Is Completeness Necessary? Penalized Estimation in Non Identified Linear Models

Andrii Babii*   Jean-Pierre Florens†

Toulouse School of Economics   *Toulouse School of Economics*

June 9, 2017

Identification is an important issue in many econometric models. This paper studies potentially non-identified and/or weakly identified ill-posed inverse models. The leading examples are the nonparametric IV regression and the functional linear IV regression. We show that in the case of identification failures, a very general family of continuously-regularized estimators is consistent for the best approximation of the parameter of interest. We obtain L_2 and L_∞ convergence rates for this general class of regularization schemes, including Tikhonov, iterated Tikhonov, spectral cut-off, and Landweber-Fridman. Unlike in the identified case, estimation of the operator has non-negligible impact on convergence rates and inference. We develop inferential methods for linear functionals in such models. Lastly, we demonstrate the discontinuity in the asymptotic distribution in case of weak identification. In particular, the estimator has a degenerate U-statistics type behavior, in the extreme case of weak instrument.

*Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). Email: babii.andrii@gmail.com.
†Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France).

1 Introduction

Structural nonparametric or high-dimensional models may be formalized by an inverse problem which is very often linear. Among many examples, we may quote the non-parametric instrumental regression, functional linear regression models, or deconvolution. All these examples reduce to a functional linear equation

$$K\varphi = r,$$

where φ is the functional parameter of interest, r is an element of a functional space, and K is a linear operator. Usual inverse problem literature assumes that K is given and r is estimated with some error. Econometric examples lead to problems, where both K and r are estimated.

Identification is an important question in econometrics. In linear inverse problems r and K are identified by the DGP and φ is identified if the equation $K\varphi = r$ has a unique solution. We assume that the solution exists or equivalently that r is in the range of K. Unicity of the solution is equivalent to $K\varphi = 0$ implies $\varphi = 0$ or K is a one-to-one operator. Note that in most of the cases when K is unknown, the estimated operator $\hat{K}$ has a finite-rank and is not one-to-one for any finite sample size.

Maximum likelihood method when there is lack of identification leads to a flat likelihood in some regions and then some ambiguity on the choice of a maximum. It is then natural to characterize the limit of the estimator in the case of potentially non-identified model.

In non-identified models there exists a set of solutions $\varphi + \mathcal{N}(K)$, where φ is a particular solution and $\mathcal{N}(K)$ is the null space of K. In the case of usual Tikhonov estimation with K known, it is well-known that the estimator converges to the element of this set of minima with the smallest norm. This limit is also equal to the projection of the φ on the orthogonal of the null-space of K, see (Engl, Hanke, and Neubauer, 1996) or (Florens, Johannes, and Van Bellegem, 2011).

This paper recalls this result and gives some illustration of this property. Our approach can be considered as an alternative to the partial identification approach, where we focus on the whole identified set, Santos (2012).

More originally, we consider the case when K is estimated. This gives further illustration that the identification property is not crucial.

As K (or $\hat{K}$) may be not one-to-one and do not have a continuous inverse, a regularization method is needed to solve the equation $K\varphi = r$. Several methods are commonly used and we will focus our presentation on the Tikhonov method, when the solution takes form $(\alpha_n I + K^*K)^{-1}K^*r$, where K^* is the adjoint operator of K and K, K^*, and r are replaced by their estimates. However, all the results of the paper may be applied to smooth regularization, when the solution takes the form $g_\alpha(K^*K)K^*r$, see (Engl et al., 1996). Particular cases include the iterated Tikhonov, the Landweber iteration, etc. Tikhonov regularizatoin is a solution of the penalized mean-square problem

$$\min \|r - K\varphi\|^2 + \alpha\|\varphi\|^2$$

and may be extended by using a penalty $\|L\varphi\|$, when L is a differential operator. This extension is called the regularization in Hilbert scale.

All these methods lead to a well-defined estimator even if the model is not identified. This is different from the parametric context. Even if the convergence result remains the same if K is estimated, the convergence rate may be affected by the identification in the estimated case. The last new element of this paper is to consider the behavior of the regularized estimator in the L_∞ norm in the non-identified case.

2 Consistency and rates of convergence

Let us consider a linear equation

$$r = K\varphi, \tag{1}$$

where $\varphi \in \mathcal{E}$ and $r \in \mathcal{F}$ are two Hilbert spaces, and K is a linear operator from $\mathcal{E}$ to $\mathcal{F}$. We assume that the equation is well-specified in the sense that there exists a solution to (1), or equivalently that r is in the range of the operator K. This solution is unique if K is one-to-one, or if the null space of K

$$\mathcal{N}(K) = \{\varphi \in \mathcal{E} : \; K\varphi = 0\}$$

reduces to $\{0\}$. We observe a noisy version of r, say $\hat{r}_n$, depending on the sample of size n, such that $\mathbb{E}\,\|\hat{r}_n - r\|^2 = O(\delta_n)$, where $\delta_n \to 0$ if $n \to \infty$. We focus our attention to the case where K is compact. The problem is then ill-posed because even if K is one-to-one, it does not have a continuous inverse on $\mathcal{F}$ when $\mathcal{E}$ is infinite-dimensional. Then we consider a regularized solution of (1), for example, the Tikhonov solution:

$$\hat{\varphi}_n^{\alpha} = (\alpha I + K^*K)^{-1}K^*\hat{r}_n, \tag{2}$$

obtained by the minimization of the Tikhonov functional

$$\|\hat{r}_n - K\varphi\|^2 + \alpha\|\varphi\|^2. \tag{3}$$

We want to consider cases where K is not necessarily one-to-one, or in econometric terminology when the model is not identified. Let us illustrate this point by two examples.

Example 1. *Functional linear instrumental regression, see (Florens and Van Bellegem, 2015). We consider an equation $Y = \langle Z, \varphi\rangle + U$, where $Y \in \mathbf{R}, Z \in \mathcal{E}, \varphi \in \mathcal{E}$ and U is a random noise verifying $\mathbb{E}[UW] = 0$. The instrumental variable W belongs to another Hilbert space $\mathcal{F}$. For simplicity we assume that Z and W have mean zero. This model leads to the linear equation*

$$\mathbb{E}[YW] = \mathbb{E}[W\langle Z, \varphi\rangle]. \tag{4}$$

In this example $K\varphi = \mathbb{E}[W\langle Z, \varphi\rangle]$ is the second-order moment between W and Z defining an operator from $\mathcal{E}$ to $\mathcal{F}$ and the identification condition is

$$\mathbb{E}[W\langle Z, \varphi\rangle] = 0 \implies \varphi = 0. \tag{5}$$

This condition is essentially the injectivity of the cross-covariance operator of Z and W, generalizing rank condition in the linear IV model, and may be interpreted as a requirement for the sufficient linear dependence between Z and W.

Example 2. *Non-parametric instrumental variables. Let $Y \in \mathbf{R}, Z \in \mathbf{R}^p$, and $W \in \mathbf{R}^q$ be three random elements and we assume that $Y = \varphi(Z) + U$ with $\mathbb{E}[U|W] = 0$. This assumption implies the linear equation*

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

and K is the conditional expectation operator from L^2_Z to L^2_W (defined with respect to the true distribution of (Y, Z, W)). Completeness (or more precisely L^2 completeness, see (Florens, Mouchart, Rolin, et al., 1990)) is defined by $\mathbb{E}[\varphi(Z)|W] = 0 \implies \varphi = 0$ and is a (non-linear) dependence condition between Z and W.

We claim two fundamental properties of the Tikhonov regularized estimation:

1. The estimator is well-defined even if K is not injective.
2. If α is suitably chosen, $\hat{\varphi}_n^\alpha$ converges to the best approximation of the true φ by identified element.

Let us precise these two points. First, let us consider the family of singular values of K, denoted (λ_j). By compactness of K this family is discrete, $j = 1, 2, \dots,$ $\lambda_j \in [0, \|K\|]$, and $\lambda_j \to 0$ if $j \to \infty$. The singular values of $\alpha I + K^*K$ are $\alpha + \lambda_j^2$ and they don't vanish for $\alpha \neq 0$, even if $\lambda_j = 0$ for some j. Moreover

$$\hat{\varphi}_n^\alpha = \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2} \langle \hat{r}_n, \psi_j \rangle \varphi_j,$$

where φ_j and ψ_j are singular vectors of K^*K and KK^*, and $\hat{\varphi}_n^\alpha$ is always well-defined, because

$$\|\hat{\varphi}_n^\alpha\|^2 \leq \sum_{j=1}^{\infty} \langle \hat{r}_n, \psi_j \rangle^2 < \infty.$$

Second, let us consider the limit of $\hat{\varphi}_n^\alpha$. Recall that the null space of a bounded operator is a closed linear subspace. This allows us to decompose the parameter of interest uniquely as

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

where φ_0 is the orthogonal projection of φ on $\mathcal{N}(K)$ and φ_1 is the orthogonal projection on $\mathcal{N}(K)^\perp$, the orthogonal complement to the null space of K, equal to the closure of the range of

K^*, denoted $\overline{\mathcal{R}(K^*)}$, (Luenberger, 1997, p.157).

If K is not one-to-one, we are faced to the problem of a set identified model. The identified set has the form of linear manifold $\varphi_1 + \mathcal{N}(K)$, where $\varphi_1 \in \overline{\mathcal{R}(K^*)}$. Equivalently, the identified parametric space is $\mathcal{E}/\mathcal{N}(K)$, the quotient space of $\mathcal{E}$ by the linear subspace $\mathcal{N}(K)$. A set estimation is then given by the linear manifold $\hat{\varphi}_n^\alpha + \mathcal{N}(K)$, which is an estimator of $\mathcal{E}/\mathcal{N}(K)$, converging to the identified parameter in $\mathcal{E}/\mathcal{N}(K)$.

While it may be interesting from the theoretical point of view to develop inferential methods for $\varphi_1 + \mathcal{N}(K)$, this identified set is usually not very tractable. For example, if the null space of K is the set of all polynomials of degree m, the identified set is simply too big to be useful in empirical applications. In this case, what we can only hope to learn is the best approximation φ_1.

The following two regularity conditions are needed for the first result of this paper.

Assumption 1. *Suppose that* $\varphi_1 \in \mathcal{R}(K^*K)^{\beta/2}$ *for some* $\beta > 0$.

Assumption 2. *Suppose that* K *and* K^* *are estimated by* $\hat{K}$ *and* $\hat{K}^*$ *so that for all* $\phi \in L_2$

$$\mathbb{E}\left\|(\hat{K} - K)\phi\right\| = O(\rho_{1,n}), \qquad \mathbb{E}\left\|\hat{r} - \hat{K}\varphi\right\|^2 = O\left(\delta_n\right),$$

and that

$$\mathbb{E}\left\|\hat{K} - K\right\|^2 = O(\rho_{2,n}), \qquad \mathbb{E}\left\|\hat{K}^* - K^*\right\|^2 = O(\rho_{2,n}).$$

We characterize convergence rates for the mean-integrated squared error in the following result. Unlike in the identified case treated in (Darolles, Fan, Florens, and Renault, 2011), the noise coming from the estimation of the operator is now important and the convergence rate is also driven by the rate at which we estimate the operator K.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied, then*

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

Assumption 3. *Suppose that for some bounded sequence* ξ_n*, we have*

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

where $\|K^*\|_{2,\infty} = \sup_{\|\psi\|\leq 1} \|K^*\psi\|_\infty < \infty$.

Theorem 2. *Suppose that Assumption 1 is satisfied with $\beta > 1$. Suppose also that Assumptions 2 and 3 hold and that ξ_n is bounded for all $n \in \mathbf{N}$. Then*

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

2.1 Functional linear IV regression

Assume for simplicity that the data are i.i.d.[1] In this example $r = \mathbb{E}[YW]$ is estimated by $\frac{1}{n}\sum_{i=1}^n Y_i W_i$ and $\delta_n = \frac{1}{n}$, whenever $\mathbb{E}\|UW\|^2 < \infty$. Moreover, K and K^* are respectively estimated by $\hat{K} = \frac{1}{n}\sum_{i=1}^n W_i\langle Z_i, .\rangle$, and $\hat{K}^* = \frac{1}{n}\sum_{i=1}^n Z_i\langle W_i, .\rangle$, and $\rho_{1,n} = \rho_{2,n} = \frac{1}{n}$, whenever $\mathbb{E}\|ZW\|^2 < \infty$. The risk becomes

$$\mathbb{E}\left\|\hat{\varphi}_{\alpha_n} - \varphi_1\right\|^2 = O\left(\frac{1}{\alpha_n n} + \alpha_n^{\beta\wedge 2}\right).$$

Then conditions $\alpha \to 0$ and $\alpha n \to \infty$ are sufficient to ensure the convergence of φ to φ_1 in the mean-square risk and so in probability.

For uniform convergence we need additionally to assume that trajectories of Z and W are sufficiently smooth, e.g. Hölder continuous, to ensure that $\delta_n = \rho_{1,n} = \rho_{2,n} = \frac{1}{n}$. This smoothness assumption combined with Hoffman-Jørgensen's inequality[2], see (Giné and Nickl, 2015, Theorem 3.1.16), gives

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

Therefore, assuming that trajectories of Z_i and W_i are bounded, we obtain $\xi_n = \frac{1}{n}$ and

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

Then conditions $\alpha_n \to 0$ and $\alpha_n n^{1/2} \to \infty$ are sufficient to ensure uniform convergence of φ to φ_1.

[1]This can be easily relaxed to weakly dependent data, e.g. covariance stationarity in the L_2 sense.
[2]Notice that continuity of trajectories ensures that the supremum is actually countable.

2.2 Non-parametric IV

We rewrite the model as

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

where now K is an operator from $L_2(\mathbf{R}^p, \mathrm{d}z)$ to $L_2(\mathbf{R}^q, \mathrm{d}w)$. In this example r and K are estimated by

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

If $h_z = h_w = h_n$, by Proposition 1, $\delta_n = \rho_{1,n} = \frac{1}{nh_n^q} + h_n^{2s}$ and $\rho_{2,n} = \frac{1}{nh_n^{p+q}} + h_n^{2s}$, where s is the regularity of the joint density of Z and W. So we have in that case

$$\mathbb{E}\left\|\hat{\varphi}_{\alpha_n} - \varphi_1\right\|^2 = O\left(\frac{1}{\alpha_n}\left(\frac{1}{nh_n^q} + h_n^{2s}\right) + \frac{1}{nh_n^{p+q}}\alpha_n^{\beta\wedge 1 - 1} + \alpha_n^{\beta\wedge 2}\right).$$

We also know that $\xi_n^{1/2} = \sqrt{\frac{\log h_n^{-1}}{nh_n^{p+q}}} + h_n^s$, see (Babii, 2016b, Proposition 5) under the assumption that f_{ZW} is in the Hölder class $B^s_{\infty,\infty}$, and so

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

3 Inference for functional linear regression

3.1 Inference for linear functionals

In many economic applications, the object of interest is not necessary a function φ, but rather its linear functional. By Riesz representation theorem any continuous linear functional can be represented as an inner product with some function $\mu \in L_2$. In this section we show that in case of identification failures we will still have convergence of suitably normalized plug-in estimator of linear functionals. Decompose $\mu = \mu_0 + \mu_1$ for $\mu_0 \in \mathcal{N}(K)$ and $\mu_1 \in \mathcal{N}(K)^\perp$. Put $\eta_{n,i} = (\alpha_n I + K^*K)^{-1}K^*W_i(U_i + \langle Z_i, \varphi_0\rangle)$.

Under mild assumptions, suitably normalized inner products with any $\mu_0 \in \mathcal{N}(K)$ have U-statistics type behavior as illustrated below.

Assumption 4. *(i) the data $X_i = (Y_i, Z_i, W_i) \in L_2(\mathcal{X}, \mathscr{X}, P), i = 1, \dots, n$ are i.i.d.; (ii) $\mathbb{E}\left[U_1^2\|W_1\|^2\right] < \infty$ and $\mathbb{E}\left[\|Z_1\|^2\|W_1\|^2\right] < \infty$; (iii) $\mathbb{E}\|W_1\|\|Z_1\||U| < \infty$, $\mathbb{E}\|W_1\|\|Z_1\|^2 < \infty$, and $\mathbb{E}\|W_1\|^2\|Z_1\|^2(U_1 + \langle Z_1, \varphi_0\rangle)^2 < \infty$.*

Let $(\lambda_j, \varphi_j, \psi_j)_{j\geq 1}$ denote the SVD decomposition of the covariance operator K, i.e. $K\varphi_j = \lambda_j\psi_j, K^*\psi_j = \lambda_j\varphi_j, j \geq 1$. To state the first result of this section, notice that there exists a unique orthogonal decomposition $W_i = W_i^0 + W_i^1$, where W_i^0 is the orthogonal projection of W_i on the null set of K^*, and W_i^1 is the projection on $\mathcal{N}(K^*)^\perp$.

Theorem 3. *Suppose that Assumptions 1 and 4 (i), (iii) are satisfied and that the sequence of regularization parameters $\alpha_n \to 0$ is such that $n\alpha_n^{1+\beta\wedge 2} \to 0$ while $n\alpha_n \to \infty$. Suppose that the instrumental variable W_i is such that W_i^0 is a non-degenerate random variable such that Assumption 4, and $W_i^1 \in \mathcal{R}\left[(KK^*)^\kappa\right], \kappa > 0$. Then for any $\mu_0 \in \mathcal{N}(K)$, we have*

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle \xrightarrow{d} \mathbb{E}\left[\|W_1\|(U_1 + \langle Z_1, \varphi_0\rangle)\langle Z_1, \mu_0\rangle\right] + \sum_{j=1}^{\infty}\lambda_j^{\mu_0}(\chi_{1,j}^2 - 1),$$

where $(\lambda_j^{\mu_0})_{j\geq 1}$ are eigenvalues of the conditional expectation operator

$$\begin{aligned} T^{\mu_0} :\ & L_2(X_1) \to L_2(X_1) \\ & f \mapsto \mathbb{E}\left[h_{\mu_0}(X_1, X_2)f(X_2)|X_1 = x_1\right], \end{aligned}$$

and for $X_i = (W_i, Z_i, U_i)$,

$$h_{\mu_0}(X_1, X_2) = \frac{\langle W_1^0, W_2^0\rangle}{2}\left\{\langle Z_1, \mu_0\rangle(U_2 + \langle Z_2, \varphi_0\rangle) + \langle Z_2, \mu_0\rangle(U_1 + \langle Z_1, \mu_0\rangle)\right\}.$$

Otherwise, if W_i^0 is degenerate, $n\alpha_n\langle\hat{\varphi}-\varphi_1,\mu_0\rangle \xrightarrow{d} 0$.

For inner products with μ_1, we need some additional assumptions.

Assumption 5. *For all $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[|\langle\eta_{n,i},\mu_1\rangle|^2 \mathbb{1}_{\left\{|\langle\eta_{n,i},\mu_1\rangle|\geq\epsilon n^{1/2}\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|\right\}}\right]}{\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|^2} = 0,$$

Since for any $\delta > 0$

$$\mathbb{E}\left[|\langle\eta_{n,i},\mu_1\rangle|^2 \mathbb{1}_{\left\{|\langle\eta_{n,i},\mu_1\rangle|\geq\epsilon n^{1/2}\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|\right\}}\right] \leq \frac{\mathbb{E}\left|\langle\eta_{n,i},\mu_1\rangle\right|^{2+\delta}}{\epsilon^\delta n^{\delta/2}\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|^\delta},$$

a sufficient condition for Assumption 5 is a Lyapunov-type restriction

$$\frac{\mathbb{E}\left|\langle\eta_{n,i},\mu_1\rangle\right|^{2+\delta}}{\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|^{2+\delta}} = O(1). \tag{6}$$

Notice that this condition is satisfied when $W_i \in \mathcal{R}\left[(K^*K)^{\tilde{\gamma}}\right]$, $\mathbb{E}\left|U_i\|W_i\|\right|^{2+\delta} < \infty$, $\mathbb{E}\left|\|Z_i\|\|W_i\|\right|^{2+\delta} < \infty$, and the following assumption is satisfied with $\gamma \geq 1/2-\tilde{\gamma}$.

Assumption 6. *For any $\mu_1 \in \mathcal{N}(K)^\perp$, let $\gamma \geq 0$ be such that $\mu_1 \in \mathcal{R}\left[(K^*K)^\gamma\right]$.*

To see that above assumptions are sufficient for Lyapynov's condition in Eq. (6), notice that

$$|\langle\eta_{n,i},\mu_1\rangle| \leq |U_i+\langle Z_i,\varphi_0\rangle| \left\|(K^*K)^\gamma(\alpha_n I+K^*K)^{-1}K^*(K^*K)^{\tilde{\gamma}}\right\| \|W_i\|\|\mu_1\|.$$

Assumption 7. *Suppose that $\beta,\gamma > 0$ and the sequence of tuning parameters $\alpha_n \to 0$ are such that (i) $\pi_n\alpha_n^{\frac{\beta}{2}\wedge 1} \to 0$ for $\pi_n = n^{1/2}\left\|\Sigma^{1/2}K(\alpha_n I+K^*K)^{-1}\mu_1\right\|^{-1}$ and (ii) $\frac{\pi_n\alpha_n^{\gamma\wedge 1/2}}{n\alpha_n} \to 0$.*

Notice that this assumption is the most restrictive when $\pi_n = O\left(n^{1/2}\right)$. In this case we need $n\alpha_n^{\beta\wedge 2} \to 0$ and $n\alpha_n^{2-2\gamma\wedge 1} \to \infty$, or $\beta\wedge 2 > 2-2\gamma\wedge 1$. For smooth functions μ_1 with $\gamma \geq 1/2$, this requirement holds when $\beta > 1$, while for less smooth functions μ_1 with $\gamma < 1/2$, we will need $\beta > 2-2\gamma$, i.e. more smoothness of φ. Therefore, having $\beta > 2$ will always ensure existence of the sequence of tuning parameters $\alpha_n \to 0$ satisfying Assumption 7.

Theorem 4. *Suppose that Assumptions 1, 4, 5, 6, and 7 are satisfied. Then for any $\mu_1 \in \mathcal{N}(K)^\perp$*

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

3.2 Asymptotic distribution in the case of irrelevant instrument

In this section we illustrate that there is a discontinuity in the asymptotic distribution when instrumental variable becomes weak. We look at the extreme case of the irrelevant instrumental variable. Let $\mathcal{S}_2$ be the space of Hilbert-Schmidt operators.

Theorem 5. *Suppose that Assumption 4 is satisfied,* $\mathbb{E}[\langle Z, \delta\rangle W] = 0,\ \forall \delta \in L_2$ *and* $\alpha_n n \to \infty$. *Then*

$$\alpha_n n(\hat{\varphi} - \varphi_1) \xrightarrow{d} \mathbb{G} g$$

in the product topology of $\mathcal{S}_2 \times L_2([0,1], \mathrm{d}s)$*, where* $\mathbb{G}$ *is a zero-mean Gaussian random element in* $\mathcal{S}_2$ *with covariance operator* $A \mapsto \mathbb{E}\left[\mathrm{trace}(A^*\langle W, .\rangle Z)\langle W, .\rangle Z\right]$ *and* g *is a zero-mean Gaussian random element in* $L_2([0,1], \mathrm{d}s)$ *with covariance operator* $\phi \mapsto \mathbb{E}[Y_1\langle W_1, \phi\rangle W_1]$. *Alternatively,*

$$\alpha_n n(\hat{\varphi} - \varphi_1) \xrightarrow{d} \mathbb{E}\left[Z_1\|W_1\|^2 Y_1\right] + J_2(h)$$

under the topology of $L_2([0,1], \mathrm{d}t)$*, where* $h(X_1, X_2) = \frac{1}{2}\langle W_1, W_2\rangle(Z_1Y_2 + Z_2Y_1)$ *and* $J_2 : L_2(\mathcal{X}, \mathscr{X}, P) \to L_2([0,1], \mathrm{d}t)$ *is a two-fold Wiener-Itô integral with respect to the Gaussian random measure on* $\mathcal{X}$.

For any orthonormal basis $(\varphi_j)_{j\geq 1}$ of $L_2(\mathcal{X}, \mathscr{X}, P)$, the multiple Wiener-Itô integral has the following representation

$$J_2(h) =_d \sum_{(i_1,i_2)=1}^{\infty} \mathbb{E}\left[h(X_1, X_2)\varphi_{i_1}(X_1)\varphi_{i_2}(X_2)\right]\left\{(\chi^2_{1,j} - 1)\delta_{i_1,i_2} + \xi_{i_1}\xi_{i_2}(1 - \delta_{i_1,i_2})\right\},$$

where $\xi_j, j = 1, 2, \ldots$ are i.i.d. $N(0,1)$ and $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise, see Appendix C for more details.

For $\mu \in L_2([0,1], \mathrm{d}t)$, consider the following operator $T^\mu : L_2(\mathcal{X}, \mathscr{X}, P) \to L_2(\mathcal{X}, \mathscr{X}, P), f \mapsto \mathbb{E}[\langle h(X_1, X_2), \mu\rangle f(X_2)|X_1 = x_1]$. This operator is Hilbert-Schmidt and then compact. Let $(\lambda^\mu_j)_{j\geq 1}$ and $(\varphi^\mu_j)_{j\geq 1}$ be eigenvalues and eigenfunctions of T^μ. Then

$$\begin{aligned}\mathbb{E}\left[\langle h(X_1, X_2), \mu\rangle\varphi^\mu_{i_1}(X_1)\varphi^\mu_{i_2}(X_2)\right] &= \mathbb{E}\left[\varphi^\mu_{i_1}(X_1)(T^\mu\varphi^\mu_{i_2})(X_1)\right] \\ &= \lambda^\mu_{i_2}\delta_{i_1,i_2}\end{aligned}$$

and we obtain the following characterization of marginals

$$\langle J_2(h), \mu\rangle =_d \sum_{j=1}^{\infty} \lambda^\mu_j(\chi^2_{1,j} - 1),$$

were $(\lambda_j)_{j\geq 0}$ are solutions to the following eigenvalue problem

$$S\begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix},$$

with $a, b \in L_2$ and $S : L_2 \times L_2 \to L_2 \times L_2$ is a matrix of operators

$$S = \frac{1}{2}\begin{pmatrix} \mathbb{E}\left[Y_1 W_1 \langle Z_1, \mu\rangle \langle W_1, .\rangle\right] & \mathbb{E}\left[W_1 \langle Z_1, \mu\rangle^2 \langle W_1, .\rangle\right] \\ \mathbb{E}\left[Y_1^2 W_1 \langle W_1, .\rangle\right] & \mathbb{E}\left[Y_1 W_1 \langle Z_1, \mu\rangle \langle W_1, .\rangle\right] \end{pmatrix}.$$

Remark 1. *Assuming Hölder smoothness of the process Z_i, we can also obtain functional convergence under the uniform topology. The limiting distribution can be expressed in terms of Gaussian functionals, known as Gaussian chaos, see De la Pena and Giné (2012).*

4 Inference for NPIV model

4.1 Inference for linear functionals

Assumption 8. *(i) The data $(Y_i, Z_i, W_i)_{i=1}^n$ are i.i.d. realizations of (Y, Z, W); (ii) $\mathbb{E}[|Y||Z] < \infty$, $\mathbb{E}[|Y|^2|W] < \infty$ a.s.; (iii) $\bar{K} \in L_\infty$, where $\bar{K}(x) = \int K_w(u)K_w(x-u)\mathrm{d}u$ is a convolution kernel and K_w is symmetric and bounded function; (iv) $f_Z \in H^s(\mathbf{R}^p) \cap L_\infty(\mathbf{R}^p), s > 0$, where $H^s(\mathbf{R}^p)$ denotes Sobolev space;*

Similarly as for the linear IV model, we decompose the function $k(w) := h_w^{-q} K_w\left(h_w^{-1}(W_i - w)\right) = k_i^0 + k_i^1$, where k_i^0 is the projection of k on $\mathcal{N}(K)$, while k_i^1 is the projection of k on $\mathcal{N}^\perp(K)$.

Theorem 6. *Suppose that Assumption 8 is satisfied, k_i^0 is a non-degenerate random variable, and $k_i^1 \in \mathcal{R}\left[(KK^*)^\kappa\right], \kappa > 0$. Then if the sequence of tuning parameters is such that $n\alpha_n h_z^p \to \infty$, $n\alpha_n^{1+\beta\wedge 2} \to 0$, while h_w fixed, for any $\mu_0 \in \mathcal{N}(K)$, we have*

$$n\alpha_n\langle\hat{\varphi} - \varphi_1, \mu_0\rangle \xrightarrow{d} \mathbb{E}\left[(U_1 + \varphi_0(Z_1))\mu_0(Z_1)\right]\|K_w\| + \sum_{j=1}^{\infty} \lambda_j^{\mu_0}(\chi_{1,j}^2 - 1),$$

where $\lambda_j^{\mu_0}, j = 1, 2\ldots$ are eigenvalues of the operator $T^{\mu_0} : L_2(X_1) \to L_2(X_1)$, $f \mapsto \mathbb{E}[h_{\mu_0}(X_1, X_2)f(X_2)|X_1 = x_1]$, where $X_i = (Y_i, Z_i, W_i)$, and

$$h_{\mu_0}(X_i, X_j) = \frac{1}{2}\left\{(U_i + \varphi_0(Z_i))\mu_0(Z_j) + (U_j + \varphi_0(Z_j))\mu_0(Z_i)\right\}\left\langle k_i^0, k_j^0\right\rangle.$$

Otherwise, if k_i^0 is degenerate

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

4.2 Asymptotic distribution when the instrument is irrelevant

In the linear IV model, the strength of the association between the instrument and the regressor is described by the covariance operator. In the nonparametric IV regression, it is described by the conditional expectation operator. Consider extreme case of violation of completeness condition, i.e. when $\mathbb{E}[\phi(Z)|W] = 0$ for all $\phi \in L_Z^{2,0} = \left\{\phi : \mathbb{E}|\phi(Z)|^2 < \infty, \mathbb{E}\phi(Z) = 0\right\}$. One reason why this may happen is that $Z \perp\!\!\!\perp W$. In this case the operator K becomes a degenerate integral operator

$$(K\phi)(w) = \int \phi(z)f_Z(z)\mathrm{d}z f_W(w),$$

and the operator K^*K has only one non-zero eigenvalue $\lambda_1 = \|f_Z\|^2\|f_W\|^2$ corresponding to the eigenvector f_Z. As a result, the data contain almost no information on about the structural function φ. We define K_0 to be a restriction of K to $L_Z^{2,0}$. Then the adjoint operator $K_0^* = P_0 K^*$, where P_0 is the projection on $L_Z^{2,0}$. Then $K_0 = K_0^* = 0$, $\varphi_1 = 0$, and obtain the following result. In what follows, we will use K and K^* to denote K_0 and K_0^*.

Theorem 7. *Suppose that Assumption 8 is satisfied. Then for any $\mu \in L_2(\mathbf{R}^p) \cap C(\mathbf{R}^p)$ if $n\alpha_n h_z^p \to \infty$, while h_w is fixed*

$$\alpha_n n\langle\hat{\varphi} - \varphi_1, \mu\rangle \xrightarrow{d} \mathbb{E}\left[Y_1\mu(Z_1)\right] h_w^{-q}\bar{K}(0) + \sum_{j=1}^{\infty} \lambda_j^{\mu}(\chi_{j,1}^2 - 1),$$

where λ_j^{μ} are eigenvalues of the operator $T^{\mu} : L_X^2 \to L_X^2$, $f \mapsto \mathbb{E}[h(X_1, X_2)f(X_2)|X_1 = x_1]$.

Unlike in the linear IV model, in the NPIV model it is not possible to obtain weak convergence of $\alpha_n n(\hat{\varphi} - \varphi_1)$ as a process in the Hilbert space. The situation is similar to the kernel density estimator, for which, despite the fact that it is possible to show root-n convergence of inner products, the underlying process is not tight.

5 Conclusion

This paper investigates non-identified high-dimensional linear and non-parametric IV models. Identification failures can occur due to the non-injectivity of covariance or conditional expectation operator. We show that if the operator is not injective, a very general class of estimators converges to the best approximation of the structural function. On the other hand, in the case of non-identification, the Tikhonov-regularized estimator exhibits U-statistics type behavior.

Appendix A: General regularization schemes

Consider an ill-posed operator equation

$$K\varphi = r,$$

where K is an operator between separable real Hilbert spaces $\mathcal{E}$ and $\mathcal{F}$. The operator K is assumed to be bounded, but it need not to be compact. Then $K^*K : \mathcal{E} \to \mathcal{E}$ is a self-adjoint operator and so it admits spectral decomposition

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

with respect to the resolution of identity E, see (Rudin, 1991, Theorem 12.23). For a bounded Borel function $g : \sigma(K^*K) \to \mathbf{R}$, we can define functions of the operator K^*K using its spectral decomposition

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

If the operator K is compact, the spectrum of K^*K is countable and the above formula reduces to

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

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

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

In econometric applications, the operator K is typically not known. The estimate of convergence rates in this cases depends crucially on the following inequality. A real function $s : (0, a) \to \mathbf{R}$ with $s(0) = 0$ is called *operator monotone* if for any pair of self-adjoint operators A and B with spectrum in $[0, a)$ and such that $A \leq B$, we have $s(A) \leq s(B)$[3]. If s is operator monotone, the following inequality holds

$$\|s(A) - s(B)\| \leq Cs(\|A - B\|) \tag{7}$$

for some constant $C < \infty$, see Mathé and Pereverzev (2002) and Lu and Pereverzev (2013).

[3] We say that for self-adjoint operators on the Hilbert space H, $A \leq B$ if and only if $\langle Ax, x\rangle \leq \langle Bx, x\rangle, \forall x \in H$.

Assumption 9. *The function* φ_1 *is such that for some* $\beta \in (0, \beta_0)$ *there exists* ψ *such that*

$$\varphi_1 = s_\beta(K^*K)\psi, \qquad \|\psi\|^2 \leq C,$$

where $s_\beta : [0, \infty) \to \mathbf{R}$ *is one of the following two families of functions indexed by* $\beta > 0$

1. *mildly ill-posed case:* $s_\beta(\lambda) = \lambda^{\beta/2}$.
2. *severely ill-posed case:* $s_\beta(\lambda) = \log^{-\beta/2}\left(\frac{1}{\lambda}\right)$ *with* $s_\beta(0) = 0$

As discussed in Mathé and Pereverzev (2002) $\lambda \mapsto \lambda^{\beta/2}$ is operator monotone on $(0, \infty)$ for all $\beta \in (0, 2]$ and $\lambda \mapsto \log^{-\beta/2}\left(\frac{1}{\lambda}\right)$ is operator monotone[4] on $(0, 1)$ for any $\beta > 0$.

Consider spectral regularization schemes, described by the family of bounded Borel functions $g_\alpha : [0, \infty) \to \mathbf{R}, \alpha > 0$ such that $\lim_{\alpha \to 0} g_\alpha(\lambda) = \lambda^{-1}$. We assume that the operator norms of K^*K and $\hat{K}^*\hat{K}$ are bounded by some constant Λ and define regularized estimator as

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

Assumption 10. *There exists positive constants* c_1, c_2*, and* c_3 *such that for all* $\beta \in (0, \beta_0], \beta_0 < \infty$

$$(i) \quad \sup_{\lambda \in [0,\Lambda]} \left| g_\alpha(\lambda)\lambda^{1/2} \right| \leq \frac{c_1}{\alpha_n^{1/2}}, \qquad (ii) \quad \sup_{\lambda \in [0,\Lambda]} |g_\alpha(\lambda)\lambda - 1| \leq 1,$$

$$(iii) \quad \sup_{\lambda \in [0,\Lambda]} |(g_\alpha(\lambda)\lambda - 1)s_\beta(\lambda)| \leq c_2 s_\beta(\alpha), \qquad (iv) \quad \sup_{\lambda \in [0,\Lambda]} |g_\alpha(\lambda)| \leq \frac{c_3}{\alpha},$$

where $s_\beta(\lambda)$ *equals either to* $\lambda^{\beta/2}$ *or to* $\log^{-\beta/2}\left(\frac{1}{\lambda}\right)$.

It is easy to verify that the following regularization schemes satisfy Assumption 10 for mildly and severely ill-posed cases, see e.g. Vainikko and Veretennikov (1986) and Lu and Pereverzev (2013)

1. Tikhonov:

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

   Assumption 10 is satisfied with $c_1 = 1/2$, $c_2 = c_3 = 1$, and $\beta_0 = 2$.

2. Principal components (spectral cut-off):

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

   Assumption 10 is satisfied with $c_1 = c_2 = c_3 = 1$ and any $\beta_0 > 0$.

[4] In this case we assume that operator norms are scaled properly.

3. Iterated Tikhonov:

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

for $m = 2, 3, \ldots$ Assumption 10 is satisfied with $c_1 = m^{1/2}, c_2 = 1$ and $c_3 = \beta_0 = m$.

4. Landweber-Fridman:

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

for $\alpha = 1/m, m = 1, 2, \ldots$ and some $c \in (0, 1/\Lambda)$. Assumption 10 is satisfied with $c_1^2 = c_2 = cm, c_2 = \left[\left(\frac{\beta_0}{ce}\right)^{\beta_0}\right] \vee 1$, and any $\beta_0 \in \mathbf{R}$.

The constant β_0 is the so-called *qualification* of the regularization scheme. It is well-known that Tikhonov regularization exhibits saturation effect and the bias can't converge faster than at the rate α_n^2. This effect is somewhat similar to the saturation of convergence rate for the bias of the kernel density estimator. Iterated Tikhonov regularization allows to improve on the rate of the bias, once sufficiently high number of iterations m is selected, similarly to selecting higher-order kernels for the kernel density estimator.

Assumption 11. *Suppose that for all $\phi \in L_2$*

$$(i) \quad \mathbb{E}\left\|\hat{r} - \hat{K}\varphi\right\|^2 = O(\delta_n), \qquad (ii) \quad \mathbb{E}\left\|(\hat{K} - K)\phi\right\|^2 = O\left(\rho_{1,n}\right),$$

and

$$(iii) \quad \mathbb{E}\left[s_\beta^2\left(\left\|\hat{K}^*\hat{K} - K^*K\right\|\right)\right] = O(\rho_{2,n}), \qquad (iv) \quad \mathbb{E}\left\|\hat{K}^* - K^*\right\|_{2,\infty}^2 = O(\xi_n),$$

where s_β is the same as in Assumption 9 and ξ_n is some bounded sequence.

The following result tells us that the estimator converges to the best approximation to the function φ_1 for a general class of regularization schemes. Typically for the principal components approach, convergence rates are obtained under assumptions on the spacing between eigenvalues of the operator K^*K, see (Hall, Horowitz, et al., 2007, Assumption 3.2). The interesting feature of the result stated below is that, it does not require such assumptions. Moreover, it allows us to cover cases when eigenvalues of K^*K decay to zero exponentially fast, including cases when Fourier coefficients of φ_1 decay polynomially.

Theorem 8. *Under Assumptions 9, 10 (i)-(iii), and 11 (i)-(iii) if* $\beta \leq 2$

$$\mathbb{E}\left\|\hat{\varphi}-\varphi_1\right\|^2 = O\left(\frac{\delta_n+\rho_{1,n}}{\alpha_n}+\rho_{2,n}+s_\beta^2(\alpha_n)\right).$$

Proof. Decompose

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

with

$$\begin{aligned}
I_n &= g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*(\hat{r}-\hat{K}\varphi)\\
II_n &= g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K}\varphi_0\\
III_n &= \left[g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K}-I\right] s_\beta(\hat{K}^*\hat{K})s_\beta^{-1}(K^*K)\varphi_1\\
IV_n &= \left[g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K}-I\right]\left\{s_\beta(K^*K)-s_\beta(\hat{K}^*\hat{K})\right\} s_\beta^{-1}(K^*K)\varphi_1.
\end{aligned}$$

By properties of functional calculus

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

giving under Assumptions 10 and 11

$$\mathbb{E}\|I_n\|^2 = O\left(\frac{\delta_n}{\alpha_n}\right).$$

Similarly,

$$\mathbb{E}\|II_n\|^2 = O\left(\frac{\rho_{1,n}}{\alpha_n}\right).$$

Likewise, under Assumptions 9 and 10

$$\begin{aligned}
\|III_n\|^2 &\leq \left\|\left[g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K}-I\right]s_\beta(\hat{K}^*\hat{K})\right\|^2\left\|s_\beta^{-1}(K^*K)\varphi_1\right\|^2\\
&\leq \sup_{\lambda\in\sigma(\hat{K}^*\hat{K})}\left|(g_{\alpha_n}(\lambda)\lambda-1)s_\beta(\lambda)\right|^2\|\psi\|^2\\
&= O\left(s_\beta^2(\alpha_n)\right).
\end{aligned}$$

Lastly, under Assumptions 9 and 10 in light of Eq. (7) operator monotonicity of s_β gives

$$\begin{aligned}
\|IV_n\|^2 &\leq \left\|g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K}-I\right\|^2\left\|s_\beta(\hat{K}^*\hat{K})-s_\beta(K^*K)\right\|^2\left\|s_\beta^{-1}(K^*K)\varphi_1\right\|^2\\
&\leq C\left\|s_\beta(\hat{K}^*\hat{K})-s_\beta(K^*K)\right\|^2\\
&\leq Cs_\beta^2\left(\left\|\hat{K}^*\hat{K}-K^*K\right\|\right),
\end{aligned}$$

and the result follows under Assumption 11. □

In the numerical ill-posed inverse literature, the investigation of uniform convergence of Tikhonov regularization dates back to Khudak (1966) and Ivanov (1967). The idea of using functional calculus and spectral families to describe general regularization schemes in Hilbert spaces is due to Bakushinskii (1967). Groetsch (1985) investigated uniform convergence rates in the case of the general spectral regularization when the operator K is known. Rajan (2003) studied uniform convergence rates for the Tikhonov regularization when there is numerical error in the operator. Whether we can have uniform convergence for general spectral regularization schemes with deterministic or stochastic error in the operator remained an open question.

The following result is the first to describe uniform convergence rates for a general family of spectrally-regularized estimators when the operator K is not known and is estimated from the data. This setting is the most relevant to econometrics and statistics.

Theorem 9. *Suppose that Assumption 9 is satisfied with $S_\beta(K^*K) = K^* s_\beta(KK^*)$. Suppose also that Assumptions 10 (iv), 11 (i)-(iv) are satisfied. Then if $\|K^*\|_{2,\infty} < \infty$ and $\beta \leq 2$*

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

Proof. Consider decomposition as in the proof of the Theorem 8. We bound the first term as

$$\begin{aligned}
\|I_n\|_\infty &= \left\|\hat{K}^* g_{\alpha_n}(\hat{K}\hat{K}^*)(\hat{r} - \hat{K}\varphi)\right\| \\
&\leq \left\|\hat{K}^*\right\|_{2,\infty} \left\|g_{\alpha_n}(\hat{K}\hat{K}^*)\right\| \left\|\hat{r} - \hat{K}\varphi\right\| \\
&\leq \left(\left\|\hat{K}^* - K^*\right\|_{2,\infty} + \|K^*\|_{2,\infty}\right) \left\|\hat{r} - \hat{K}\varphi\right\| \sup_{\lambda \in \sigma(\hat{K}\hat{K}^*)} |g_{\alpha_n}(\lambda)|\,.
\end{aligned}$$

Whence by Cauchy-Schwartz inequality

$$\mathbb{E}\|I_n\|_\infty = O\left(\frac{\delta_n^{1/2}}{\alpha_n}\right).$$

Similarly

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

The third term is treated as

$$\begin{aligned}\|III_n\|_\infty &\leq \left\|\hat{K}^*\left[g_{\alpha_n}(\hat{K}\hat{K}^*)\hat{K}\hat{K}^* - I\right] s_\beta(\hat{K}\hat{K}^*)\right\|_{2,\infty} \left\|S_\beta^{-1}(K^*K)\varphi\right\| \\ &\leq \left(\left\|\hat{K}^* - K^*\right\|_{2,\infty} + \|K^*\|_{2,\infty}\right) \sup_{\lambda\in\sigma(\hat{K}\hat{K}^*)} \left|[g_{\alpha_n}(\lambda)\lambda - 1]\, s_\beta(\lambda)\right| C^{1/2}\end{aligned}$$

whence

$$\mathbb{E}\|III_n\|_\infty = O\left(s_\beta(\alpha_n)\right).$$

Lastly,

$$\|IV_n\|_\infty \leq \left\|g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right\|_\infty \left\|S_\beta(\hat{K}^*\hat{K}) - S_\beta(K^*K)\right\|_{2,\infty} C^{1/2},$$

where

$$\begin{aligned}\left\|g_{\alpha_n}(\hat{K}^*\hat{K})\hat{K}^*\hat{K} - I\right\|_\infty &\leq \left\|\hat{K}^*\right\|_{2,\infty}\left\|g_{\alpha_n}(\hat{K}\hat{K}^*)\hat{K}\right\| + 1 \\ &\leq \left(\left\|\hat{K}^* - K^*\right\|_{2,\infty} + \|K^*\|_{2,\infty}\right) \sup_{\lambda\in\sigma(\hat{K}\hat{K}^*)} \left|g_{\alpha_n}(\lambda)\lambda^{1/2}\right| + 1 \\ &\leq \frac{c_1}{\alpha_n^{1/2}}\left(\left\|\hat{K}^* - K^*\right\|_{2,\infty} + \|K^*\|_{2,\infty}\right) + 1\end{aligned}$$

and

$$\left\|S_\beta(\hat{K}^*\hat{K}) - S_\beta(K^*K)\right\|_{2,\infty} \leq \left\|\hat{K}^* - K^*\right\|_{2,\infty}\left\|s_\beta(\hat{K}\hat{K}^*)\right\| + \|K^*\|_{2,\infty}\left\|s_\beta(\hat{K}^*\hat{K}) - s_\beta(K^*K)\right\|.$$

Therefore, by Cauchy-Schwartz inequality

$$\mathbb{E}\|IV_n\|_\infty = O\left(\frac{\rho_{2,n}^{1/2}}{\alpha_n^{1/2}}\right)$$

□

Appendix B: Proofs

Proof of Theorem 1. Decompose

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

The bias term is treated exactly in the same way as in the identified case using now a source condition on φ_1 in Assumption 1

$$\|IV_n\|^2 = \left\|\alpha_n(\alpha_n I + K^*K)^{-1}\varphi_1\right\|^2 = O\left(\alpha_n^{\beta\wedge 2}\right).$$

The first term under Assumption 2 is treated as

$$\mathbb{E}\|I_n\|^2 \leq \mathbb{E}\left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|^2\left\|\hat{r} - \hat{K}\varphi\right\|^2 \leq \frac{1}{4\alpha_n}\mathbb{E}\left\|\hat{r} - \hat{K}\varphi\right\|^2 = O\left(\frac{\delta_n}{\alpha_n}\right).$$

The second term is a new component that comes from the fact that there is identification failure

$$\mathbb{E}\|II_n\|^2 \leq \mathbb{E}\left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|^2\left\|(\hat{K} - K)\varphi_0\right\|^2 \leq \frac{1}{4\alpha_n}\mathbb{E}\left\|(\hat{K} - K)\varphi_0\right\|^2 = O\left(\frac{\rho_{1,n}}{\alpha_n}\right).$$

The third term is decomposed further into

$$\begin{aligned}
III_n &= -\left[\alpha_n(\alpha_n I + \hat{K}^*\hat{K})^{-1} - \alpha_n(\alpha_n I + K^*K)^{-1}\right]\varphi_1 \\
&= -(\alpha_n I + \hat{K}^*\hat{K})^{-1}\alpha_n\left[K^*K - \hat{K}^*\hat{K}\right](\alpha_n I + K^*K)^{-1}\varphi_1 \\
&= (\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\left[\hat{K} - K\right]\alpha_n(\alpha_n I + K^*K)^{-1}\varphi_1 \\
&\qquad + (\alpha_n I + \hat{K}^*\hat{K})^{-1}\left[\hat{K}^* - K^*\right]\alpha_n K(\alpha_n I + K^*K)^{-1}\varphi_1 \\
&= V_n + VI_n,
\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}\|V_n\|^2 &\leq \frac{1}{4\alpha_n}\mathbb{E}\left\|\hat{K}-K\right\|^2\left\|\alpha_n(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\ &= O\left(\frac{\rho_{2,n}}{\alpha_n}\alpha_n^{\beta\wedge 2}\right)\\ \mathbb{E}\|VI_n\|^2 &\leq \frac{1}{\alpha_n^2}\mathbb{E}\left\|\hat{K}^*-K^*\right\|^2\left\|\alpha K(\alpha_n I+K^*K)^{-1}\varphi_1\right\|^2\\ &= O\left(\frac{\rho_{2,n}}{\alpha_n}\alpha_n^{\beta\wedge 1}\right).\end{aligned}$$

□

Proof of the Theorem 2. Consider the same decomposition as in the proof of Theorem 1. Notice that the assumption that $\varphi_1 \in \mathcal{R}(K^*K)^{\beta/2}$ for $\beta > 1$ can be re-parametrized as $\varphi_1 \in \mathcal{R}\left[(K^*K)^{\tilde{\beta}}K^*\right]$ for $\tilde{\beta} = \frac{\beta-1}{2} > 0$. Then the fourth term is treated similarly to the identified case in Babii (2016a)

$$\|IV_n\|_\infty \leq \|K^*\|_{2,\infty}\left\|\alpha_n(\alpha_n I+KK^*)^{-1}(KK^*)^{\tilde{\beta}}\psi\right\| = O\left(\alpha_n^{\tilde{\beta}\wedge 1}\right)$$

The first term is treated using Cauchy-Schwartz inequality

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

The new term coming from the non-identification is handled similarly

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

The third term is decomposed further similarly as in the proof of Theorem 1, but to bound

$\mathbb{E}\|V_n\|_\infty$ and $\mathbb{E}\|VI_n\|_\infty$ we will use slightly different strategy. First,

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

Second, using the inequality $\|(\alpha I + K^*K)^{-1}\|_\infty \leq \frac{\|K^*\|_{2,\infty}/2+\alpha^{1/2}}{\alpha^{3/2}}$ from Rajan (2003)

$$
\begin{aligned}
\mathbb{E}\|VI_n\|_\infty &= \mathbb{E}\left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\right\|_\infty\left\|\hat{K}^*-K^*\right\|_{2,\infty}\|\alpha_n K(\alpha_n I + K^*K)^{-1}\varphi_1\| \\
&= \left(\frac{1}{2\alpha_n^{3/2}}\left(\|K^*\|_{2,\infty}\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty} + \mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}^2\right) + \frac{1}{\alpha_n}\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}\right) O(\alpha_n) \\
&= O\left(\frac{1}{\alpha_n^{1/2}}\sqrt{\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}^2} + \sqrt{\mathbb{E}\left\|\hat{K}^*-K^*\right\|_{2,\infty}^2}\right) \\
&= O\left(\frac{\xi_n^{1/2}+\xi_n}{\alpha_n^{1/2}} + \xi_n^{1/2}\right).
\end{aligned}
$$

Collecting all estimates together, we obtain the result. □

The following proposition provides some supplementary results for the NPIV model.

Proposition 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} is in the Nikol'ski class $B_{2,\infty}^s$; (iii) kernel functions $K_z : \mathbf{R}^p \to \mathbf{R}$ and $K_w : \mathbf{R}^q \to \mathbf{R}$ are such that for $l \in \{w, z\}$, $K_l \in L_2(\mathbf{R})$, $\int K_l(u)\mathrm{d}u = 1$, $\int \|u\|^s K_l(u)\mathrm{d}u < \infty$, and $\int u^k K_l(u)\mathrm{d}u = 0$ for all multindices $|k| = 1, \ldots, \lfloor s \rfloor$. Then for all $\phi \in L_2$*

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

and

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

Proof. Decompose

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

By 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 $f_{ZW} \in B^s_{2,\infty}$, see (Giné and Nickl, 2015, p.404).

For the variance put

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

with $[\phi * K_z](Z_i) = \int \phi(z) h_n^{-p} K_z\left(h_n^{-1}(Z_i - z)\right) \mathrm{d}z$, and notice that

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

Then

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

where the last line follows, since by change of variables, Cauchy-Schwartz inequality, and by translation invariance of Lebesgue measure

$$\int f_Z(z) |[\phi * K_z](z)|^2 \mathrm{d}z \leq \|K_z\|^2\|\phi\|^2.$$

This establishes the first claim and since

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

the second claim follows if we can show 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

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

By Cauchy-Schwartz inequality

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

where $[f_{ZW}*K_{w,h}](w)=\int f_{ZW}(z,\tilde{w})h^{-q}K_w\left(h^{-1}(w-\tilde{w})\right)\mathrm{d}\tilde{w}$. Since $f_{ZW}\in B^s_{2,\infty}$ we obtain

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

see e.g. (Giné and Nickl, 2015, 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 function and standard computations for the risk of the joint density, (Giné and Nickl, 2015, Chapter 5). □

Proof of Theorem 3. Put $b_n=\alpha_n(\alpha_nI+K^*K)^{-1}\varphi_1$ and notice that $(\alpha_nI+\hat{K}^*\hat{K})^{-1}\hat{K}^*=\hat{K}^*(\alpha_nI+\hat{K}\hat{K}^*)^{-1}$. Using this, similarly to the proof of Theorem 1, decompose

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

Then the first term can be written as

$$\begin{aligned}\alpha_n I_n &= \left\langle \alpha_n(\alpha_n I + KK^*)^{-1}(\hat{r} - \hat{K}\varphi_1), \hat{K}\mu_0 \right\rangle \\ &\equiv I_n^0 + I_n^1,\end{aligned}$$

with

$$\begin{aligned}I_n^0 &= \frac{1}{n^2}\sum_{i,j=1}^n (U_i + \langle Z_i, \varphi_0\rangle \langle Z_j, \mu_0\rangle \left\langle W_i^0, W_j^0 \right\rangle \\ I_n^1 &= \frac{1}{n^2}\sum_{i,j=1}^n (U_i + \langle Z_i, \varphi_0\rangle)\langle Z_j, \mu_0\rangle \left\langle \alpha_n(\alpha_n I + KK^*)^{-1} W_i^1, W_j^1 \right\rangle.\end{aligned}$$

Since projection is a bounded linear operator, it commutes with expectation and it is easy to see that

$$\begin{aligned}\mathbb{E}\left[W_1^0(U_1 + \langle Z_1, \varphi_0\rangle)\right] &= 0 \\ \mathbb{E}\left[W_1^0\langle Z_1, \mu_0\rangle\right] &= K\mu_0 = 0.\end{aligned}$$

Then

$$nI_n^0 =: \zeta_n + nU_n,$$

with

$$\begin{aligned}\zeta_n &= \frac{1}{n}\sum_{i=1}^n (U_i + \langle Z_i, \varphi_0\rangle)\langle Z_i, \mu_0\rangle \left\|W_i^0\right\| \\ nU_n &= \frac{1}{n}\sum_{i<j}\left\{\langle Z_i, \mu_0\rangle(U_j + \langle Z_j, \varphi_0\rangle) + \langle Z_j, \mu_0\rangle(U_i + \langle Z_i, \mu_0\rangle)\right\}\left\langle W_i^0, W_j^0\right\rangle\end{aligned}$$

Under Assumption 4

$$\zeta_n \xrightarrow{a.s.} \mathbb{E}\left[\left\|W_i^0\right\|(U_i + \langle Z_i, \mu_0\rangle)\langle Z_i, \mu_0\rangle\right],$$

while nU_n is a degenerate U-statistics with kernel function h_{μ_0}, since

$$\mathbb{E}_{X_2}\left[h_{\mu_0}(X_1, X_2)\right] = \frac{1}{2}\left\{\langle Z_1, \mu_0\rangle \left\langle W_1^0, K\varphi_0\right\rangle + (U_1 + \langle Z_1, \varphi_0\rangle)\left\langle W_1^0, K\mu_0\right\rangle\right\} = 0.$$

Under Assumption 4 by the standard CLT for degenerate U-statistics, see Gregory (1977) or Serfling (1980)

$$nU_n \xrightarrow{d} \sum_{j=1}^{\infty}\lambda_j^{\mu_0}(\chi_{1,j}^2 - 1),$$

where $(\lambda_j^{\mu_0})_{j\geq 1}$ are eigenvalues of T_{μ_0}.

It remains to show that all other terms after normalization with $n\alpha_n$ go to zero. It is easy to verify that the variance of $n\alpha_n I_n^1 \to 0$, since $W_i^1 \in \mathcal{R}\left[(KK^*)^{\kappa}\right]$. Notice also that $\mu_0 \in \mathcal{N}(K)$

implies that

$$(\alpha_n I + K^*K)^{-1}\mu_0 = \frac{1}{\alpha_n}\mu_0. \tag{8}$$

Using this fact

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

By Cauchy-Schwartz inequality

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

Therefore as long as $n\alpha_n \to \infty$, we will have $n\alpha_n II_n \xrightarrow{p} 0$.

Next, under Assumption 1 there exists some $\tilde{\psi} \in L_2$

$$\|b_n\| = \left\|\alpha_n(\alpha_n I + K^*K)^{-1}(K^*K)^{\frac{\beta}{2}}\tilde{\psi}\right\| = O\left(\alpha_n^{\frac{\beta}{2}\wedge 1}\right),$$

whence, for III_n and IV_n, we have

$$\begin{aligned} III_n &\leq \left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\right\| \left\|\hat{K} - K\right\| \|b_n\|\|\mu_0\| = O_p\left(\frac{\alpha_n^{\frac{\beta}{2}\wedge 1}}{\sqrt{n\alpha_n}}\right) \\ IV_n &\leq \left\|(\alpha_n I + \hat{K}^*\hat{K})^{-1}\right\| \left\|\hat{K}^* - K^*\right\| \|Kb_n\|\|\mu_0\| = O_p\left(\frac{\alpha_n^{\left(\frac{\beta}{2}+\frac{1}{2}\right)\wedge\frac{1}{2}}}{\sqrt{n\alpha_n}}\right). \end{aligned}$$

Then $n\alpha_n III_n \xrightarrow{p} 0$, since $n\alpha_n^{1+\beta\wedge 2} \to 0$ and $\alpha_n n IV_n \xrightarrow{p} 0$, since $n\alpha_n^2 \to 0$.

Lastly, notice that bias is identically zero by Eq. (8) and orthogonality between φ_1 and μ_0

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

□

Proof of Theorem 4. The proof is similar to the proof of Theorem 3 and we omit steps discussed there. Decompose

$$
\begin{aligned}
\langle\hat{\varphi}-\varphi_1,\mu_1\rangle = & \left\langle \frac{1}{n}\sum_{i=1}^{n}(\alpha_n I+K^*K)^{-1}K^*W_i(U_i+\langle Z_i,\varphi_0\rangle),\mu_1\right\rangle \\
& +\left\langle\left((\alpha_n I+\hat{K}^*\hat{K})^{-1}-(\alpha_n I+K^*K)^{-1}\right)\hat{K}^*\frac{1}{n}\sum_{i=1}^{n}W_i(U_i+\langle Z_i,\varphi_0\rangle),\mu_1\right\rangle \\
& +\left\langle(\alpha_n I+K^*K)^{-1}(\hat{K}^*-K^*)\frac{1}{n}\sum_{i=1}^{n}W_i(U_i+\langle Z_i,\varphi_0\rangle),\mu_1\right\rangle \\
& +\left\langle\left\{(\alpha_n I+\hat{K}^*\hat{K})^{-1}-(\alpha_n I+K^*K)^{-1}\right\}\hat{K}^*(\hat{K}-K)b_n,\mu_1\right\rangle \\
& +\left\langle\left\{(\alpha_n I+\hat{K}^*\hat{K})^{-1}-(\alpha_n I+K^*K)^{-1}\right\}(\hat{K}^*-K^*)Kb_n,\mu_1\right\rangle \\
& +\left\langle(\alpha_n I+K^*K)^{-1}K^*(\hat{K}-K)b_n,\mu_1\right\rangle \\
& +\left\langle(\alpha_n I+K^*K)^{-1}(\hat{K}^*-K^*)Kb_n,\mu_1\right\rangle \\
& +\left\langle(\alpha_n I+K^*K)^{-1}(\hat{K}^*-K^*)(\hat{K}-K)b_n,\mu_1\right\rangle \\
& +\langle b_n,\mu_1\rangle \\
\equiv\ & I_n+II_n+III_n+IV_n+V_n+VI_n+VII_n+VIII_n+IX_n.
\end{aligned}
$$

Under Assumption 5 by the Lindeberg-Feller's central limit theorem

$$\pi_n I_n\xrightarrow{d}N(0,1).$$

It remains to show that all other terms after normalization with π_n go to zero. For II_n we have

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

Since $\mu_1\in\mathcal{R}\left[(K^*K)^\gamma\right]$, there exists some $\psi\in L_2$ such that $\mu_1=(K^*K)^\gamma\psi$ and so

$$
\begin{aligned}
II_n &\leq O_p\left(\frac{1}{\sqrt{n}}\right)\left\|\hat{K}^*(\alpha_n I+\hat{K}^*\hat{K})^{-1}\hat{K}^*\right\|\left\|\hat{K}-K\right\|\left\|(\alpha_n I+K^*K)^{-1}(K^*K)^\gamma\psi\right\| \\
&\quad+O_p\left(\frac{1}{\sqrt{n}}\right)\left\|\hat{K}^*(\alpha_n I+\hat{K}^*\hat{K})^{-1}\right\|\left\|\hat{K}^*-K^*\right\|\left\|K(\alpha_n I+K^*K)^{-1}(K^*K)^\gamma\psi\right\| \\
&=O_p\left(\frac{\alpha_n^{\gamma\wedge 1}+\alpha_n^{\gamma\wedge\frac{1}{2}}}{n\alpha_n}\right).
\end{aligned}
$$

while

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

For IV_n and V_n we have

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

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

For VI_n and VII_n we obtain

$$\begin{aligned}
VI_n &\leq \left\|\hat{K} - K\right\| \|b_n\| \left\|K^*(\alpha_n I + K^*K)^{-1}\mu_1\right\| = O_p\left(\frac{\alpha_n^{\frac{\beta}{2}\wedge 1 + \gamma\wedge\frac{1}{2}}}{\sqrt{n\alpha_n}}\right) \\
VII_n &\leq \left\|\hat{K}^* - K^*\right\| \|Kb_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| = O_p\left(\frac{\alpha_n^{\frac{\beta}{2}\wedge\frac{1}{2} + \gamma\wedge 1}}{\sqrt{n\alpha_n}}\right).
\end{aligned}$$

Lastly,

$$VIII_n \leq \left\|\hat{K}^* - K^*\right\| \left\|\hat{K} - K\right\| \|b_n\| \left\|(\alpha_n I + K^*K)^{-1}\mu_1\right\| = O_p\left(\frac{\alpha_n^{\frac{\beta}{2}\wedge 1 + \gamma\wedge 1}}{n\alpha_n}\right).$$

Notice that Assumption 7 (i) ensures that $\pi_n IX_n \to 0$, while (ii) ensures that all other terms except for I_n, multiplied by π_n converge in probability to zero. □

Proof of Theorem 5. Since $K = K^* = 0$, $\varphi = \varphi_0$, $\varphi_1 = 0$, and

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

Notice that under Assumption 4

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

and $\mathbb{E}\|\hat{K}^*\|^2 = O\left(\frac{1}{n}\right)$, implying $\hat{K}^*\hat{K} = O_p\left(\frac{1}{n}\right)$ in the space of bounded linear operators. Therefore, as $\alpha_n n \to \infty$, by the continuous mapping theorem in metric spaces, (Van Der Vaart and Wellner, 2000, Theorem 1.3.6), $\left(I + \frac{1}{\alpha_n}\hat{K}^*\hat{K}\right)^{-1} \xrightarrow{p} I$ and

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

By Slutsky's theorem in metric spaces, (Van Der Vaart and Wellner, 2000, Example 1.4.7), it suffices to analyze the weak convergence of $n\hat{K}^*\hat{r}$.

Notice that $\phi \mapsto \langle W, \phi\rangle Z$ is a random element in the space of Hilbert-Schmidt operators, denoted by $\mathcal{S}_2$. This space is a Hilbert space with respect to the inner product $\langle A, B\rangle_{HS} = \mathrm{trace}(B^*A), \forall A, B \in \mathcal{S}_2$. Under Assumption 4

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

converges weakly to zero-mean Gaussian random operator $\mathbb{G}$ in $\mathcal{S}_2$ with covariance operator $A \mapsto \mathbb{E}\left[\mathrm{trace}(A^*\langle W, .\rangle Z)\langle W, .\rangle Z\right]$. On the other hand,

$$\sqrt{n}\hat{r} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i Y_i$$

converges weakly to zero-mean Gaussian random vector g in L_2 with covariance operator $\phi \mapsto \mathbb{E}\left[Y_i^2\langle W_i, \phi\rangle W_i\right]$. Therefore, under the product topology of $\mathcal{S}_2 \times L_2$ by the continuous mapping theorem

$$n\hat{K}^*\hat{r} \xrightarrow{d} \mathbb{G}g,$$

establishing the first statement.

For the second statement, notice that

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

Assuming $\mathbb{E}\left|\|W_i\|^2\|Z_i\|Y_i\right| < \infty$, by the Mourier law of large numbers

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

The second term is a normalized degenerate U-statistics in L_2 with kernel function $h(X_1, X_2) = \frac{1}{2}\langle W_1, W_2\rangle\,(Z_1Y_2 + Z_2Y_1)$

$$nU_n = \frac{2}{n}\sum_{i<j}\frac{Z_iY_j + Z_jY_i}{2}\langle W_i, W_j\rangle.$$

Under the Assumption 4 (ii), by the Borovskich CLT for Hilbert-space valued U-statistics, see Theorem 10

$$nU_n \xrightarrow{d} J_2(h),$$

where $J_2(h)$ is a two-fold Wiener-Itô integral with respect to the Gaussian random measure on $\mathcal{X}$. □

Proof of Theorem 6. Put $b_n = \alpha_n(\alpha_n I + K^*K)^{-1}\varphi_1$ and notice that $(\alpha_n I + \hat{K}^*\hat{K})^{-1}\hat{K}^* = \hat{K}^*(\alpha_n I + \hat{K}\hat{K}^*)^{-1}$. Using this, similarly to the proof of Theorem 1, decompose

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

Then the first term can be written as

$$\begin{aligned}
\alpha_n I_n &= \left\langle \alpha_n(\alpha_n I + KK^*)^{-1}(\hat{r} - \hat{K}\varphi_1), \hat{K}\mu_0\right\rangle \\
&\equiv I_n^0 + I_n^1 + I_n^2 + I_n^3,
\end{aligned}$$

with

$$I_n^0 = \frac{1}{n^2}\sum_{i,j=1}^{n}(U_i+\varphi_0(Z_i))\mu_0(Z_j)\left\langle k_i^0, k_j^0\right\rangle$$

$$I_n^1 = \frac{1}{n^2}\sum_{i,j=1}^{n}(U_i+\varphi_0(Z_i))\left([\mu_0 * K_z](Z_j)-\mu_0(Z_j)\right)\left\langle k_i^0, k_j^0\right\rangle$$

$$I_n^2 = \frac{1}{n^2}\sum_{i,j=1}^{n}(\varphi_1(Z_i)-[\varphi_1 * K_z](Z_i))[\mu_0 * K_z](Z_j)\left\langle k_i^0, k_j^0\right\rangle$$

$$I_n^3 = \frac{1}{n}\sum_{i,j=1}^{n}(Y_i-[\varphi * K_z](Z_i))[\mu_0 * K_z](Z_j)\left\langle \alpha_n(\alpha_n I + KK^*)^{-1}k_i^1, k_j^1\right\rangle$$

$$[f * K_z](z) = \int_{\mathbf{R}^p} f(u)h_z^{-p}K_z\left(h_z^{-1}(Z_i-u)\right)\mathrm{d}u$$

We decompose the first term further as $nI_n^0 = \zeta_n + nU_n$, where

$$\zeta_n = \frac{1}{n}\sum_{i=1}^{n}(U_i+\varphi_0(Z_i))\mu_0(Z_i)\left\|k_i^0\right\| \xrightarrow{a.s.} \mathbb{E}\left[(U_1+\varphi_0(Z_1))\mu_0(Z_1)\right]\|K_w\|$$

and

$$nU_n = \frac{2}{n}\sum_{i<j}\frac{1}{2}\left\{(U_i+\varphi_0(Z_i))\mu_0(Z_j)+(U_j+\varphi_0(Z_j))\mu_0(Z_i)\right\}\left\langle k_i^0, k_j^0\right\rangle$$

is a degenerate U-statistics, since $\mu_0, \varphi_0 \in \mathcal{N}(K)$, whence

$$\mathbb{E}_{X_2}\left[h^{\mu_0}(X_1,X_2)\right] = \frac{1}{2}\left\{(U_1+\varphi_0(Z_1))\left\langle k_1^0, \mathbb{E}\left[\mu_0(Z_2)k_2^0\right]\right\rangle + \mu_0(Z_1)\left\langle k_1^0, \mathbb{E}\left[(U_2+\varphi_0(Z_2))k_2^0\right]\right\rangle\right\} = 0$$

By the CLT for degenerate U-statistics, Gregory (1977)

$$nU_n \xrightarrow{d} \sum_{j=1}^{\infty}\lambda_j^{\mu_0}(\chi_{1,j}^2-1).$$

Similarly to the proof of Theorem 3, it is possible to show that after the normalization by $n\alpha_n$ all other terms tend to zero. □

Proof of Theorem 7.

$$\alpha_n n(\hat{\varphi}_{\alpha_n}-\varphi_1) = \left(I+\frac{1}{\alpha_n}\hat{P}_0\hat{K}^*\hat{K}\right)^{-1} n\hat{P}_0\hat{K}^*\hat{r}$$

We first show that if $n\alpha_n h_z^p \to \infty$, while the bandwidth h_w is fixed, then $\frac{1}{\alpha_n}\hat{P}_0\hat{K}^*\hat{K} \xrightarrow{p} 0$ in the operator norm. To that end, bounding operator norm by the Hilbert-Schmidt norm, we obtain

$$\begin{aligned}
\mathbb{E}\left\|\hat{K}^*\hat{K}\right\|^2 &\leq \mathbb{E}\left\|\int_{\mathbf{R}^p}\hat{f}_{ZW}(z_1,w)\hat{f}_{ZW}(z_2,w)\mathrm{d}w\right\|^2 \\
&= \mathbb{E}\left\|\frac{1}{n^2h_z^{2p}h_w^q}\sum_{i,j=1}^n K_z\left(h_z^{-1}(Z_i-z_1)\right)K_z\left(h_z^{-1}(Z_j-z_2)\right)\bar{K}\left(h_w^{-1}(W_i-W_j)\right)\right\|^2 \\
&\leq 2T_1+2T_2,
\end{aligned}$$

where the norm in the right-side is that of $L_2(\mathbf{R}^p\times\mathbf{R}^q,\mathrm{d}z_1\times\mathrm{d}z_2)$, $\bar{K}(x)=\int K_w(u)K_w(x-u)\mathrm{d}u$ is a convolution kernel (assuming that K_w is symmetric), and

$$\begin{aligned}
T_1 &= \mathbb{E}\left\|\frac{1}{n^2h_z^{2p}h_w^q}\sum_{i=1}^n K_z\left(h_z^{-1}(Z_i-z_1)\right)K_z\left(h_z^{-1}(Z_i-z_2)\right)\bar{K}(0)\right\|^2, \\
T_2 &= \mathbb{E}\left\|\frac{1}{n^2h_z^{2p}h_w^q}\sum_{i<j}\left\{K_z\left(h_z^{-1}(Z_i-z_1)\right)K_z\left(h_z^{-1}(Z_j-z_2)\right)+K_z\left(h_z^{-1}(Z_j-z_1)\right)K_z\left(h_z^{-1}(Z_i-z_2)\right)\right\}\bar{K}\left(h_w^{-1}(W_i-W_j)\right)\right\|^2.
\end{aligned}$$

The first term is treated as a sum of i.i.d. Hilbert space valued random elements

$$\begin{aligned}
T_1 &\lesssim \frac{1}{n^3}\mathbb{E}\left\|h_z^{-2p}K_z\left(h_z^{-1}(Z_i-z_1)\right)K_z\left(h_z^{-1}(Z_i-z_2)\right)\right\|^2+\frac{1}{n^2}\left\|h_z^{-2p}\mathbb{E}\left[K_z\left(h_z^{-1}(Z_i-z_1)\right)K_z\left(h_z^{-1}(Z_i-z_2)\right)\right]\right\|^2 \\
&= O\left(\frac{1}{n^3h_z^{2p}}+\frac{1}{n^2h_z^{2p}}\right),
\end{aligned}$$

where we assume that $K_z\in L_2(\mathbf{R})$ is symmetric.

The second term is treated as a degenerate Hilbert space valued U-statistics. To that end, using the moment inequality in (Korolyuk and Borovskich, 1994, Theorem 2.1.6), under the assumption that $\mathbb{E}[\phi(Z)|W]=0,\forall\phi\in L_Z^2$, we obtain

$$\begin{aligned}
T_2 &\lesssim \frac{1}{n^2}\mathbb{E}\left\|\frac{1}{h_z^{2p}h_w^q}K_z\left(h_z^{-1}(Z_1-z_1)\right)K_z\left(h_z^{-1}(Z_2-z_2)\right)\bar{K}\left(h_w^{-1}(W_1-W_2)\right)\right\|^2 \\
&= O\left(\frac{1}{n^2h_z^{2p}}\right),
\end{aligned}$$

assuming additionally that $\bar{K}\in L_\infty(\mathbf{R})$. Therefore, as $n\alpha_nh_z^p\to\infty$, $\frac{1}{\alpha_n}\left\|\hat{K}^*\hat{K}\right\|\xrightarrow{p}0$, whence by the continuous mapping theorem and by the Slutsky's theorem in metric spaces, Van Der Vaart and Wellner (2000) it is sufficient to analyze the weak convergence of

$$n\hat{P}_0\hat{K}^*\hat{r}=n(\hat{P}_0-P_0)\hat{K}^*\hat{r}+\frac{1}{nh_z^ph_w^q}\sum_{i,j}Y_iP_0K_z\left(h_z^{-1}(Z_j-z)\right)\bar{K}_w\left(h_w^{-1}(W_i-W_j)\right),$$

where the first term is negligible comparing to the second one. Therefore, putting

$$
\begin{aligned}
\left\langle n\hat{P}_0\hat{K}^*\hat{r},\mu\right\rangle &= \zeta_n + U_n + R_n + o_p(1),\\
\zeta_n &= \frac{1}{n}\sum_{i=1}^n Y_i\mu^0(Z_i)h_w^{-q}\bar{K}(0),\\
U_n &= \frac{2}{n}\sum_{i<j}\frac{1}{2}\left\{Y_i\mu^0(Z_j)+Y_j\mu^0(Z_i)\right\}h_w^{-q}\bar{K}\left(h_w^{-1}(W_i-W_j)\right),\\
R_n &= \frac{1}{nh_w^q}\sum_{i,j=1}^n Y_i\bar{K}\left(h_n^{-1}(W_i-W_j)\right)\left\{[K_z*\mu^0](Z_j)-\mu^0(Z_j)\right\},
\end{aligned}
$$

where $\mu^0 = P_0\mu$ and $[K_z*\mu^0](z) = h_n^{-p}\int K\left(h_z^{-1}(z-u)\right)\mu^0(u)\mathrm{d}u$. By the law of large numbers

$$
\zeta_n \xrightarrow{a.s} \mathbb{E}\left[Y_1\mu^0(Z_1)\right]h_w^{-q}\bar{K}(0).
$$

Since $\mathbb{E}[\phi(Z)|W]=0,\forall\phi\in L_Z^2$, U_n is a degenerate U-statistics. By the central limit theorem, Gregory (1977),

$$
U_n = \frac{2}{n}\sum_{i<j}h(X_i,X_j)\xrightarrow{d}\sum_{j=1}^\infty\lambda_j^\mu(\chi_{1,j}^2-1),
$$

Lastly, we show that $R_n\xrightarrow{p}0$. To that end, put $R_n = R_{1n}+R_{2n}$ with

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

My Markov's inequality it is sufficient to control the first or the second moment. Assuming that $\mathbb{E}[|Y||Z]<\infty$ a.s., that $f_Z\in H^s(\mathbf{R}^p)$ for some $s>0$, and that $\mu^0\in L_2(\mathbf{R}^p)\cap C(\mathbf{R}^p)$, by (Giné and Nickl, 2015, Lemma 4.3.18)

$$
\mathbb{E}|R_{1n}|\lesssim\int_{\mathbf{R}^p}\left|[K_z*\mu^0](z)-\mu^0(z)\right|f_Z(z)\mathrm{d}z = o(1).
$$

Similarly if $\mathbb{E}\left[|Y|^2|W\right]<\infty$ a.s. and $\bar{K}\in L_\infty$, by the moment inequality in (Korolyuk and Borovskich, 1994, Theorem 2.1.3)]

$$
\begin{aligned}
\mathbb{E}|R_{2n}|^2 &\lesssim \mathbb{E}\left|Y_1\left\{[K_z*\mu^0](Z_2)-\mu^0(Z_2)\right\}h_w^{-q}\bar{K}\left(h_w^{-1}(W_1-W_2)\right)\right|^2\\
&\lesssim\int_{\mathbf{R}^p}\left|[K_z*\mu^0](z)-\mu^0(z)\right|f_Z(z)\mathrm{d}z = o(1).
\end{aligned}
$$

□

Appendix C: CLT for degenerate U-statistics in Hilbert space

Gaussian random measures and Wiener-Itô integrals

Let $(\mathcal{X}, \mathscr{X}, P)$ be a probability measure space and H a separable Hilbert space. Let $L_2(\mathcal{X}^m, H)$ be the space of all functions $f : \mathcal{X}^m \to H$ such that $\mathbb{E}\|f(X_1, \ldots, X_m)\|^2 < \infty$. For $\mathscr{X}_P = \{A \in \mathscr{X} : P(A) < \infty\}$, the stochastic process $\{\mathbb{W}(A), A \in \mathscr{X}_P\}$ is called the *Gaussian random measure* if

1. for all $A \in \mathscr{X}_P$
$$\mathbb{W}(A) \sim N(0, P(A));$$
2. for any collection of disjoint sets $(A_k)_{k=1}^n$ in $\mathscr{X}_P$, $\mathbb{W}(A_k), k = 1, \ldots, n$ are independent and
$$\mathbb{W}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbb{W}(A_k).$$

Take a sequence of pairwise disjoint sets $(A_k)_{k=1}^n$ in $\mathscr{X}_P$ and let S_m be a set of simple functions of the form
$$f(x_1, \ldots, x_m) = \sum_{i_1,\ldots,i_m=1}^n c_{i_1,\ldots,i_m} \mathbb{1}_{A_{i_1}}(x_1) \times \cdots \times \mathbb{1}_{A_{i_m}}(x_m),$$
where $c_{i_1,\ldots,i_m}$ are zero if any of two indices in the set $i_1, \ldots, i_m$ are equal, i.e. f vanishes on the diagonal. For a Gaussian random measure $\mathbb{W}$ corresponding to P, we define the following random operator
$$S_m \ni f \mapsto J_m(f) = \sum_{i_1,\ldots,i_m}^n c_{i_1,\ldots,i_m} \mathbb{W}(A_{i_1}) \ldots \mathbb{W}(A_{i_m}) \in H.$$

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 3. *Let $B_t \in \mathbf{R}$ be a Brownian motion on $[0, \infty)$. Then for any $(t, s] \subset [0, \infty)$, we can define a Gaussian random measure $\mathbb{W}((t, s]) = B_s - B_t$ and a Wiener-Itô integral $J : L_2([0, \infty), \mathrm{d}t) \to \mathbf{R}$ as $J(f) = \int f(t)\mathrm{d}B_t$.*

Central limit theorem

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

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

Similarly to the real case if $\mathbb{E}h(x_1, X_2) = 0$, the U-statistics is called degenerate. The following result provides the limiting distribution of the degenerate H-valued U-statistics.

Theorem 10 (Borovskich (1986))**.** *Suppose that the kernel function h is such that $\mathbb{E}h(X_1, X_2) = 0$, $\mathbb{E}\|h(X_1, X_2)\|^2 < \infty$, and that the U-statistics is degenerate. Then*

$$nU_n \xrightarrow{d} J_2(h),$$

where $J_2(h) = \int_{\mathcal{X}\times\mathcal{X}} h(x_1, x_2)\mathbb{W}(\mathrm{d}x_1)\mathbb{W}(\mathrm{d}x_2)$.

Proof. See (Korolyuk and Borovskich, 1994, Theorem 4.10.2) for more general result for U-statistics of arbitrary degree. □

If $H = \mathbf{R}$, $(\varphi_j)_{j\ge1}$ is arbitrary orthonormal system in $L_2(\mathcal{X}, \mathbf{R})$ and ξ_i are i.i.d. $N(0, 1)$, the Wiener-Itô integral has the following representation

$$J_2(h) =_d \sum_{(i_1,i_2)=1}^{\infty} \mathbb{E}\left[h(X_1, X_2)\varphi_{i_1}(X_1)\varphi_{i_2}(X_2)\right] \left\{(\xi_{i_1}^2 - 1)\delta_{i_1,i_2} + \xi_{i_1}\xi_{i_2}(1 - \delta_{i_1,i_2})\right\},$$

This follows from the fact that multiple Wiener-Itô integrals have representation in terms of Hermite polynomials, see Itô (1951) and Korolyuk and Borovskich (1994). Alternatively, it is possible to show directly that the limiting distribution of the degenerate U-statistics of degree 2 is the expression in the right-side.

References

Andrii Babii. Honest confidence sets in nonparametric iv regression and other ill-posed models. *Working paper*, 2016a.

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

Anatolii Borisovich Bakushinskii. 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, 1967.

Yu. V. Borovskich. *Theory of U-statistics in Hilbert space (Russian)*, volume 86.78. Institute of Mathematics, Ukrain. Acad. of Sci., Kiev, 1986.

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

Victor De la Pena and Evarist Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

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

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

Jean-Pierre Florens, Michel Mouchart, Jean-Marie Rolin, et al. Elements of bayesian statistics. 1990.

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

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

Gavin G Gregory. Large sample theory for u-statistics and tests of fit. *The annals of statistics*, pages 110–123, 1977.

CW Groetsch. Uniform convergence of regularization methods for fredholm equations of the first kind. *Journal of the Australian Mathematical Society (Series A)*, 39(02):282–286, 1985.

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

Kiyosi Itô. Multiple wiener integral. *Journal of the Mathematical Society of Japan*, 3(1):157–169, 1951.

VK Ivanov. On fredholm integral equations of the first kind. *Differents. ur-niya*, 3(3):410–421, 1967.

Yu I Khudak. On the regularization of solutions of integral equations of the first kind. *USSR Computational Mathematics and Mathematical Physics*, 6(4):217–221, 1966.

V. S. Korolyuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273. Springer Science & Business Media, 1994.

Rainer Kress. *Linear integral equations*, volume 82. Springer, 2014.

Shuai Lu and Sergei V Pereverzev. *Regularization theory for ill-posed problems: selected topics*, volume 58. Walter de Gruyter, 2013.

David G Luenberger. *Optimization by vector space methods.* John Wiley & Sons, 1997.

Peter Mathé and Sergei V Pereverzev. Moduli of continuity for operator valued functions. 2002.

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

Walter Rudin. Functional analysis. international series in pure and applied mathematics, 1991.

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

Robert J Serfling. Approximation theorems of mathematical statistics. Technical report, 1980.

GM Vainikko and A Yu Veretennikov. Iteration procedures in ill-posed problems, 1986.

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