

Is completeness necessary? Estimation in nonidentified linear models*

Andrii Babii[†]

UNC Chapel Hill

Jean-Pierre Florens[‡]

Toulouse School of Economics

April 20, 2020

Abstract

This paper documents the consequences of the identification failures in a class of linear ill-posed inverse models. The Tikhonov-regularized estimator converges to a well-defined limit equal to the best approximation of the structural parameter in the orthogonal complement to the null space of the operator. We illustrate that in many instances the best approximation may coincide with the structural parameter or at least may reasonably approximate it. We obtain new nonasymptotic risk bounds in the uniform and the Hilbert space norms for the best approximation. Nonidentification has important implications for the large sample distribution of the Tikhonov-regularized estimator, and we document the transition between the Gaussian and the weighted chi-squared limits. The theoretical results are illustrated for the nonparametric IV and the functional linear IV regressions and are further supported by the Monte Carlo experiments.

Keywords: nonidentified linear models, weak identification, nonparametric IV regression, functional linear IV regression, Tikhonov regularization.

JEL Classifications: C14, C26

*We acknowledge support from ANR under grant ANR-17-EURE-0010 (Investissements d’Avenir program). We are 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 the Duke workshop, Triangle Econometrics Conference, 4th ISNPS, 2018 NASMES, 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 – esplanade de l’Université, 31080 Toulouse, France.

1 Introduction

Structural nonparametric and high-dimensional econometric models often lead to ill-posed inverse problems. Among many examples, we may quote the nonparametric instrumental regression, functional linear regressions, measurement errors, and random coefficients models. All these examples generate the functional linear equation

$$K\varphi = r,$$

where φ and r are some functions, and K is a linear operator. Classical numerical inverse problems literature, see [Engl, Hanke, and Neubauer \(1996\)](#), studies the *deterministic* ill-posed inverse 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* linear ill-posed inverse problem.

Identification is an integral part of the econometric analysis going back to [Koopmans \(1949\)](#), [Koopmans and Reiersol \(1950\)](#), and [Rothenberg \(1971\)](#) in the parametric case. In nonparametric ill-posed inverse problems, r and K 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. Unicity of the solution is equivalent to assuming that K is a one-to-one operator, or equivalently to $K\phi = 0 \implies \phi = 0$ by the linearity of K . Note that in econometric applications, the operator K is usually unknown and the estimated operator \hat{K} has a finite rank and is not one-to-one for any finite sample size.

The maximum likelihood estimator when there is a lack of identification leads to a flat likelihood in some regions of the parameter space and then to some ambiguity on the choice of a maximum. It is then natural to characterize the limit of the estimator for such a potentially nonidentified model. In the nonidentified inverse model, the identified set is a 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 . As K or \hat{K} may fail to be one-to-one and do not have a continuous generalized inverse, a regularization method is needed to estimate φ .

Several regularization methods are commonly used and we will focus our presentation on the Tikhonov-regularized estimator

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

where \hat{K}^* is the adjoint operator of \hat{K} , \hat{r} is an estimator of r , and α_n is a regularization parameter. For deterministic problems, it is well-known that the Tikhonov regularization allows us to recover accurately the best approximation to φ . The first original contribution of this paper, is to show that this is also the case in generic and realistic

econometric settings with the estimated operator K . To that end, we obtain novel non-asymptotic uniform and Hilbert space risk bounds for the Tikhonov-regularized estimator of the best approximation to φ in the orthogonal complement of $\mathcal{N}(K)$, denoted $\mathcal{N}(K)^\perp$. Note that the space $\mathcal{N}(K)^\perp$ may often be spanned by a large number of basis functions (infinitely many if $\mathcal{N}(K)$ is finite-dimensional), which allows for a reasonable approximation to φ even when $\varphi \notin \mathcal{N}(K)^\perp$.¹ This leads to an attractive projection interpretation for the nonparametric IV estimator of φ when $\varphi \notin \mathcal{N}(K)^\perp$, similar to the projection interpretation of the regression function estimated with the ordinary least-squares. In contrast, the parametric linear IV estimator does have the projection interpretation; see the recent work of [Escanciano and Li \(2018\)](#) for an IV estimator that enjoys projection interpretation.

Our second original contribution is to illustrate that the nonidentification has important implications for the large-sample approximation to the distribution of the Tikhonov-regularized estimator and its linear functionals. We find that in the extremely nonidentified case, the asymptotic distribution is driven by the degenerate U-statistics, while in the intermediate cases, we observe a certain transition between the Gaussian and the weighted chi-squared limits.

The paper is organized as follows. Section 2 discusses the identification in the nonparametric IV and the functional linear IV regressions. In Section 3, we obtain the non-asymptotic risk bounds in the uniform and the Hilbert space norms for a class of Tikhonov-regularized estimators.² Section 4 shows that in the extreme case of identification failures, the Tikhonov-regularized estimator is driven by the degenerate U-statistics in large samples. Section 5 illustrates the transition between the Gaussian and the weighted chi-squared asymptotics in the intermediate cases for the functional linear IV regression. 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 A.1. We also review several relevant for us

¹Indeed, even when φ has an infinite series expansion, most of the nonlinearities can usually be captured by a fairly small number of basis functions. For this reason, the best approximation plays an important role in the numerical analysis and the engineering literature, see [Engl, Hanke, and Neubauer \(1996\)](#). It is also worth stressing that the Tikhonov-regularized estimator converges to the best approximation without knowing the basis of $\mathcal{N}(K)^\perp$.

²These results can be understood as a generalization of [Carrasco, Florens, and Renault \(2007\)](#), [Darolles, Fan, Florens, and Renault \(2011\)](#), [Florens, Johannes, and Van Belleghem \(2011\)](#), and [Carrasco, Florens, and Renault \(2014\)](#) who largely focus on L_2 convergence rates and/or identified cases; see also [Gagliardini and Scaillet \(2012\)](#) for the uniform convergence rates for the Tikhonov regularization in the identified case and [Babii \(2020b\)](#) for the uniform inference for Tikhonov-regularized estimators that rely on the results obtained in this paper.

results from the theory of the generalized inverse operators and the theory of the Hilbert space valued U-statistics in the Online Appendices [B.1](#) and [B.2](#).

2 Identification

Consider the functional linear equation

$$K\varphi = r,$$

where $K : \mathcal{E} \rightarrow \mathcal{H}$ is a bounded 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

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

We illustrate the statistical interpretation of the one-to-one property of K in the nonparametric instrumental regression and the functional linear IV regression.

Example 2.1 (Nonparametric instrumental regression). *Consider*

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

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 conditional mean-independence of the unobservable U from the instrumental variable W 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 $K : L_2(Z) \rightarrow L_2(W)$ is a conditional expectation operator.³ 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

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

³For a random variable X , we define $L_2(X) = \{\phi : \mathbb{E}|\phi(X)|^2 < \infty\}$.

It is a (non-linear) dependence condition between the endogenous regressor Z and the instrument W .⁴

Example 2.2 (Functional linear instrumental regression). Consider

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

where $(Y, Z, W) \in \mathbf{R} \times \mathcal{E} \times \mathcal{H}$, see [Florens and Van Bellegem \(2015\)](#).⁵ The covariance restriction between the unobservable U and the Hilbert space-valued instrumental variable $W \in \mathcal{E}$ leads to the functional linear equation

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

If (Z, W) has zero mean, then the operator $K : \mathcal{E} \rightarrow \mathcal{H}$ is a covariance operator. The completeness condition requires that the covariance operator is one-to-one, or

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

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*

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

⁴It is well-known that the completeness condition is not testable. Nevertheless, the nonparametric identification in econometric models often relies on the completeness argument. Other prominent examples include: the measurement error models, 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 \(2020a\)](#); functional regression models, see [Florens and Van Bellegem \(2015\)](#) and [Benatia, Carrasco, and Florens \(2017\)](#). Nonlinear variations of the completeness condition are also exploited in quantile treatment effect models, see [Chernozhukov and Hansen \(2005\)](#) and nonlinear asset pricing models, see [Chen and Ludvigson \(2009\)](#). The importance of the completeness condition for the nonparametric identification generated some efforts aimed to understand complete and incomplete distributions, see [D'Haultfoeuille \(2011\)](#), [Andrews \(2017\)](#), and [Hu and Shiu \(2018\)](#) among others.

⁵The functional regression models are suitable for handling the high-dimensional data sampled at mixed frequencies, see [Andreou, Ghysels, and Kourtellis \(2013\)](#), [Ghysels, Sinko, and Valkanov \(2007\)](#), and [Ghysels, Santa-Clara, and Valkanov \(2006\)](#); see also [Babii \(2020a\)](#) who estimate the real-time price elasticity for spot markets and [Benatia, Carrasco, and Florens \(2017\)](#) who study the electricity consumption in Canada.

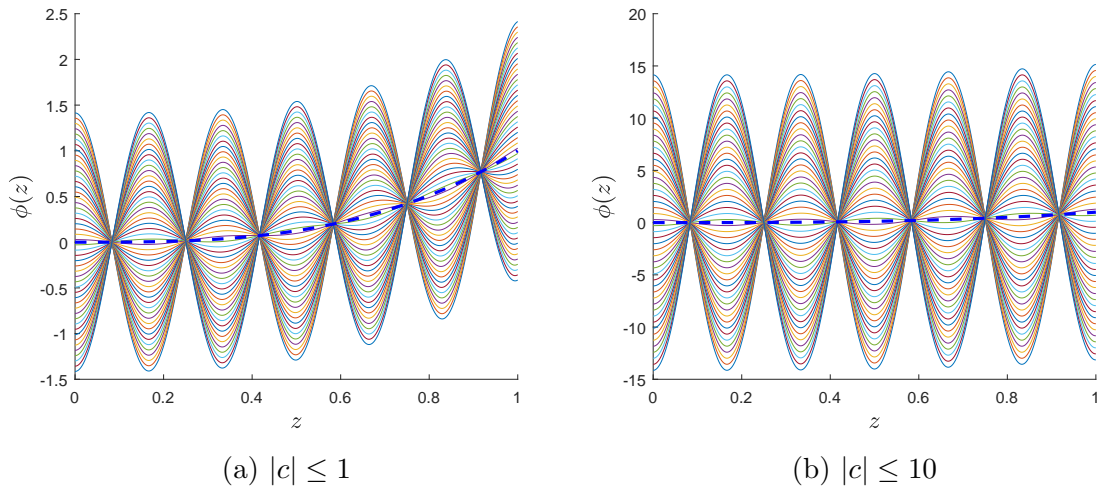


Figure 1: Functions in bounded subsets of the identified set with $k = 3$. Blue thick line is the true structural function $\varphi(z) = z^3$.

where $\mathcal{N}(K)$ is the null space of K . Note that without further restrictions, the identified set is unbounded. To visualize the identified set, suppose that $\mathcal{N}(K)$ is one-dimensional, e.g., spanned by some frequency $\varphi_k(x) = \sin(2\pi kx)$ with $k \in \mathbf{N}$, so that the identified set is $\Phi^{\text{ID}} = \varphi + \{c\varphi_k : c \in \mathbf{R}\}$. Since the scaling constant $c \in \mathbf{R}$ can be arbitrarily large, the identified set Φ^{ID} is not informative of the structural parameter φ . Imposing uniform norm bounds φ , the identified can be localized, but it still may contain enormous amount of functions and is not informative on the global shape of φ as illustrated on Figure 1.

A natural solution is to eliminate the "bad" frequency φ_k spanning $\mathcal{N}(K)$ and to use all the remaining frequencies in the orthogonal complement to the null space of K , denoted $\mathcal{N}(K)^\perp$, to approximate the structural parameter φ . This leads to the notion of the best approximation to φ , denoted φ_1 , which unlike the identified set can be informative on the global shape of φ , see Figure 2.

Formally, 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 $\mathcal{N}(K)$.

Since $\mathcal{N}(K)^\perp = \overline{\mathcal{R}(K^*)}$, see [Luenberger \(1997\)](#), p.157, the best approximation φ_1 equals to the structural parameter φ whenever the structural parameter φ belongs to $\mathcal{R}(K^*)$. This condition has also an appealing regularity interpretation known as

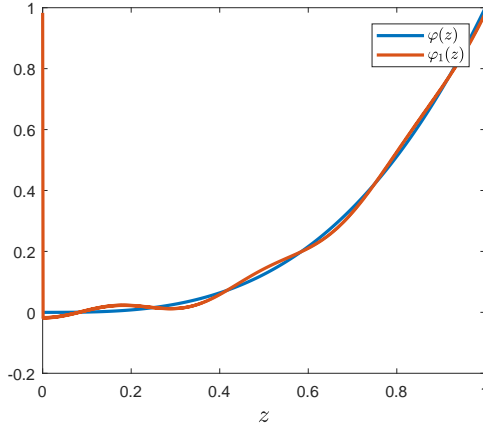


Figure 2: Structural parameter $\varphi(z) = z^3$ and its best approximation φ_1 .

the source condition.⁶ To see this, note that $\mathcal{R}(K^*) = \mathcal{R}(K^*K)^{1/2}$, cf., 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 despite the fact that the completeness condition fails.

If the function φ is not sufficiently regular compared to the ill-posedness of K , then the best approximation can still 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 instrumental regression. It is worth stressing that we do not propose to estimate φ_1 using the basis of the orthogonal complement to the null space, which is usually unknown. The attractive feature of the Tikhonov-regularized estimator is that it converges to φ_1 without knowing the basis of $\mathcal{N}(K)^\perp$.

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

$$K : L_2(Z) \rightarrow L_2(W)$$

$$\phi \mapsto \mathbb{E}[\phi(Z)|W]$$

is compact. By the spectral theorem, there exists $(\lambda_j, \varphi_j, \psi_j)_{j \geq 1}$, where $\lambda_j \rightarrow 0$ is a sequence of singular values, $(\varphi_j)_{j \geq 1}$ is a complete orthonormal system of $\mathcal{N}(K)^\perp =$

⁶The source condition requires that $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$ for some $\beta \in (0, \infty)$, see Carrasco, Florens, and Renault (2007). It describes the regularity of the function φ compared to the smoothing properties of the operator K and is in a certain sense unavoidable, cf., Engl, Hanke, and Neubauer (1996), Theorem 4.11.

$\overline{\mathcal{R}(K^*)}$, and $(\psi_j)_{j \geq 1}$ is the complete orthonormal system of $N^\perp(K^*) = \overline{\mathcal{R}(K)}$. The structural function φ can be identified 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 identified 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 $f_k(z) = f_{Z|W=w_k}(z)$ with $k \geq 1$. Then*

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

and if $\varphi \in \text{span}\{f_k : k \geq 1\}$, then we clearly have $\varphi \in \mathcal{N}(K)^\perp$. The structural parameter φ is identified if it can be represented as a linear combination of $(f_k)_{k \geq 1}$.

It is worth stressing that functions encountered in practical 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 might still be able to capture most of the nonlinearities.

3 Nonasymptotic risk bounds

In this section, we derive the nonasymptotic risk bounds for the Tikhonov-regularized estimators in the Hilbert space and the uniform norms. All these results are uniform over the relevant classes of models and do not rely on the completeness condition. We also illustrate our bounds in the special cases of the nonparametric IV and the functional linear IV regressions.

3.1 Tikhonov-regularized estimator

Estimation of the structural function φ is an ill-posed inverse problem and requires regularization at least for two reasons. First, the generalized inverse of K is typically not continuous, see Appendix B.1 for more details. 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 sequence of regularization parameters and $\|\cdot\|$ is the norm of the relevant Hilbert space. The problem admits the following closed-form solution

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

The estimator enjoys two fundamental properties:

1. It is well-defined even if K or \hat{K} is not one-to-one.
2. If $\alpha_n \rightarrow 0$ suitably fast, it converges to the best approximation of φ in $\mathcal{N}(K)^\perp$.

Indeed, the operator $\alpha_n I + \hat{K}^* \hat{K}$ has eigenvalues $\alpha_n + \hat{\lambda}_j$ that do not vanish for $\alpha_n > 0$, even if $\hat{\lambda}_j = 0$ for some j . As a result, the inverse operator and the estimator $\hat{\varphi}$ are always well-defined. In this way, the Tikhonov-regularized estimator smoothes out the discontinuities of the generalized inverse operator $(\hat{K}^* \hat{K})^\dagger$. In the following section, we also show that the Tikhonov estimator converges to the best approximation φ_1 of φ in $\mathcal{N}(K)^\perp$ and characterize the estimation accuracy in the Hilbert space and the uniform norms.⁷

3.2 Hilbert space risk bounds

The class of models consists of distributions of the data for which the best approximation φ_1 has regularity $\beta > 0$ compared to the operator K . For each particular model, this class will be augmented to include probability distributions satisfying some additional moment conditions.

Assumption 3.1. *Suppose that the distribution of the data is such that (φ, K) belongs to the following regularity class*

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

for some positive constants β, C , where $\|K\| = \sup_{\|\phi\| \leq 1} \|K\phi\|$ is the operator norm of K .

⁷To the best of our knowledge, [Florens, Johannes, and Van Bellegem \(2011\)](#) is the only study that derives formally L_2 convergence rates for the best approximation; see also [Chen and Pouzo \(2012\)](#) for a consistency result.

Note that if $\varphi \in \mathcal{F}(\beta, C)$ with $\beta \geq 1$ and $\varphi_1 = \varphi$, the condition reduces to the source condition discussed, e.g., in Carrasco, Florens, and Renault (2007) and Darolles, Fan, Florens, and Renault (2011). The source condition does not require the Hölder or the Sobolev smoothness and allows for non-differentiable and even discontinuous functions. It quantifies instead the intrinsic property of the ill-posed model – how fast the Fourier coefficients of φ_1 decrease to zero relative to the speed at which eigenvalues of the operator K^*K tend to zero. In particular, it allows for the severely ill-posed models, whenever the regularity of φ_1 matches the ill-posedness of K .

We observe estimators (\hat{r}, \hat{K}) and impose the following condition.

Assumption 3.2. *Suppose that*

$$(i) \mathbb{E} \left\| \hat{r} - \hat{K}\varphi \right\|^2 \leq C_1\delta_{1n}, \quad (ii) \mathbb{E} \left\| \hat{K}\varphi_0 \right\| \leq C_2\delta_{2n}, \quad (iii) \mathbb{E} \left\| \hat{K} - K \right\|^2 \leq C_3\rho_{1n},$$

where the constants C_1, C_2, C_3 depend only on $\mathcal{F}(\beta, C)$ and do not depend on (φ, K) .

The verification of this high-level condition will be illustrated for two models at the end of this section. Conditions (i) and (ii) are essentially assumptions on the speed of convergence of the residuals in the stochastic ill-posed inverse model. In nonidentified models, residuals may have an additional non-zero component $\hat{K}\varphi_0$ coming from the nonidentified part of the function φ_0 . Condition (iii) is a standard assumption on how well the operator K is estimated by \hat{K} in the operator norm.

The following result describes a nonasymptotic risk bound in the norm of the Hilbert space \mathcal{E} under (possible) identification failures.

Theorem 3.1. *Suppose that Assumptions 3.1 and 3.2 are satisfied. Then*

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

where $\mathcal{F} = \mathcal{F}(\beta, C)$ and the constant can be found in the proof.

The risk bound tells us the *guaranteed* expected estimation accuracy for all DGPs in the source class given that the econometrician has a sample of a particular size n and sets the tuning parameter to α_n . The convergence of $\hat{\varphi}$ to φ_1 is driven by the following elements:

- residuals of the ill-posed inverse problem of order $O \left(\frac{\delta_{1n}}{\alpha_n} \right)$;

- residuals due to the identification failure of order $O\left(\frac{\delta_{2n}}{\alpha_n}\right)$;
- estimation error of the operator of order $O\left(\frac{\rho_{1n}\alpha^{\beta\wedge 1}}{\alpha_n}\right)$;
- regularization bias of order $O\left(\alpha_n^{\beta\wedge 2}\right)$.

Note that if $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$ with $\beta \geq 1$, then $\varphi_0 = 0$, $\rho_{1n} = 0$, and we recover the same convergence rate as for the identified models. In this case, the Tikhonov regularized estimator converges to φ , although the completeness condition fails. More generally, we distinguish the following possibilities:

- identified models: $\varphi_0 = 0$ and $\rho_{1n}\alpha^{\beta\wedge 1} \lesssim \delta_{2n}$, in which case the convergence rate is driven by the residuals and the regularization bias;
- weakly identified models: $\varphi_0 = 0$ and $\delta_{2n} \lesssim \rho_{1n}\alpha^{\beta\wedge 1}$, in which case residuals and the estimation of the operator drive the convergence rate.
- nonidentified models: $\varphi_0 \neq 0$ and $\rho_{1n}\alpha^{\beta\wedge 1} \lesssim \delta_{2n}$, in which case we observe an additional effect of the estimation of the operator at φ_0 compared to the identified case.
- strongly nonidentified models: $\varphi_0 \neq 0$ and $\delta_{2n} \lesssim \rho_{1n}\alpha^{\beta\wedge 1}$, in which case we observe an additional effect of the estimation of the operator at φ_0 compared to the weakly identified case.

In the special case of the identified model, we recover the optimal convergence rate for $\mathcal{F}(\beta, C)$ for all $\beta \in (0, 2]$, see [Mair and Ruymgaart \(1996\)](#). For $\beta > 2$, the Tikhonov regularization achieves the optimal rate with some additional iterations, cf., [Carrasco, Florens, and Renault \(2007\)](#).

3.3 Uniform risk bounds

Suppose now that the space of continuous functions on a compact set $D \subset \mathbf{R}^d$, denoted $C(D)$, is embedded into the space \mathcal{E} . Suppose also that $\mathcal{R}(K^*) \subset C(D)$ and that $\varphi_1 \in C(D)$. Let also $\|K^*\|_{2,\infty} = \sup_{\|\phi\|_{\infty} \leq 1} \|K^*\phi\|_{\infty}$ be the mixed operator norm of K^* . For the uniform risk bounds we introduce an additional assumption below.

Assumption 3.3. *Suppose that $\|K^*\|_{2,\infty} \leq C_4$ and that*

$$\mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty}^2 \leq C_4 \rho_{2n}$$

for a constant C_4 that depends only on $\mathcal{F}(\beta, C)$ and not on (φ, K) .

The following result describes a nonasymptotic bound on the uniform estimation accuracy under (possible) identification failures.

Theorem 3.2. *Suppose that Assumptions 3.1, 3.2, and 3.3 are satisfied and $\alpha_n \vee \rho_{2n} = O(1)$. Then*

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

where $\mathcal{F} = \mathcal{F}(\beta, C)$, $\beta > 1$ and the constant can be found in the proof.

The uniform estimation accuracy depends on the additional residual term of the nonidentified element φ_0 . If $\varphi \in \mathcal{R}(K^*K)^{\beta/2}$, $\beta \geq 1$, then $\varphi = \varphi_1$, and the Tikhonov-regularized estimator is uniformly consistent even if the completeness condition fails.

3.4 Applications

3.4.1 Functional linear IV regression

Following, Example 2.2, the econometrician observes an i.i.d. sample⁸ $(Y_i, Z_i, W_i)_{i=1}^n$. Then

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

are estimated with

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

Under Assumption 3.1, by elementary computations

$$\mathbb{E} \|\hat{r} - \hat{K}\varphi\|^2 = \frac{\mathbb{E}\|UW\|^2}{n}, \quad \mathbb{E} \|\hat{K}\varphi_0\|^2 = \frac{\mathbb{E}\|W\langle\varphi_0, Z\rangle\|^2}{n}, \quad \mathbb{E} \|\hat{K} - K\|^2 \leq \frac{\mathbb{E}\|ZW\|^2}{n}.$$

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

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

⁸The i.i.d. assumption can be relaxed to the covariance stationarity and absolute summability of autocovariances in the L_1 sense, see Babii (2020a).

The functional linear IV regression is either identified ($\varphi_0 = 0$) or nonidentified ($\varphi_0 \neq 0$). Then conditions $\alpha_n \rightarrow 0$ and $\alpha_n n \rightarrow \infty$ as $n \rightarrow \infty$ are sufficient to guarantee the consistency in the Hilbert space norm.

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 additionally 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 Hölder ball with smoothness $s > d/2$. By the Hoffman-Jørgensen's inequality, e.g., see [Giné and Nickl \(2016\)](#), p.129

$$\begin{aligned} \left(\mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty}^2 \right)^{1/2} &\leq \left(\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i] \right\|_\infty^2 \right)^{1/2} \\ &\leq 12\sqrt{3} \left(16\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n Z_i W_i - \mathbb{E}[Z_i W_i] \right\|_\infty + \frac{C^2}{n} \right) \end{aligned}$$

and by the bracketing moment inequality, e.g., see [Giné and Nickl \(2016\)](#), p.202

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

Therefore, $\rho_{2n} = n^{-1}$, and the bound in the Theorem 3.2 becomes

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

Then conditions $\alpha_n \rightarrow 0$ and $\alpha_n n^{1/2} \rightarrow \infty$ as $n \rightarrow \infty$ ensure the uniform consistency of $\hat{\varphi}$.

3.4.2 Nonparametric IV regression

Following Example 2.1, rewrite the model as

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

where K is an operator from $L_2([0, 1]^p)$ to $L_2([0, 1]^q)$. We estimate r and K with kernel smoothing

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

where K_w, K_z are kernel functions (e.g., products of univariate kernels) and h_n is a bandwidth parameter. Under mild assumptions, by Proposition A.1.1, $\delta_{1n} = \delta_{2n} = \frac{1}{nh_n^q} + h_n^{2s}$ and $\rho_{1n} = \frac{1}{nh_n^{p+q}} + h_n^{2s}$, where s is the Hölder smoothness of the joint density of (Z, W) . It follows from Theorem 3.1 that

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

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

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 (2020b), Proposition A.3.1. Then it follows from Theorem 3.2 that

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

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

4 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. In the following section, we document a certain transition between the chi-squared and the Gaussian limits in the intermediate cases lying between the point identification and the extreme nonidentification. The case of the extreme nonidentification is also a manifestation of the weak instruments problem.⁹

4.1 Functional linear IV regression

In the functional linear IV regression, 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.

Theorem 4.1. *Suppose that Assumption 5.1 is satisfied, $\mathbb{E}[\langle Z, \delta \rangle W] = 0$, $\forall \delta \in \mathcal{E}$, and $\alpha_n n \rightarrow \infty$. Then*

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

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 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, this result does not require $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$.

4.2 Nonparametric IV regression

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

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

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

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

⁹To the best of our knowledge, a complete treatment of the weak instruments problems in the nonparametric IV and the functional linear IV regressions is not currently available. Our results on the extreme nonidentification could potentially be a useful starting point for developing such a theory.

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

Let h_z and h_w be the bandwidth parameters smoothing respectively over Z and W . For the kernel-smoothed nonparametric IV regression, in the extreme case of nonidentification, the inner products are distributed according to the weighted sum of chi-squared random variables as illustrated below.

Theorem 4.2. Suppose that Assumption 4.1 is satisfied, $\mathbb{E}[\phi(Z)|W] = 0, \forall \phi \in L_{2,0}(Z)$, and that $n\alpha_n h_z^p \rightarrow \infty$ with h_w being fixed. Then for every $\mu \in L_2([0, 1]^p)$

$$\alpha_n n \langle \hat{\varphi} - \varphi_1, \mu \rangle \xrightarrow{d} \mathbb{E}[Y P_0 \mu(Z)] h_w^{-q} \bar{K}(0) + \sum_{j=1}^{\infty} \lambda_j (\chi_j^2 - 1),$$

where $(\chi_j^2)_{j \geq 1}$ are independent chi-squared random variables with 1 degree of freedom, $(\lambda_j)_{j \geq 1}$ are eigenvalues of T , and P_0 is the orthogonal projection on $L_{2,0}(Z)$.

Note that it is not possible to obtain the weak convergence of $\alpha_n n \langle \hat{\varphi} - \varphi_1, \mu \rangle$ in the Hilbert space for the nonparametric IV regression because this process is not tight.

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 the asymptotic distribution in nonidentified models and show that the degenerate U-statistics asymptotics discovered in Section 4 can emerge even when we move from the extreme nonidentified 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 inner products with μ_0 first and introduce several assumptions below.

Assumption 5.1. (i) the data $(Y_i, Z_i, W_i)_{i=1}^n$ are the i.i.d. sample of $X = (Y, Z, W)$ 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 $\mathbb{E}[|U||Z|||W|^2] < \infty$.

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. $\alpha_n \rightarrow 0$, $n\alpha_n^{1+\beta\wedge 1} \rightarrow 0$, and $n\alpha_n \rightarrow \infty$ as $n \rightarrow \infty$.

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

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

The following result characterizes the asymptotic distribution of the inner products provided that the instrumental variable W does not concentrate in $\mathcal{N}(K^*)^\perp$ or, in other words, W^0 is a nondegenerate stochastic process.

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

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

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 . If W^0 is degenerate, then for every $\mu_0 \in \mathcal{N}(K)$

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

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

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

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

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

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

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

Since for every $\delta > 0$

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

a sufficient condition for Assumption 5.3 is the Lyapunov's condition $E |\langle \eta_n, \mu_1 \rangle|^{2+\delta} = O(1)$. The Lyapunov's condition is satisfied under the moment conditions $\mathbb{E}|U|^{2+\delta} < \infty$ and $\mathbb{E}\|Z\|^{2+\delta}$, whenever $\mu_1 \in \mathcal{R}[(K^*K)^\gamma]$ and $W \in \mathcal{R}[(K^*K)^{\tilde{\gamma}}]$ with $\gamma + \tilde{\gamma} \geq 1/2$ since in this case, we have

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

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

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

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

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

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 the 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.¹⁰ The resulting large sample distribution may be Gaussian, the weighted sum of independent chi-squared random variables, or the mixture of the two.¹¹

¹⁰Note 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.

¹¹The critical values when the normalizing sequence is unknown can be obtained with resampling methods, see, e.g., Bertail, Politis, and Romano (1999).

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

$$f_{ZW}^{\text{ID}}(z, w) = \frac{f_{ZW}(z, w)}{\int_{[0,1]^2} f_{ZW}(z, w) dz dw} \mathbf{1}\{(z, w) \in [0, 1]^2\},$$

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

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

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

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

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

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

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 focus on the Tikhonov-regularized estimator, see [Babii \(2020b\)](#) for more details on the practical implementation and the data-driven choice of tuning parameters. [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 3](#) 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 that are 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	0.0337	0.3428	0.0249	0.2560
2	0.0225	0.2935	0.0078	0.2374
∞	0.0214	0.2923	0.0076	0.2376

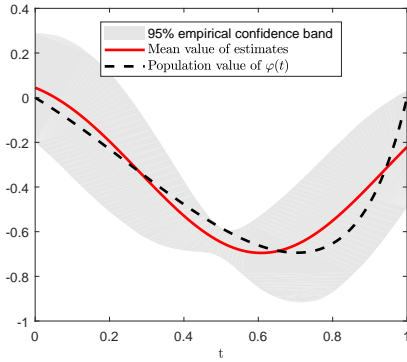
Table 1: L_2 and L_∞ errors.
5000 Monte Carlo experiments, $h_z = 0.15$, $h_w = 0.1$, $\alpha = 0.003$.

7 Conclusion

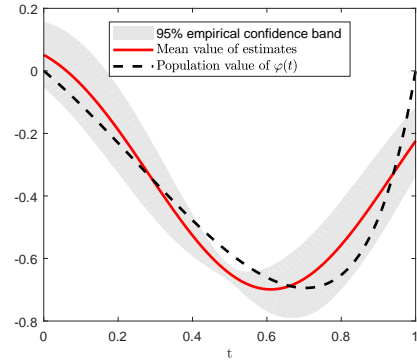
This paper investigates nonidentified linear ill-posed inverse problems using the non-parametric IV and the functional linear IV regressions as illustrating examples. Identification failures occur due to the non-injectivity of the covariance or the conditional expectation operators. We illustrate that if the operator is not injective, the Tikhonov-regularized estimator converges to the best approximation of the structural parameter in $\mathcal{N}(K)^\perp$ and derive novel uniform and Hilbert space norm bounds for the risk.

Consequently, even if the completeness condition fails, the consistent estimation of the structural parameter φ is possible whenever $\varphi \in \mathcal{N}(K)^\perp$. We show that even if this is not the case, we may still be able to learn useful information about the global shape of the structural parameter φ since in many cases it can be accurately approximated by a relatively small number of basis functions in $\mathcal{N}(K)^\perp$. This gives us an appealing projection interpretation for the nonparametric IV regression model similar to the one shared by the OLS estimator and not shared by the classical parametric linear IV regression.

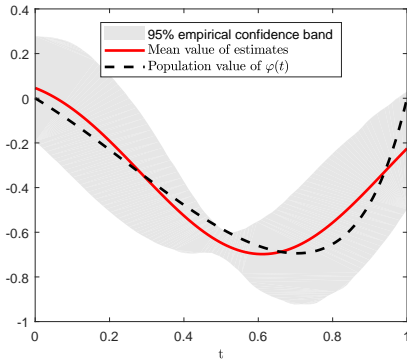
We show that the nonidentification has important implications for the large sample behavior of the Tikhonov-regularized estimator. We find that in the extreme non-identified cases, the distribution is driven by the degenerate U-statistics asymptotics, while in the intermediate cases the transition between the weighted chi-squared and the Gaussian limits is possible.



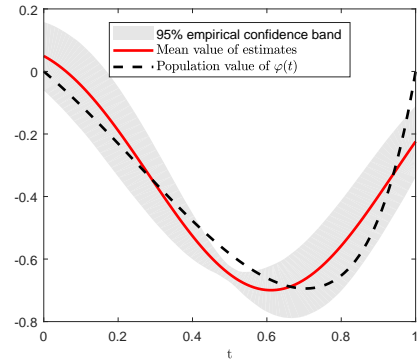
(a) $n = 1000, J_0 = \infty$



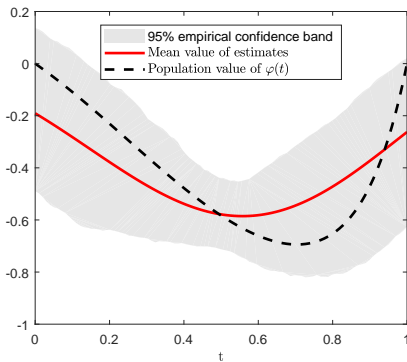
(b) $n = 5000, J_0 = \infty$



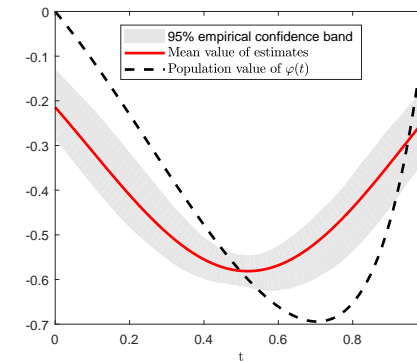
(c) $n = 1000, J_0 = 2$



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



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



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

Figure 3: Estimates averaged over 5000 experiments and empirical confidence bands.

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.
- ANDREOU, E., E. GHYSELS, AND A. KOURTELLOS (2013): “Should macroeconomic forecasters use daily financial data and how?,” *Journal of Business & Economic Statistics*, 31(2), 240–251.
- ANDREWS, D. W. (2017): “Examples of ℓ^2 -complete and boundedly-complete distributions,” *Journal of Econometrics*, 199(2), 213–220.
- ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2017): “Earnings and consumption dynamics: a nonlinear panel data framework,” *Econometrica*, 85(3), 693–734.
- BABII, A. (2020a): “High-dimensional mixed-frequency IV regression,” *UNC Working paper*.
- (2020b): “Honest confidence sets in nonparametric IV regression and other ill-posed models,” *Econometric Theory (forthcoming)*.
- 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.
- 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*.

- 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*.
- 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.
- 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*.
- ENGL, H. W., M. HANKE, AND A. NEUBAUER (1996): *Regularization of inverse problems*. Springer.
- ESCANCIANO, J. C., AND W. LI (2018): “Optimal linear instrumental variables approximations,” *arXiv preprint arXiv:1805.03275*.
- 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,” .
- FLORENS, J.-P., AND S. VAN BELLEGEM (2015): “Instrumental variable estimation in functional linear models,” *Journal of Econometrics*, 186(2), 465–476.

- GAGLIARDINI, P., AND O. SCAILLET (2012): “Tikhonov regularization for non-parametric instrumental variable estimators,” *Journal of Econometrics*, 167(1), 61–75.
- GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): “Predicting volatility: getting the most out of return data sampled at different frequencies,” *Journal of Econometrics*, 131(1-2), 59–95.
- GHYSELS, E., A. SINKO, AND R. VALKANOV (2007): “MIDAS regressions: Further results and new directions,” *Econometric Reviews*, 26(1), 53–90.
- 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,” *The 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 non-classical 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*, pp. 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.
- 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.
- NAIR, M. T. (2009): *Linear operator equations: approximation and regularization*. World Scientific Publishing Company.
- NEWHEY, 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*, pp. 577–591.
- 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.
- 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.
- VAN DER VAART, A. W., AND J. A. WELLNER (2000): *Weak convergence and empirical processes*. Springer.

APPENDIX

A.1 Proofs

Proof of Theorem 3.1. Decompose

$$\begin{aligned}
\hat{\varphi} - \varphi_1 &= (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* (\hat{r} - \hat{K} \varphi) \\
&\quad + (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \hat{K} \varphi_0 \\
&\quad + (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \hat{K} \varphi_1 - (\alpha_n I + K^* K)^{-1} K^* K \varphi_1 \\
&\quad + ((\alpha_n I + K^* K)^{-1} K^* K - I) \varphi_1 \\
&\triangleq I_n + II_n + III_n + IV_n.
\end{aligned}$$

IV_n is the regularization bias that can be controlled under Assumption 3.1

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

see Babii (2020a). The first term is controlled under Assumption 3.2 (i)

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

The second term is controlled under Assumption 3.2 (ii)

$$\begin{aligned}
\mathbb{E} \|II_n\|^2 &\leq \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \right\|^2 \left\| \hat{K} \varphi_0 \right\|^2 \\
&\leq \frac{C_2 \delta_{2n}}{4\alpha_n}.
\end{aligned}$$

The third term is decomposed further

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

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

$$\begin{aligned}
\mathbb{E} \|III_n^a\|^2 &= \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \left[\hat{K} - K \right] \alpha_n (\alpha_n I + K^* K)^{-1} \varphi_1 \right\|^2 \\
&\leq \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \right\|^2 \left\| \hat{K} - K \right\|^2 \left\| \alpha_n (\alpha_n I + K^* K)^{-1} \varphi_1 \right\|^2 \\
&\leq \sup_{\lambda \geq 0} \left| \frac{\lambda^{1/2}}{\alpha_n + \lambda} \right|^2 \mathbb{E} \left\| \hat{K} - K \right\|^2 C^{(2\beta-3)\vee 1} \alpha_n^{\beta \wedge 2} \\
&\leq \frac{C_3 \rho_{1n}}{4\alpha_n} C^{(2\beta-3)\vee 1} \alpha_n^{\beta \wedge 2}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} \|III_n^b\|^2 &= \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \left[\hat{K}^* - K^* \right] \alpha_n K (\alpha_n I + K^* K)^{-1} \varphi_1 \right\|^2 \\
&\leq \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \right\|^2 \left\| \hat{K}^* - K^* \right\|^2 C \left\| \alpha_n K (\alpha_n I + K^* K)^{-1} (K^* K)^{\beta/2} \right\|^2 \\
&\leq \sup_{\lambda \geq 0} \left| \frac{1}{\alpha_n + \lambda} \right|^2 \mathbb{E} \left\| \hat{K}^* - K^* \right\|^2 C \sup_{\lambda \in [0, C^2]} \left| \frac{\alpha_n \lambda^{(\beta+1)/2}}{\alpha_n + \lambda} \right|^2 \\
&\leq \frac{C_3 \rho_{1n}}{\alpha_n^2} C^{(2\beta-1)\vee 1} \alpha_n^{(\beta+1) \wedge 2},
\end{aligned}$$

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

Proof of the Theorem 3.2. Consider the same decomposition as in the proof of Theorem 3.1. Note that $\varphi_1 = \left[(K^* K)^{\frac{\beta-1}{2}} K^* \right] \psi$ with $\beta > 1$. Then the bias term is

treated similarly to the identified case, see Babii (2020b), Proposition 3.1

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

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

$$\begin{aligned}
\mathbb{E}\|I_n\|_\infty &= \mathbb{E} \left\| \hat{K}^* (\alpha_n I + \hat{K} \hat{K}^*)^{-1} (\hat{r} - \hat{K} \varphi) \right\|_\infty \\
&\leq \mathbb{E} \left\| \hat{K}^* \right\|_{2,\infty} \left\| (\alpha_n I + \hat{K} \hat{K}^*)^{-1} \right\| \left\| \hat{r} - \hat{K} \varphi \right\| \\
&\leq \frac{1}{\alpha_n} \left(\|K^*\|_{2,\infty} \mathbb{E} \left\| \hat{r} - \hat{K} \varphi \right\| + \mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty} \left\| \hat{r} - \hat{K} \varphi \right\| \right) \\
&\leq \frac{1}{\alpha_n} \left(C_4 + \left(\mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty}^2 \right)^{1/2} \right) \left(\mathbb{E} \left\| \hat{r} - \hat{K} \varphi \right\|^2 \right)^{1/2} \\
&\leq C_1^{1/2} \left(C_4 + C_4^{1/2} \rho_{2n}^{1/2} \right) \frac{\delta_{1n}^{1/2}}{\alpha_n}.
\end{aligned}$$

The second term is controlled as

$$\begin{aligned}
\mathbb{E}\|II_n\|_\infty &= \left\| \hat{K}^* (\alpha_n I + \hat{K} \hat{K}^*)^{-1} K \varphi_0 \right\|_\infty \\
&\leq \left\| \hat{K}^* \right\|_{2,\infty} \left\| (\alpha_n I + \hat{K} \hat{K}^*)^{-1} \right\| \left\| \hat{K} \varphi_0 \right\| \\
&\leq \frac{1}{\alpha_n} \left(C_4 + \left(\mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty}^2 \right)^{1/2} \right) \left(\mathbb{E} \left\| \hat{K} \varphi_0 \right\|^2 \right)^{1/2} \\
&\leq C_2^{1/2} \left(C_4 + C_4^{1/2} \rho_{2n}^{1/2} \right) \frac{\delta_{2n}^{1/2}}{\alpha_n}.
\end{aligned}$$

The third term is decomposed further similarly as in the proof of Theorem 3.1 in

III_n^a and III_n^b . We bound each of the two terms separately. First,

$$\begin{aligned}
\mathbb{E}\|III_n^a\|_\infty &= \left\| \hat{K}^*(\alpha_n I + \hat{K} \hat{K}^*)^{-1} [\hat{K} - K] \alpha_n (\alpha_n I + K^* K)^{-1} \varphi_1 \right\|_\infty \\
&\leq \mathbb{E} \left\| \hat{K}^* \right\|_{2,\infty} \left\| (\alpha_n I + \hat{K} \hat{K}^*)^{-1} \right\| \left\| \hat{K} - K \right\| \left\| \alpha_n (\alpha_n I + K^* K)^{-1} \varphi_1 \right\| \\
&\leq \frac{1}{\alpha_n} \left(C_4 + \left(\mathbb{E} \left\| \hat{K}^* - K^* \right\|_{2,\infty}^2 \right)^{1/2} \right) \left(\mathbb{E} \left\| \hat{K} - K \right\|^2 \right)^{1/2} C^{(\beta-1.5)\vee 0.5} \alpha_n^{\frac{\beta}{2} \wedge 1} \\
&\leq \left(C_4 + C_4^{1/2} \rho_{2n}^{1/2} \right) \frac{C_3^{1/2} \rho_{1n}^{1/2}}{\alpha_n} C^{(\beta-1.5)\vee 0.5} \alpha_n^{\frac{\beta}{2} \wedge 1}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\|III_n^b\|_\infty &= \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} [\hat{K}^* - K^*] \alpha_n K (\alpha_n I + K^* K)^{-1} \varphi_1 \right\|_\infty \\
&= \mathbb{E} \left\| (\alpha_n I + \hat{K}^* \hat{K})^{-1} \right\|_\infty \left\| \hat{K}^* - K^* \right\|_{2,\infty} \left\| \alpha_n K (\alpha_n I + K^* K)^{-1} \varphi_1 \right\| \\
&= \frac{1}{2\alpha_n^{3/2}} \mathbb{E} \left(\left\| \hat{K}^* \right\|_{2,\infty} + 2\alpha_n^{1/2} \right) \left\| \hat{K}^* - K^* \right\|_{2,\infty} C^{(\beta-1/2)\vee 1/2} \alpha_n^{\frac{\beta+1}{2} \wedge 1} \\
&\leq \frac{1}{2\alpha_n^{3/2}} \left(C_4 + C_4^{1/2} \rho_{2n}^{1/2} + 2\alpha_n^{1/2} \right) C_4^{1/2} \rho_{2n}^{1/2} C^{(\beta-1/2)\vee 1/2} \alpha_n^{\frac{\beta+1}{2} \wedge 1}.
\end{aligned}$$

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_C^s denote the the Hölder class.

Proposition A.1.1. *Suppose that (i) $(Y_i, Z_i, W_i)_{i=1}^n$ are i.i.d. and $\mathbb{E}|Y_1|^2 < \infty$; (ii) $f_{ZW} \in C_C^s$; (iii) kernel functions $K_z : \mathbf{R}^p \rightarrow \mathbf{R}$ and $K_w : \mathbf{R}^q \rightarrow \mathbf{R}$ are such that for $l \in \{w, z\}$, $K_l \in L_1 \cap L_2$, $\int K_l(u) du = 1$, $\int \|u\|^s K_l(u) du < \infty$, and $\int u^k K_l(u) du = 0$ for all multindices $|k| = 1, \dots, \lfloor s \rfloor$. Then for all ϕ with $\|\phi\| \leq C$*

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

and

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

Proof. Decompose

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

with

$$\begin{aligned} V_n &= \int \phi(z) \left(\hat{f}_{ZW}(z, w) - \mathbb{E} \hat{f}_{ZW}(z, w) \right) dz \\ B_n &= \int \phi(z) \left(\mathbb{E} \hat{f}_{ZW}(z, w) - f_{ZW}(z, w) \right) dz. \end{aligned}$$

By the Cauchy-Schwartz inequality

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

where the right side is of order $O(h_n^s)$ under the assumption $f_{ZW} \in C_C^s$, 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

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

where $[\phi * K_z](Z_i) = \int \phi(z) h_n^{-p} K_z(h_n^{-1}(Z_i - z)) dz$. Then

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

where the second line follows by change of variables, and the last by $\|f_Z\|_\infty \leq C$, and Young's inequality. This proves the second claim. To establish, the first claim, note that

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

Therefore, we need to show additionally that $\mathbb{E} \|\hat{r} - r\|^2 = O\left(\frac{1}{nh_n^q} + h_n^{2s}\right)$. To this end decompose

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

Under the i.i.d. assumption, the variance is

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

By Cauchy-Schwartz inequality

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

where $[f_{ZW} * K_w, h](w) = \int f_{ZW}(z, W') h^{-q} K_w(h_n^{-1}(w - W')) dW'$. Since $f_{ZW} \in C_C^s$, we obtain

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

see, e.g., [Giné and Nickl \(2016\)](#), Proposition 4.3.8. The third claim follows from the fact that the operator norm can be bounded by the L_2 norm of the joint density of (Z, W) , and the standard computations for the risk of the joint density, [Giné and Nickl \(2016\)](#), Chapter 5. \square

Proof of Theorem 4.1. Since $\mathbb{E}[\langle Z, \delta \rangle W] = 0$ for all $\delta \in \mathcal{E}$, we have $\varphi_1 = 0$. Then

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

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 $\alpha_n n \rightarrow \infty$, by the continuous mapping theorem, see [van der Vaart and Wellner \(2000\)](#), Theorem 1.3.6

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

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

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

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 $J(h) = \iint_{\mathcal{X} \times \mathcal{X}} h(x_1, x_2) \mathbb{W}(dx_1) \mathbb{W}(dx_2)$ 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 $X = (Y, Z, W)$. \square

Proof of Theorem 4.2. 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 4.1 (i) since $\mathbb{E}[\phi(Z)|W] = 0$ for

all $\phi \in L_{2,0}(Z)$

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

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

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

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

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

where $[K_z * P_0 \mu](z) = h_n^{-p} \int K(h_z^{-1}(z - u)) P_0 \mu(u) du$. Under Assumption 4.1, by the strong law of large numbers

$$\zeta_n \xrightarrow{a.s.} \mathbb{E}[Y P_0 \mu(Z)] h_w^{-q} \bar{K}(0).$$

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

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

Note that

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

where the first two lines follow under Assumption 4.1 (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 $\mathbb{E}[|Y|^2|W] < \infty$ a.s. and $\bar{K} \in L_\infty$, by the moment inequality in [Korolyuk and Borovskich \(1994\)](#), Theorem 2.1.3

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

□

Proof of Theorem 5.1. Put $b_n = \alpha_n(\alpha_n I + K^* K)^{-1} \varphi_1$ and note that $(\alpha_n I + \hat{K}^* \hat{K})^{-1} \hat{K}^* = \hat{K}^* (\alpha_n I + \hat{K} \hat{K}^*)^{-1}$. Then, similarly to the proof of Theorem 3.1, decompose

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

Since

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

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

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

Then

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

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

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

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

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

and

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

Lastly, the bias is zero by Eq. A.1 and the orthogonality between φ_1 and μ_0

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

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 $n\alpha_n^{1+\beta\wedge 1} \rightarrow 0$ and $n\alpha_n \rightarrow \infty$,

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

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

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

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

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

Under Assumption 5.1 by the strong law of large numbers

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

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 $\mathbb{E}[W^0 \langle Z, \mu_0 \rangle] = P_0 K \mu_0 = 0$ and $\mathbb{E}[W^0 (Y - \langle Z, \varphi_1 \rangle)] = P_0 \mathbb{E}[WU] + P_0 K \varphi_0 = 0$. 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

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

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

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

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

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

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

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

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

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

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 $\mu_1 \in \mathcal{R}[(K^* K)^\gamma]$, 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

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

with

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

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

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

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 $K : \mathcal{E} \rightarrow \mathcal{H}$ 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 $\Phi^{\text{ID}} = \varphi + \mathcal{N}(K)$, 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 the 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 $K^\dagger : \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp \rightarrow \mathcal{E}$ defined by $K^\dagger r = \varphi_1$, where $\varphi_1 \in \Phi^{\text{ID}}$ is a unique solution to*

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

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 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), p.117-118 for more details.

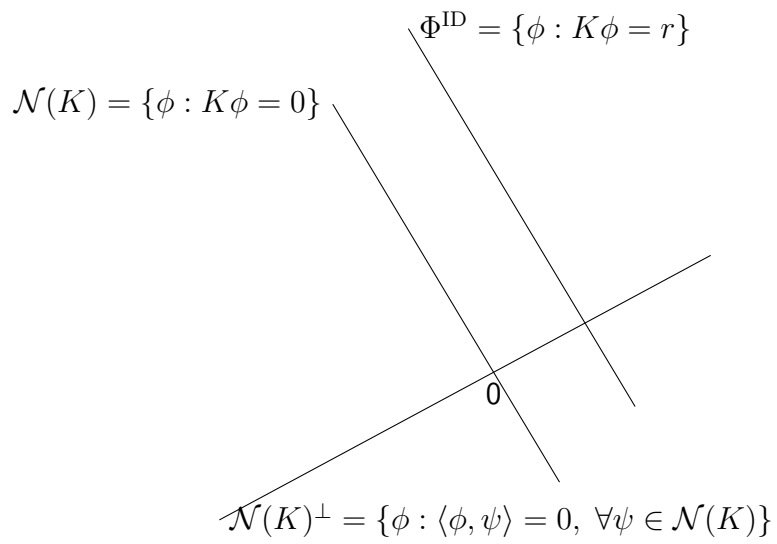


Figure B.1: Fundamental subspaces of \mathcal{E} .

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*

$$K : L_2 \rightarrow L_2$$

$$\phi \mapsto (K\phi)(w) = \int \phi(z)k(z, w)dz$$

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 .¹² In other words, we are faced with an ill-posed inverse problem. Tikhonov regularization can be understood as a method that smoothes out the discontinuities of the generalized inverse $(K^*K)^\dagger$.¹³

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 $f : \mathcal{X}^m \rightarrow H$ 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

¹²In practice the situation is even more complex, because the operator K is also estimated from the data.

¹³By 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 $J_m : S_m \rightarrow H$

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

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 $J : L_2([0, \infty), dt) \rightarrow \mathbf{R}$ defined as $J(f) = \int f(t) dB_t$.

B.2.2 Central limit theorem

Let (\mathcal{X}, Σ, 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}, Σ, P) . Consider some symmetric function $h : \mathcal{X} \times \mathcal{X} \rightarrow H$, 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 $J(h) = \iint_{\mathcal{X} \times \mathcal{X}} h(x_1, x_2) \mathbb{W}(dx_1) \mathbb{W}(dx_2)$ is a stochastic Wiener-Itô integral and \mathbb{W} is a Gaussian random measure on H .