Is completeness necessary? Estimation in nonidentified linear models*

Andrii Babii† Jean-Pierre Florens‡
UNC Chapel Hill *Toulouse School of Economics*

November 9, 2021

Abstract

We show that estimators based on spectral regularization converge to the best approximation of a structural parameter in a class of nonidentified linear ill-posed inverse models. Importantly, this convergence holds in the uniform and Hilbert space norms. We describe several circumstances when the best approximation coincides with a structural parameter, or at least reasonably approximates it, and discuss how our results can be useful in the partial identification setting. Lastly, we document that identification failures have important implications for the asymptotic distribution of a linear functional of regularized estimators, which can have a weighted chi-squared component. The theory is illustrated for various high-dimensional and nonparametric IV regressions.

*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). We are grateful to the Co-Editor Ivan Fernández-Val and three anonymous referees whose comments helped us to improve the paper. We are also grateful to Alex Belloni, Irene Botosaru, Christoph Breunig, Federico Bugni, Eric Ghysels, Joel Horowitz, Pascal Lavergne, Thierry Magnac, Matt Masten, Nour Meddahi, Whitney Newey, Jia Li, Adam Rosen, and other participants of Duke workshop, TSE Econometrics seminar, Triangle Econometrics Conference, 4th ISNPS Conference, 2018 NASMES Conference, and Bristol Econometric Study Group for helpful comments and conversations. All remaining errors are ours.

†University of North Carolina at Chapel Hill - Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. Email: babii.andrii@gmail.com.

‡Toulouse School of Economics – 1, Esplanade de l'Université, 31080 Toulouse, France. Email: jean-pierre.florens@tse-fr.eu

Keywords: completeness condition, weak identification, nonparametric IV regression, high-dimensional regressions, spectral regularization.
JEL Classifications: C14, C26

1 Introduction

Structural nonparametric and high-dimensional models are often ill-posed. Among many examples, we may quote the nonparametric IV regression, various high-dimensional regressions, measurement errors, and random coefficient models. All these examples lead to the ill-posed functional equation

$$K\varphi = r,$$

where φ is a structural parameter of interest, r is a function, and K is a linear operator. The classical numerical literature on ill-posed inverse problems, see Engl, Hanke, and Neubauer (1996), studies *deterministic* problems, where the operator K is usually known and r is measured with a deterministic numerical error. In econometric applications, both K and r are estimated from the data, and we are faced with the *statistical* ill-posed inverse problems.

Identification is an integral part of econometric analysis, going back to Koopmans (1949), Koopmans and Reiersol (1950), and Rothenberg (1971) in the parametric case. In nonparametric and high-dimensional ill-posed inverse problems, K and r are directly identified from the data-generating process, while the structural parameter φ is identified if the equation $K\varphi = r$ has a unique solution. The uniqueness is equivalent to assuming that K is a one-to-one operator, or in other words that $K\phi = 0 \implies \phi = 0$ for all ϕ in the domain of K. It is worth stressing that the operator K is usually unknown in econometric applications and the estimated operator $\hat{K}$ has a finite rank and is not one-to-one for every finite sample size.

The maximum likelihood estimator when there is a lack of identification leads to a flat likelihood in some regions of a parameter space and then to the ambiguity on the choice of a maximum. It is then natural to characterize the limit of an estimator for such a potentially nonidentified model. In the nonidentified ill-posed inverse models, the identified set is an linear manifold $\phi + \mathcal{N}(K)$, where ϕ is any solution to $K\phi = r$, and $\mathcal{N}(K)$ is the null space of K. Note that the identified set is in general unbounded and is not informative on the structural parameter φ without additional constraints.

As K or $\hat{K}$ may fail to be one-to-one and typically have a discontinuous generalized inverse, some regularization is needed to estimate consistently the structural parameter. In this paper, we focus on spectral regularization methods consisting of modifying the spectrum of the operator $\hat{K}$. Tikhonov regularization, also known as functional ridge regression, is one prominent example; see Tikhonov (1963). Other important instances of spectral regularizations include the iterated Tikhonov; the spectral cut-off, which is also related to the functional principal component analysis, see Yao, Müller, and Wang (2005); and the Landweber-Fridman, which is also related to the functional gradient descent.[1] We show that estimators based on the spectral regularization are uniformly consistent for the best approximation to the structural parameter in the orthogonal complement to the null space of the operator and study the distributional consequences of identification failures.

In some cases, the best approximation may coincide with the structural parameter or at least may reasonably approximate it, even when the operator K has a non-trivial null space. This provides an attractive interpretation for the nonparametric IV regression under identification failures similar in a way to the best approximation property of least-squares under misspecifications; see Angrist and Pischke (2008), Chapter 3.[2] Lastly, the best approximation can also be used in the partial identification approach.

Contribution and related literature. There exists a large literature on the spectral regularization of statistical inverse problems; see Carrasco, Florens, and Renault (2007, 2014), Darolles, Fan, Florens, and Renault (2011), Florens, Johannes, and Van Bellegem (2011), Gagliardini and Scaillet (2012), and Babii (2020) among others.[3] With an exception for Florens, Johannes, and Van Bellegem (2011) and Chen and Pouzo (2012), the existing literature does not study the consequences of identification failures. In particular, related to our work, Florens, Johannes, and Van Bellegem (2011) derive the convergence rates of Tikhonov regularization to the best approximation in the L_2 norm when the operator is *known* and extend this result to the case of unknown operator under some high-level conditions on the regularization parameter and moments of the estimation error. However, these conditions have not been verified for specific models, such as the nonparametric IV or various high-dimensional regressions.

Our original contributions to this literature are to show that 1) the convergence

[1] The functional gradient descent is also related to the boosting procedure, see Friedman (2001).

[2] In contrast, the parametric 2SLS estimator does have the best approximation interpretation; see Escanciano and Li (2021) for an example of the IV estimator that has such an interpretation.

[3] There also exists an alternative approach to regularize the statistical inverse problems consisting of sieve approximations; see Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2012), and Chen and Christensen (2018) among others.

to the best approximation holds in the L_∞ norm when Tikhonov regularization is used; 2) the convergence holds for *general* spectral regularizations in the L_∞ and Hilbert space norms under low-level and easy-to-verify conditions; 3) the asymptotic distribution of linear functionals can transition between the Gaussian and chi-squared limits under identification failures. These results are illustrated for various high-dimensional regressions and the nonparametric IV regression.

Our paper also relates to the identification literature in the special case of nonparametric IV regression, see Newey and Powell (2003), D'Haultfoeuille (2011), Canay, Santos, and Shaikh (2013), Andrews (2017), Freyberger (2017), Hu and Shiu (2018), and more generally other ill-posed inverse models encountered in econometrics. In particular, our results on the consistency to the best approximation in the L_∞ norm in conjunction with Freyberger (2017) can be used to obtain the set estimators under identification failures.

It is also worth emphasizing that our general theory is not limited by the nonparametric IV regression and can potentially be applied to other ill-posed inverse models such as measurement errors, see Hu and Schennach (2008); dynamic models with unobserved state variables, see Hu and Shum (2012); demand models, see Berry and Haile (2014) and Dunker, Hoderlein, and Kaido (2017); neoclassical trade models, see Adao, Costinot, and Donaldson (2017); models of earnings and consumption dynamics, see Arellano, Blundell, and Bonhomme (2017) and Botosaru (2019); structural random coefficient models, see Hoderlein, Nesheim, and Simoni (2017); discrete games, see Kashaev and Salcedo (2020); models of two-sided markets, see Sokullu (2016); high-dimensional mixed-frequency IV regressions, see Babii (2021) and various functional regressions, see Florens and Van Bellegem (2015) and Benatia, Carrasco, and Florens (2017); and endogenous sample selection models, see Breunig, Mammen, and Simoni (2018). Related identification issues also appear in quantile treatment effect models, see Chernozhukov and Hansen (2005), and nonlinear asset pricing models, see Chen and Ludvigson (2009) and Chen, Pelger, and Zhu (2020).

The paper is organized as follows. Section 2 discusses the identification in linear ill-posed inverse models using the nonparametric IV and various high-dimensional regressions as leading examples. In Section 3, we obtain the non-asymptotic risk bounds in the L_∞ and the Hilbert space norms for a class of Tikhonov-regularized estimators, which are extended to general regularization schemes in the Appendix Section A.1. Section 4 illustrates how these results can be applied to the partial identification. In the Appendix Section A.2, we show that in the extreme case of identification failures, the Tikhonov-regularized estimator is driven by the degenerate U-statistics in large samples. Building on these results, we illustrate in Section 5 the transition between the Gaussian and the weighted chi-squared asymptotics in the intermediate cases. We report on a Monte Carlo study in Section 6 which provides further insights about the validity of our asymptotic results in finite sample settings typically encountered in empirical applications. Section 7 concludes. All proofs are collected in the Appendix Section A.3. We also review several relevant results from the theory of the generalized inverse operators and the theory of the Hilbert space valued U-statistics in the Online Appendix Sections B.1 and B.2.

2 Identification

Consider the functional linear equation

$$K\varphi = r,$$

where $K : \mathcal{E} \to \mathcal{H}$ is a compact linear operator, defined on some Hilbert spaces $\mathcal{E}$ and $\mathcal{H}$, and $\varphi \in \mathcal{E}$ is a structural parameter of interest. The structural parameter φ is point identified if the operator K is one-to-one, or in other words if the null space of K, denoted $\mathcal{N}(K) = \{\phi \in \mathcal{E} : K\phi = 0\}$, reduces to $\{0\}$. Equivalently, the point identification of φ requires that

$$K\phi = 0 \implies \phi = 0, \qquad \forall \phi \in \mathcal{E}.$$

We illustrate the statistical interpretation of the one-to-one property of K in the nonparametric IV regression and the high-dimensional regressions.

Example 2.1 (Nonparametric IV regression)**.** *Consider*

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

where $(Y, Z, W) \in \mathbf{R} \times \mathbf{R}^p \times \mathbf{R}^q$ *is a random vector, see Darolles, Fan, Florens, and Renault (2011). The exclusion restriction leads to the functional linear equation*

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

where $K : L_2(Z) \to L_2(W)$ *is a conditional expectation operator.*[4] *The completeness condition, or more precisely the* L_2*-completeness, see Florens, Mouchart, and Rolin (1990) and Newey and Powell (2003), is the one-to-one property of the conditional expectation operator*

$$\mathbb{E}[\phi(Z)|W] = 0 \implies \phi = 0, \qquad \forall \phi \in L_2(Z).$$

It is a (non-linear) dependence condition between the endogenous regressor Z *and the instrument* W*.*

[4] For a random variable X, we define $L_2(X) = \{\phi : \mathbb{E}|\phi(X)|^2 < \infty\}$.

Example 2.2 (High-dimensional regressions)**.** *Consider*

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

where $(Y, Z, W) \in \mathbf{R} \times \mathcal{E} \times \mathcal{H}$*, see Florens and Van Bellegem (2015).*[5] *The exclusion restriction leads to the functional linear equation*

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

where $K : \mathcal{E} \to \mathcal{H}$ *is a covariance operator. The completeness condition is a one-to-one property of the covariance operator*

$$\mathbb{E}[\langle Z, \phi \rangle W] = 0 \implies \phi = 0, \qquad \forall \phi \in \mathcal{E}.$$

It generalizes the rank condition used in the linear IV regression and requires a sufficient (linear) dependence between Z *and* W*.*

If the completeness condition fails, then the null space of the operator K is a non-trivial closed linear subspace of $\mathcal{E}$ and the structural parameter φ is only set identified. The identified set is a closed *linear manifold*

$$\Phi^{\mathrm{ID}} = \varphi + \mathcal{N}(K),$$

where $\mathcal{N}(K)$ is the null space of K. The identified set is in general unbounded, however, practical applications also involve some smoothness restrictions under which the identified set can be bounded.

Since $\mathcal{N}(K)$ is a closed linear subspace of $\mathcal{E}$, decompose

$$\varphi = \varphi_1 + \varphi_0,$$

where φ_1 is the unique projection of φ on $\mathcal{N}(K)^\perp$ and φ_0 is the orthogonal projection of φ on $\mathcal{N}(K)$. By orthogonality, $\|\varphi\|^2 = \|\varphi_1\|^2 + \|\varphi_0\|^2 \geq \|\varphi_1\|^2$. Therefore, φ_1 has a smaller norm than φ, with the two being equal when $\varphi_0 = 0$. Since $\mathcal{N}(K)^\perp = \overline{\mathcal{R}(K^*)}$, see Luenberger (1997), p.157, the best approximation φ_1 equals to the structural

[5] For instance, when $W = Z$ and $\mathcal{E} = L_2$ with counting measure, we obtain a high-dimensional regression model, suitable for non-sparse data

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

When $\mathcal{E} = L_2$ with Lebesgue measure, this model is also sometimes called the functional regression; see Florens and Van Bellegem (2015) and Babii (2021).

parameter φ whenever the structural parameter φ belongs to $\overline{\mathcal{R}(K^*)}$. This condition has also an appealing regularity interpretation known as the source condition. To see this, note that $\mathcal{R}(K^*) = \mathcal{R}(K^*K)^{1/2}$; see Engl, Hanke, and Neubauer (1996), Proposition 2.18. Therefore, if the ill-posed inverse problem has a sufficiently high regularity, so that $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$ with $\beta \in [1, \infty)$, then $\varphi_1 = \varphi$, and the structural function φ is point identified although the completeness condition fails.

The best approximation is also be informative, whenever the structural function φ can be well approximated by the family of basis functions of $\mathcal{N}(K)^{\perp}$. The following example illustrates this further for the nonparametric IV regression.

Example 2.3 (Nonparametric IV regression)**.** *Suppose that the conditional expectation operator*

$$\begin{aligned} K : L_2(Z) &\to L_2(W) \\ \phi &\mapsto \mathbb{E}[\phi(Z)|W] \end{aligned}$$

is compact. By the spectral theorem, there exists $(\lambda_j, \varphi_j, \psi_j)_{j\geq 1}$*, where* $\lambda_j \to 0$ *is a sequence of singular values,* $(\varphi_j)_{j\geq 1}$ *is a complete orthonormal system of* $\mathcal{N}(K)^{\perp} = \overline{\mathcal{R}(K^*)}$*, and* $(\psi_j)_{j\geq 1}$ *is the complete orthonormal system of* $\mathcal{N}^{\perp}(K^*) = \overline{\mathcal{R}(K)}$*. The best approximation coincides with the structural function* φ*, whenever it can be represented in terms of the family* $(\varphi_j)_{j\geq 1}$*.*

It is also known that the completeness condition fails in the nonparametric IV regression when Z has the Lebesgue density while the instrumental variable is a discrete random variable. The following example illustrates that if the instrumental variable W takes a sufficiently large number of discrete values, then the function φ might be accurately estimated even if the completeness condition fails.

Example 2.4 (Nonparametric IV with discrete instrument)**.** *Consider the nonparametric IV regression with a discrete instrumental variable* $W \in \{w_k : \ k \geq 1\}$*. Put* $f_k(z) = f_{Z|W=w_k}(z)$ *for every* $k \geq 1$*. Then*

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

and if $\varphi \in \operatorname{span}\{f_k : \ k \geq 1\}$*, then* $\varphi \in \mathcal{N}(K)^{\perp}$*. The best approximation coincides with the structural parameter* φ *whenever it can be represented in terms of the family* $(f_k)_{k\geq 1}$*.*

It is worth stressing that functions encountered in empirical settings can typically be well-approximated by a fairly small number of series terms, in which case even if φ cannot be exactly represented by families $(\varphi_j)_{j\geq 1}$ or $(f_k)_{k\geq 1}$, these families could capture most of the nonlinearities; see also Section 4 for a partial identification perspective.

3 Nonasymptotic risk bounds

In this section, we derive the nonasymptotic risk bounds for the Tikhonov-regularized estimator in the Hilbert space and the uniform norms. All results will be stated uniformly over the relevant class of models without relying on the completeness condition.

3.1 Tikhonov-regularized estimator

Estimation of the structural function φ is an ill-posed inverse problem and requires regularization for two reasons. First, the generalized inverse of K is typically not continuous; see Appendix B.1. Second, the estimator $\hat{K}$ is typically a finite-rank operator and is not one-to-one for any finite sample size.

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

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

where $\alpha_n > 0$ is a regularization parameter. It is easy to see that the solution to this problem is

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

The estimator enjoys two fundamental properties:

1. It is well-defined even when K or $\hat{K}$ is not one-to-one.
2. If $\alpha_n \to 0$ at an appropriate rate, it converges to the best approximation to φ in $\mathcal{N}(K)^\perp$.

Indeed, even if $\hat{K}^*\hat{K}$ had an eigenvalue $\hat{\lambda}_j = 0$, the corresponding eigenvalue of $\alpha_n I + \hat{K}^*\hat{K}$ would be $\alpha_n + \hat{\lambda}_j > 0$ with a well-defined inverse. In the following two sections, we show that 2. holds in the Hilbert space and the uniform norms.

3.2 Hilbert space risk

First, we describe the relevant class of structural functions and operators:

Assumption 3.1. *The structural parameter $\varphi = \varphi_1 + \varphi_0$ and the operator K are in*

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

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

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

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

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

We estimate (r, K) with $(\hat{r}, \hat{K})$ and make 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* $C_1, C_2 < \infty$ *do not depend on* (φ, K) *and* $\delta_n, \rho_{1n} \to 0$ *as* $n \to \infty$*.*

Assumption 3.2 describes the convergence rate of residuals and the estimated operator. Note also that the residuals in the nonidentified model, $\hat{r} - \hat{K}\varphi_1$, can be written as a sum of identified residuals $\hat{r} - \hat{K}\varphi$ and $\hat{K}\varphi_0$ which can be controlled separately.

The following result holds for the Hilbert space norm:

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

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

where $a \wedge b = \max\{a, b\}$*.*

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

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

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

- identified case: $\varphi_0 = 0$ and $\rho_{1n}\alpha^{\beta\wedge 1} \lesssim \delta_n$, where the convergence rate to φ is driven by residuals $\hat{r} - \hat{K}\varphi$;
- weakly identified case: $\varphi_0 = 0$ and $\delta_n \lesssim \rho_{1n}\alpha^{\beta\wedge 1}$, where the convergence rate to φ is driven by the estimated operator $\hat{K}$;
- nonidentified case: $\varphi_0 \neq 0$ and $\rho_{1n}\alpha^{\beta\wedge 1} \lesssim \delta_n$, where the convergence rate to φ_1 is driven by residuals $\hat{r} - \hat{K}\varphi_1$;
- strongly nonidentified models: $\varphi_0 \neq 0$ and $\delta_n \lesssim \rho_{1n}\alpha^{\beta\wedge 1}$, where the convergence rate to φ_1 is driven by the estimated operator $\hat{K}$.

In the identified case, the optimal choice of the regularization parameter is $\alpha_n \sim \delta_n^{1/(\beta+1)}$, which leads to the optimal convergence rate of order $O\left(\delta_n^{\beta/(\beta+1)}\right)$ for the class $\mathcal{F}(\beta, C)$ with $\beta \in (0, 2]$; see Mair and Ruymgaart (1996). The rate is not optimal for $\beta > 2$, but we show in the Appendix Section A.1 that the optimal rate can be achieved with additional iterations or some alternative regularization schemes.

Lastly, it is worth mentioning that in the special case of L_2 spaces, Theorem 3.1 provides a sharper result than Florens, Johannes, and Van Bellegem (2011), Theorem 2.2, where the rate is always driven by ρ_{1n}.

3.3 L_∞ risk

Suppose now that the space of continuous functions on a compact set $D \subset \mathbf{R}^p$, denoted $(C(D), \|.\|_\infty)$, is embedded into the space $\mathcal{E}$, where $\|.\|_\infty$ is the uniform norm. Suppose also that $\mathcal{R}(K^*) \subset C(D)$ and that $\varphi_1 \in C(D)$. Let $\|K^*\|_{2,\infty} = \sup_{\|\phi\|\leq 1}\|K^*\phi\|_\infty$ be the mixed operator norm of K^*. The following assumption describes how well the operator K^* is estimated by $\hat{K}^*$ in the $\|.\|_{2,\infty}$ norm.

Assumption 3.3. *Suppose that $\|K^*\|_{2,\infty} \leq C_3$ and that $\mathbb{E}\|\hat{K}^* - K^*\|^2_{2,\infty} \leq C_3\rho_{2n}$, where $C_3 < \infty$ does not depend on (φ, K) and $\rho_{2n} \to 0$.*

The following result holds:

Theorem 3.2. *Suppose that Assumptions 3.1, 3.2, and 3.3 are satisfied with $\varphi_1 = (K^*K)^{\beta/2}K^*\psi$. Then 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 describes the uniform convergence rates for generic ill-posed inverse problems. In constrast to Theorem 3.1, the L_∞ convergence rate is also driven by the estimation error in K^* measured in the $\|.\|_{2,\infty}$ norm. We also obtain the L_∞ convergence rate for more general spectral regularization schemes in the Appendix Section A.1.

3.4 Applications

3.4.1 High-dimensional regressions

Following, Example 2.2, the econometrician observes an i.i.d. sample $(Y_i, Z_i, W_i)_{i=1}^n$.[6] Then

$$r = \mathbb{E}[YW], \qquad K\phi = \mathbb{E}[W\langle \phi, Z\rangle], \qquad K^*\psi = \mathbb{E}[Z\langle \psi, W\rangle]$$

are estimated with

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

Then the elementary computations give

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

Let $\mathcal{F}(\beta, C)$ be the class of models as in the Assumption 3.1 and suppose also that $\mathbb{E}\|UW\|^2 \vee \mathbb{E}\|ZW\|^2 \leq C$ for all models in this class. Then $\delta_n = \rho_{1n} = n^{-1}$ and Theorem 3.1 shows that the Hilbert space risk is of order

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

The high-dimensional regression is either identified ($\varphi_0 = 0$) or nonidentified ($\varphi_0 \neq 0$). Then conditions $\alpha_n \to 0$ and $\alpha_n n \to \infty$ as $n \to \infty$ are sufficient to guarantee the consistency in the Hilbert space norm. In particular, the optimal choice $\alpha_n \sim n^{-1/(\beta+1)}$ leads to the convergence rate of order $O(n^{-\beta/(\beta+1)})$.

For the uniform convergence, suppose that $\mathcal{E} = L_2(S)$, i.e. a set of functions on some bounded set $S \subset \mathbf{R}^d$, square-integrable with respect to the Lebesgue measure. To verify the Assumption 3.3, we also assume that models in $\mathcal{F}(\beta, C)$ are such that $\|Z\|_\infty \vee \|W\|_\infty \leq C < \infty$ and that stochastic processes Z and W are in some

[6] The i.i.d. assumption can be relaxed to the covariance stationarity and absolute summability of autocovariances; see Babii (2021).

Hölder ball with smoothness $s > d/2$. Then by the Hoffman-Jørgensen and moment inequalities, see Giné and Nickl (2016), p.129 and p.202,

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

Therefore, Assumption 3.3 is satisfied with $\rho_{2n} = n^{-1}$, and Theorem 3.2 shows that

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

Then conditions $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ as $n \to \infty$ ensure the uniform consistency of $\hat{\varphi}$. In particular, the optimal choice $\alpha_n \sim n^{-1/(\beta+2)}$ leads to the convergence rate of order $O(n^{-\beta/(2\beta+4)})$.

3.4.2 Nonparametric IV regression

Following Example 2.1, rewrite 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 $K : L_2([0,1]^p) \to L_2([0,1]^q)$. We can estimate r and K via kernel smoothing:

$$\begin{aligned}
\hat{r}(w) &= \frac{1}{nh_n^q}\sum_{i=1}^n Y_i K_w\left(h_n^{-1}(W_i - w)\right), \\
(\hat{K}\phi)(w) &= \int \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z, \\
\hat{f}_{ZW}(z,w) &= \frac{1}{nh_n^{p+q}}\sum_{i=1}^n K_z\left(h_n^{-1}(Z_i - z)\right)K_w\left(h_n^{-1}(W_i - w)\right),
\end{aligned}$$

where K_w and K_z are symmetric kernel functions and $h_n \to 0$ is a bandwidth parameter. Under mild assumptions, by Proposition A.3.1, $\delta_n = \frac{1}{nh_n^q} + h_n^{2s}$ and $\rho_{1n} = \frac{1}{nh_n^{p+q}} + h_n^{2s}$, where s is the Hölder smoothness of f_{ZW}. Therefore, Theorem 3.1 shows that 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 value of the regularity parameter β. For consistency of $\hat{\varphi}$ to φ_1, we need $\alpha_n nh_n^q \to \infty$, $\alpha_n^{(1-\beta)\wedge 0} nh_n^{p+q} \to \infty$, and $h_n^{2s}/\alpha_n \to 0$ as $n \to \infty, \alpha_n \to 0$, and $h_n \to 0$.

We also know that $\rho_{2n}^{1/2} = \sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s$, see e.g., Babii (2020), Proposition A.3.1. Therefore, Theorem 3.2 shows that

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

For the uniform consistency of $\hat{\varphi}$ to the best approximation φ_1, 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 $h_n \to 0$.

4 Partial identification

In this section, we consider an application of our results to partial identification in the nonparametric IV regression following Freyberger (2017). Let

$$\left|\Phi^{\mathrm{ID}}\right| = \sup_{\phi_1,\phi_2\in\Phi^{\mathrm{ID}}} \|\phi_1 - \phi_2\|_\infty$$

be the L_∞-diameter of the identified set. While the identified set is in general an unbounded linear manifold, additional smoothness restrictions are typically imposed on the structural function φ which can make it bounded. To that end, suppose that φ and φ_1 belong to the Lipschitz smoothness class

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

We would like to infer the L_∞ diameter of the identified set. Note given the best approximation $\varphi_1 \in \Phi^{\mathrm{ID}}$ and the diameter of the identified set $|\Phi^{\mathrm{ID}}|$, we can use the tube $\varphi_1 \pm |\Phi^{\mathrm{ID}}|$ to partially identify the structural function $\varphi \in \Phi^{\mathrm{ID}}$. Formally, to infer the size of the identified set, consider the following testing problem for a fixed $\varepsilon > 0$,

$$H_0 :\ \left|\Phi^{\mathrm{ID}}\right| \geq \varepsilon \qquad \text{vs.} \qquad H_1 :\ \left|\Phi^{\mathrm{ID}}\right| < \varepsilon.$$

Consider the following statistics

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

where $\mathscr{F}_\varepsilon = \{\phi \in L_\infty[0,1] :\ \|\phi\|_s \leq 2C/\varepsilon\}$. Among other things, Freyberger (2017) shows that under the null hypothesis $T = 0$, and establishes bounds on critical values for a test that controls the size uniformly in the special case of the nonparametric IV estimator. We will focus on the pointwise results, which can easily be made uniform in light of our previous results. To describe the test, let

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

be a sample counterpart to T, where $\hat{K}$ is an estimator of K. Then we aim to construct $q_{1-\alpha}$ such that under H_0,

$$\lim_{n\to\infty} \Pr(n\hat{T} > q_{1-\alpha}) \leq \alpha.$$

The test rejects when $n\hat{T} > q_{1-\alpha}$ and the above requirement ensures that the test has level α. Inverting the test, we can estimate the diameter of the identified as

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

Therefore, in conjunction with our results on the L_∞ norm consistency of regularized estimators, the tube $[\hat{\varphi} - \hat{\varepsilon}, \hat{\varphi} + \hat{\varepsilon}]$ is a set estimator of φ; see also Freyberger (2017), Remark 2.[7]

In the following two subsection, we discuss how to obtain the critical value $q_{1-\alpha}$ in the special cases of high-dimensional and nonparametric IV regressions.

4.1 High-dimensional regressions

Recall that for high-dimensional regressions, the operator $K : \mathcal{E} \to \mathcal{H}$ is estimated as

$$\hat{K}\phi = \frac{1}{n}\sum_{i=1}^{n} W_i \langle Z_i, \phi \rangle.$$

Under H_0, there exists ϕ such that $K\phi = 0$, $\|\phi\|_\infty = 1$, and $\|\phi\|_s \leq 2C/\varepsilon$. Let $q_{1-\alpha}$ be the quantile of order $1-\alpha$ of $\sum_{j\geq 1} \lambda_{\phi,j}\chi_j^2$, where $(\lambda_{\phi,j})_{j\geq 1}$ are the eigenvalues of the operator

$$\begin{aligned} C_\phi &: \mathcal{H} \to \mathcal{H} \\ &h \mapsto \mathbb{E}[W\langle W, h\rangle \langle Z, \phi\rangle^2], \end{aligned}$$

and $(\chi_j^2)_{j\geq 1}$ are i.i.d. chi-squared random variables. In practice, one can use the eigenvalues of an estimator of C_ϕ to simulate $q_{1-\alpha}$. The following result holds:

[7] We are grateful to the anonymous referee for pointing out this connection.

Theorem 4.1. *Suppose that* $(Z_i, W_i)_{i=1}^n$ *is an i.i.d. sample of* (Z, W) *and* $\mathbb{E}\|ZW\|^2 < \infty$*. Then under* H_0*,*

$$\lim_{n\to\infty} \Pr(n\hat{T} > q_{1-\alpha}) \leq \alpha.$$

The proof of this result appears in the appendix. Alternatively, one could apply the multiplier bootstrap to the statistics $\|(\hat{K} - K)\phi\|$. The validity of the multiplier bootstrap procedure would follow from the standard empirical process results; see van der Vaart and Wellner (2000).

4.2 Nonparametric IV regression

Recall that the estimated operator is

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

where $\hat{f}_{ZW}$ is an estimator of f_{ZW}. While establishing the exact asymptotic distribution of the test statistics $\hat{T}$ is fairly involved, in light of the proof of Theorem 4.1, it suffices to approximate the distribution of $\|\sqrt{n}(\hat{K} - K)\phi\|$. For the latter, one could reply on the multiplier bootstrap, which is expected to be valid, at least when $\hat{f}_{ZW}$ is estimated with series; see Chernozhukov, Newey, and Santos (2015) and Freyberger (2017), Supplement S.2.

5 Linear functionals

In some economic applications, the object of interest is a linear functional of the structural function φ, e.g., the consumer surplus or the deadweight loss functionals. Note that the consistency of the continuous linear functional in the nonidentified model follows from our results in Section 3. In this section, we focus on asymptotic distributions in nonidentified models and show that the degenerate U-statistics asymptotics documented in Section A.2 can emerge in the intermediate cases.

By the Riesz representation theorem any continuous linear functional on a Hilbert space $\mathcal{E}$ can be represented as an inner product with some $\mu \in \mathcal{E}$. The asymptotic distribution of the linear functional depends crucially on whether the Riesz representer is in $\mathcal{N}(K)$ or in $\mathcal{N}(K)^\perp$. To understand how $\langle\hat{\varphi} - \varphi, \mu\rangle$ behaves asymptotically, consider the unique orthogonal decomposition $\mu = \mu_0 + \mu_1$, where μ_0 is the orthogonal projection on $\mathcal{N}(K)$ and μ_1 is the orthogonal projection on $\mathcal{N}(K)^\perp$. We focus on the distribution of $\langle\hat{\varphi} - \varphi, \mu_0\rangle$. The following assumption is a mild restriction on the distribution of the data:

Assumption 5.1. *(i) the data* $(Y_i, Z_i, W_i)_{i=1}^n$ *are the i.i.d. sample of* (Y, Z, W) *with* $\mathbb{E}|U|^2 < \infty$ *and* $\mathbb{E}\|Z\|^2 < \infty$*; (ii)* $\mathbb{E}\|ZW\|^2 < \infty$*,* $\mathbb{E}\|UW\|^2 < \infty$*,* $\mathbb{E}\|UZW\| < \infty$*,* $\mathbb{E}\|Z\|^2\|W\| < \infty$*, and* $\mathbb{E}\left[|U|\|Z\|\|W\|^2\right] < \infty$*.*

Decompose $W = W^0 + W^1$, where W^0 is the orthogonal projection of W on $\mathcal{N}(K^*)$ and W^1 is the orthogonal projection of W on $\mathcal{N}(K^*)^\perp$.

Assumption 5.2. $\alpha_n \to 0$*,* $n\alpha_n^{1+\beta\wedge 1} \to 0$*, and* $n\alpha_n \to \infty$ *as* $n \to \infty$*.*

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

$$h(x, x') = \frac{\langle w^0, w^{0\prime}\rangle}{2}\left\{\langle z, \mu_0\rangle(y' - \langle z', \varphi_1\rangle) + \langle z', \mu_0\rangle(y - \langle z, \varphi_1\rangle)\right\}.$$

The following result holds:

Theorem 5.1. *Suppose that Assumptions 3.1, 5.1, and 5.2 are satisfied. Then if* $W^0 = 0$*, for every* $\mu_0 \in \mathcal{N}(K)$*,*

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

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*. On the other hand, if* $W^0 = 0$*, then for every,* $\mu_0 \in \mathcal{N}(K)$

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle \xrightarrow{d} 0.$$

Note that the convergence to a non-degenerate distribution is possible whenever W is not entirely concentrated in $\mathcal{N}(K^*)^\perp$, i.e. $W^0 \neq 0$.

For the asymptotic distribution of inner products with $\mu_1 \in \mathcal{N}(K)^\perp$, put $\eta_n = (Y - \langle Z, \varphi_1\rangle)(\alpha_n I + K^*K)^{-1}K^*W$ and note that

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

where Σ is the variance operator of $(Y - \langle Z, \varphi_1\rangle)W$. The following assumption is a Lindeberg's condition.

Assumption 5.3. *Suppose that for all $\epsilon > 0$*

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

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

Since for every $\delta > 0$

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

a sufficient condition for Assumption 5.3 is the Lyapunov's condition $\mathbb{E}\left|\langle \eta_n, \mu_1\rangle\right|^{2+\delta} = O(1)$. The following assumptions are sufficient for the Lyapunov's condition: $\mathbb{E}|U|^{2+\delta} < \infty$, $\mathbb{E}\|Z\|^{2+\delta}$, $\mu_1 \in \mathcal{R}\left[(K^*K)^\gamma\right]$ and $W \in \mathcal{R}\left[(K^*K)^{\tilde{\gamma}}\right]$ with $\gamma + \tilde{\gamma} \geq 1/2$. To see that this is the case, note that

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

Also, we assume the following:

Assumption 5.4. *Suppose that (i) $\mu_1 \in \mathcal{R}[(K^*K)^\gamma]$ for some $\gamma > 0$; (ii) $\alpha_n \to 0$, $\pi_n \alpha_n^{(\gamma+\beta/2)\wedge 1} \to 0$, $\frac{\pi_n \alpha_n^{\gamma \wedge \frac{1}{2}}}{n\alpha_n} \to 0$, and $n\alpha_n^{1+\beta\wedge 1} \to 0$ as $n \to \infty$.*

Note that the Assumption 5.4 is the most restrictive when $\pi_n = O\left(n^{1/2}\right)$. In this case we need $n\alpha_n^{(2\gamma+\beta)\wedge 2} \to 0$, $n\alpha_n^{2-2\gamma\wedge 1} \to \infty$, and $n\alpha_n^{1+\beta\wedge 1} \to 0$. If the function μ_1 is smooth enough in the sense that $\gamma \geq 1/2$ and $\beta > 0$, then this condition reduces to $n\alpha_n \to \infty$ and $n\alpha_n^{1+\beta\wedge 1} \to 0$ as $n \to \infty$.

Theorem 5.2. *Suppose that Assumptions 3.1, 5.1, 5.3, and 5.4 are satisfied. Then for every $\mu_1 \in \mathcal{N}(K)^\perp$*

$$\pi_n \langle \hat{\varphi} - \varphi_1, \mu_1 \rangle \xrightarrow{d} N(0, 1).$$

For the inner products with μ_1, the speed of convergence is $O(n^{-c}), c \in (0, 1/2]$, depending on the mapping properties of the operators K and Σ, and the smoothness of μ_1. Consequently, in light of Theorems 5.1 and 5.2, for the inner product $\langle \hat{\varphi} - \varphi_1, \mu\rangle$ with $\mu = \mu_0 + \mu_1$, the normalizing sequence can be π_n or $\alpha_n n$ depending on their relative speed.[8] The resulting large sample distribution may be Gaussian, the

[8]Note that the root-n estimability of inner products for linear ill-posed inverse problems in the identified case is studied, e.g,. in Carrasco, Florens, and Renault (2007), Carrasco, Florens, and Renault (2014); see also Severini and Tripathi (2012) for the nonparametric IV regression.

weighted sum of independent chi-squared random variables, or the mixture of the two.[9]

6 Monte Carlo experiments

In this section we study the validity of our asymptotic theory using Monte Carlo experiments. To construct the DGP with a non-trivial null space of the operator K, consider a Gaussian density truncated to the unit square

$$f_{ZW}^{\mathrm{ID}}(z,w) = \frac{f_{ZW}(z,w)}{\int_{[0,1]^2} f_{ZW}(z,w)\mathrm{d}z\mathrm{d}w}\mathbb{1}\{(z,w)\in[0,1]^2\},$$

where f_{ZW} is the density of

$$\begin{pmatrix} Z \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 0.05 & 0.01 \\ 0.01 & 0.05 \end{pmatrix}\right).$$

Put $J=\{1,2,\dots,J_0\}$ for some $J_0 \in \mathbf{N}$ and let $(\varphi_j)_{j\geq 1}$ be a trigonometric basis of $L_2[0,1]$. Define

$$f_{ZW}^{\mathrm{NID}} = C\sum_{j=1}^{J_0}\sum_{k=1}^{\infty}\langle f_{Z,W}, \varphi_j\otimes\varphi_k\rangle \varphi_j\otimes\varphi_k,$$

where C is a normalizing constant, ensuring that f_{ZW}^{NID} integrates to 1. Let K be an integral operator with the kernel f_{ZW}^{NID}. Then the null space of K is infinite-dimensional

$$\mathcal{N}(K) \subset \mathrm{span}\{\varphi_j : \ j \geq J_0+1\}$$

and the identified set is not tractable. We also set $J_0=\infty$, if $J=\mathbf{N}$, in which case $\varphi_0 = 0$ and $\varphi=\varphi_1$.

We use the rejection sampling to simulate the data from f_{ZW}^{NID}. The rest of the DGP is

$$Y=\varphi(Z)+U, \qquad U=\varepsilon Z, \quad \varepsilon\sim N(0,1) \perp\!\!\!\perp (Z,W),$$

where $\varphi(z) = z^3 - z^2 - z + \sum_{j=4}^{10}(-1)^j z^j$. Note that the function φ exhibits non-trivial nonlinearities and, at the same time, it has an infinite series representation in the trigonometric basis.

For simplicity, we consider the Tikhonov-regularized estimator; see Babii (2020) for more details on the practical implementation. As in Babii (2020) we use the

[9] The critical values when the normalizing sequence is unknown can be obtained with resampling methods, see, e.g. Bertail, Politis, and Romano (1999).

cross-validation to select the tuning parameters, but without making additional adjustments needed for inference; see also Centorrino (2014). Table 1 displays the empirical L_2 and L_∞ errors for three different degrees of the identification. When $J_0 = 1$ or $J_0 = 2$, the operator K has the infinite-dimensional null space, while for $J_0 = \infty$, the model is point identified. For $J_0 = 1$, we can only recover the information related to the first basis vector, and Figure 1 illustrates significant distortions in this case. However, when the function φ is point identified, we do not do significantly better compared to the nonidentified case with $J_0 = 2$, in which case the first two basis vectors are used to approximate φ. Therefore, even for cases close to the extreme failures of the completeness condition, we may still be able to learn a lot about the global shape properties of φ.

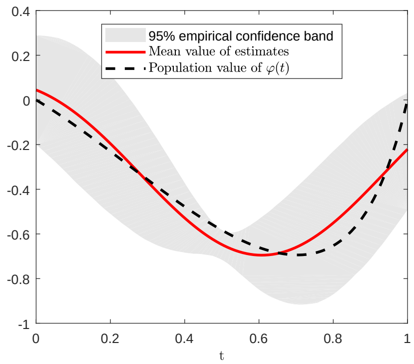
	$n = 1000$		$n = 5000$	
J_0	L_2	L_∞	L_2	L_∞
1	0.0386	0.3913	0.0489	0.2835
2	0.0364	0.3414	0.0121	0.2612
∞	0.0293	0.3166	0.0113	0.2604

Table 1: L_2 and L_∞ errors. 5000 Monte Carlo experiments.

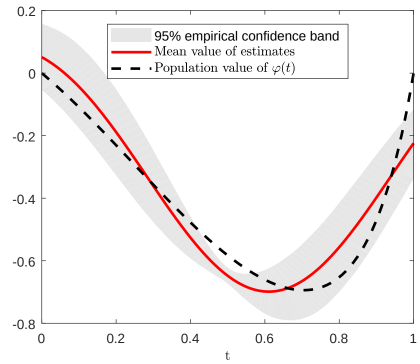
7 Conclusion

This paper develops a theory of nonidentified ill-posed inverse problems using the nonparametric IV and the high-dimensional regressions as illustrating examples. Identification failures occur due to the non-injectivity of the covariance or the conditional expectation operators. We show that if these operators are not injective, the estimators based on the spectral regularization converge to the best approximation of the structural parameter in the orthogonal complement to the null space of the operator and derive new uniform and Hilbert space norm bounds for the risk.

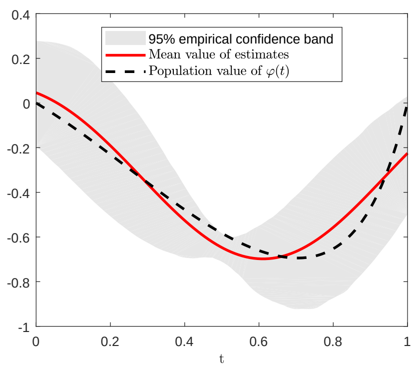
This provides us an appealing projection interpretation for the nonparametric IV regression under identification failures similar to the one shared by the ordinary least-squares under misspecification. We describe several circumstances when the best approximation coincides with a structural parameter, or at least reasonably approximates it, and discuss how our results can be useful in the partial identification setting under smoothness constraints. It is also worth mentioning that the smoothnes constraints can also be incorporated in the regularization procedure, see Babii and Florens (2021), leading to more precise estimates in small samples.



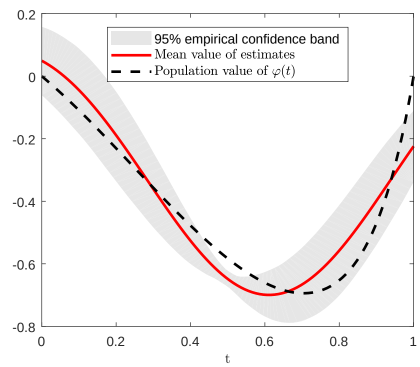
(a) $n = 1000, J_0 = \infty$



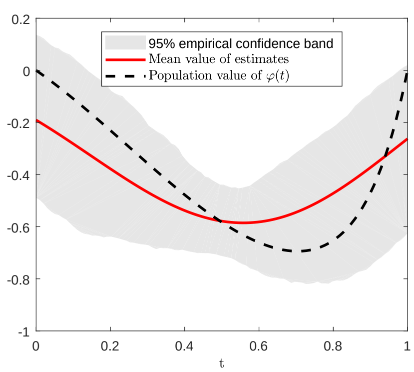
(b) $n = 5000, J_0 = \infty$



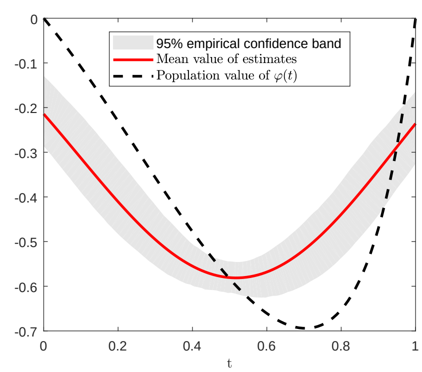
(c) $n = 1000, J_0 = 2$



(d) $n = 5000, J_0 = 2$



(e) $n = 1000, J_0 = 1$



(f) $n = 5000, J_0 = 1$

Figure 1: Estimates averaged over 5000 experiments with 95% pointwise empirical confidence bands.

Lastly, we illustrate that under identification failures, the asymptotic distribution of linear functionals can transition between the weighted chi-squared and Gaussian limits.

References

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric counterfactual predictions in neoclassical models of international trade," *American Economic Review*, 107(3), 633–689.

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

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

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85(3), 693–734.

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

——— (2021): "High-dimensional mixed-frequency IV regression," *Journal of Business & Economic Statistics*, (forthcoming).

Babii, A., and J.-P. Florens (2021): "Are unobservables separable?," *arXiv preprint arXiv:1705.01654*.

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.

Benatia, D., M. Carrasco, and J.-P. Florens (2017): "Functional linear regression with functional response," *Journal of Econometrics*, 201(2), 269–291.

Berry, S. T., and P. A. Haile (2014): "Identification in differentiated products markets using market-level data," *Econometrica*, 82(5), 1749–1797.

Bertail, P., D. N. Politis, and J. P. Romano (1999): "On subsampling estimators with unknown rate of convergence," *Journal of the American Statistical Association*, 94(446), 569–579.

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

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

Botosaru, I. (2019): "Identifying distributions in a panel model with heteroskedasticity: an application to earnings volatility," *Bristol University Working Paper*.

Breunig, C., E. Mammen, and A. Simoni (2018): "Nonparametric estimation in case of endogenous selection," *Journal of Econometrics*, 202(2), 268–285.

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

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

——— (2014): "Asymptotic normal inference in linear inverse problems," *Handbook of applied nonparametric and semiparametric econometrics and statistics*, pp. 65–96.

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

Chen, L., M. Pelger, and J. Zhu (2020): "Deep learning in asset pricing," *Available at SSRN 3350138*.

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

Chen, X., and S. C. Ludvigson (2009): "Land of addicts? An empirical investigation of habit-based asset pricing models," *Journal of Applied Econometrics*, 24(7), 1057–1093.

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

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., W. K. Newey, and A. Santos (2015): "Constrained conditional moment restriction models," *arXiv preprint arXiv:1509.06311*.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (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.

Dunker, F., S. Hoderlein, and H. Kaido (2017): "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models," *Cemmap Working Paper*.

Egger, H. (2005): "Accelerated Newton-Landweber iterations for regularizing nonlinear inverse problems," *SFB-Report*, 3, 2005.

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

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

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

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990): *Elements of bayesian statistics*. CRC Press.

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

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

Giné, E., and R. Nickl (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.

Groetsch, C. W. (1977): *Generalized inverses of linear operators*. Chapman & Hall Pure and Applied Mathematics.

Hoderlein, S., L. Nesheim, and A. Simoni (2017): "Semiparametric estimation of random coefficients in structural economic models," *Econometric Theory*, 33(6), 1265–1305.

Hu, Y., and S. M. Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76(1), 195–216.

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

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1), 32–44.

Kashaev, N., and B. Salcedo (2020): "Discerning solution concepts for discrete games," *Journal of Business & Economic Statistics (forthcoming)*.

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

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

Korolyuk, V. S., and Y. V. Borovskich (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.

Mair, B. A., and F. H. Ruymgaart (1996): "Statistical inverse estimation in Hilbert scales," *SIAM Journal on Applied Mathematics*, 56(5), 1424–1444.

Mathé, P., and S. V. Pereverzev (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., and J. L. Powell (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.

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

Shorack, G. R., and J. A. Wellner (2009): *Empirical processes with applications to statistics*. SIAM.

Sokullu, S. (2016): "A semi-parametric analysis of two-sided markets: an application to the local daily newspapers in the USA," *Journal of Applied Econometrics*, 31(5), 843–864.

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

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

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

APPENDIX

A.1 General regularization schemes

Consider a linear operator equation

$$K\varphi = r,$$

where $K : \mathcal{E} \to \mathcal{H}$ is a linear operator and $\mathcal{E}, \mathcal{H}$ are Hilbert spaces. The operator K is assumed to be bounded with $\|K\|^2 \leq \Lambda$ for some $\Lambda < \infty$, but not necessarily compact. Then $K^*K : \mathcal{E} \to \mathcal{E}$ is a normal operator with spectral decomposition

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

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

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

If additionally the operator K is compact, the spectrum of K^*K is countable, and

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

where E_j is a projection operator on the eigenspace corresponding to λ_j. If $(\varphi_j, \psi_j)_{j \geq 1}$ is a sequence of eigenvectors of K^*K, then for all $\varphi \in \mathcal{E}$

$$g(K^*K)\varphi = \sum_{j=1}^{\infty} g(\lambda_j) \langle \varphi, \varphi_j \rangle \psi_j.$$

We are interested in recovering the best approximation to the structural parameter φ when estimates $(\hat{K}, \hat{r})$ of (K, r) are available with $\|\hat{K}\|^2 \leq \Lambda$ a.s. To that end, we consider a slightly more general version of Assumption 3.1:

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 nondecreasing positive function such that* $\lambda \mapsto s_\beta^2(\lambda)/\lambda^\beta$ *is nonincreasing.*

The following two cases are of 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 $s_\beta(0) = 0$.

It is worth mentioning that the mildly ill-posed case allows for the exponential decline of eigenvalues of K^*K provided that the Fourier coefficients of φ_1 also decline exponentially fast. On the other hand, the severely ill-posed case allows for less regular φ_1.

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

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

Our theoretical results require that the regularization scheme satisfies the following assumption:

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)* $\sup_{\lambda\leq\Lambda}|g_\alpha(\lambda)| \leq c_3/\alpha$*.*

It is easy to verify that the following regularization schemes satisfy Assumption A.1.2 in the mildly and the severely ill-posed cases:[10]

1. Tikhonov regularization:
$$g_\alpha(\lambda) = \frac{1}{\alpha+\lambda}.$$
Assumption A.1.2 holds with $c_1 = 1/2$, $c_2 = c_3 = 1$, and $\beta_0 = 2$.

2. Spectral cut-off regularization (principal components):
$$g_\alpha(\lambda) = \lambda^{-1}\mathbb{1}\{\lambda \geq \alpha\}.$$
Assumption A.1.2 holds with $c_1 = c_2 = c_3 = 1$ and every $\beta_0 > 0$.

[10] 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 get around these problems, we can assume that the norm on $\mathcal{H}$ is scaled, so that $\|K\|^2 \leq 1/e$.

3. Iterated Tikhonov regularization:

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

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

4. Landweber-Fridman regularization:

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

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 called the qualification of the regularization scheme. It is well known that the simple Tikhonov regularization exhibits a saturation effect and its bias cannot converge faster than at the rate $O(\alpha^2)$. This can be fixed with the iterated Tikhonov regularization similarly to using higher-order kernels for nonparametric kernel estimators.

The following result describes the rate of convergence of $\hat{\varphi}$ to φ_1 for general regularization schemes:

Theorem A.1. *Suppose that Assumptions A.1.1, A.1.2, and 3.2 are satisfied with* $\beta \leq \beta_0$.[11] *Then in the mildly ill-posed case*

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

while in the severely ill-posed case

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

Before we state the proof, note that

1. In the mildly ill-posed case, the optimal choice of regularization parameter is $\alpha_n \sim \delta_n^{\frac{1}{\beta+1}}$ provided that $\rho_{1n}^{\beta\wedge 1}\log^2 n \lesssim \delta_n/\alpha_n$. Then the convergence rate is

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

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

2. In the severelly ill-posed case, one can choose $\alpha_n \sim \delta_n^{1/2}$. Then the convergence rate is
$$\|\hat{\varphi} - \varphi_1\|^2 = O_P\left(\frac{1}{\log^\beta n}\right),$$
provided that $\delta_n, \rho_n \sim n^{-c}$ for some $c > 0$.

Proof of Theorem A.1. Decompose

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

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

To see that this decomposition holds, note that $\varphi = \varphi_1 + \varphi_0$ and that under Assumption A.1.1, $\varphi_1 = s_\beta(K^*K)\psi$. By the isometry of functional calculus

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

which under Assumptions A.1.2 (i) and 3.2 shows that $\mathbb{E}\|I_n\|^2 = O\left(\delta_n/\alpha_n\right)$.

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

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

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

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

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

Therefore, $\|II_n\|^2 \leq Cc_2^2 s_\beta^2(\alpha_n)$.

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

$$\begin{aligned}
\|III_n\|^2 &\leq \left\| g_\alpha(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I \right\|^2 \left\| s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K) \right\|^2 \|\psi\|^2 \\
&\leq C \sup_{\lambda \leq \Lambda} |(g_\alpha(\lambda)\lambda - 1)|^2 \left\| s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K) \right\|^2 \\
&\leq Cc_2^2 \left\| s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K) \right\|^2 .
\end{aligned}$$

The result follows by Lemma A.1.1 appearing at the end of the section. □

The following result provides the uniform convergence rates for a generic 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*

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

while in the severely ill-posed case

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

Proof. Consider the following decomposition

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

with

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

We bound the first term as

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

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

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

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

Lastly, 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 Lemma A.1.1 as in the proof of Theorem A.1. □

Lemma A.1.1. *Suppose that Assumption 3.2 (iii) is satisfied 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. By Egger (2005), Lemma 3.2,

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

Then the first statement follows since $\|\hat{K} - K\| = O_P(\rho_{1n}^{1/2})$ 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 in the neighborhood of zero.

For the second statement, by Mathé and Pereverzev (2002), Theorem 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 second line follows under Assumption 3.2 (iii) and since $x \mapsto \log^{-\beta/2}(1/x)$ is strictly increasing in the neighborhood of zero. □

A.2 Extreme nonidentification

In this section, we obtain an approximation of the large sample distribution of the Tikhonov-regularized estimators in extremely nonidentified cases. Interestingly, we show that the asymptotic distribution is a weighted sum of independent chi-squared random variables. This result will serve as a starting point for Section 5, where we document a certain transition between the chi-squared and the Gaussian limits in the intermediate cases.[12]

A.2.1 High-dimensional regressions

In the high-dimensional regressions, the identification strength is described by the covariance operator of Z and W. In the extremely nonidentified case, the covariance operator is degenerate and we obtain the following result.

[12]The extreme nonidentification also relates to the weak problem of weak instruments. To the best of our knowledge, a complete treatment of the weak instruments problems in the nonparametric IV and the high-dimensional regressions is not currently available. Our results might be a useful starting point for developing such a theory.

Theorem A.1. *Suppose that Assumption 5.1 is satisfied,* $\mathbb{E}[\langle Z, \delta \rangle W] = 0, \ \forall \delta \in \mathcal{E}$, *and* $\alpha_n n \to \infty$. *Then*

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

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.*

Note that the theorem states the weak convergence in the topology of the Hilbert space $\mathcal{E}$, which is impossible to achieve in the regular identified case. It can be shown that the distribution of inner products of $J(h)$ with $\mu \in \mathcal{E}$ is a weighted sum of chi-squared random variables. Also, interestingly, Theorem A.1 does not require that $\alpha_n \to 0$ as $n \to \infty$.

A.2.2 Nonparametric IV regression

In the nonparametric IV regression, the identification strength is described by the conditional expectation operator. In the extreme non-identified case,

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

where $L_{2,0}(Z) = \{\phi \in L_2(Z) : \ \mathbb{E}\phi(Z) = 0\}$, so that 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 $\mathbb{E}_X$ is 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 and later, 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. The following assumption is a set of mild restrictions on the distribution of the data:

Assumption A.2.1. *(i)* $(Y_i, Z_i, W_i)_{i=1}^n$ *is an i.i.d. sample of* (Y, Z, W); *(ii)* $\mathbb{E}\left[|Y||Z\right] < \infty$, $\mathbb{E}[|Y|^2|W] < \infty$ *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 smoothing respectively over Z and W. The following result holds:

Theorem A.2. *Suppose that Assumption A.2.1 is satisfied,* $\mathbb{E}[\phi(Z)|W] = 0, \forall \phi \in L_{2,0}(Z)$*, and* $n\alpha_n h_z^p \to \infty$ *with* h_w *being 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*.*

It is worth mentioning that obtaining the functional convergence of $\alpha_n n(\hat{\varphi} - \varphi_1)$ is impossible in the case of the nonparametric IV regression. Note that when $h_w \to 0$, then with a suitable normalization we can obtain only convergence to a fixed constant.

A.3 Proofs of main results

Proof of Theorem 3.1. Decompose

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

III_n is the regularization bias that can controlled under Assumption 3.1

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

see Babii (2021). The first term is controlled under Assumption 3.2 (i)

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

The second term is decomposed further

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

It follows from the previous computations and Assumption 3.2 (iii) that

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

and

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

where we use $\|\hat{K}^* - K^*\| = \|\hat{K} - K\|$. Combining all estimates together, we obtain the result. □

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

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

Next, by the Cauchy-Schwartz inequality and Assumption 3.2 (iii) and Assumption 3.3, the first term is

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

The second term is decomposed further similarly as in the proof of Theorem 3.1 in II_n^a and II_n^b. We bound each of the two terms separately. First,

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

Second, under Assumption 3.3, by the inequality in Babii (2020), Lemma A.4.1, see also Nair (2009), Problem 5.8

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

Collecting all estimates together, we obtain the result. $\square$

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

Proposition A.3.1. *Suppose that (i) $(Y_i, Z_i, W_i)_{i=1}^n$ are i.i.d. and $\mathbb{E}|Y_1|^2 \leq C\infty$; (ii) $f_{ZW} \in C^s_M$; (iii) kernel functions $K_z : \mathbf{R}^p \to \mathbf{R}$ and $K_w : \mathbf{R}^q \to \mathbf{R}$ are such that for $l \in \{w, z\}$, $K_l \in L_1 \cap L_2$, $\int K_l(u)\mathrm{d}u = 1$, $\int \|u\|^s K_l(u)\mathrm{d}u < \infty$, and $\int u^k K_l(u)\mathrm{d}u = 0$ for all multindices $|k| = 1, \dots, \lfloor s \rfloor$. 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

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

By the Cauchy-Schwartz inequality

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

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

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

see, e.g., Giné and Nickl (2016), Proposition 4.3.8. Therefore,

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

Next, decompose

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

with

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

By the Cauchy-Schwartz inequality

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

where 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

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

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

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

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

The second claim follows from the inequality

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

and the standard resultson the L_2 error of the kernel density estimator, Giné and Nickl (2016), Chapter 5. □

Proof of Theorem 4.1. Note that under H_0

$$\begin{aligned}\Pr(n\hat{T} > q_{1-\alpha}) &= \Pr\left(\inf_{\phi\in\mathscr{F}_\varepsilon:\,\|\phi\|_\infty=1}\|\sqrt{n}\hat{K}\phi\|^2 > q_{1-\alpha}\right)\\ &\leq \Pr\left(\|\sqrt{n}\hat{K}\phi\|^2 > q_{1-\alpha}\right)\\ &= \Pr\left(\|\sqrt{n}(\hat{K}-K)\phi\|^2 > q_{1-\alpha}\right),\end{aligned}$$

where the last line follows since under H_0, $K\phi = 0$. Since $\mathbb{E}\|WZ\|^2 < \infty$, by the Hilbert space central limit theorem, see Bosq (2000), Theorem 2.7,

$$\begin{aligned}\sqrt{n}(\hat{K}-K)\phi &= \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\langle Z_i,\phi\rangle - \mathbb{E}[W_i\langle Z_i,\phi\rangle]\\ &\xrightarrow{d} \mathbb{G}_\phi,\end{aligned}$$

where $\mathbb{G}_\phi$ is a zero-mean Gaussian process with covariance operator C_ϕ. Next, by the Karhunen-Loève decomposition

$$\|\mathbb{G}_\phi\|^2 = \sum_{j\geq 1}\lambda_{\phi,j}\chi_j^2,$$

where $(\lambda_{\phi,j})_{j\geq 1}$ are eigenvalues of C_ϕ and $(\chi_j^2)_{j\geq 1}$ are i.i.d. chi-squared random variables; see Shorack and Wellner (2009), Chapter 5. Therefore, under H_0

$$\lim_{n\to\infty}\Pr(n\hat{T} > q_{1-\alpha}) \leq \alpha.$$

□

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

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

Under Assumption 5.1

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

Then $\|\hat{K}^*\hat{K}\| \leq \|\hat{K}\|^2 = O_P(n^{-1})$. Therefore, as $\alpha_n n \to \infty$, by the continuous mapping theorem; see van der Vaart and Wellner (2000), Theorem 1.3.6

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

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

Note that

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

Under Assumption 5.1, by the Mourier law of large numbers

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

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

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

Under the Assumption 5.1, by the Borovskich CLT, see Theorem B.1

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

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 $X = (Y, Z, W)$. □

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

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

where $\hat{P}_0$ is the estimator of P_0. Under Assumption A.2.1 (i) since $\mathbb{E}[\phi(Z)|W] = 0$ for all $\phi \in L_{2,0}(Z)$

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

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

$$n P_0 \hat{K}^* \hat{r} = \frac{1}{n h_z^p h_w^q} \sum_{i,j} Y_i P_0 K_z \left(h_z^{-1}(Z_j - z) \right) \bar{K} \left(h_w^{-1}(W_i - W_j) \right).$$

To that end, for every $\mu \in L_2([0,1]^p)$

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

with

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

where $[K_z * P_0\mu](z) = h_n^{-p}\int K\left(h_z^{-1}(z-u)\right)P_0\mu(u)\mathrm{d}u$. Under Assumption A.2.1, by the strong law of large numbers

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

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

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

Lastly, decompose $R_n = R_{1n}+R_{2n}$ with

$$\begin{aligned}
R_{1n} &= \frac{1}{n}\sum_{i=1}^{n}Y_i\left\{[K_z * P_0\mu](Z_i)-P_0\mu(Z_i)\right\}h_w^{-q}\bar{K}(0)\\
R_{2n} &= \frac{1}{n}\sum_{i<j}Y_i\left\{[K_z * P_0\mu](Z_j)-P_0\mu(Z_j)\right\}h_w^{-q}\bar{K}\left(h_w^{-1}(W_i-W_j)\right).
\end{aligned}$$

Note that

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

where the first two lines follow under Assumption A.2.1 (i)-(ii), the third by the Cauchy-Schwartz inequality and $\|P_0\|\leq 1$, and the last by Giné and Nickl (2016), Proposition 4.1.1. (iii). Similarly, since $\mathbb{E}\left[|Y|^2|W\right]<\infty$ a.s. and $\bar{K}\in L_\infty$, by the moment inequality in Korolyuk and Borovskich (1994), Theorem 2.1.3

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

□

Proof of Theorem 5.1. Put $b_n=\alpha_n(\alpha_nI+K^*K)^{-1}\varphi_1$ and note that $(\alpha_nI+\hat{K}^*\hat{K})^{-1}\hat{K}^*=$

$\hat{K}^*(\alpha_n I + \hat{K}\hat{K}^*)^{-1}$. Then, similarly to the proof of Theorem 3.1, decompose

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

Since

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

it remains to show that all other terms are asymptotically negligible. Note that since $\mu_0 \in \mathcal{N}(K)$,

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

Then

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

By the Cauchy-Schwartz inequality and computations similar to those in the proof of Theorem 3.1

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

Next, under Assumption 3.1 from the proof of Theorem 3.1 we also know that $\|b_n\| =$

$O\left(\alpha_n^{\frac{\beta}{2}\wedge 1}\right)$ and that $\|Kb_n\| = O\left(\alpha_n^{\frac{\beta+1}{2}\wedge 1}\right)$. Therefore

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

and

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

Lastly, the bias is zero by Eq. A.1 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 that under Assumption 5.1

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

Therefore, since under Assumption 5.2 $n\alpha_n^{1+\beta\wedge 1} \to 0$ and $n\alpha_n \to \infty$,

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

with

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

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

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

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

$$
\begin{aligned}
\zeta_n^0 &= \frac{1}{n}\sum_{i=1}^n (Y_i - \langle Z_i, \varphi_1\rangle)\langle Z_i, \mu_0\rangle\left\|W_i^0\right\|^2 \\
\mathbf{U}_n^0 &= \frac{1}{n}\sum_{i<j}\left\{\langle Z_i, \mu_0\rangle(Y_j - \langle Z_j, \varphi_0\rangle) + \langle Z_j, \mu_0\rangle(Y_i - \langle Z_i, \mu_0\rangle)\right\}\left\langle W_i^0, W_j^0\right\rangle.
\end{aligned}
$$

Under Assumption 5.1 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 $\mathbb{E}\left[W^0(Y - \langle Z, \varphi_1\rangle)\right] = P_0\mathbb{E}[WU] + P_0 K\varphi_0 = 0$. Therefore, $\mathbf{U}_n^0$ is a centered degenerate U-statistics with a kernel function h. Under Assumption 5.1 by the CLT for the 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

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

It follows from Bakushinskii (1967) that $\|\alpha_n(\alpha_n I + KK^*)^{-1}W^1\| = o(1)$. Then under Assumption 5.1 by the dominated convergence theorem

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

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

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

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

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

where the last line follows under Assumptions 5.1 and 5.2, and previous discussions. Finally, if W^0 degenerates to zero, then $S_n^0 = 0$ and

$$\alpha_n n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle \xrightarrow{d} 0.$$

□

Proof of Theorem 5.2. Similarly to the proof of Theorem 5.1, decompose

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

Under Assumption 5.3 by the Lindeberg-Feller central limit theorem

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

It remains to show that all other terms normalized with π_n tend to zero. For II_n, by the Cauchy-Schwartz inequality

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

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 5.4. 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

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

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

$$
\begin{aligned}
IV_n^c &\leq \left\|\hat{K} - K\right\| \|b_n\| \left\|K(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1 + \gamma\wedge\frac{1}{2}}}{\sqrt{n\alpha_n}} \\
IV_n^d &\leq \left\|\hat{K}^* - K^*\right\| \left\|\hat{K} - K\right\| \|b_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge 1 + \gamma\wedge 1}}{n\alpha_n} \\
IV_n^e &\leq \left\|\hat{K}^* - K^*\right\| \|Kb_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| \lesssim_P \frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2} + \gamma\wedge 1}}{\sqrt{n\alpha_n}}.
\end{aligned}
$$

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

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

and

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

Lastly,

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

Therefore, under Assumption 5.4, all terms but I_n are $o_P(1)$. □

ONLINE APPENDIX

B.1 Generalized inverse

In this section we collect some facts about the generalized inverse operator from the operator theory; see also Carrasco, Florens, and Renault (2007) for a comprehensive review of different aspects of the theory of ill-posed inverse models in econometrics. Let $\varphi \in \mathcal{E}$ be a structural parameter in a Hilbert space $\mathcal{E}$ and let $K : \mathcal{E} \to \mathcal{H}$ be a bounded linear operator mapping to a Hilbert spaces $\mathcal{H}$. Consider the functional equation

$$K\varphi = r.$$

If the operator K is not one-to-one, then structural parameter φ is not point identified and the identified set is a closed linear manifold described as $\Phi^{\mathrm{ID}} = \varphi + \mathcal{N}(K)$, where $\mathcal{N}(K) = \{\phi : K\phi = 0\}$ is the null space of K; see Figure B.1. The following result offers equivalent characterizations of the identified set; see Groetsch (1977), Theorem 3.1.1 for a formal proof.

Proposition B.1.1. *The identified set Φ^{ID} equals to the set of solutions to*

(i) the least-squares problem: $\min_{\phi \in \mathcal{E}} \|K\phi - r\|$;

(ii) the normal equations: $K^*K\phi = K^*r$, where K^* is the adjoint operator to K.

The generalized inverse is formally defined below.

Definition B.1.1. *The generalized inverse of the operator K is a unique linear operator $K^\dagger : \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp \to \mathcal{E}$ defined by $K^\dagger r = \varphi_1$, where $\varphi_1 \in \Phi^{\mathrm{ID}}$ is a unique solution to*

$$\min_{\phi \in \Phi^{\mathrm{ID}}} \|\phi\|. \tag{A.1}$$

For nonidentified linear models, the generalized inverse maps r to the unique minimal norm element of Φ^{ID}. It follows from Eq. A.1 that φ_1 is a projection of 0 on the identified set. Therefore, φ_1 also equals to the projection of the structural parameter φ on the orthogonal complement to the null space $\mathcal{N}(K)^\perp$, see Figure B.1 and we call φ_1 the best approximation to the structural parameter φ. The generalized inverse operator is typically a discontinuous map as illustrated in the following proposition; see Groetsch (1977), pp.117-118 for more details.

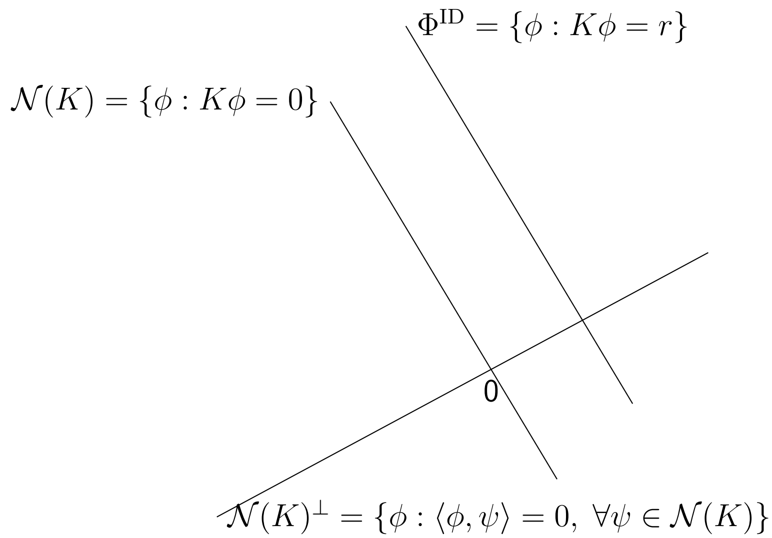


Figure B.1: Fundamental subspaces of $\mathcal{E}$.

Proposition B.1.2. *Suppose that the operator K is compact. Then the generalized inverse $K^\dagger$ is continuous if and only if $\mathcal{R}(K)$ is finite-dimensional.*

The following example illustrates this when K is an integral operator on spaces of square-integrable functions.

Example B.1.1. *Suppose that K is an integral operator*

$$\begin{aligned} K : L_2 &\to L_2 \\ \phi &\mapsto (K\phi)(w) = \int \phi(z)k(z,w)\mathrm{d}z \end{aligned}$$

Then K is compact whenever the kernel function k is square integrable. In this case the generalized inverse is continuous if and only if k is a degenerate kernel function

$$k(z,w) = \sum_{j=1}^{m} \phi_j(z)\psi_j(w).$$

It is worth stressing that in the NPIV model, the kernel function k is typically a non-degenerate probability density function. Moreover, in econometric applications r is usually estimated from the data, so that $K^\dagger \hat{r} \xrightarrow{p} K^\dagger r = \varphi_1$ may not hold even when $\hat{r} \xrightarrow{p} r$ due to the discontinuity of $K^\dagger$.[13] In other words, we are faced with an ill-posed inverse problem. Tikhonov regularization can be understood as a method that smooths out the discontinuities of the generalized inverse $(K^*K)^\dagger$.[14]

B.2 Degenerate U-statistics in Hilbert spaces

B.2.1 Wiener-Itô integral

In this section, we review relevant for us theory of the degenerate U-statistics in Hilbert spaces. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space and let H be a separable Hilbert space. We use $L_2(\mathcal{X}^m, H)$ to denote the space of all functions $f : \mathcal{X}^m \to H$ such that $\mathbb{E}\|f(X_1, \dots, X_m)\|^2 < \infty$. The stochastic process $\{\mathbb{W}(A), A \in \Sigma_\mu\}$ indexed by the sigma-field $\Sigma_\mu = \{A \in \Sigma : \ \mu(A) < \infty\}$ is called the *Gaussian random measure* if

[13] In practice the situation is even more complex, because the operator K is also estimated from the data.

[14] By Proposition B.1.1 solving $K\varphi = r$ is equivalent to solving $K^*K\varphi = K^*r$. The latter is more attractive to work with because the spectral theory of self-adjoint operators in Hilbert spaces applies to K^*K.

1. For all $A \in \Sigma_\mu$
$$\mathbb{W}(A) \sim N(0, \mu(A));$$

2. For any collection of disjoint sets $(A_k)_{k=1}^K$ in Σ_μ, $\mathbb{W}(A_k), k = 1, \dots, K$ are independent and
$$\mathbb{W}\left(\bigcup_{k=1}^K A_k\right) = \sum_{k=1}^K \mathbb{W}(A_k).$$

Let $(A_k)_{k=1}^K$ be pairwise disjoint sets in Σ_μ and let S_m be a set of simple functions $f \in L_2(\mathcal{X}^m, H)$ such that

$$f(x_1, \dots, x_m) = \sum_{i_1,\dots,i_m=1}^K c_{i_1,\dots,i_m} \mathbb{1}_{A_{i_1}}(x_1) \times \dots \times \mathbb{1}_{A_{i_m}}(x_m),$$

where $c_{i_1,\dots,i_m}$ is zero if any of two indices $i_1, \dots, i_m$ are equal, i.e., f vanishes on the diagonal. For a Gaussian random measure $\mathbb{W}$ corresponding to P, consider the following random operator $J_m : S_m \to H$

$$J_m(f) = \sum_{i_1,\dots,i_m=1}^K c_{i_1,\dots,i_m} \mathbb{W}(A_{i_1}) \dots \mathbb{W}(A_{i_m}).$$

The following three properties are immediate from the definition of J_m:

1. Linearity;

2. $\mathbb{E} J_m(f) = 0$;

3. Isometry: $\mathbb{E}\langle J_m(f), J_m(g)\rangle_H = \langle f, g\rangle_{L_2(\mathcal{X}^m, H)}$.

The set S_m is dense in $L_2(\mathcal{X}^m, H)$ and J_m can be extended to a continuous linear isometry on $L_2(\mathcal{X}^m, H)$, called the Wiener-Itô integral.

Example B.2.1. *Let $(B_t)_{t\geq 0}$ be a real-valued Brownian motion. Then for any $(t, s] \subset [0, \infty)$, $\mathbb{W}((t, s]) = B_s - B_t$ is a Gaussian random measure (μ is the Lebesgue measure) with the Wiener-Itô integral $J : L_2([0, \infty), \mathrm{d}t) \to \mathbf{R}$ defined as $J(f) = \int f(t)\mathrm{d}B_t$.*

B.2.2 Central limit theorem

Let $(\mathcal{X}, \Sigma, P)$ be a probability space, where $\mathcal{X}$ is a separable metric space and Σ is a Borel σ-algebra. Let $(X_i)_{i=1}^n$ be i.i.d. random variables taking values in $(\mathcal{X}, \Sigma, P)$. Consider some symmetric function $h : \mathcal{X} \times \mathcal{X} \to H$, where H is a separable Hilbert space. The H-valued U-statistics of degree 2 is defined as

$$\mathbf{U}_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j).$$

The U-statistics is called degenerate if $\mathbb{E}h(x_1, X_2) = 0$. The following result provides the limiting distribution of the degenerate H-valued U-statistics; see Korolyuk and Borovskich (1994), Theorem 4.10.2 for a formal proof.

Theorem B.1. *Suppose that $\mathbf{U}_n$ is a degenerate U-statistics such that $\mathbb{E}h(X_1, X_2) = 0$ and $\mathbb{E}\|h(X_1, X_2)\|^2 < \infty$. Then*

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

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 and $\mathbb{W}$ is a Gaussian random measure on H.