional on a Hilbert space $\mathcal{E}$ can be represented as an inner product with some $\mu \in \mathcal{E}$. Thus, we focus on the asymptotic distribution of $\langle \hat{\varphi}, \mu \rangle$. We let $\mu = \mu_0 + \mu_1$ denote the orthogonal decomposition of μ, where μ_0 and μ_1 are the orthogonal projections on $\mathcal{N}(K)$ and $\mathcal{N}(K)^\perp$, respectively. Similarly, we decompose $W = W^0 + W^1$, where W^0 and W^1 are the orthogonal projections of W on $\mathcal{N}(K^*)$ and $\mathcal{N}(K^*)^\perp$, respectively.

We define the variance operator $\Sigma = \mathbb{E}[(Y - \langle Z, \varphi_1 \rangle)^2 W \otimes W]$ and the sequence

$$\pi_n = n^{1/2} \left\| \Sigma^{1/2} K (\alpha_n I + K^* K)^{-1} \mu_1 \right\|^{-1},$$

with $\mu_1 = \mu - \mu_0$. The following assumptions place mild restrictions on the data distribution and tuning parameters.

Assumption 4.1. (i) The data $(Y_i, Z_i, W_i)_{i=1}^n$ are an i.i.d. sample of (Y, Z, W) with $\mathbb{E}|U|^{2+\delta} < \infty$ and $\mathbb{E}\|Z\|^{2+\delta} < \infty$ for some $\delta > 0$; (ii) $\mathbb{E}\|ZW\|^2 < \infty$, $\mathbb{E}\|UW\|^2 < \infty$, $\mathbb{E}\|UZW\| < \infty$, $\mathbb{E}\|Z\|^2\|W\| < \infty$, and $\mathbb{E}\left[|U|\|Z\|\|W\|^2\right] < \infty$; and (iii) $W^0 \neq 0$.

Assumption 4.2. (i) $\mu_1 \in \mathcal{R}(K^*K)^\gamma$ and $W \in \mathcal{R}(K^*K)^{\tilde{\gamma}}$ for some $\gamma, \tilde{\gamma} > 0$ with $\tilde{\gamma} + \gamma \geq 1/2$ and (ii) $\alpha_n \to 0$, $n\alpha_n \to \infty$, $\pi_n \alpha_n^{(\gamma+\beta/2)\wedge 1} \to 0$, $\frac{\pi_n \alpha_n^{\gamma \wedge \frac{1}{2}}}{n\alpha_n} \to 0$, and $n\alpha_n^{1+\beta\wedge 1} \to 0$ as $n \to \infty$.

Assumption 4.2 (ii) is an undersmoothing condition. Its requirements are the most stringent when $\pi_n \sim n^{1/2}$, in which case $n\alpha_n^{(2\gamma+\beta)\wedge 2} \to 0$, $n\alpha_n^{2-2\gamma\wedge 1} \to \infty$, and $n\alpha_n^{1+\beta\wedge 1} \to 0$ are needed. The last two conditions are not binding whenever $\gamma \geq 1/2$ and $\beta > 0$, in which case the first condition is automatically satisfied.

We let $(\chi_j^2)_{j\geq 1}$ be i.i.d. chi-square random variables with 1 degree of freedom, and we let $(\lambda_j)_{j\geq 1}$ be the eigenvalues of the operator $T : \phi \mapsto \mathbb{E}_X[\phi(X)h(X, X')]$ on $L_2(X)$ with

$$h(X, X') = \frac{\langle W^0, W^{0'} \rangle}{2} \left\{ \langle Z, \mu_0 \rangle (Y' - \langle Z', \varphi_1 \rangle) + \langle Z', \mu_0 \rangle (Y - \langle Z, \varphi_1 \rangle) \right\},$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

where $\mathbb{E}_X$ denotes the expectation with respect to $X = (Y, Z, W)$ and where X' is an independent copy of X.

The following result holds.

THEOREM 4.1. *Under Assumptions 3.1, 4.1, and 4.2, we have*

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle \xrightarrow{d} \mathbb{E}\left[\|W\|^2(Y - \langle Z, \varphi_1\rangle)\langle Z, \mu_0\rangle\right] + \sum_{j\geq 1}\lambda_j(\chi_j^2 - 1)$$

and

$$\pi_n\langle\hat{\varphi} - \varphi, \mu_1\rangle \xrightarrow{d} N(0,1).$$

Since

$$\langle\hat{\varphi} - \varphi_1, \mu\rangle = \langle\hat{\varphi} - \varphi_1, \mu_1\rangle + \langle\hat{\varphi} - \varphi_1, \mu_0\rangle,$$

the following corollary is a trivial consequence of Theorem 4.1.

COROLLARY 4.1. *Under Assumptions 3.1, 4.1, and 4.2, we have the following:*

1. *If $\pi_n = o(n\alpha_n)$, then*

$$\pi_n\langle\hat{\varphi} - \varphi_1, \mu\rangle \xrightarrow{d} N(0,1).$$

2. *If $n\alpha_n = o(\pi_n)$, then*

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu\rangle \xrightarrow{d} \left(\mathbb{E}\left[\|W\|^2(Y - \langle Z, \varphi_1\rangle)\langle Z, \mu_0\rangle\right] + \sum_{j\geq 1}\lambda_j(\chi_j^2 - 1)\right).$$

3. *If $\pi_n = n\alpha_n$, then*

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu\rangle \xrightarrow{d} N(0,1) + \left(\mathbb{E}\left[\|W\|^2(Y - \langle Z, \varphi_1\rangle)\langle Z, \mu_0\rangle\right] + \sum_{j\geq 1}\lambda_j(\chi_j^2 - 1)\right).$$

The normalizing sequence in Corollary 4.1 can be either π_n or $n\alpha_n$, and the asymptotic distribution may feature a Gaussian component, a chi-square component, or both, depending on the mapping properties of the operators K and Σ. Specifically, if the completeness condition holds, the asymptotic distribution consists solely of a Gaussian component. However, if the completeness condition fails, a chi-square component may also appear.

To illustrate this, we let $W = (W(t))_{t\in[0,1]}$ be a Brownian bridge with a Karhunen–Loève decomposition $W(t) = \sum_{j=1}^{\infty}\frac{\xi_j}{j\pi}h_j(t)$, where $\xi_j \sim_{i.i.d.} N(0,1)$ and $h_j(t) = \sqrt{2}\sin(\pi jt)$ (see Shorack and Wellner, 1986, p. 33). We let $Z(s) = \sum_{j=1}^{\infty}\frac{\zeta_j}{j\pi}e_j(s)$ be a Gaussian process with $\zeta_j \sim_{i.i.d.} N(0,1)$, $\mathrm{Cov}(\zeta_j, \xi_k) = \rho\mathbf{1}_{j\neq k}$, and

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

$e_j(s) = \sqrt{2}\cos(\pi j s)$. In the high-dimensional IV regression, we have

$$Y = \int_0^1 \varphi(s)Z(s)\mathrm{d}s + U, \qquad \mathbb{E}[UW(t)] = 0, \qquad \forall t \in [0,1],$$

with $\varphi(s) = \sum_{j=1}^{\infty} j^{-a} e_j(s), a > 1/2$ and $e_j(s) = \sqrt{2}\cos(\pi j s)$. We suppose that the object of interest is the average structural function

$$\int_0^1 \varphi(s)\mu(s)\mathrm{d}s = \langle \varphi, \mu \rangle$$

for $\mu \in L_2[0,1]$ with $\mu(s) = 1 + \sum_{j=1}^{\infty} j^{-b} e_j(s)$ and $b > 1/2$.[8] The covariance operator $K = \mathbb{E}[W \otimes Z] : L_2[0,1] \to L_2[0,1]$ simplifies to $K = \sum_{j=1}^{\infty} \frac{\rho}{j^2\pi^2} h_j \otimes e_j$ and its adjoint to $K^* = \sum_{j=1}^{\infty} \frac{\rho}{j^2\pi^2} e_j \otimes h_j$. It can be observed that $1 \in \mathcal{N}(K)$, $\varphi_0 = 0$, and $\mu_0 = 1$, so the structural function φ and average structural function $\langle \varphi, \mu \rangle$ are identified even though K has a nontrivial null space.

To verify Assumption 4.2, we observe that by Parseval's identity,

$$\left\| (K^*K)^{-\beta/2} \varphi_1 \right\|^2 = \frac{\pi^{4\beta}}{\rho^{2\beta}} \sum_{j=1}^{\infty} \frac{1}{j^{2a-4\beta}} < \infty \iff a - 2\beta > 1/2$$

and

$$\left\| (K^*K)^{-\gamma} \mu_1 \right\|^2 = \frac{\pi^{8\gamma}}{\rho^{4\gamma}} \sum_{j=1}^{\infty} \frac{1}{j^{2b-8\gamma}} < \infty \iff b - 4\gamma > 1/2;$$

cf. Section 3. Under homoskedasticity, we have $\Sigma = \sigma^2 \mathbb{E}[W \otimes W] = \sum_{j=1}^{\infty} \frac{\sigma^2}{j^2\pi^2} h_j \otimes h_j$, which implies that

$$\| \Sigma^{1/2} K (\alpha_n + K^*K)^{-1} \mu_1 \|^2 = \frac{\sigma^2}{\pi^2} \sum_{j=1}^{\infty} \frac{1}{j^{2(b+1)}} \frac{\sigma_j^2}{(\alpha_n + \sigma_j^2)^2}$$

with $\sigma_j = \frac{\rho}{j^2\pi^2}$. Therefore, we obtain

$$\pi_n^2 \sim \frac{n}{\sum_{j=1}^{\infty} j^{-2(b+3)} (\alpha_n + j^{-4})^{-2}}.$$

For $\gamma = 1/2$, we require that the function μ_1 has sufficient regularity, with $b > 5/2$. For larger values of β, more regularity in φ_1 is needed, which corresponds to larger values of a. In this case, $\pi_n^2 \sim n$, and Assumption 4.2 (ii) requires $n\alpha_n^2 \to 0$, ensuring that $n\alpha_n = o(\pi_n)$. Consequently, Corollary 4.1 shows that

$$n\alpha_n \langle \hat{\varphi} - \varphi, \mu \rangle \xrightarrow{d} \mathbb{E}\left[U \langle Z, \mu_0 \rangle \| W \|^2 \right] + \sum_{j=1}^{\infty} \lambda_j (\chi_j^2 - 1).$$

[8] Note that $\int_0^1 \mu(s)\mathrm{d}s = 1$ and that $\varphi \in L_2[0,1]$.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

5. MONTE CARLO EXPERIMENTS

In this section, we examine the validity of our asymptotic theory via Monte Carlo simulations. The data-generating process (DGP) is specified as follows:

$$\varphi(z) = \sum_{j=1}^{\infty} (-1)^j j^{-6} e_j(z), \qquad Y = \mathbb{E}[\varphi(Z)|W] + U, \qquad U \sim 0.1N(0,1),$$

$$f_{ZW}^J(z,w) = C_J \sum_{j=1}^{J} j^{-3} e_j(z) e_j(w), \qquad \forall z, w \in [0,1],$$

where we use the basis $e_j(x) = \sqrt{2}\sin(j\pi x), j \geq 1$ and set C_J such that the density integrates to 1 (see also Hall and Horowitz, 2005). The parameter J controls the degree of identification, with $K : L_2[0,1] \to L_2[0,1]$ having an infinite-dimensional null space whenever $J < \infty$. Specifically, K has a J-dimensional range and

$$\mathcal{N}(K) \subset \text{span}\{e_j : j > J\}.$$

Accordingly, we consider three cases: $J \in \{1, 2, \infty\}$, where $J = \infty$ corresponds to the fully identified model.

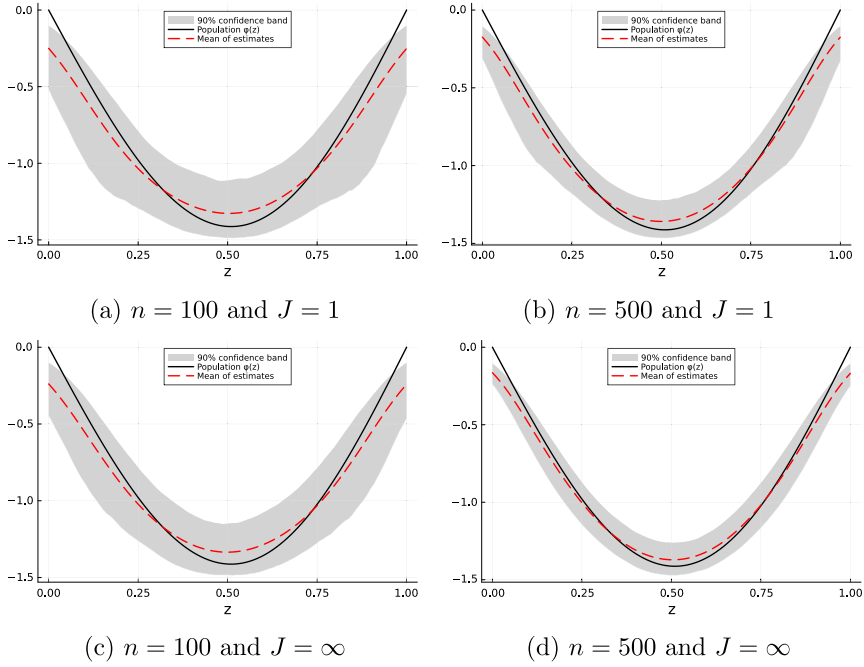
We simulate samples of size $n \in \{100, 500\}$ with 5,000 replications. For simplicity, we focus on the Tikhonov-regularized estimator with a product of Gaussian kernels to estimate the best approximation φ_1. The bandwidth parameter is selected via Silverman's rule of thumb, and the regularization parameter is chosen via leave-one-out cross-validation (see Babii (2020) and Centorrino, 2014 for details on the numerical implementation). All the integrals are approximated by Riemann sums on a uniform grid in $[0,1]$, and infinite sums are truncated at a large finite value beyond which the numerical results remain unchanged.

Table 1 shows the empirical L_2 and L_∞ errors for the three degrees of identification, as measured by $J \in \{1, 2, \infty\}$. The estimation errors decrease as the dimension of the null space shrinks, for both $n = 100$ and $n = 500$. Larger sample sizes also result in smaller errors, which is consistent with our asymptotic theory. Figure 1 displays the true structural function φ, the mean of the estimates $\mathbb{E}\hat{\varphi}$, and the pointwise 90% empirical confidence bands based on 5,000 replications. The results align with those in Table 1. Note that for $J = 1$, we can only recover information related to the first basis vector, illustrating an extreme failure of the

TABLE 1. L_2 and L_∞ errors calculated from 5,000 experiments

	$n = 100$		$n = 500$	
J	L_2	L_∞	L_2	L_∞
1	0.0378	0.3295	0.0162	0.2156
2	0.0345	0.3199	0.0130	0.2054
∞	0.0311	0.3109	0.0107	0.1974

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press



(a) $n = 100$ and $J = 1$ (b) $n = 500$ and $J = 1$

(c) $n = 100$ and $J = \infty$ (d) $n = 500$ and $J = \infty$

FIGURE 1. Mean estimates with pointwise 90% empirical confidence bands calculated from 5,000 experiments.

completeness condition. Nevertheless, even in this case, a significant amount of information about the structural function can be learned, especially when $n = 500$.

6. CONCLUSIONS

This article develops a theory for nonidentified linear models, using nonparametric IV regression and high-dimensional regressions as illustrative examples. Identification failures arise because of the noninjectivity of covariance or conditional expectation operators. We show that when these operators are noninjective, spectral-regularized estimators converge to the best approximation of the structural parameter in the orthogonal complement of the operator's null space, and we derive new bounds on the risk in both uniform and Hilbert space norms.

This framework offers an appealing projection interpretation for nonparametric IV regression under identification failure, which is analogous to that of ordinary least squares under misspecification. We describe cases where the best approximation coincides with or closely approximates the structural parameter, highlighting the usefulness of these results in partial identification settings. Additionally, incorporating smoothness constraints into the regularization procedure can yield

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

more precise estimates in small samples (see Babii and Florens, 2024 for an approach using Sobolev-type penalties). Finally, under identification failure, the asymptotic distribution of a continuous linear functional can transition between weighted chi-square and Gaussian components.

To conclude, despite a large amount of research over the last two decades, there is only a small amount of empirical work using nonparametric IV regression. One reason is that it involves identifying conditions that are difficult to validate empirically and our article shows that numerous results may be obtained without these conditions. Economic analysis often hinges on strong functional restrictions that may lead to erroneous counterfactual conclusions and policy recommendations. Nonparametric IV regression is useful for obtaining robust and reliable estimates (see e.g., Compiani, 2022).

APPENDIX

A.1. SPECTRAL REGULARIZATION

In this section, we extend our results to encompass a range of estimators based on spectral regularization. This framework includes iterated ridge (Tikhonov) regression, functional gradient descent (Landweber–Fridman), and functional PCA (spectral cut-off) methods. Consider the linear operator equation:

$$K\varphi = r,$$

where $K : \mathcal{E} \to \mathcal{H}$ is a linear operator between Hilbert spaces $\mathcal{E}$ and $\mathcal{H}$. We assume that K is bounded, with $\|K\|^2 \leq \Lambda$ for some $\Lambda < \infty$, though it may not be compact. Then $K^*K : \mathcal{E} \to \mathcal{E}$ is a normal operator with the spectral decomposition:

$$K^*K = \int_{\sigma(K^*K)} \lambda \mathrm{d}E(\lambda),$$

where $\sigma(K^*K)$ denotes the spectrum of K^*K and E is the resolution of the identity (see Rudin, 1991, Thm. 12.23). For a bounded Borel function $g : [0, \Lambda] \to \mathbf{R}$, we define:

$$g(K^*K) = \int_{\sigma(K^*K)} g(\lambda)\mathrm{d}E(\lambda).$$

If K is compact, the spectrum of K^*K is countable, and we have

$$g(K^*K) = \sum_{j=1}^{\infty} g(\sigma_j^2)E_j,$$

where $E_j = h_j \otimes e_j$ projects on the eigenspace corresponding to the eigenvalue σ_j^2.

We aim to recover the best approximation to the structural parameter φ when estimates $(\hat{K}, \hat{r})$ of (K, r) are available with $\|\hat{K}\|^2 \leq \Lambda$ almost surely. To do this, we consider an extended version of Assumption 3.1.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Assumption A.1.1. Suppose that (φ, K) belongs to

$$\mathcal{F}(\beta, C) = \left\{(\varphi, K) : \ \varphi_1 = s_\beta(K^*K)\psi, \ \|\psi\|^2 \vee \|\varphi_0\| \vee \|K\| \leq C\right\},$$

where $s_\beta : [0, \Lambda] \to \mathbf{R}$ is a non-decreasing positive function such that $\lambda \mapsto s_\beta^2(\lambda)/\lambda^\beta$ is non-increasing.

The following two cases are of particular interest:

1. Mildly ill-posed problem: $s_\beta(\lambda) = \lambda^{\beta/2}$.
2. Severely ill-posed problem: $s_\beta(\lambda) = \log^{-\beta/2}\left(\frac{1}{\lambda}\right)$ with $s_\beta(0) = 0$.

In the mildly ill-posed case, the eigenvalues of K^*K can decline exponentially, provided that the Fourier coefficients of φ_1 decay at the same rate. In contrast, the severely ill-posed case accommodates less regular φ_1 and K.

The spectral regularization scheme is defined by a family of bounded Borel functions $g_\alpha : [0, \infty) \to \mathbf{R}$, where $\alpha > 0$ is a regularization parameter such that $\lim_{\alpha \to 0} g_\alpha(\lambda) = \lambda^{-1}$. When $\hat{K}$ is bounded, the regularized estimator is defined by

$$\hat{\varphi} = g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{r}.$$

Our theoretical results require the following assumption on the regularization scheme.

Assumption A.1.2. There exist $c_1, c_2, c_3, \beta_0 > 0$ such that: (i) $\sup_{\lambda \leq \Lambda} |g_\alpha(\lambda)\lambda^{1/2}| \leq c_1/\sqrt{\alpha}$; (ii) $\sup_{\lambda \leq \Lambda} |(g_\alpha(\lambda)\lambda - 1)\lambda^r| \leq c_2\alpha^r, \forall r \in [0, 2\beta_0]$; and (iii) $\sup_{\lambda \leq \Lambda} |g_\alpha(\lambda)| \leq c_3/\alpha$.

The following regularization schemes satisfy Assumption A.1.2 in both the mildly and severely ill-posed cases:[9]

1. Ridge (Tikhonov):

   $$g_\alpha(\lambda) = \frac{1}{\alpha + \lambda}.$$

   Assumption A.1.2 holds with $c_1 = 1/2$, $c_2 = c_3 = 1$, and $\beta_0 = 2$.
2. PCA (spectral cut-off):

   $$g_\alpha(\lambda) = \lambda^{-1}\mathbb{1}\{\lambda \geq \alpha\}.$$

   Assumption A.1.2 holds with $c_1 = c_2 = c_3 = 1$ and every $\beta_0 > 0$.
3. Iterated Tikhonov/ridge regularization:

   $$g_{\alpha,m}(\lambda) = \sum_{j=0}^{m-1} \frac{\alpha^j}{(\alpha + \lambda)^{j+1}} = \frac{1}{\lambda}\left(1 - \left(\frac{\alpha}{\lambda + \alpha}\right)^m\right), \qquad m = 2, 3, \ldots$$

   Assumption A.1.2 holds with $c_1 = m^{1/2}$, $c_2 = 1$, and $c_3 = \beta_0 = m$.

[9] In the severely ill-posed case, the function $\lambda \mapsto s_\beta^2(\lambda)/\lambda^\beta$ is nonincreasing only on $(0, 1/e]$ and s_β is not defined at $\lambda = 1$. To address these issues, we assume that the norm on $\mathcal{H}$ is scaled such that $\|K\|^2 \leq 1/e$.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

4. Gradient descent (Landweber–Fridman):

$$g_{\alpha,c}(\lambda) = \sum_{j=0}^{1/\alpha-1} (1-c\lambda)^j = \frac{1}{\lambda}\left(1-(1-c\lambda)^{1/\alpha}\right), \tag{A.1}$$

where $\alpha = 1/m$ for some $m \in \mathbf{N}$ and $c \in (0, 1/\Lambda)$. Assumption A.1.2 is satisfied with $c_1^2 = c$, $c_2 = c \vee 1$, $c_3 = \left(\frac{\beta}{ce}\right)^{\beta/2} \vee 1$, and every $\beta_0 > 0$.

The constant β_0 is known as the qualification of the regularization scheme. It is well-established that simple Tikhonov regularization experiences a saturation effect, limiting bias convergence to $O(\alpha^2)$. This issue can be resolved using iterated Tikhonov regularization, similar to higher-order kernels in nonparametric kernel estimation.

To provide a spectral interpretation of gradient descent in point 4, consider the minimization problem:

$$\min_{\phi} \underbrace{\frac{1}{2}\|K\phi - r\|^2}_{=:Q(\phi)},$$

where the gradient is $\nabla Q(\phi) := K^*(K\phi - r)$. The gradient descent with a constant step size $c \in (0, 1/\Lambda)$, leads to the update rule:

$$\begin{aligned}\phi_m &= \phi_{m-1} - c\nabla Q(\phi_{m-1}) \\ &= [I - cK^*K]\phi_{m-1} + cK^*r.\end{aligned}$$

Iterating this update rule recursively, we obtain

$$\phi_m = [I - cK^*K]^m\phi_0 + \sum_{j=0}^{m-1}[I - cK^*K]^jK^*r.$$

Assuming that the initial value is $\phi_0 = 0$ and replacing (K, r) by $(\hat{K}, \hat{r})$, we obtain

$$\hat{\varphi}_\alpha = g_{\alpha,c}(\hat{K}^*\hat{K})\hat{K}^*\hat{r}$$

with $g_{\alpha,c}$ and $\alpha = 1/m$ as in equation (A.1).

The following theorem establishes the convergence rate of $\hat{\varphi}$ to φ_1 for general regularization schemes.

THEOREM A.1. *Suppose Assumptions A.1.1, A.1.2, and 3.2 hold with $\beta \le \beta_0$.*[10] *Then, in the mildly ill-posed case,*

$$\left\|\hat{\varphi} - \varphi_1\right\|^2 = O_P\left(\frac{\delta_n}{\alpha_n} + \rho_{1n}^{\beta\wedge 1}\left(1 + \mathbb{1}_{\beta=1}\log^2\rho_{1n}^{-1}\right) + \alpha_n^\beta\right),$$

and in the severely ill-posed case,

$$\left\|\hat{\varphi} - \varphi_1\right\|^2 = O_P\left(\frac{\delta_n}{\alpha_n} + \log^{-\beta}\rho_{1n}^{-1} + \log^{-\beta}\alpha_n^{-1}\right).$$

[10] Note that we can always take $\beta = \beta_0$ if $\beta > \beta_0$.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Before presenting the proof, note that:

1. For the mildly ill-posed case, the optimal choice for the regularization parameter is $\alpha_n \sim \delta_n^{\frac{1}{\beta+1}}$, assuming that $\rho_{1n}^{\beta\wedge 1}\log^2 n \lesssim \delta_n/\alpha_n$. This selection results in the convergence rate:

$$\|\hat{\varphi}-\varphi_1\|^2 = O_P\left(\delta_n^{\frac{\beta}{\beta+1}}\right).$$

2. For the severely ill-posed case, one can select $\alpha_n \sim \delta_n^{1/2}$. This choice yields the convergence rate:

$$\|\hat{\varphi}-\varphi_1\|^2 = O_P\left(\frac{1}{\log^{\beta} n}\right),$$

assuming $\delta_n, \rho_n \sim n^{-c}$ for some $c > 0$.

Proof of Theorem A.1. Decompose

$$\hat{\varphi}-\varphi_1 = I_n + II_n + III_n,$$

where

$$\begin{aligned} I_n &= g_\alpha(\hat{K}^*\hat{K})\hat{K}^*(\hat{r}-\hat{K}\varphi), \\ II_n &= \left[g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right] s_\beta(\hat{K}^*\hat{K})\psi, \\ III_n &= \left[g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right]\left\{s_\beta(K^*K) - s_\beta(\hat{K}^*\hat{K})\right\}\psi. \end{aligned}$$

To verify this decomposition, observe that $\varphi = \varphi_1 + \varphi_0$ and, under Assumption A.1.1, $\varphi_1 = s_\beta(K^*K)\psi$. Using the isometry of functional calculus, we bound each term as follows:

$$\begin{aligned} \|I_n\|^2 &\le \left\|g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\right\|^2 \left\|\hat{r}-\hat{K}\varphi_1\right\|^2 \\ &\le \sup_{\lambda\le\Lambda}\left|g_\alpha(\lambda)\lambda^{1/2}\right|^2 \left\|\hat{r}-\hat{K}\varphi_1\right\|^2. \end{aligned}$$

Under Assumptions A.1.2 (i) and 3.2, this shows that $\mathbb{E}\|I_n\|^2 = O(\delta_n/\alpha_n)$.

Note also that under Assumption A.1.1, s_β is nondecreasing, whence under Assumption A.1.2 (ii) with $r = 0$, we obtain

$$\left|(g_\alpha(\lambda)\lambda - 1)s_\beta(\lambda)\right| \le c_2|s_\beta(\lambda)| \le c_2 s_\beta(\alpha_n), \qquad \forall\lambda\in[0,\alpha_n].$$

Similarly, since $\lambda \mapsto s_\beta(\lambda)/\lambda^{\beta/2}$ is nonincreasing, under Assumption A.1.2 (ii) with $r = \beta/2$

$$\left|(g_\alpha(\lambda)\lambda - 1)s_\beta(\lambda)\right| \le \left|(g_\alpha(\lambda)\lambda - 1)\lambda^{\beta/2}\right| \left|\frac{s_\beta(\lambda)}{\lambda^{\beta/2}}\right| \le c_2\alpha_n^{\beta/2} s_\beta(\alpha_n)/\alpha_n^{\beta/2}, \qquad \forall\lambda\ge\alpha_n.$$

Thus,

$$\begin{aligned} \|II_n\|^2 &\le \left\|\left[g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right] s_\beta(\hat{K}^*\hat{K})\right\|^2 \|\psi\|^2 \\ &\le \sup_{\lambda\le\Lambda}\left|(g_\alpha(\lambda)\lambda - 1)s_\beta(\lambda)\right|^2 C. \end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Finally, under Assumption A.1.2 (ii) with $r = 0$

$$\begin{aligned}\|III_n\|^2 &\leq \left\|g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right\|^2 \left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\|^2 \|\psi\|^2 \\ &\leq C \sup_{\lambda \leq \Lambda} |(g_\alpha(\lambda)\lambda - 1)|^2 \left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\|^2 \\ &\leq C c_2^2 \left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\|^2.\end{aligned}$$

The result follows by applying Lemma A.1.1 which appears at the end of this section. □

The following theorem establishes the uniform convergence rates for a broad class of regularized estimators:

THEOREM A.2. *Suppose that Assumptions A.1.1, A.1.2, 3.2, and 3.3 are satisfied with* $\varphi_1 = s_\beta(K^*K)K^*\psi$ *and* $\beta_0 \geq \beta$. *Then, in the mildly ill-posed case,*

$$\left\|\hat{\varphi} - \varphi_1\right\|_\infty = O_P\left(\frac{\delta_n^{1/2}}{\alpha_n} + \frac{1}{\alpha_n^{1/2}}\left(\rho_{2n}^{1/2} + \rho_{1n}^{\frac{\beta \wedge 1}{2}}\left(1 + \mathbb{1}_{\beta=1}\log \rho_{1n}^{-1}\right)\right) + \alpha_n^{\beta/2}\right),$$

and in the severely ill-posed case,

$$\left\|\hat{\varphi} - \varphi_1\right\|_\infty = O_P\left(\frac{\delta_n^{1/2}}{\alpha_n} + \frac{1}{\alpha_n^{1/2}}\left(\rho_{2n}^{1/2} + \log^{-\beta/2}\rho_{1n}^{-1}\right) + \log^{-\beta/2}\alpha_n^{-1}\right).$$

Proof. We start with the decomposition

$$\hat{\varphi} - \varphi_1 = I_n + II_n + III_n,$$

where

$$\begin{aligned}I_n &= g_\alpha(\hat{K}^*\hat{K})\hat{K}^*(\hat{r} - \hat{K}\varphi_1), \\ II_n &= \left[g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right]s_\beta(\hat{K}^*\hat{K})\hat{K}^*\psi, \\ III_n &= \left[g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right]\left\{s_\beta(K^*K)K^* - s_\beta(\hat{K}^*\hat{K})\hat{K}^*\right\}\psi.\end{aligned}$$

We proceed by bounding each term individually. For the first term, we have

$$\begin{aligned}\|I_n\|_\infty &= \left\|\hat{K}^* g_\alpha(\hat{K}\hat{K}^*)(\hat{r} - \hat{K}\varphi_1)\right\| \\ &\leq \|\hat{K}^*\|_{2,\infty} \sup_{\lambda \leq \Lambda} |g_\alpha(\lambda)| \left\|\hat{r} - \hat{K}\varphi_1\right\| \\ &= O_P\left(\frac{\delta_n^{1/2}}{\alpha_n}\right),\end{aligned}$$

where the last line follows under Assumptions 3.3 and A.1.2 (iii).

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

For the second term, note that

$$\begin{aligned}\|II_n\|_\infty &= \left\|\hat{K}^*\left[g_\alpha(\hat{K}\hat{K}^*)\hat{K}\hat{K}^* - I\right]s_\beta(\hat{K}\hat{K}^*)\psi\right\| \\ &\leq \|\hat{K}^*\|_{2,\infty}\sup_{\lambda\leq\Lambda}|(g_\alpha(\lambda)\lambda - 1)s_\beta(\lambda)|\|\psi\| \\ &= O_P(s_\beta(\alpha_n)),\end{aligned}$$

where the last line follows by the same argument as in the proof of Theorem A.1 under Assumptions 3.3 and A.1.2 (ii).

Next, for the third term,

$$\|III_n\|_\infty \leq \left\|g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right\|_\infty \left\|s_\beta(K^*K)K^* - s_\beta(\hat{K}^*\hat{K})\hat{K}^*\right\|_{2,\infty}\|\psi\|.$$

Under Assumptions A.1.2 (i) and 3.3

$$\begin{aligned}\left\|g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right\|_\infty &\leq \|\hat{K}^*\|_{2,\infty}\left\|g_\alpha(\hat{K}\hat{K}^*)\hat{K}\right\| + 1 \\ &\leq \|\hat{K}^*\|_{2,\infty}\sup_{\lambda\leq\Lambda}\left|g_\alpha(\lambda)\lambda^{1/2}\right| + 1 \\ &= O_P\left(\frac{1}{\alpha_n^{1/2}}\right).\end{aligned}$$

Finally, under Assumption 3.3

$$\begin{aligned}\left\|s_\beta(K^*K)K^* - s_\beta(\hat{K}^*\hat{K})\hat{K}^*\right\|_{2,\infty} &= \left\|K^*s_\beta(KK^*) - \hat{K}^*s_\beta(\hat{K}\hat{K}^*)\right\|_{2,\infty} \\ &\lesssim \left\|\hat{K}^* - K^*\right\|_{2,\infty} + \left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\| \\ &\lesssim \rho_{2n}^{1/2} + \left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\|.\end{aligned}$$

The result now follows by applying Lemma A.1.1 similar to the proof of Theorem A.1. □

LEMMA A.1.1. *Assume that Assumption 3.2 (iii) holds with $\rho_{1n} \sim n^{-c}$ for some $c > 0$. Then*

$$\left\|(\hat{K}^*\hat{K})^{\beta/2} - (K^*K)^{\beta/2}\right\| = O_P\left(\left(1 + \mathbb{1}_{\beta=1}\log\rho_{1n}^{-1}\right)\rho_{1n}^{\frac{\beta\wedge 1}{2}}\right)$$

and

$$\left\|\log^{-\beta/2}(\hat{K}^*\hat{K})^{-1} - \log^{-\beta/2}(K^*K)^{-1}\right\| = O_P\left(\log^{-\beta/2}\rho_{1n}^{-1}\right).$$

Proof. Using Egger (2005, Lem. 3.2), we have:

$$\left\|(\hat{K}^*\hat{K})^{\beta/2} - (K^*K)^{\beta/2}\right\| \lesssim \begin{cases}\|\hat{K} - K\|^\beta, & \beta < 1 \\ \|\hat{K} - K\|\left\{1 + \|\hat{K}\| + \|K\| + \log\|\hat{K} - K\|^{-1}\right\}, & \beta = 1 \\ \|\hat{K} - K\|(\|\hat{K}\| + \|K\|)^{\beta/2}, & \beta > 1.\end{cases}$$

The first statement follows because $\|\hat{K} - K\| = O_P(\rho_{1n}^{1/2})$, with $\rho_{1n} \to 0$ as $n \to \infty$ under Assumption 3.2 (iii), and since $x \mapsto x\log(1/x)$ is strictly increasing near zero.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

For the second statement, by Mathé and Pereverzev (2002, Thm. 4),

$$\begin{aligned}\left\|\log^{-\beta/2}(\hat{K}^*\hat{K})^{-1}-\log^{-\beta/2}(K^*K)^{-1}\right\| &= O_P\left(\log^{-\beta/2}\|\hat{K}^*\hat{K}-K^*K\|^{-1}\right)\\ &= O_P\left(\log^{-\beta/2}\rho_{1n}^{-1}\right),\end{aligned}$$

where the last line uses Assumption 3.2 (iii) and the fact that $x \mapsto \log^{-\beta/2}(1/x)$ is strictly increasing near zero. □

A.2. PROOFS OF MAIN RESULTS

Proof of Theorem 3.1. Decompose

$$\begin{aligned}\hat{\varphi}-\varphi_1 &= (\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*(\hat{r}-\hat{K}\varphi_1)\\ &\quad+(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{K}\varphi_1-(\alpha_n I+K^*K)^{-1}K^*K\varphi_1\\ &\quad+\left((\alpha_n I+K^*K)^{-1}K^*K-I\right)\varphi_1\\ &\triangleq I_n+II_n+III_n.\end{aligned}$$

The term III_n represents the regularization bias, which can be controlled under Assumption 3.1 as follows:

$$\begin{aligned}\|III_n\|^2 &= \left\|\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\ &\le C\left\|\alpha_n(\alpha_n I+K^*K)^{-1}(K^*K)^{\beta/2}\right\|^2\\ &\le C\sup_{\lambda\in[0,\,\|K\|^2]}\left|\frac{\alpha_n\lambda^{\beta/2}}{\alpha_n+\lambda}\right|^2\\ &\le C^{(2\beta-3)\vee 1}\alpha_n^{\beta};\end{aligned}$$

see Babii (2022) for more details. The first term I_n is controlled under Assumption 3.2 (i)

$$\begin{aligned}\mathbb{E}\|I_n\|^2 &\le \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|^2\left\|\hat{r}-\hat{K}\varphi_1\right\|^2\\ &\le \sup_{\lambda\ge 0}\left|\frac{\lambda^{1/2}}{\alpha_n+\lambda}\right|^2\mathbb{E}\left\|\hat{r}-\hat{K}\varphi_1\right\|^2\\ &\le \frac{1}{4\alpha_n}\mathbb{E}\left\|\hat{r}-\hat{K}\varphi_1\right\|^2\\ &\le \frac{C_1\delta_n}{4\alpha_n}.\end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Next, decompose II_n further as follows:

$$\begin{aligned}
II_n &= -\left[\alpha_n(\alpha_n I+\hat{K}^*\hat{K})^{-1}-\alpha_n(\alpha_n I+K^*K)^{-1}\right]\varphi_1\\
&= -(\alpha_n I+\hat{K}^*\hat{K})^{-1}\alpha_n\left[K^*K-\hat{K}^*\hat{K}\right](\alpha_n I+K^*K)^{-1}\varphi_1\\
&= (\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\left[\hat{K}-K\right]\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\\
&\quad +(\alpha_n I+\hat{K}^*\hat{K})^{-1}\left[\hat{K}^*-K^*\right]\alpha_n K(\alpha_n I+K^*K)^{-1}\varphi_1\\
&= II_n^a+II_n^b.
\end{aligned}$$

Now, using previous results and Assumption 3.2 (iii), we have

$$\begin{aligned}
\mathbb{E}\|II_n^a\|^2 &= \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\left[\hat{K}-K\right]\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\
&\le \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|^2\left\|\hat{K}-K\right\|^2\left\|\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\
&\le \sup_{\lambda\ge 0}\left|\frac{\lambda^{1/2}}{\alpha_n+\lambda}\right|^2\mathbb{E}\left\|\hat{K}-K\right\|^2 C^{(2\beta-3)\vee 1}\alpha_n^{\beta\wedge 2}\\
&\le \frac{C_2\rho_{1n}}{4\alpha_n}C^{(2\beta-3)\vee 1}\alpha_n^{\beta\wedge 2}.
\end{aligned}$$

For II_n^b, we obtain

$$\begin{aligned}
\mathbb{E}\|II_n^b\|^2 &= \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\left[\hat{K}^*-K^*\right]\alpha_n K(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\
&\le \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\right\|^2\left\|\hat{K}^*-K^*\right\|^2 C\left\|\alpha_n K(\alpha_n I+K^*K)^{-1}(K^*K)^{\beta/2}\right\|^2\\
&\le \sup_{\lambda\ge 0}\left|\frac{1}{\alpha_n+\lambda}\right|^2\mathbb{E}\left\|\hat{K}^*-K^*\right\|^2 C\sup_{\lambda\in[0,C^2]}\left|\frac{\alpha_n\lambda^{(\beta+1)/2}}{\alpha_n+\lambda}\right|^2\\
&\le \frac{C_2\rho_{1n}}{\alpha_n^2}C^{(2\beta-1)\vee 1}\alpha_n^{(\beta+1)\wedge 2},
\end{aligned}$$

where the third line uses $\|\hat{K}^*-K^*\|=\|\hat{K}-K\|$. Combining all estimates, we obtain the final result. □

Proof of the Theorem 3.2. Consider the same decomposition as in the proof of Theorem 3.1. Given that $\varphi_1=(K^*K)^{\beta/2}K^*\psi$, the bias term is handled similarly to the identified case (see Babii (2020, Prop. 3.1)):

$$\begin{aligned}
\|III_n\|_\infty &= \left\|\alpha_n K^*(\alpha_n I+KK^*)^{-1}(KK^*)^{\frac{\beta}{2}}\psi\right\|_\infty\\
&\le \left\|K^*\right\|_{2,\infty}\left\|\alpha_n(\alpha_n I+KK^*)^{-1}(KK^*)^{\frac{\beta}{2}}\right\|\|\psi\|\\
&= O(\alpha_n^{\beta/2}).
\end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Next, using the Cauchy–Schwartz inequality, Assumption 3.2 (iii), and Assumption 3.3, we can bound the first term as follows:

$$\begin{aligned}\mathbb{E}\|I_n\|_\infty &= \mathbb{E}\left\|\hat{K}^*(\alpha_n I+\hat{K}\hat{K}^*)^{-1}(\hat{r}-\hat{K}\varphi_1)\right\|_\infty \\ &\le \mathbb{E}\|\hat{K}^*\|_{2,\infty}\left\|(\alpha_n I+\hat{K}\hat{K}^*)^{-1}\right\|\left\|(\hat{r}-\hat{K}\varphi_1)\right\| \\ &\le \frac{1}{\alpha_n}\left(\|K^*\|_{2,\infty}\mathbb{E}\left\|(\hat{r}-\hat{K}\varphi_1)\right\|+\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}\left\|(\hat{r}-\hat{K}\varphi_1)\right\|\right) \\ &\le C_1^{1/2}\left(C_3+C_3^{1/2}\rho_{2n}^{1/2}\right)\frac{\delta_n^{1/2}}{\alpha_n}.\end{aligned}$$

The second term can be decomposed into two parts, II_n^a and II_n^b, similarly to the proof of Theorem 3.1, and we bound each part separately. First, for II_n^a:

$$\begin{aligned}\mathbb{E}\|II_n^a\|_\infty &= \left\|\hat{K}^*(\alpha_n I+\hat{K}\hat{K}^*)^{-1}\left[\hat{K}-K\right]\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\|_\infty \\ &\le \mathbb{E}\left\|\hat{K}^*\right\|_{2,\infty}\left\|(\alpha_n I+\hat{K}\hat{K}^*)^{-1}\right\|\left\|\hat{K}-K\right\|\left\|\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\| \\ &\le \frac{1}{\alpha_n}\left(C_3+\left(\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}^2\right)^{1/2}\right)\left(\mathbb{E}\left\|\hat{K}-K\right\|^2\right)^{1/2}C^{(\beta-1.5)\vee 0.5}\alpha_n^{\beta/2} \\ &\le \left(C_3+C_3^{1/2}\rho_{2n}^{1/2}\right)\frac{C_2^{1/2}\rho_{1n}^{1/2}}{\alpha_n}C^{(\beta-1.5)\vee 0.5}\alpha_n^{\beta/2}.\end{aligned}$$

For the second part, II_n^b, applying Assumption 3.3, and using the inequality in Babii (2020, Lem. A.4.1) (see also Nair (2009, Problem 5.8), we get:

$$\begin{aligned}\mathbb{E}\|II_n^b\|_\infty &= \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\left[\hat{K}^*-K^*\right]\alpha_n K(\alpha_n I+K^*K)^{-1}\varphi_1\right\|_\infty \\ &= \mathbb{E}\left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\right\|_\infty\left\|\hat{K}^*-K^*\right\|_{2,\infty}\left\|\alpha_n K(\alpha_n I+K^*K)^{-1}\varphi_1\right\| \\ &= \frac{1}{2\alpha_n^{3/2}}\mathbb{E}\left(\|\hat{K}^*\|_{2,\infty}+2\alpha_n^{1/2}\right)\left\|\hat{K}^*-K^*\right\|_{2,\infty}C^{(\beta-1/2)\vee 1/2}\alpha_n^{\frac{\beta+1}{2}\wedge 1} \\ &\le \frac{1}{2\alpha_n^{3/2}}\left(C_3+C_3^{1/2}\rho_{2n}^{1/2}+2\alpha_n^{1/2}\right)C_3^{1/2}\rho_{2n}^{1/2}C^{(\beta-1/2)\vee 1/2}\alpha_n^{\frac{\beta+1}{2}\wedge 1}.\end{aligned}$$

Combining all these estimates yields the final result. □

The following proposition provides low-level conditions for Assumptions 3.2 and 3.3 in the context of nonparametric IV regression estimated with kernel smoothing. Let C_M^s denote the Hölder class.

PROPOSITION A.2.1. *Suppose that (i) $(Y_i,Z_i,W_i)_{i=1}^n$ are i.i.d. and $\mathbb{E}|Y_1|^2\le C\infty$; (ii) $f_{ZW}\in C_M^s$; (iii) kernel functions $K_z:\mathbf{R}^p\to\mathbf{R}$ and $K_w:\mathbf{R}^q\to\mathbf{R}$ are such that for $l\in\{w,z\}$,*

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

$K_l \in L_1 \cap L_2$, $\int K_l(u)\mathrm{d}u = 1$, $\int \|u\|^s K_l(u)\mathrm{d}u < \infty$, and $\int u^k K_l(u)\mathrm{d}u = 0$ for all multindices $|k| = 1, \ldots, \lfloor s \rfloor$. Then

$$\mathbb{E}\left\|\hat{r} - \hat{K}\varphi_1\right\|^2 = O\left(\frac{1}{nh_n^q} + h_n^{2s}\right) \qquad \text{and} \qquad \mathbb{E}\left\|\hat{K} - K\right\|^2 = O\left(\frac{1}{nh_n^{p+q}} + h_n^{2s}\right),$$

where the constants do not depend on (K, φ).

Proof. For the first claim, note that

$$\mathbb{E}\left\|\hat{r} - \hat{K}\varphi_1\right\|^2 \leq 2\mathbb{E}\left\|\hat{r} - r\right\|^2 + 2\mathbb{E}\left\|(\hat{K} - K)\varphi_1\right\|^2.$$

Decompose

$$\mathbb{E}\left\|\hat{r} - r\right\|^2 = \mathbb{E}\left\|\hat{r} - \mathbb{E}\hat{r}\right\|^2 + \left\|\mathbb{E}\hat{r} - r\right\|^2.$$

Under the i.i.d. assumption,

$$\begin{aligned}
\mathbb{E}\left\|\hat{r} - \mathbb{E}\hat{r}\right\|^2 &= \mathbb{E}\left\|\frac{1}{nh_n^q}\sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right) - \mathbb{E}\left[Y_i h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right)\right]\right\|^2 \\
&= \frac{1}{n}\mathbb{E}\left\|Y_i h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right) - \mathbb{E}\left[Y_i h_n^{-q} K_w\left(h_n^{-1}(W_i - w)\right)\right]\right\|^2 \\
&\leq \frac{1}{nh_n^q}\mathbb{E}|Y_1|^2\|K_w\|^2 \\
&= O\left(\frac{1}{nh_n^q}\right).
\end{aligned}$$

Applying the Cauchy–Schwartz inequality, we obtain

$$\begin{aligned}
\mathbb{E}\hat{r} - r &= \mathbb{E}\left[\varphi(Z_i)h_n^{-q}K_w\left(h_n^{-1}(W_i - w)\right)\right] - \int \varphi(z) f_{ZW}(z, w)\mathrm{d}z \\
&= \int \varphi(z)\left\{[f_{ZW} * K_w](z, w) - f_{ZW}(z, w)\right\}\mathrm{d}z \\
&\leq \|\varphi\|\,\|f_{ZW} * K_w - f_{ZW}\|,
\end{aligned}$$

where $[f_{ZW} * K_{w,h}](z, w) = \int f_{ZW}(z, w')h_n^{-q}K_w\left(h_n^{-1}(w - w')\right)\mathrm{d}w'$. Since $f_{ZW} \in C_M^s$, we obtain

$$\left\|\mathbb{E}\hat{r} - r\right\| = O(h^s),$$

as shown in Giné and Nickl (2016, Prop. 4.3.8). Consequently,

$$\mathbb{E}\|\hat{r} - r\|^2 = O\left(\frac{1}{nh_n^q} + h_n^{2s}\right).$$

Next, decompose

$$(\hat{K}\varphi_1 - K\varphi_1)(w) \triangleq V_n(w) + B_n(w),$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

where

$$V_n = \int \varphi_1(z)\left(\hat{f}_{ZW}(z,w) - \mathbb{E}\hat{f}_{ZW}(z,w)\right)\mathrm{d}z,$$
$$B_n = \int \varphi_1(z)\left(\mathbb{E}\hat{f}_{ZW}(z,w) - f_{ZW}(z,w)\right)\mathrm{d}z.$$

By the Cauchy–Schwartz inequality,

$$\|B_n\| \leq \|\varphi_1\| \left\|\mathbb{E}\hat{f}_{ZW} - f_{ZW}\right\|,$$

and the right side is of order $O(h_n^s)$ under the assumption $f_{ZW} \in C_M^s$ (see Giné and Nickl, 2016, p. 404).

Next, note that

$$V_n(w) = \frac{1}{nh_n^q}\sum_{i=1}^n \eta_{n,i}(w)$$

with

$$\eta_{n,i}(w) = K_w\left(h_n^{-1}(W_i - w)\right)\left[\varphi_1 * K_z\right](Z_i) - \mathbb{E}\left[K\left(h_n^{-1}(W_i - w)\right)\left[\varphi_1 * K_z\right](Z_i)\right],$$

where $\left[\varphi_1 * K_z\right](Z_i) = \int \varphi_1(z)h_n^{-p}K_z\left(h_n^{-1}(Z_i - z)\right)\mathrm{d}z$. Then

$$\begin{aligned}\mathbb{E}\|V_n\|^2 &\leq \frac{1}{nh_n^{2q}}\int\int\int \left|K_w(h_n^{-1}(w' - w))\right|^2 \left|\left[\varphi_1 * K_z\right](z')\right|^2 \mathrm{d}w f_{ZW}(z', w')\mathrm{d}w'\mathrm{d}z' \\ &= \frac{1}{nh_n^q}\|K_w\|^2 \int \left|\left[\varphi_1 * K_z\right](z)\right|^2 f_Z(z)\mathrm{d}z \\ &= O\left(\frac{1}{nh_n^q}\right),\end{aligned}$$

where the second line follows from a change of variables, and the last by $\|f_Z\|_\infty \leq C$ and Young's inequality. Combining all estimates, we establish the first claim.

The second claim follows from

$$\mathbb{E}\left\|\hat{K} - K\right\|^2 \leq \mathbb{E}\left\|\hat{f}_{ZW} - f_{ZW}\right\|^2,$$

and standard results on the L_2 error of kernel density estimators (see Giné and Nickl (2016, Chap. 5)). □

Proof of Theorem 4.1. We first examine the distribution of $n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle$. Define $b_n = \alpha_n(\alpha_n I + K^*K)^{-1}\varphi_1$ and observe that $(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^* = \hat{K}^*(\alpha_n I + \hat{K}\hat{K}^*)^{-1}$. Then,

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

as in the proof of Theorem 3.1, decompose

$$\begin{aligned}\langle\hat{\varphi}-\varphi_1,\mu_0\rangle &= \left\langle \hat{K}^*(\alpha_n I+KK^*)^{-1}(\hat{r}-\hat{K}\varphi_1),\mu_0\right\rangle \\ &\quad+\left\langle \hat{K}^*\left((\alpha_n I+\hat{K}\hat{K}^*)^{-1}-(\alpha_n I+KK^*)^{-1}\right)(\hat{r}-\hat{K}\varphi_1),\mu_0\right\rangle \\ &\quad+\left\langle (\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*(\hat{K}-K)b_n,\mu_0\right\rangle \\ &\quad+\left\langle (\alpha_n I+\hat{K}^*\hat{K})^{-1}(\hat{K}^*-K^*)Kb_n,\mu_0\right\rangle \\ &\quad+\langle b_n,\mu_0\rangle \\ &\triangleq I_n+II_n+III_n+IV_n+V_n.\end{aligned}$$

We will show that

$$I_n=\left\langle \alpha_n(\alpha_n I+KK^*)^{-1}(\hat{r}-\hat{K}\varphi_1),\hat{K}\mu_0\right\rangle$$

is the leading term in this decomposition and that all other terms are asymptotically negligible. Since $\mu_0\in\mathcal{N}(K)$, we have

$$(\alpha_n I+K^*K)^{-1}\mu_0=\frac{1}{\alpha_n}\mu_0. \tag{A.2}$$

Thus,

$$\begin{aligned}II_n &= \left\langle (\alpha_n I+\hat{K}\hat{K}^*)^{-1}(KK^*-\hat{K}\hat{K}^*)(\alpha_n I+KK^*)^{-1}(\hat{r}-\hat{K}\varphi_1),\hat{K}\mu_0\right\rangle \\ &= \left\langle (\alpha_n I+\hat{K}\hat{K}^*)^{-1}\hat{K}(K^*-\hat{K}^*)(\alpha_n I+KK^*)^{-1}(\hat{r}-\hat{K}\varphi_1),\hat{K}\mu_0\right\rangle \\ &\quad+\left\langle (\alpha_n I+\hat{K}\hat{K}^*)^{-1}(K-\hat{K})K^*(\alpha_n I+KK^*)^{-1}(\hat{r}-\hat{K}\varphi_1),\hat{K}\mu_0\right\rangle.\end{aligned}$$

Using the Cauchy–Schwartz inequality and similar computations to those in the proof of Theorem 3.1

$$\begin{aligned}II_n &\le \left\|(\alpha_n I+\hat{K}\hat{K}^*)^{-1}\hat{K}\right\|\left\|K^*-\hat{K}^*\right\|\left\|(\alpha_n I+KK^*)^{-1}\right\|\left\|\hat{r}-\hat{K}\varphi_1\right\|\left\|\hat{K}\mu_0\right\| \\ &\quad+\left\|(\alpha_n I+\hat{K}\hat{K}^*)^{-1}\right\|\left\|K-\hat{K}\right\|\left\|K^*(\alpha_n I+KK^*)^{-1}\right\|\left\|\hat{r}-\hat{K}\varphi_1\right\|\left\|(\hat{K}-K)\mu_0\right\| \\ &\le \frac{1}{\alpha_n^{3/2}}\left\|\hat{K}-K\right\|^2\left\|\hat{r}-\hat{K}\varphi_1\right\|\|\mu_0\|.\end{aligned}$$

Next, under Assumption 3.1 from the proof of Theorem 3.1, we also know that $\|b_n\|=O(\alpha_n^{(\beta/2)\wedge 1})$ and $\|Kb_n\|=O(\alpha_n^{((\beta+1)/2)\wedge 1})$, so

$$\begin{aligned}III_n &\le \left\|(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|\left\|\hat{K}-K\right\|\|b_n\|\|\mu_0\| \\ &\lesssim \frac{1}{\alpha_n^{1/2}}\left\|\hat{K}-K\right\|\alpha_n^{\frac{\beta}{2}\wedge 1}\end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

and

$$IV_n \leq \left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\right\| \left\|\hat{K}^* - K^*\right\| \|Kb_n\| \|\mu_0\|$$
$$\lesssim \frac{1}{\alpha_n^{1/2}} \left\|\hat{K} - K\right\| \alpha_n^{\frac{\beta\wedge 1}{2}}.$$

Lastly, the bias term $\langle b_n, \mu_0\rangle$ is zero due to equation (A.2) and the orthogonality between φ_1 and μ_0

$$\langle b_n, \mu_0\rangle = \left\langle \varphi_1, \alpha_n(\alpha_n I + K^*K)^{-1}\mu_0 \right\rangle = \langle \varphi_1, \mu_0\rangle = 0.$$

It follows from the discussion in Section 3 under Assumption 4.1 that

$$\left\|\hat{K} - K\right\| = O_P\left(\frac{1}{n^{1/2}}\right) \quad \text{and} \quad \left\|\hat{r} - \hat{K}\varphi_1\right\| = O_P\left(\frac{1}{n^{1/2}}\right).$$

Thus, since $n\alpha_n^{1+\beta\wedge 1} \to 0$ and $n\alpha_n \to \infty$ under Assumption 4.2,

$$n\alpha_n\langle \hat{\varphi} - \varphi_1, \mu_0\rangle = n\alpha_n \left\langle (\alpha_n I + KK^*)^{-1}(\hat{r} - \hat{K}\varphi_1), \hat{K}\mu_0 \right\rangle + o_P(1)$$
$$\triangleq S_n + o_P(1),$$

where

$$S_n = n\alpha_n \left\langle (\alpha_n I + KK^*)^{-1}(\hat{r} - \hat{K}\varphi_1), \hat{K}\mu_0 \right\rangle.$$

Next, decompose $S_n = S_n^0 + S_n^1$ with

$$S_n^0 = \frac{1}{n}\sum_{i,j=1}^n (Y_i - \langle Z_i, \varphi_1\rangle)\langle Z_j, \mu_0\rangle \left\langle \alpha_n(\alpha_n I + KK^*)^{-1} W_i^0, W_j^0 \right\rangle,$$

$$S_n^1 = \frac{1}{n}\sum_{i,j=1}^n (Y_i - \langle Z_i, \varphi_1\rangle)\langle Z_j, \mu_0\rangle \left\langle \alpha_n(\alpha_n I + KK^*)^{-1} W_i^1, W_j^1 \right\rangle.$$

Since $W_i^0 \in \mathcal{N}(K^*)$, we have $(\alpha_n I + KK^*)^{-1} W_i^0 = \frac{1}{\alpha_n} W_i^0$. Using this fact, decompose further $S_n^0 \triangleq \zeta_n^0 + \mathbf{U}_n^0$ with

$$\zeta_n^0 = \frac{1}{n}\sum_{i=1}^n (Y_i - \langle Z_i, \varphi_1\rangle)\langle Z_i, \mu_0\rangle \left\|W_i^0\right\|^2,$$

$$\mathbf{U}_n^0 = \frac{1}{n}\sum_{i<j} \left\{\langle Z_i, \mu_0\rangle(Y_j - \langle Z_j, \varphi_0\rangle) + \langle Z_j, \mu_0\rangle(Y_i - \langle Z_i, \mu_0\rangle)\right\} \left\langle W_i^0, W_j^0 \right\rangle.$$

Under Assumption 4.1 by the strong law of large numbers

$$\zeta_n^0 \xrightarrow{a.s.} \mathbb{E}\left[\left\|W^0\right\|^2 (Y - \langle Z, \varphi_1\rangle)\langle Z, \mu_0\rangle\right].$$

Next, note that $W^0 = P_0 W$ and $W^1 = (I - P_0)W$, where P_0 is the projection operator on $\mathcal{N}(K^*)$. Since projection is a bounded linear operator, it commutes with the expectation, cf., Bosq (2000, p. 29), whence $\mathbb{E}\left[W^0\langle Z, \mu_0\rangle\right] = P_0 K\mu_0 = 0$ and $\mathbb{E}\left[W^0(Y - \langle Z, \varphi_1\rangle)\right] =$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

$P_0\mathbb{E}[WU]+P_0K\varphi_0=0$. Therefore, $\mathbf{U}_n^0$ is a centered degenerate U-statistic with the kernel function h. Under Assumption 4.1 by the CLT for degenerate U-statistics, see Gregory (1977),

$$\mathbf{U}_n^0 \xrightarrow{d} \sum_{j=1}^{\infty}\lambda_j(\chi_j^2-1).$$

It remains to show that $S_n^1=o_P(1)$. To that end decompose $S_n^1=\zeta_n^1+\mathbf{U}_n^1$ with

$$\zeta_n^1=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\langle Z_i,\varphi_1\rangle)\langle Z_i,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_i^1,W_i^1\right\rangle,$$
$$\mathbf{U}_n^1=\frac{1}{n}\sum_{i\neq j}(Y_i-\langle Z_i,\varphi_1\rangle)\langle Z_j,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_i^1,W_j^1\right\rangle.$$

It follows from Bakushinskii (1967) that $\left\|\alpha_n(\alpha_nI+KK^*)^{-1}W^1\right\|=o(1)$. Then under Assumption 4.1 by the dominated convergence theorem

$$\begin{aligned}\mathbb{E}\left|\zeta_n^1\right| &\le \|\mu_0\|\mathbb{E}\left[(|U|+\|Z\|\|\varphi_0\|)\|Z\|\left\|\alpha_n(\alpha_nI+KK^*)^{-1}W^1\right\|\right]\\ &\lesssim \mathbb{E}\left[(\|UZ\|+\|Z\|^2)\|W\|\left\|\alpha_n(\alpha_nI+KK^*)^{-1}W^1\right\|\right]\\ &= o(1),\end{aligned}$$

whence by Markov's inequality $\zeta_n^1=o_P(1)$. Lastly, note that

$$\begin{aligned}\mathbf{U}_n^1=\frac{1}{n}\sum_{i<j}\Big\{&(Y_i-\langle Z_i,\varphi_1\rangle)\langle Z_j,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_i^1,W_j^1\right\rangle\\ &+(Y_j-\langle Z_j,\varphi_1\rangle)\langle Z_i,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_j^1,W_i^1\right\rangle\Big\}\end{aligned}$$

is a centered degenerate U-statistic. Then by the moment inequality in Korolyuk and Borovskich (1994, Thm. 2.1.3),

$$\begin{aligned}\mathbb{E}\left|\mathbf{U}_n^1\right|^2 &\le 2^{-1}\mathbb{E}\left|(U_1+\langle Z_1,\varphi_0\rangle)\langle Z_2,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_1^1,W_2^1\right\rangle\right|^2\\ &\quad+2^{-1}\mathbb{E}\left|(U_2+\langle Z_2,\varphi_0\rangle)\langle Z_1,\mu_0\rangle\left\langle\alpha_n(\alpha_nI+KK^*)^{-1}W_2^1,W_1^1\right\rangle\right|^2\\ &\lesssim \mathbb{E}\left[\|Z\|^2\left\|\alpha_n(\alpha_nI+KK^*)^{-1}W^1\right\|^2\right]=o(1),\end{aligned}$$

where the last line follows under Assumptions 4.1 and 4.2, and previous discussions. This shows that

$$n\alpha_n\langle\hat{\varphi}-\varphi_1,\mu_0\rangle \xrightarrow{d} \mathbb{E}\left[\|W\|^2(Y-\langle Z,\varphi_1\rangle)\langle Z,\mu_0\rangle\right]+\sum_{j\ge1}\lambda_j(\chi_j^2-1).$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Next, we focus on the distribution of $\pi_n\langle\hat{\varphi}-\varphi_1,\mu_1\rangle$. Decompose

$$\begin{aligned}\langle\hat{\varphi}-\varphi_1,\mu_1\rangle &= \left\langle(\alpha_n I+K^*K)^{-1}K^*(\hat{r}-\hat{K}\varphi_1),\mu_1\right\rangle\\ &\quad+\left\langle\left\{(\alpha_n I+\hat{K}^*\hat{K})^{-1}-(\alpha_n I+K^*K)^{-1}\right\}\hat{K}^*(\hat{r}-\hat{K}\varphi_1),\mu_1\right\rangle\\ &\quad+\left\langle(\alpha_n I+K^*K)^{-1}(\hat{K}^*-K^*)(\hat{r}-\hat{K}\varphi_1),\mu_1\right\rangle\\ &\quad+\left\langle(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{K}\varphi_1-(\alpha_n I+K^*K)^{-1}K^*K\varphi_1,\mu_1\right\rangle\\ &\quad+\langle b_n,\mu_1\rangle\\ &\triangleq I_n'+II_n'+III_n'+IV_n'+\langle b_n,\mu_1\rangle.\end{aligned}$$

Put $\eta_n=(Y-\langle Z,\varphi_1\rangle)(\alpha_n I+K^*K)^{-1}K^*W$ and note that

$$\begin{aligned}\mathrm{Var}(\langle\eta_n,\mu_1\rangle)&=\mathbb{E}\left|\langle(Y-\langle Z,\varphi_1\rangle W,K(\alpha_n I+K^*K)^{-1}\mu_1\rangle\right|^2\\ &=\langle\Sigma K(\alpha_n I+K^*K)^{-1}\mu_1,K(\alpha_n I+K^*K)^{-1}\mu_1\rangle\\ &=\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\|^2.\end{aligned}$$

Suppose that for all $\epsilon>0$, the following Lindeberg's condition is satisfied

$$\lim_{n\to\infty}\frac{\pi_n^2}{n}\mathbb{E}\left[|\langle\eta_n,\mu_1\rangle|^2\mathbb{1}_{\{\pi_n|\langle\eta_n,\mu_1\rangle|\geq\epsilon n\}}\right]=0 \tag{A.3}$$

with $\pi_n=n^{1/2}\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|^{-1}$. Then by the Lindeberg–Feller central limit theorem

$$\begin{aligned}\pi_n I_n'&=\frac{\pi_n}{n}\sum_{i=1}^n(U_i+\langle Z_i,\varphi_0\rangle)\left\langle(\alpha_n I+K^*K)^{-1}K^*W_i,\mu_1\right\rangle\\ &\xrightarrow{d}N(0,1).\end{aligned}$$

and the result follows provided that all other terms are asymptotically negligible. To see that Lindeberg's condition in equation (A.3) is satisfied, note that for every $\delta>0$,

$$\mathbb{E}\left[|\langle\eta_n,\mu_1\rangle|^2\mathbb{1}_{\{\pi_n|\langle\eta_n,\mu_1\rangle|\geq\epsilon n\}}\right]\leq\frac{\pi_n^\delta}{\epsilon^\delta n^\delta}\mathbb{E}|\langle\eta_n,\mu_1\rangle|^{2+\delta}$$

and that $\pi_n\sim n^{-c}$ with $c\in(0,1/2]$ depending on the mapping properties of operators K and Σ. Therefore, the Lindeberg condition is satisfied provided that $\mathbb{E}|\langle\eta_n,\mu_1\rangle|^{2+\delta}=O(1)$. This is easily verified under Assumption 4.1 since

$$\begin{aligned}|\langle\eta_n,\mu_1\rangle|&\lesssim|U+\langle Z,\varphi_0\rangle|\left\|(K^*K)^{\tilde{\gamma}}(\alpha_n I+K^*K)^{-1}K^*(K^*K)^{\gamma}\right\|\\ &\lesssim|U|+|\langle Z,\varphi_0\rangle|.\end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Therefore, it remains to show that all other terms normalized with π_n are asymptotically negligible. For II_n', by the Cauchy–Schwartz inequality

$$II_n' = \left\langle \frac{1}{n}\sum_{i=1}^n W_i(U_i + \langle Z_i, \varphi_0\rangle), \hat{K}^* \left((\alpha_n I + \hat{K}^*\hat{K})^{-1} - (\alpha_n I + K^*K)^{-1} \right) \mu_1 \right\rangle$$
$$\leq \left\| \frac{1}{n}\sum_{i=1}^n W_i(U_i + \langle Z_i, \varphi_0\rangle) \right\| \left\| \hat{K}^*(\alpha_n I + \hat{K}^*\hat{K})^{-1}(\hat{K}^*\hat{K} - K^*K)(\alpha_n I + K^*K)^{-1}\mu_1 \right\|.$$

Since $\mu_1 \in \mathcal{R}\left[(K^*K)^\gamma\right]$, there exists some $\psi \in \mathcal{E}$ such that $\mu_1 = (K^*K)^\gamma \psi$ and so

$$\begin{aligned} II_n' &\lesssim_P n^{-1/2} \left\| \hat{K}^*(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^* \right\| \left\| \hat{K} - K \right\| \left\| (\alpha_n I + K^*K)^{-1}(K^*K)^\gamma \psi \right\| \\ &\quad + n^{-1/2} \left\| \hat{K}^*(\alpha_n I + \hat{K}^*\hat{K})^{-1} \right\| \left\| \hat{K}^* - K^* \right\| \left\| K(\alpha_n I + K^*K)^{-1}(K^*K)^\gamma \psi \right\| \\ &\lesssim_P n^{-1} \left\| (\alpha_n I + K^*K)^{-1}(K^*K)^\gamma \psi \right\| + n^{-1}\alpha_n^{-1/2} \left\| K(\alpha_n I + K^*K)^{-1}(K^*K)^\gamma \psi \right\| \\ &\lesssim_P n^{-1}\alpha_n^{\gamma \wedge 1 - 1} + n^{-1}\alpha_n^{\gamma \wedge 1/2 - 1} = o_P(\pi_n^{-1}), \end{aligned}$$

where the last line follows under Assumption 4.2. Similarly,

$$\begin{aligned} III_n' &\leq \left\| \hat{K}^* - K^* \right\| \left\| \frac{1}{n}\sum_{i=1}^n W_i(U_i + \langle Z_i, \varphi_0\rangle) \right\| \left\| (\alpha_n I + K^*K)^{-1}(K^*K)^\gamma \psi \right\| \\ &\lesssim_P n^{-1}\alpha_n^{\gamma \wedge 1 - 1} \\ &= o_P(\pi_n^{-1}). \end{aligned}$$

Next, decompose

$$\begin{aligned} IV_n' &= \left\langle (\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{K}\varphi_1 - (\alpha_n I + K^*K)^{-1}K^*K\varphi_1, \mu_1 \right\rangle \\ &= \left\langle \alpha_n(\alpha_n I + \hat{K}^*\hat{K})^{-1}\left[\hat{K}^*\hat{K} - K^*K\right](\alpha_n I + K^*K)^{-1}\varphi_1, \mu_1 \right\rangle \\ &= \left\langle (\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*(\hat{K} - K)b_n, \mu_1 \right\rangle \\ &\quad + \left\langle (\alpha_n I + \hat{K}^*\hat{K})^{-1}(\hat{K}^* - K^*)Kb_n, \mu_1 \right\rangle \\ &\triangleq IV_n^a + IV_n^b + IV_n^c + IV_n^d + IV_n^e \end{aligned}$$

with

$$IV_n^a = \left\langle \left\{ (\alpha_n I + \hat{K}^*\hat{K})^{-1} - (\alpha_n I + K^*K)^{-1} \right\} \hat{K}^*(\hat{K} - K)b_n, \mu_1 \right\rangle$$
$$IV_n^b = \left\langle \left\{ (\alpha_n I + \hat{K}^*\hat{K})^{-1} - (\alpha_n I + K^*K)^{-1} \right\} (\hat{K}^* - K^*)Kb_n, \mu_1 \right\rangle$$
$$IV_n^c = \left\langle (\alpha_n I + K^*K)^{-1}K^*(\hat{K} - K)b_n, \mu_1 \right\rangle$$
$$IV_n^d = \left\langle (\alpha_n I + K^*K)^{-1}(\hat{K}^* - K^*)(\hat{K} - K)b_n, \mu_1 \right\rangle$$
$$IV_n^e = \left\langle (\alpha_n I + K^*K)^{-1}(\hat{K}^* - K^*)Kb_n, \mu_1 \right\rangle.$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

We bound the last three terms by the Cauchy–Schwartz inequality

$$IV_n^c \le \left\|\hat{K}-K\right\| \|b_n\| \left\|K(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1+\gamma\wedge\frac{1}{2}}}{\sqrt{n\alpha_n}}$$

$$IV_n^d \le \left\|\hat{K}^*-K^*\right\| \left\|\hat{K}-K\right\| \|b_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1+\gamma\wedge 1}}{n\alpha_n}$$

$$IV_n^e \le \left\|\hat{K}^*-K^*\right\| \|Kb_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2}+\gamma\wedge 1}}{\sqrt{n\alpha_n}}.$$

Next, for the first two terms, by the Cauchy–Schwartz inequality, we have

$$\begin{aligned} IV_n^a &\le \left\|\hat{K}-K\right\|^2 \left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\right\| \|b_n\| \left\|K(\alpha_n I + K^*K)^{-1}\mu_1\right\| \\ &\quad + \left\|\hat{K}^*-K^*\right\| \left\|\hat{K}(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\right\| \left\|\hat{K}-K\right\| \|b_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \\ &\lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1+\gamma\wedge\frac{1}{2}}}{n\alpha_n} + \frac{\alpha_n^{\frac{\beta}{2}\wedge 1+\gamma\wedge 1}}{n\alpha_n} \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1+\gamma\wedge\frac{1}{2}}}{n\alpha_n} \end{aligned}$$

and

$$\begin{aligned} IV_n^b &\le \left\|\hat{K}-K\right\| \left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\right\| \left\|\hat{K}^*-K^*\right\| \|Kb_n\| \left\|K(\alpha_n I + K^*K)^{-1}\mu_1\right\| \\ &\quad + \left\|\hat{K}^*-K^*\right\|^2 \left\|\hat{K}(\alpha_n I + \hat{K}^*\hat{K})^{-1}\right\| \|Kb_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \\ &\lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2}+\gamma\wedge\frac{1}{2}}}{n\alpha_n} + \frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2}+\gamma\wedge 1}}{n\alpha_n} \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2}+\gamma\wedge\frac{1}{2}}}{n\alpha_n}. \end{aligned}$$

Lastly,

$$\pi_n\langle b_n, \mu_1\rangle \lesssim \pi_n\|(K^*K)^\gamma b_n\| \lesssim \pi_n\alpha_n^{(\gamma+\beta/2)\wedge 1}.$$

This completes the proof. □

A.3. PARTIAL IDENTIFICATION

In this section, we discuss how our results can be applied to partial identification, drawing on ideas from Freyberger (2017) and Santos (2012). This section is intended primarily for illustrative purposes and a detailed exploration of the partial identification approach is left for future research.

Define the L_∞-diameter of the identified set as

$$|I_0| \triangleq \sup_{\phi_1,\phi_2\in I_0} \|\phi_1-\phi_2\|_\infty.$$

Suppose that both φ and φ_1 belong to the Lipschitz smoothness class

$$\mathscr{F} \triangleq \left\{\phi\in L_\infty([0,1]^p):\ \|\phi\|_s \triangleq \|\phi\|_\infty + \sup_{z_1\neq z_2}\frac{|\phi(z_1)-\phi(z_2)|}{\|z_1-z_2\|} \le C\right\}.$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

If the best approximation φ_1 and the diameter of the identified set $|I_0|$ are known, then $\varphi_1 \pm |I_0|$ is a valid identified set for the structural function φ. Our results establish that we can estimate the best approximation consistently. The remaining task is to infer the diameter of the identified set.

To this end, for a fixed $\varepsilon > 0$, we consider the following hypotheses:

$$H_0 : |I_0| \geq \varepsilon \qquad \text{vs.} \qquad H_1 : |I_0| < \varepsilon$$

following the approach in Freyberger (2017) to construct a test statistic. Under H_0, there exist functions $\phi_1, \phi_2 \in I_0$ such that $\|\phi_1 - \phi_2\|_\infty \geq \varepsilon$. For $\phi = \phi_1 - \phi_2$, we have $K\phi = 0$, $\|\phi\|_s \leq 2C$, and $\|\phi\|_\infty \geq \varepsilon$. This motivates the following statistic

$$T_\varepsilon \triangleq \inf_{\phi : \|\phi\|_s \leq 2C,\, \|\phi\|_\infty \geq \varepsilon} \|K\phi\|^2 = \inf_{\phi \in \mathscr{F}_\varepsilon,\, \|\phi\|_\infty = 1} \|K\phi\|^2.$$

where $\mathscr{F}_\varepsilon = \left\{\phi \in L_\infty([0,1]^p) : \|\phi\|_s \leq 2C/\varepsilon\right\}$. Clearly, under H_0, we have $T_\varepsilon = 0$. Define the sample counterpart as

$$\hat{T}_\varepsilon = \inf_{\phi \in \mathscr{F}_\varepsilon :\, \|\phi\|_\infty = 1} \|\hat{K}\phi\|^2,$$

where $\hat{K}$ is a consistent estimator of K.

For NPIV regression, suppose that the density $\hat{f}_{ZW}$ is estimated using a series estimator, and the statistic $\hat{T}_\varepsilon$ is computed over a J-dimensional sieve. Freyberger (2017, Thm. 2) shows that

$$\hat{\varepsilon} = \sup\left\{\varepsilon \in [0, \bar{C}] : n\hat{T}_\varepsilon \leq q_{1-\alpha}\right\}$$

consistently estimates the upper bound on $|I_0|$, where $\bar{C}$ is the largest ε for which $\{\phi : \|\phi\|_s \leq 2C, \|\phi\|_\infty \geq \varepsilon\} \neq \emptyset$ and $q_{1-\alpha}$ is a critical value.

Combining this result with our result on the L_∞ consistency for the best approximation, the interval $[\hat{\varphi} - \hat{\varepsilon}, \hat{\varphi} + \hat{\varepsilon}]$ serves as a valid set estimator for φ.[11]

A.4. EXTREME NONIDENTIFICATION

In this section, we derive an approximation of the large sample distribution of the Tikhonov-regularized estimators in cases of extreme nonidentification. Remarkably, we show that the asymptotic distribution is a weighted sum of independent chi-squared random variables. This result serves as a foundation for Section 4, where we analyze a transition between chi-squared and Gaussian limits in intermediate cases.[12]

A.4.1. High-dimensional Regressions

In high-dimensional regressions, the strength of identification is determined by the covariance operator of Z and W. In cases of extreme nonidentification, this operator is degenerate, leading to the following result.

[11] We thank the anonymous referee for highlighting this connection.

[12] Extreme nonidentification also relates to the weak instruments problem.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

THEOREM A.1. *Suppose that Assumption 4.1 holds,* $\mathbb{E}[\langle Z,\delta\rangle W]=0$, $\forall\delta\in\mathcal{E}$, *and* $\alpha_n n\to\infty$. *Then*

$$\alpha_n n(\hat{\varphi}-\varphi_1)\xrightarrow{d}\mathbb{E}\left[\|W\|^2YZ\right]+J(h),$$

where $h(X,X')=\frac{1}{2}\langle W,W'\rangle(ZY'+Z'Y)$, $X'=(Y',Z',W')$ *is an independent copy of* $X=(Y,Z,W)$, *and* J *is a stochastic Wiener–Itô integral.*

The theorem states weak convergence in the topology of the Hilbert space $\mathcal{E}$, which cannot be achieved in cases of regular identification. Furthermore, it can be shown that the distribution of the inner products of $J(h)$ with $\mu\in\mathcal{E}$ is a weighted sum of chi-squared random variables. Interestingly, Theorem A.1 does not require that $\alpha_n\to 0$ as $n\to\infty$.

Proof of Theorem A.1. Given that $\mathbb{E}[\langle Z,\delta\rangle W]=0$ for all $\delta\in\mathcal{E}$, we have $\varphi_1=0$. Thus,

$$\alpha_n n\left(\hat{\varphi}-\varphi_1\right)=\left(I+\frac{1}{\alpha_n}\hat{K}^*\hat{K}\right)^{-1}n\hat{K}^*\hat{r}.$$

Under Assumption 4.1

$$\mathbb{E}\|\hat{K}\|^2=\mathbb{E}\|\hat{K}-K\|^2\le\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n Z_iW_i-\mathbb{E}[ZW]\right\|^2=O\left(\frac{1}{n}\right).$$

Thus, $\|\hat{K}^*\hat{K}\|\le\|\hat{K}\|^2=O_P\left(n^{-1}\right)$. Since $\alpha_n n\to\infty$, the continuous mapping theorem (see van der Vaart and Wellner, (2000, Thm. 1.3.6)) implies that

$$\alpha_n n\left(\hat{\varphi}-\varphi_1\right)=(I+o_P(1))^{-1}n\hat{K}^*\hat{r}.$$

Applying Slutsky's theorem (see van der Vaart and Wellner, (2000, Example 1.4.7)), it suffices to determine the asymptotic distribution of $n\hat{K}^*\hat{r}$.

Observe that

$$\begin{aligned}n\hat{K}^*\hat{r}&=\frac{1}{n}\sum_{i,j=1}^n\langle W_i,W_j\rangle Z_iY_j\\&=\frac{1}{n}\sum_{i=1}^n\|W_i\|^2Z_iY_i+\frac{1}{n}\sum_{i\neq j}\langle W_i,W_j\rangle Z_iY_j.\end{aligned}$$

Under Assumption 4.1, by the Mourier law of large numbers

$$\frac{1}{n}\sum_{i=1}^n\|W_i\|^2Z_iY_i\xrightarrow{a.s.}\mathbb{E}\left[\|W\|^2ZY\right].$$

Since $\mathbb{E}[\langle Z,\delta\rangle W]=0,\forall\delta\in\mathcal{E}$, the second term forms a Hilbert space-valued degenerate U-statistic

$$\begin{aligned}n\mathbf{U}_n&\triangleq\frac{1}{n}\sum_{i\neq j}\langle W_i,W_j\rangle Z_iY_j\\&=\frac{2}{n}\sum_{i<j}\frac{Z_iY_j+Z_jY_i}{2}\langle W_i,W_j\rangle.\end{aligned}$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Under the Assumption 4.1, the Borovskich CLT, see Theorem S.2.1 in the Supplementary Material, gives

$$n\mathbf{U}_n \xrightarrow{d} J(h),$$

where $J(h) = \iint_{\mathcal{X}\times\mathcal{X}} h(x_1,x_2)\mathbb{W}(\mathrm{d}x_1)\mathbb{W}(\mathrm{d}x_2)$ is a stochastic Wiener–Itô integral, $\mathbb{W}$ is a Gaussian random measure on $\mathcal{X}$, $h(X,X') = \frac{ZY'+Z'Y}{2}\langle W,W'\rangle$, and $X' = (Y',Z',W')$ is an independent copy of $X = (Y,Z,W)$. □

A.4.2. Nonparametric IV Regression

In nonparametric IV regression, the strength of identification is described by the conditional expectation operator. In the extreme nonidentified case,

$$\mathbb{E}[\phi(Z)|W] = 0, \qquad \forall\phi \in L_{2,0}(Z),$$

where $L_{2,0}(Z) = \{\phi \in L_2(Z) : \mathbb{E}\phi(Z) = 0\}$, and K is a degenerate conditional expectation operator. Consider the operator $T : \phi \mapsto \mathbb{E}_X[\phi(X)h(X,X')]$ on $L_2(X)$, where $\mathbb{E}_X$ denotes expectation with respect to $X = (Y,Z,W)$ only, X' is an independent copy of X, and

$$h(x,x') = \frac{1}{2}\left\{yP_0\mu(z') + y'P_0\mu(z)\right\}h_w^{-q}\bar{K}\left(h_w^{-1}(w-w')\right).$$

Here, P_0 is the projection operator on $L_{2,0}$ and $\bar{K}(v) = \int K_w(v-u)K_w(u)\mathrm{d}u$ is the convolution kernel. We assume the following mild conditions on the distribution of the data:

Assumption A.4.1. (i) $(Y_i,Z_i,W_i)_{i=1}^n$ is an i.i.d. sample of (Y,Z,W); (ii) $\mathbb{E}[|Y||Z] < \infty$, $\mathbb{E}[|Y|^2|W] < \infty$ a.s.; (iii) $K_j \in L_1 \cap L_2, j \in \{z,w\}$ and K_w is a symmetric and bounded function; (iv) $f_Z \in L_\infty$.

Let h_z and h_w be the bandwidth parameters for smoothing over Z and W, respectively. The following result holds.

THEOREM A.2. *Suppose that Assumption A.4.1 holds,* $\mathbb{E}[\phi(Z)|W] = 0, \forall\phi \in L_{2,0}(Z)$, *and* $n\alpha_n h_z^p \to \infty$ *with* h_w *fixed. Then for every* $\mu \in L_2([0,1]^p)$,

$$\alpha_n n\langle\hat{\varphi} - \varphi_1, \mu\rangle \xrightarrow{d} \mathbb{E}\left[YP_0\mu(Z)\right]h_w^{-q}\bar{K}(0) + \sum_{j=1}^{\infty}\lambda_j(\chi_j^2 - 1),$$

where $(\chi_j^2)_{j\geq 1}$ *are independent chi-squared random variables with 1 degree of freedom and* $(\lambda_j)_{j\geq 1}$ *are eigenvalues of* T.

Proof of Theorem A.2. Since $\mathbb{E}[\phi(Z)|W] = 0, \forall\phi \in L_{2,0}(Z)$, we find that $\varphi_1 = 0$. Additionally, the adjoint operator of K is P_0K^*, where P_0 is the orthogonal projection on $L_{2,0}(Z)$. Thus,

$$\alpha_n n(\hat{\varphi} - \varphi_1) = \left(I + \frac{1}{\alpha_n}P_0\hat{K}^*\hat{K}\right)^{-1} nP_0\hat{K}^*\hat{r},$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

where $\hat{P}_0$ is the estimator of P_0. Since $\mathbb{E}[\phi(Z)|W] = 0$ for all $\phi \in L_{2,0}(Z)$, under Assumption A.4.1 (i),

$$\begin{aligned}\mathbb{E}\left\|P_0\hat{K}^*\hat{K}\right\| &\leq \mathbb{E}\|P_0\hat{K}\|^2 \leq \mathbb{E}\|P_0\hat{f}_{ZW}\|^2 \\ &= \mathbb{E}\left\|\frac{1}{nh_z^p h_w^q}\sum_{i=1}^n P_0K_z\left(h_z^{-1}(Z_i - z)\right)K_w\left(h_z^{-1}(W_i - w)\right)\right\|^2 \\ &\leq \frac{1}{nh_z^{2p}h_w^{2q}}\mathbb{E}\left\|P_0K_z\left(h_z^{-1}(Z_i - z)\right)K_w\left(h_z^{-1}(W_i - w)\right)\right\|^2 \\ &= \frac{1}{nh_z^p h_w^q}\|P_0K_z\|\|K_w\| = O\left(\frac{1}{nh_z^p}\right).\end{aligned}$$

Therefore, as $n\alpha_n h_z^p \to \infty$, we have $\frac{1}{\alpha_n}\left\|\hat{P}_0\hat{K}^*\hat{K}\right\| = o_P(1)$. By the continuous mapping and Slutsky's theorems, it then suffices to determine the asymptotic distribution of

$$nP_0\hat{K}^*\hat{r} = \frac{1}{nh_z^p h_w^q}\sum_{i,j} Y_i P_0K_z\left(h_z^{-1}(Z_j - z)\right)\bar{K}\left(h_w^{-1}(W_i - W_j)\right).$$

For every $\mu \in L_2([0,1]^p)$, we have

$$\begin{aligned}\left\langle nP_0\hat{K}^*\hat{r}, \mu\right\rangle &= \left\langle n\hat{K}^*\hat{r}, P_0\mu\right\rangle \\ &\triangleq \zeta_n + \mathbf{U}_n + R_n,\end{aligned}$$

where

$$\begin{aligned}\zeta_n &= \frac{1}{n}\sum_{i=1}^n Y_i P_0\mu(Z_i)h_w^{-q}\bar{K}(0), \\ \mathbf{U}_n &= \frac{2}{n}\sum_{i<j}\frac{1}{2}\left\{Y_i P_0\mu(Z_j) + Y_j P_0\mu(Z_i)\right\}h_w^{-q}\bar{K}\left(h_w^{-1}(W_i - W_j)\right), \\ R_n &= \frac{1}{nh_w^q}\sum_{i,j=1}^n Y_i\left\{[K_z * P_0\mu](Z_j) - P_0\mu(Z_j)\right\}\bar{K}\left(h_n^{-1}(W_i - W_j)\right)\end{aligned}$$

with $[K_z * P_0\mu](z) = h_n^{-p}\int K\left(h_z^{-1}(z-u)\right)P_0\mu(u)\mathrm{d}u$. Under Assumption A.4.1, the strong law of large numbers gives

$$\zeta_n \xrightarrow{a.s} \mathbb{E}\left[YP_0\mu(Z)\right]h_w^{-q}\bar{K}(0).$$

Since $\mathbb{E}[\phi(Z)|W] = 0, \forall\phi \in L_{2,0}(Z)$, $\mathbf{U}_n$ is a centered degenerate U-statistic. By the central limit theorem for degenerate U-statistics, see Gregory (1977),

$$\mathbf{U}_n = \frac{2}{n}\sum_{i<j} h(X_i, X_j) \xrightarrow{d} \sum_{j=1}^\infty \lambda_j(\chi_j^2 - 1).$$

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Finally, decompose $R_n = R_{1n} + R_{2n}$, where

$$R_{1n} = \frac{1}{n}\sum_{i=1}^{n} Y_i \{[K_z * P_0\mu](Z_i) - P_0\mu(Z_i)\} h_w^{-q} \bar{K}(0),$$

$$R_{2n} = \frac{1}{n}\sum_{i<j} Y_i \left\{[K_z * P_0\mu](Z_j) - P_0\mu(Z_j)\right\} h_w^{-q} \bar{K}\left(h_w^{-1}(W_i - W_j)\right).$$

Note that

$$\begin{aligned}
\mathbb{E}|R_{1n}| &\leq \mathbb{E}\,|Y\{[K_z * P_0\mu](Z) - P_0\mu(Z)\}|\, h_w^{-q}\bar{K}(0) \\
&\lesssim \int |[K_z * P_0\mu](z) - P_0\mu(z)| f_Z(z)\mathrm{d}z \\
&\leq \|K_z * \mu - \mu\|^2 \|f_Z\|^2 \\
&= o(1),
\end{aligned}$$

where the first two lines follow from Assumption A.4.1 (i)–(ii), the third by the Cauchy–Schwartz inequality and $\|P_0\| \leq 1$, and the last by Giné and Nickl (2016, Prop. 4.1.1(iii)). Similarly, since $\mathbb{E}\left[|Y|^2|W\right] < \infty$ a.s. and $\bar{K} \in L_\infty$, we use the moment inequality in Korolyuk and Borovskich (1994, Thm. 2.1.3), to obtain

$$\begin{aligned}
\mathbb{E}|R_{2n}|^2 &\lesssim \mathbb{E}\left|Y\left\{[K_z * P_0\mu](Z') - P_0\mu(Z')\right\} h_w^{-q}\bar{K}\left(h_w^{-1}(W - W')\right)\right|^2 \\
&\lesssim \int |[K_z * P_0\mu](z) - P_0\mu(z)| f_Z(z)\mathrm{d}z = o(1).
\end{aligned}$$

This completes the proof. □

SUPPLEMENTAL MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466625000039.

REFERENCES

Andrews, D. W. (2017). Examples of ℓ^2-complete and boundedly-complete distributions. *Journal of Econometrics*, 199(2), 213–220.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Babii, A. (2020). Honest confidence sets in nonparametric iv regression and other ill-posed models. *Econometric Theory*, 36(4), 658–706.

Babii, A. (2022). High-dimensional mixed-frequency iv regression. *Journal of Business & Economic Statistics*, 4(40), 1470–1483.

Babii, A., & Florens, J.-P. (2024). Are unobservables separable? *Econometric Theory*, first published online Jan 26, 1–33.

Babii, A., Carrasco, M., & Tsafack, I. (2024). Functional partial least-squares: optimal rates and adaptation. arXiv preprint arXiv:2402.11134.

Bakushinskii, A. B. (1967). A general method of constructing regularizing algorithms for a linear incorrect equation in hilbert space. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(3), 672–677.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

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

Bontemps, C., Florens, J.-P., & Meddahi, N. (2025). Functional ecological inference. *Journal of Econometrics*, 105918. https://doi.org/10.1016/j.jeconom.2024.105918.

Bosq, D. (2000). *Linear processes in function spaces*. Springer.

Canay, I. A., Santos, A., & Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6), 2535–2559.

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

Carrasco, M., Florens, J.-P., & Renault, E. (2014). Asymptotic normal inference in linear inverse problems. In J. S. Racine (ed.), *Handbook of applied nonparametric and semiparametric econometrics and statistics* (pp. 65–96). Oxford University Press. https://10.1093/oxfordhb/9780199857944.013.003.

Centorrino, S. (2014). Data driven selection of the regularization parameter in additive nonparametric instrumental regressions. In *Economics department*. Stony Brook University.

Chen, X., & Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, 9(1), 39–84.

Chen, X., & Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1), 277–321.

Chernozhukov, V., Newey, W. K., & Santos, A. (2023). Constrained conditional moment restriction models. *Econometrica*, 91(2), 709–736.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics*, 13(2), 545–591.

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

D'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3), 460–471.

Egger, H. (2005). Accelerated newton-landweber iterations for regularizing nonlinear inverse problems. RICAM-Report, 1.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Springer.

Escanciano, J. C., & Li, W. (2021). Optimal linear instrumental variables approximations. *Journal of Econometrics*, 221(1), 223–246.

Florens, J.-P., & Van Bellegem, S. (2015). Instrumental variable estimation in functional linear models. *Journal of Econometrics*, 186(2), 465–476.

Florens, J.-P., Johannes, J., & Van Bellegem, S. (2011). Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(3), 472–496.

Freyberger, J. (2017). On completeness and consistency in nonparametric instrumental variable models. *Econometrica*, 85(5), 1629–1644.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.

Gagliardini, P., & Scaillet, O. (2012). Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167(1), 61–75.

Giné, E., & Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press.

Gregory, G. G. (1977). Large sample theory for u-statistics and tests of fit. *Annals of Statistics*, 5(1), 110–123.

Hall, P., & Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*, 33(6), 2904–2929.

Hu, Y., & Shiu, J.-L. (2018). Nonparametric identification using instrumental variables: Sufficient conditions for completeness. *Econometric Theory*, 34(3), 659–693.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica*, 17(2), 125–144.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press

Koopmans, T. C., & Reiersol, O. (1950). The identification of structural characteristics. *Annals of Mathematical Statistics*, 21(2): 165–181.

Korolyuk, V. S., & Borovskich, Y. V. (1994). *Theory of U-statistics*. Springer Science & Business Media.

Kress, R. (2014). *Linear integral equations*. Springer Science & Business Media.

Luenberger, D. G. (1997). *Optimization by vector space methods*. John Wiley & Sons.

Mathé, P., & Pereverzev, S. V. (2002). Moduli of continuity for operator valued functions. *Numerical Functional Analysis and Optimization*, 23(5–6), 623–631.

Nair, M. T. (2009). *Linear operator equations: Approximation and regularization*. World Scientific Publishing Company.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5), 1565–1578.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, 39(3), 577–591.

Rudin, W. (1991). *Functional analysis*. McGraw-Hill.

Santos, A. (2011). Instrumental variable methods for recovering continuous linear functionals. *Journal of Econometrics*, 161(2), 129–146.

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

Severini, T. A., & Tripathi, G. (2006). Some identification issues in nonparametric linear models with endogenous regressors. *Econometric Theory*, 22(2), 258–278.

Severini, T. A., & Tripathi, G. (2012). Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors. *Journal of Econometrics*, 170(2), 491–498.

Shorack, G. R., & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons.

Tikhonov, A. N. (1963). On the solution of ill-posed problems and the method of regularization. *Doklady Akademii Nauk*, 151(3), 501–504.

Tsybakov, A., Bickel, P., & Ritov, Y. (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37(4), 1705–1732.

van der Vaart, A. W., & Wellner, J. A. (2000). *Weak convergence and empirical processes*. Springer.

Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470), 577–590.

https://doi.org/10.1017/S0266466625000039 Published online by Cambridge University Press