University of Mannheim / Department of Economics

Working Paper Series

Goodness-of-fit tests based on series estimators in nonparametric instrumental regression

Christoph Breunig

Working Paper 13-12

May 2012

Goodness-of-fit tests based on series estimators in nonparametric instrumental regression.*

Christoph Breunig⋆

Universität Mannheim

May 10, 2012

This paper proposes several tests of restricted specification in nonparametric instrumental regression. Based on series estimators, test statistics are established that allow for tests of the general model against a parametric or nonparametric specification as well as a test of exogeneity of the vector of regressors. The tests are asymptotically normally distributed under correct specification and consistent against any alternative model. Under a sequence of local alternative hypotheses, the asymptotic distribution of the tests is derived. Moreover, uniform consistency is established over a class of alternatives whose distance to the null hypothesis shrinks appropriately as the sample size increases.

Keywords: Nonparametric regression, instrument, linear operator, orthogonal series estimation, hypothesis testing, local alternative, uniform consistency.

JEL classification: C12, C14.

1. Introduction

While parametric instrumental variables estimators are widely used in econometrics, its nonparametric extension has not been introduced until the last decade. The study of nonparametric instrumental regression models was initiated by Darolles et al. [2011] and Newey and Powell [2003]. In these models, given a scalar dependent variable Y, a vector of regressors Z, and a vector of instrumental variables W, the structural function φ satisfies

$$Y = \varphi(Z) + U \quad \text{with} \quad \mathbb{E}[U|W] = 0 \tag{1.1}$$

for an error term U. Here, Z contains potentially endogenous entries, i.e., $\mathbb{E}[U|Z]$ may not be zero. Model (1.1) does not involve the *a priori* assumption that the structural

*This work was supported by the DFG-SNF research group FOR916. I thank Enno Mammen, James Stock, Jan Johannes, Federico Crudo, and Petyo Bonev for helpful comments.

⋆Lehrstuhl für Statistik, Abteilung Volkswirtschaftslehre, L7, 3-5, 68131 Mannheim, Germany, e-mail: cbreunig@staff.mail.uni-mannheim.de

function is known up to finitely many parameters. Hence, by considering a nonparametric model, we minimize the likelihood of misspecification. On the other hand, implementing the nonparametric instrumental regression model can be challenging.
Nonparametric instrumental regression models have attracted increasing attention in the econometric literature. For example, Ai and Chen [2003], Blundell et al. [2007], Chen and Reiß [2011] or Newey and Powell [2003] consider sieve minimum distance estimators of φ, while Darolles et al. [2011], Hall and Horowitz [2005], Gagliardini and Scaillet [2011] or Florens et al. [2011] study penalized least squares estimators. A linear Galerkin approach to construct an estimator of φ has been proposed by Johannes and Schwarz [2010]. When the methods of analysis are widened to include nonparametric techniques, one must confront two mayor challenges. First, identification in model (1.1) requires far stronger assumptions about the instrumental variables than for the parametric case (cf. Newey and Powell [2003]). Second, the accuracy of any estimator of φ can be low, even for large sample sizes. More precisely, Chen and Reiß [2011] showed that for a large class of joint distributions of (Z, W) only logarithmic rates of convergence can be obtained. The reason for this slow convergence is that model (1.1) leads to an inverse problem which is *ill posed* in general, i.e., the solution does not depend continuously on the data.
In light of the difficulties of estimating the nonparametric function φ in model (1.1), the need for statistically justified model simplifications is paramount. We do not face an ill posed inverse problem if a parametric structure of φ or exogeneity of Z can be justified. If these model simplifications are not supported by the data, one might still be interested in whether a smooth solution to model (1.1) exists and if some regressors could be omitted from the structural function φ. These model simplifications have important potential since they might increase the accuracy of estimators of φ or lower the required conditions imposed on the instrumental variables to ensure identification.
In this work we present a new family of goodness-of-fit statistics which allows for several restricted specification tests of the model (1.1). Our method can be used for testing either a parametric or nonparametric specification. In addition, we perform a test of exogeneity and of dimension reduction of the vector of regressors Z, i.e., whether certain regressors can be omitted from the structural function φ. By a withdrawal of regressors which are only weakly correlated with the instrument, identification in the restricted model might be possible although φ is not identified in the original model (1.1). Note that model (1.1) leads to the conditional moment equation $\mathbb{E}[Y - \varphi(Z)|W] = 0$. Multiplying both sides with an m-dimensional vector of functions $f_j(W)$, $1 \leqslant j \leqslant m$, and taking expectations leads to the unconditional moment equation $\mathbb{E}[(Y - \varphi(Z))f_j(W)] = 0$, $1 \leqslant j \leqslant m$. Our test statistic is based on the Euclidean norm of the vector $\mathbb{E}[(Y - \varphi_0(Z))f_j(W)] = 0$, $1 \leqslant j \leqslant m$, where the hypothesis function φ_0 is replaced by an estimator under the restriction. So, only the estimator of φ_0 differs for the different specification test considered in this paper. It is worth noting that by our methodology we can omit some assumptions typically found in related literature, such as smoothness conditions on the joint distribution of (Z, W).
There is a large literature concerning hypothesis testing of restricted specification of regression. In the context of conditional moment equation, Donald et al. [2003] and Tripathi and Kitamura [2003] make use of empirical likelihood methods to test parametric restrictions of the structural function. In addition, Santos [2012] allows for different hypothesis tests, such as a test of homogeneity. Based on kernel techniques, Horowitz [2006], Blundell and Horowitz [2007], and Horowitz [2011] propose test statistics in which an additional smoothing step (on the exogenous entries of Z) is carried out. Horowitz [2006] considers a parametric specification test. Blundell and Horowitz [2007] establish a consistent test of

exogeneity of the vector of regressors Z, whereas Horowitz [2011] tests whether the endogenous part of Z can be omitted from φ. Gagliardini and Scaillet [2007] and Horowitz [2012] develop nonparametric specification tests in an instrumental regression model. We like to emphasize that their test cannot be applied to model (1.1) where some entries of Z might be exogenous.

Our method is also applicable when an additional smoothing step is carried out. It is shown that the asymptotic behavior of our test relies crucially on the behavior of the smoothing operator. We study the power of the test against a sequence of alternatives that tend to zero at a certain rate as the sample size increases. If the eigenvalues of this operator have a sufficiently fast decay, then our test can detect linear alternatives at a distance of $n^{-1/2}$ (as in Horowitz [2006], Blundell and Horowitz [2007], Horowitz [2012], and Horowitz [2011]). In contrast, estimating the structural function φ nonparametrically leads to far slower polynomial or even logarithmic convergence rates (cf. Chen and Reiß [2011]). Applying the additional smoothing step, however, changes the function class over which uniform consistency can be obtained.

The paper is organized as follows. In Section 2 we start with a simple hypothesis test, i.e., whether φ coincides with a known function φ_0. We obtain asymptotic normality and consistency of our proposed test statistic. Moreover, we judge its power by considering linear local alternatives and establish uniform consistency over a class of functions. In Sections 3–6 we consider a parametric specification test, a test of exogeneity, and a nonparametric specification test. The goodness-of-fit statistics are obtained by replacing φ_0 in the statistic of Section 2 by an appropriate estimator. Under modified assumptions, the asymptotic results of Section 2 still remain valid. All proofs can be found in the appendix.

2. A simple hypothesis test

In this section we propose a goodness-of-fit statistic for testing the hypothesis $H_0 : \varphi = \varphi_0$, where φ_0 is a known function, against the alternative $\varphi \neq \varphi_0$. We develop a test statistic based on $\mathcal{L}^2$ distance. As we will see in the following chapters, it is sufficient to replace φ_0 by an appropriate estimator to allow for tests of the general model against other specifications. We first give basic assumptions, then obtain the asymptotic distribution of the proposed statistic, and further discuss its power and consistency properties.

2.1. Assumptions and notations.

The model revisited The nonparametric instrumental regression model (1.1) leads to a linear operator equation. To be more precise, let us introduce the conditional expectation operator $T\phi := \mathbb{E}[\phi(Z)|W]$ mapping $\mathcal{L}^2_Z = \{\phi : \mathbb{E}\,|\phi(Z)|^2 < \infty\}$ to $\mathcal{L}^2_W = \{\psi : \mathbb{E}\,|\psi(W)|^2 < \infty\}$ (which are endowed with the usual inner products $\langle \cdot, \cdot \rangle_Z$ and $\langle \cdot, \cdot \rangle_W$, respectively). Consequently, model (1.1) can be written as

$$g = T\varphi \tag{2.1}$$

where the function $g := \mathbb{E}[Y|W]$ belongs to $\mathcal{L}^2_W$ (which can be assured by assuming $\mathbb{E}[Y^2] < \infty$). Throughout the paper we assume that an iid. n-sample of (Y, Z, W) from the model (1.1) is available.

Moment assumptions. Let us introduce pre-specified orthonormal basis $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ in $\mathcal{L}^2_Z$ and $\mathcal{L}^2_W$, respectively. We need moment conditions on the basis, more specific, on the random variables $e_j(Z)$ and $f_l(W)$ for $j,l \geqslant 1$, which we summarize in the next assumption.

Assumption 1. *There exists some constant $\eta \geqslant 1$ such that*

(i) $\sup_{j\geqslant 1} \mathbb{E}\,|e_j(Z)|^4 \leqslant \eta^4$,

(ii) $\sup_{l\geqslant 1} \mathbb{E}\,|f_l(W)|^4 \leqslant \eta^4$.

Assumption 1 holds for sufficiently large η if the basis $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are uniformly bounded, such as trigonometric bases or B-splines that have been orthogonalized. Moreover, this assumption is satisfied by the Hermite polynomials.
The results derived below involve assumptions on the conditional moments of the random variables U given W gathered in the following assumption.

Assumption 2. *There exists $\sigma > 0$ such that $\mathbb{E}[U^4|W] \leqslant \sigma^4$.*

Mapping properties of the operators We will see below that the power of our test can be increased by carrying out an additional smoothing step. Therefore, we introduce the smoothing operator L on $\mathcal{L}^2_W$. In contrast to the unknown conditional expectation operator T, which has to be estimated, the operator L can be chosen by the statistician. The following assumption ensures identification of φ in the model (2.1).

Assumption 3. *The conditional expectation operator T is nonsingular.*

If Assumption 3 is violated we rather test of the operator equation $g = T\varphi_0$ and hence consider a conditional moment restriction test. We discuss the implications of our results also in this case. Let L have an eigenvalue decomposition given by $\{\tau_j^{1/2}, f_j\}_{j\geqslant 1}$. We allow in this paper for a wide range of smoothing operators. We also permit for L being the identity operator, i.e., no smoothing step is carried out. We only require the following condition on the operator L determined by the sequence of eigenvalues $\tau = (\tau_j)_{j\geqslant 1}$.

Assumption 4. *The weighting sequence τ is positive, nonincreasing, and satisfies $\tau_1 = 1$.*

Assumption 4 ensures that the operator L is nonsingular.

Remark 2.1. Horowitz [2006], Blundell and Horowitz [2007], and Horowitz [2011] consider as a smoothing operator a Fredholm integral operator, i.e., $L\phi(s) = \int_0^1 \ell(s,t)\phi(t)dt$ for some function $\phi \in \mathcal{L}^2[0,1]$ and some kernel function $\ell : [0,1]^2 \to \mathbb{R}$. In order to ensure $L\phi \in \mathcal{L}^2[0,1]$ it is typically assumed that $\int_0^1\int_0^1 |\ell(s,t)|^2 dsdt < \infty$. Let $\{\tau_j^{1/2}, f_j\}_{j\geqslant 1}$ be the eigenvalue decomposition of L. By Parseval's identity

$$\int_0^1\int_0^1 |\ell(s,t)|^2 dsdt = \int_0^1 \sum_{j=1}^\infty \tau_j |f_j(s)|^2 ds = \sum_{j=1}^\infty \tau_j$$

where the right hand side is only finite if the sequence τ decays sufficiently fast. □

Matrix and operator notations. Given $m \geqslant 1$, $\mathcal{E}_m$ and $\mathcal{F}_m$ denote the subspace of $\mathcal{L}^2_Z$ and $\mathcal{L}^2_W$ spanned by the functions $\{e_j\}_{j=1}^m$ and $\{f_l\}_{l=1}^m$, respectively. E_m and $E_m^\perp$ (resp. F_m and $F_m^\perp$) denote the orthogonal projections on $\mathcal{E}_m$ (resp. $\mathcal{F}_m$) and its orthogonal complement $\mathcal{E}_m^\perp$ (resp. $\mathcal{F}_m^\perp$), respectively. If we restrict a linear operator $K : \mathcal{L}^2_Z \to \mathcal{L}^2_W$ to an operator from $\mathcal{E}_m$ to $\mathcal{F}_m$, then it can be represented by a matrix $[K]_{\underline{m}}$ with entries $[K]_{l,j} = \langle Ke_j, f_l\rangle_W$

for $1 \leqslant j, l \leqslant m$. Its spectral norm is denoted by $\|[K]_{\underline{m}}\|$ and its transposed by $[K]_{\underline{m}}^t$. The adjoint operator of K is denoted by K^*. We write Id for the identity operator and ∇_v for the diagonal operator with singular value decomposition $\{v_j, e_j, f_j\}_{j\geqslant 1}$. Respectively, given functions $\phi \in \mathcal{L}_Z^2$ and $\psi \in \mathcal{L}_W^2$ we define by $[\phi]_{\underline{m}}$ and $[\psi]_{\underline{m}}$ m-dimensional vectors with entries $[\phi]_j = \langle \phi, e_j\rangle_Z$ and $[\psi]_l = \langle \psi, f_l\rangle_W$ for $1 \leqslant j, l \leqslant m$. Moreover, $e_{\underline{m}}(Z)$ and $f_{\underline{m}}(W)$ denote random vectors with entries $e_j(Z)$ and $f_j(W)$, $1 \leqslant j \leqslant m$, respectively. For any weighting sequence w we introduce vectors $e_{\underline{m}}^w(Z)$ and $f_{\underline{m}}^w(W)$ with entries $e_j^w(Z) = \sqrt{w_j}e_j(Z)$ and $f_j^w(W) = \sqrt{w_j}f_j(W)$, $1 \leqslant j \leqslant m$. In addition, the weighted norm is denoted by $\|\phi\|_w^2 = \sum_{j=1}^\infty w_j[\phi]_j^2$. In the following we write $a_n \lesssim b_n$ when there exists a generic constant $C > 0$ such that $a_n \leqslant C b_n$ for sufficiently large n and $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously.

2.2. The test statistic and its asymptotic distribution

Under Assumptions 3 and 4 the hypothesis H_0 is equivalent to $Lg = LT\varphi_0$. We project the function $L(g - T\varphi_0)$ on the finite dimensional subspace $\mathcal{F}_{m_n}$ for some integer m_n which tends to infinity as the sample size n increases to infinity. Then our test statistic is the empirical counterpart of $\|F_{m_n}L(g - T\varphi_0)\|_W^2$, i.e.,

$$S_n := \big\|n^{-1}\sum_{i=1}^n (Y_i - \varphi_0(Z_i)) f_{\underline{m_n}}^\tau(W_i)\big\|^2. \tag{2.2}$$

When no additional smoothing is carried out, i.e., $L = \mathrm{Id}$, then $\tau_j = 1$ for all $j \geqslant 1$. To achieve asymptotic normality we need to standardize our test statistic S_n by appropriate mean and variance, which we introduce in the following definition.

Definition 2.1. *For all $m \geqslant 1$ let Σ_m be the covariance matrix of the random vector $U f_{\underline{m}}^\tau(W)$ with entries $\varsigma_{jj'} = \mathbb{E}\left[U^2 f_j^\tau(W) f_{j'}^\tau(W)\right]$, $1 \leqslant j, j' \leqslant m$. Then the trace and the Frobenius norm of Σ_m are respectively denoted by*

$$\mu_m := \mu(\Sigma_m) := \sum_{j=1}^m \varsigma_{jj} \quad \text{and} \quad \varsigma_m := \varsigma(\Sigma_m) := \Big(\sum_{j,j'=1}^m \varsigma_{jj'}^2\Big)^{1/2}.$$

In addition, if $\varsigma_m = O(1)$ as $m \to \infty$ we define

$$\mathcal{V} := \mathcal{V}(\Sigma_m) := 1 + \frac{4}{3\varsigma_\infty^4} \sum_{j,j',l,l'=1}^\infty \varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}$$

where $\varsigma_\infty = \lim_{m\to\infty} \varsigma(\Sigma_m)$.

Indeed the next result shows that S_n after standardization is asymptotically normally distributed if m_n increases appropriately as the the sample size n tends to infinity.

Theorem 2.1. *Let Assumptions 1–4 hold true. If m_n satisfies*

$$\varsigma_{m_n}^{-1} = o(1) \quad \text{and} \quad m_n \Big(\sum_{j=1}^{m_n} \tau_j^2\Big)^2 = o(n) \tag{2.3}$$

then we have for all $\varphi \in \mathcal{L}_Z^2$ under H_0

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n S_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

Remark 2.2. If there exists some constant $\sigma_o > 0$ such that $\mathbb{E}[U^2|W] \geqslant \sigma_o^2$ then we have $\varsigma_{m_n}^2 \geqslant \sigma_o^4 \sum_{j=1}^{m_n} \tau_j^2$. Thus, condition $\varsigma_{m_n}^{-1} = o(1)$ is satisfied for any positive sequence τ such that $\sum_{j=1}^{m_n} \tau_j^2$ is unbounded as n increases. When no additional smoothing is carried out, i.e., $L = \mathrm{Id}$, then condition (2.3) holds if $\mathbb{E}[U^2|W] \geqslant \sigma_o^2$ and $m_n^3 = o(n)$. Moreover, from condition (2.3) we see that by the choice of a stronger decaying sequence τ the parameter m_n may be chosen larger. From the following theorem we see that if $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$ only $m_n = o(n)$ is required. □

In the following result we establish asymptotic normality of our test when the sequence of weights τ may have a stronger decay than in Theorem 2.1, i.e., we consider the case where τ satisfies $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$. This condition together with the Assumption 2 implies $\varsigma_{m_n} \leqslant \sigma^4 \sum_{j=1}^{m_n} \tau_j^2 = O(1)$ in contrast to condition (2.3) in Theorem 2.1. Still asymptotic normality can be obtained, but an additional additive term occurs in the variance.

Theorem 2.2. *Let Assumptions 1–4 hold true. If m_n satisfies*

$$\sum_{j=1}^{m_n} \tau_j^2 = O(1) \quad \text{and} \quad m_n = o(n) \tag{2.4}$$

then for all $\varphi \in \mathcal{L}_Z^2$ under H_0 we have

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\, S_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}) \quad \text{as } n \to \infty.$$

Remark 2.3. Theorem 2.1 and 2.2 continue to hold if we replace ς_{m_n}, μ_{m_n}, and $\mathcal{V}$ by the estimators $\widehat{\varsigma}_{m_n}^2 := \sum_{j,j'=1}^{m_n} \widehat{\varsigma}_{jj'}^2$, $\widehat{\mu}_{m_n} := \sum_{j=1}^{m_n} \widehat{\varsigma}_{jj}$, and $\widehat{\mathcal{V}}_{m_n} = 1{+}4\sum_{j,j',l,l'=1}^{m_n} \widehat{\varsigma}_{jj'}\widehat{\varsigma}_{ll'}\widehat{\varsigma}_{jl}\widehat{\varsigma}_{j'l'}/(3\widehat{\varsigma}_{m_n}^4)$, respectively, where

$$\widehat{\varsigma}_{jj'} := n^{-1}\sum_{i=1}^{n} \sqrt{\tau_j \tau_{j'}}|Y_i - \varphi_0(Z_i)|^2 f_j(W_i) f_{j'}(W_i).$$

In the following sections where φ_0 is unknown and has to be estimated we might simply replace φ_0 in $\widehat{\varsigma}_{jj'}$ by the proposed estimators. □

2.3. Limiting behavior under local alternatives.

Let us study the power of the test, i.e., the probability to reject a false hypothesis, against a sequence of linear local alternatives that tends to zero as $n \to \infty$. It is shown that the power of our tests essentially relies on the choice of the weighting sequence τ.
Let us start with the case $\varsigma_{m_n}^{-1} = o(1)$. We consider the following sequence of linear local alternatives

$$Y = \varphi_0(Z) + \varsigma_{m_n}^{1/2} n^{-1/2} \delta(Z) + U \tag{2.5}$$

for some function $\delta \in L_Z^4 := \{\phi : \mathbb{E}\,|\phi(Z)|^4 < \infty\}$. The next result establishes asymptotic normality for the standardized test statistic S_n.

Proposition 2.3. *Given the conditions of Theorem 2.1 it holds under (2.5)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\, S_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}\big(2^{-1/2}\|T\delta\|_\tau^2, 1\big) \quad \text{as } n \to \infty.$$

As we see below the test statistic S_n has power advantages if $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$. Let us consider the sequence of linear local alternatives

$$Y = \varphi_0(Z) + n^{-1/2}\delta(Z) + U \tag{2.6}$$

for some function $\delta \in L_Z^4$.

Proposition 2.4. *Given the conditions of Theorem 2.2 it holds under (2.6)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\,S_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma_\infty)^{-1}\|T\delta\|_\tau^2, \mathcal{V}\big) \quad \textit{as } n \to \infty.$$

Remark 2.4. Under homoscedasticity, i.e., $\mathbb{E}[U^2|W] = \sigma_o^2$, we see from Proposition 2.3 that our test can detect linear alternatives at a rate $\big(\sum_{j=1}^{m_n}\tau_j^2\big)^{1/4}n^{-1/2}$. On the other hand, if $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ then S_n can detect local linear alternatives at a rate $n^{-1/2}$. But still our test with $L = \mathrm{Id}$ can have better power against certain smooth classes of alternatives as illustrated by Hong and White [1995] and Horowitz and Spokoiny [2001]. Indeed, in the next subsection we show that additional smoothing changes the class of alternatives over which uniform consistency can be obtained. □

2.4. Consistency

In this subsection we establish consistency against a fixed alternative and uniform consistency of our test over appropriate function classes. Let us first consider the case of a fixed alternative. We assume that H_0 does not hold, i.e., $\mathbb{P}(\varphi = \varphi_0) < 1$. The following proposition shows that our test has the ability to reject a false null hypothesis with probability 1 as the sample size grows to infinity.

Proposition 2.5. *Assume that H_0 does not hold. Let $\mathbb{E}\,|Y - \varphi_0(Z)|^4 < \infty$. Consider the sequence λ satisfying $\lambda_n = o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 2.1 or 2.2 we have*

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(nS_n - \mu_{m_n}\big) > \lambda_n\Big) = 1 + o(1).$$

In the following we show that our tests are consistent uniformly over the function class

$$\mathcal{G}_n^\rho = \Big\{\varphi \in \mathcal{L}_Z^2 : \|F_{m_n}T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n}n^{-1} \text{ and } \sup_{z\in\mathrm{Supp}(Z)}|\varphi(z) - \varphi_0(z)|^2 \leqslant \rho\Big\}$$

where $\mathrm{Supp}(Z)$ denotes the support of Z. Clearly, if H_0 is false then $\|F_{m_n}T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n}n^{-1}$ for n sufficiently large. By Assumption 4 the sequence τ is nonincreasing sequence with $\tau_1 = 1$ and hence $\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2 \leqslant \|T(\varphi-\varphi_0)\|_W^2 \leqslant \|\varphi-\varphi_0\|_Z^2$ by Jensen's inequality. We conclude that $\mathcal{G}_n^\rho$ contains all functions whose L_Z^2-distance to the structural function φ is at least $n^{-1}\varsigma_{m_n}$ within a constant. If the coefficients $[T(\varphi-\varphi_0)]_j$ fluctuate for large j then φ does not belong to $\mathcal{G}_n^\rho$ if the decay of τ is too strong. On the other hand, if $[T(\varphi-\varphi_0)]_j$ is sufficiently small for j up to a finite constant than φ does not necessarily belong to $\mathcal{G}_n^\rho$ with τ having a slow decay. For the next result let $q_{1-\alpha}$ denote the $1-\alpha$ quantile of $\mathcal{N}(0,1)$ in case of $\varsigma_{m_n}^{-1} = o(1)$ or $\mathcal{N}(0,\mathcal{V})$ in case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$.

Proposition 2.6. *Under the conditions of Theorem 2.1 or 2.2 we have for any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ that*

$$\lim_{n\to\infty}\inf_{\varphi\in\mathcal{G}_n^\rho}\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\,S_n - \mu_{m_n}\big) > q_{1-\alpha}\Big) \geqslant 1 - \varepsilon.$$

3. A parametric specification test

The method of orthogonal series estimation involves the choice of basis functions. Thereby, the natural question arises whether, given the specified basis $\{e_j\}_{j\geqslant 1}$, a say k dimensional vector of generalized Fourier coefficients is sufficient to develop the function φ. Let $\{e_j\}_{j\geqslant 1}$ satisfy Assumption 1. Then we consider the hypothesis $H_{\mathrm{p}} : \varphi = \varphi_0$ where in this section $\varphi_0 = \sum_{j=1}^{k}[\varphi_0]_j\, e_j$. The alternative hypothesis is that $\varphi \notin \mathcal{E}_k$. Under H_p standard parametric estimation techniques can be used to estimate the unknown k dimensional vector of coefficients $[\varphi_0]_{\underline{k}}$.

3.1. The test statistic and its asymptotic distribution

Let $\breve{\varphi}_k$ be an estimator of the parametric function φ_0. The k dimensional vector of generalized Fourier coefficients of φ_0 can be estimated $\sqrt{n}$-consistently by applying, for example, the generalized method of moments. Thereby, we may assume $\|\breve{\varphi}_k - \varphi_0\|_Z = O_p(n^{-1/2})$. We obtain our test statistic by replacing φ_0 in the definition of S_n given in (2.2) by the estimator $\breve{\varphi}_k$, i.e.,

$$S_n^{\mathrm{p}} := \Big\|n^{-1}\sum_{i=1}^{n}\big(Y_i - \breve{\varphi}_k(Z_i)\big)f_{\underline{m_n}}^{\tau}(W_i)\Big\|^2.$$

The following proposition establishes asymptotic normality of S_n^{p} after standardization given the same conditions as Theorem 2.1 where the function φ_0 was assumed to be known.

Theorem 3.1. *Let $\breve{\varphi}_k$ be an estimator of φ_0 satisfying $\|\breve{\varphi}_k - \varphi_0\|_Z = O_p(n^{-1/2})$. Then given the conditions of Theorem 2.1 it holds under H_p*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\, S_n^{p} - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{as } n \to \infty.$$

In the following theorem we state an asymptotic distribution result for S_n^{p} when $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$. In this case, we assume that the estimator $\breve{\varphi}_k$ satisfies

$$\sqrt{n}([\breve{\varphi}_k]_{\underline{k}} - [\varphi_0]_{\underline{k}}) = n^{-1/2}\sum_{i=1}^{n}h_{\underline{k}}(V_i) + o_p(1) \tag{3.1}$$

where $V_i := (Y_i, Z_i, W_i, \varphi_0)$ and $h_{\underline{k}}(V_i) = (h_1(V_i), \dots, h_k(V_i))$ where h_j, $1 \leqslant j \leqslant k$, are real valued functions. It is well known that this representation holds if $[\breve{\varphi}_k]_{\underline{k}}$ is the generalized method of moments estimator. In case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ we have to modify the standardization of the statistic S_n as follows. For $m \geqslant 1$ let Σ_m^{p} denote the covariance matrix of the centered random vector $U f_{\underline{m}}^{\tau}(W) + \mathbb{E}[f_{\underline{m}}^{\tau}(W)e_{\underline{k}}(Z)^t]h_{\underline{k}}(V)$. Then we define $\varsigma_m^{\mathrm{p}} = \varsigma(\Sigma_m^{\mathrm{p}})$, $\mu_m^{\mathrm{p}} = \mu(\Sigma_m^{\mathrm{p}})$, and $\mathcal{V}^{\mathrm{p}} = \mathcal{V}(\Sigma_m^{\mathrm{p}})$ where $\varsigma(\cdot)$, $\mu(\cdot)$, and $\mathcal{V}(\cdot)$ were introduced in Definition 2.1. Clearly, $\varsigma_{m_n}^{\mathrm{p}} = O(1)$.

Theorem 3.2. *Let $\breve{\varphi}_k$ be an estimator of φ_0 satisfying condition (3.1) with $\mathbb{E}\, h_j(V) = 0$ and $\mathbb{E}\,|h_j(V)|^4 < \infty$ for $1 \leqslant j \leqslant k$. Then given the conditions of Theorem 2.2 it holds under H_p*

$$(\sqrt{2}\varsigma_{m_n}^{p})^{-1}\big(n\, S_n^{p} - \mu_{m_n}^{p}\big) \xrightarrow{d} \mathcal{N}(0,\mathcal{V}^{p}) \qquad \text{as } n \to \infty.$$

Remark 3.1. Santos [2012] gave examples when only partial identification in the nonparametric model (1.1) is possible. We like to emphasize that the asymptotic results remain valid if Assumption 4 is not satisfied, i.e., T is singular, as long as we replace H_0 by the hypothesis $g = T\varphi_0$. This test of conditional moment restriction has also been considered by Donald et al. [2003], Tripathi and Kitamura [2003] and Santos [2012]. □

3.2. Limiting behavior under local alternatives and consistency.

In this section we study the power of the test, i.e., the probability to reject a false hypothesis, against a sequence of linear local alternatives that tends to zero as $n \to \infty$. Moreover, we establish consistency and uniform consistency over appropriate function classes of our tests.

Proposition 3.3. *Given the conditions of Theorem 3.1 it holds under (2.5)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\,S_n^p - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}\big(2^{-1/2}\|T\delta\|_\tau^2, 1\big) \quad \text{as } n \to \infty.$$

Given the conditions of Theorem 3.2 it holds under (2.6)

$$(\sqrt{2}\varsigma_{m_n}^p)^{-1}\big(n\,S_n^p - \mu_{m_n}^p\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma_\infty^p)^{-1}\|T\delta\|_\tau^2, \mathcal{V}^p\big) \quad \text{as } n \to \infty.$$

Remark 3.2. Under homoscedasticity, i.e., $\mathbb{E}[U^2|W] = \sigma_o^2$, and $L = \mathrm{Id}$ we see from Proposition 3.3 that our test has the same power properties as the test of Hong and White [1995]. On the other hand, $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ then our test can detect local linear alternatives at a rate $n^{-1/2}$ which coincides with the findings of Horowitz [2006]. □

The next proposition establishes consistency of our test against a fixed alternative model. It is assumed that H_p is false, i.e., $\mathbb{P}(\varphi \in \mathcal{E}_k) < 1$.

Proposition 3.4. *Assume that H_p does not hold. Let $\mathbb{E}\,|Y - \varphi_0(Z)|^4 < \infty$. Consider the sequence λ satisfying $\lambda_n = o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 3.1 we have*

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(nS_n^p - \mu_{m_n}\big) > \lambda_n\Big) = 1 + o(1)$$

Given the conditions of Theorem 3.2 it holds

$$\mathbb{P}\Big((\sqrt{2}\varsigma_{m_n}^p\mathcal{V}^p)^{-1}\big(n\,S_n^p - \mu_{m_n}^p\big) > \lambda_n\Big) = 1 + o(1).$$

In the following we show that our tests are consistent uniformly over the function class

$$\mathcal{H}_n^\rho = \Big\{\varphi \in \mathcal{L}_Z^2 : \inf_{\varphi_0 \in \mathcal{E}_k} \|F_{m_n}T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n}n^{-1} \text{ and } \sup_{z \in \mathrm{Supp}(Z)} |\varphi(z)| \leqslant \rho\Big\}.$$

If H_p is false then $\inf_{\varphi_0 \in \mathcal{E}_k}\|F_{m_n}T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n}n^{-1}$ for n sufficiently large. Similarly as in the previous section it can be seen that on $\mathcal{H}_n^\rho$ it holds $\inf_{\varphi_0 \in \mathcal{E}_k}\|F_{m_n}T(\varphi - \varphi_0)\|_\tau^2 \leqslant \inf_{\varphi_0 \in \mathcal{E}_k}\|\varphi - \varphi_0\|_Z^2$. Hence, $\mathcal{H}_n^\rho$ only contains functions whose L_Z^2 distance to any function in $\mathcal{E}_k$ is at least $\varsigma_{m_n}n^{-1}$ within a constant. In the next result, $q_{1-\alpha}$ denotes the $1-\alpha$ quantile of standard normal distribution.

Proposition 3.5. *Let $\sup_{j \geqslant 1}\mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$. For any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ we have under the conditions of Theorem 3.1 that*

$$\lim_{n\to\infty}\inf_{\varphi \in \mathcal{H}_n^\rho}\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\,S_n^p - \mu_{m_n}\big) > q_{1-\alpha}\Big) \geqslant 1 - \varepsilon,$$

whereas under the conditions of Theorem 3.2 it holds

$$\lim_{n\to\infty}\inf_{\varphi \in \mathcal{H}_n^\rho}\mathbb{P}\Big((\sqrt{2}\varsigma_{m_n}^p\mathcal{V}^p)^{-1}\big(n\,S_n^p - \mu_{m_n}^p\big) > q_{1-\alpha}\Big) \geqslant 1 - \varepsilon.$$

4. A test of exogeneity

Endogeneity of regressors is a common problem in econometric applications. Falsely assuming exogeneity of the regressors leads to inconsistent estimators and moreover, treating exogenous regressors as if they were endogenous can lower the accuracy of estimation dramatically. In this section we propose a test whether the vector of regressors Z is exogenous, i.e., $\mathbb{E}[U|Z]=0$. In this section let $\varphi_0(Z)=\mathbb{E}[Y|Z]$ then the hypothesis under consideration is given by $H_{\mathrm{e}}:\varphi=\varphi_0$. The alternative hypothesis is that $\varphi\neq\varphi_0$.

4.1. The test statistic and its asymptotic distribution

Since the eigenvalues of $[\widehat{\mathrm{Id}}]_{\underline{k}}:=n^{-1}\sum_{i=1}^n e_{\underline{k}}(Z_i)e_{\underline{k}}(Z_i)^t$ could be arbitrarily close to zero we propose the following least square estimator of φ_0 with additional thresholding, i.e.,

$$\overline{\varphi}_k(\cdot):=\begin{cases} e_{\underline{k}}(Z)^t[\widehat{\mathrm{Id}}]_{\underline{k}}^{-1}\frac{1}{n}\sum_{i=1}^n Y_i e_{\underline{k}}(Z_i), & \text{if } [\widehat{\mathrm{Id}}]_{\underline{k}} \text{ is nonsingular, } \|[\widehat{\mathrm{Id}}]_{\underline{k}}^{-1}\|\leqslant 2,\\ 0, & \text{otherwise.}\end{cases} \tag{4.1}$$

In contrast to the parametric case we need to allow for k tending to infinity as $n\to\infty$ in order to ensure consistency of the estimator $\overline{\varphi}_k$. The proposed goodness-of-fit statistic is given by S_n introduced in (2.2) where φ_0 is replaced by $\overline{\varphi}_{k_n}$, i.e.,

$$S_n^{\mathrm{e}}=\|n^{-1}\sum_{i=1}^n(Y_i-\overline{\varphi}_{k_n}(Z_i))f_{\underline{m_n}}^{\tau}(W_i)\|^2$$

where k_n and m_n tend to infinity as $n\to\infty$. Moreover, as typically in nonparametric statistics it is necessary to make some *a priori* assumption on the unknown function φ_0. Let $\gamma=(\gamma_j)_{j\geqslant1}$ be a nondecreasing sequence with $\gamma_1=1$. We assume that φ_0 belongs to the ellipsoid $\mathcal{F}_\gamma^\rho:=\big\{\phi\in\mathcal{F}_\gamma:\|\phi\|_\gamma^2=\sum_{j\geqslant1}\gamma_j[\phi]_j^2\leqslant\rho\big\}$ for some constant ρ where $\mathcal{F}_\gamma$ denotes the completion of $\mathcal{L}_Z^2$ with respect to the norm $\|\cdot\|_\gamma$. Roughly speaking, the sequence of weights γ measures the quality of approximation of φ_0 given the pre-specified basis $\{e_j\}_{j\geqslant1}$.

Theorem 4.1. *Let Assumptions 1–4 be satisfied. In addition assume $\mathbb{E}|Y|^4<\infty$. Let $\varphi_0\in\mathcal{F}_\gamma^\rho$ with γ satisfying $j^2=o(\gamma_j)$. If*

$$n=o(\gamma_{k_n}\varsigma_{m_n}),\ k_n=o(\varsigma_{m_n}),\quad \textit{and}\quad m_n\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2=o(n) \tag{4.2}$$

then under H_e it holds

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\,S_n^e-\mu_{m_n}\big)\xrightarrow{d}\mathcal{N}(0,1)\qquad \textit{as } n\to\infty.$$

Remark 4.1. If Z and W are uniformly distributed on $[0,1]$ and a trigonometric basis is considered then φ_0 belongs to $\mathcal{F}_\gamma^\rho$ with $j^2=o(\gamma_j)$ only if it is differentiable. In contrast to Blundell and Horowitz [2007] no smoothness assumptions on the joint distribution of (Z,W) is required here. □

Example 4.1. Let Z be continuously distributed with $\dim(Z)=r$ and set $L=\mathrm{Id}$. Consider the polynomial case where $\gamma_j\sim j^{2p/r}$ with $p>1$ and let $m_n\sim n^\nu$ with $0<\nu<1/3$. If $\mathbb{E}[U^2|W]\geqslant\sigma_o^2>0$ then condition (4.2) is satisfied if $k_n\sim n^\kappa$ with

$$r(1-\nu/2)/(2p)<\kappa<\nu/2. \tag{4.3}$$

Note that condition (4.3) requires $p>r(2-q)/(2q)$. □

The next result states an asymptotic distribution result for the statistic S_n^{e} after standardization if $\sum_{j=1}^{m_n}\tau_j^2=O(1)$. Let us denote $U_Z:=Y-E_{k_n}\varphi_0(Z)$ then clearly $\mathbb{E}[e_j(Z)U_Z]=0$ for all $1\leqslant j\leqslant k_n$. Let $\Sigma_{m_n}^{\mathrm{e}}$ be the covariance matrix of the centered vector $U\underline{f_{m_n}^\tau}(W)-U_Z\,\mathbb{E}[\underline{f_{m_n}^\tau}(W)\underline{e_{k_n}}(Z)^t]\underline{e_{k_n}}(Z)$. Then we define $\varsigma_{m_n}^{\mathrm{e}}=\varsigma(\Sigma_{m_n}^{\mathrm{e}})$, $\mu_{m_n}^{\mathrm{e}}=\mu(\Sigma_{m_n}^{\mathrm{e}})$, and $\mathcal{V}^{\mathrm{e}}=\mathcal{V}(\Sigma_{m_n}^{\mathrm{e}})$ where $\varsigma(\cdot)$, $\mu(\cdot)$, and $\mathcal{V}(\cdot)$ are given in Definition 2.1. By imposing moment conditions on U_Z we show in the proof of the following theorem that $\varsigma_{m_n}^{\mathrm{e}}=O(1)$.

Theorem 4.2. *Let Assumptions 1–4 be satisfied. In addition assume $\mathbb{E}[U_Z^4|Z]\leqslant\sigma^4$. Let $\varphi_0\in\mathcal{F}_\gamma^\rho$ with γ satisfying $\gamma_j=o(j^2)$. If*

$$\sum_{j=1}^{m_n}\tau_j^2=O(1),\quad n=O(\gamma_{k_n}),\ \text{and}\ m_nk_n^2=o(n) \tag{4.4}$$

then under H_e it holds

$$(\sqrt{2}\varsigma_{m_n}^e)^{-1}\big(n\,S_n^e-\mu_{m_n}^e\big)\xrightarrow{d}\mathcal{N}(0,\mathcal{V}^e)\qquad\text{as }n\to\infty.$$

Example 4.2. Let Z and γ be as in in Example 4.1. Here, the eigenvalues of L satisfy $\tau_j\sim j^{-2}$. Condition (4.4) is satisfied if $m_n\sim n^\nu$ with $0<\nu<1/2$ and $k_n\sim n^\kappa$ with $r/(2p)<\kappa<(1-\nu)/2$. □

Remark 4.2. If T is singular the asymptotic results of Theorem 4.1 and 4.2 still remain valid if H_{e} is replaced by the hypothesis $\mathbb{E}[\varphi(Z)-\mathbb{E}[Y|Z]|W]=0$. In this case, however, $\mathbb{E}[\varphi(Z)-\mathbb{E}[Y|Z]|W]$ might be zero even if Z is endogenous. On the other hand, if $\mathbb{E}[\varphi(Z)-\mathbb{E}[Y|Z]|W]\neq0$ then Z cannot be exogenous. □

4.2. Limiting behavior under local alternatives and consistency.

Similar to the previous sections we study the power and consistency properties of our test.

Proposition 4.3. *Given the conditions of Theorem 4.1 with $\mathbb{E}[Y^2|Z]\leqslant\eta^2$ for some constant $\eta>1$ it holds under (2.5)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\,S_n^e-\mu_{m_n}\big)\xrightarrow{d}\mathcal{N}\big(2^{-1/2}\|T\delta\|_\tau^2,1\big)\qquad\text{as }n\to\infty.$$

Given the conditions of Theorem 4.2 it holds under (2.6)

$$(\sqrt{2}\varsigma_{m_n}^e)^{-1}\big(n\,S_n^e-\mu_{m_n}^e\big)\xrightarrow{d}\mathcal{N}\big((\sqrt{2}\varsigma_\infty^e)\|T\delta\|_\tau^2,\mathcal{V}^e\big)\qquad\text{as }n\to\infty.$$

Let us now establish consistency of our tests when H_{e} does not hold, i.e., $\mathbb{P}\big(\varphi=\varphi_0\big)<1$.

Proposition 4.4. *Assume that H_e does not hold. Let $\mathbb{E}\,|Y-\varphi_0(Z)|^4<\infty$. Consider the sequence λ satisfying $\lambda_n=o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 4.1 we have*

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(nS_n^e-\mu_{m_n}\big)>\lambda_n\Big)=1+o(1),$$

whereas in the setting of Theorem 4.2

$$\mathbb{P}\Big((\sqrt{2}\varsigma_{m_n}^e\mathcal{V}^e)^{-1}\big(n\,S_n^e-\mu_{m_n}^e\big)>\lambda_n\Big)=1+o(1).$$

In the following we show that our tests are consistent uniformly over the function classes where, in contrast to the previous sections, regularity conditions on the function φ are imposed. More precisely, we consider the class

$$\mathcal{I}_n^\rho = \Big\{\varphi \in \mathcal{F}_\gamma^\rho : \|F_{m_n} T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\, \varsigma_{m_n} n^{-1}\Big\}.$$

Again, $q_{1-\alpha}$ denotes the $1-\alpha$ quantile of $\mathcal{N}(0,1)$.

Proposition 4.5. *Let $\sup_{j\geqslant 1} \mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$. Under the conditions of Theorem 4.1 we have for any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ that*

$$\lim_{n\to\infty} \inf_{\varphi\in\mathcal{I}_n^\rho} \mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\, S_n^e - \mu_{m_n}\big) > q_{1-\alpha}\Big) \geqslant 1-\varepsilon,$$

whereas under the conditions of Theorem 4.2 it holds

$$\lim_{n\to\infty} \inf_{\varphi\in\mathcal{I}_n^\rho} \mathbb{P}\Big((\sqrt{2}\varsigma_{m_n}^e \mathcal{V}^e)^{-1}\big(n\, S_n^e - \mu_{m_n}^e\big) > q_{1-\alpha}\Big) \geqslant 1-\varepsilon.$$

5. A nonparametric specification test

A solution to the linear operator equation (2.1) only exists if g belongs to the range of T. We refer to Gagliardini and Scaillet [2007] for a detailed discussion when existence of a solution to (2.1) fails. In many econometric applications the function of interest is smooth, i.e., belongs to some function class $\mathcal{F}_\gamma^\rho$ with γ being an increasing sequence of weights. We consider the hypothesis

$H_{\rm np}$: there exists a solution $\varphi_0 \in \mathcal{F}_\gamma^\rho$ to (2.1) for some $\rho > 0$ and γ with $k_n^3 = O(\gamma_{k_n})$.

The alternative hypothesis is that there exists no function in $\mathcal{F}_\gamma^\rho$ that solves (2.1) for any constant $\rho > 0$ and any sequence γ satisfying $k_n^3 = O(\gamma_{k_n})$. In addition, we see in this section that our results allow for a test of dimension reduction of the vector of regressors Z, i.e., whether some regressors can be omitted from the structural function φ. This generalizes the result of Horowitz [2011] who tests whether the endogenous part of Z can be omitted from φ. As we point out further, by omitting regressors that are only weakly correlated to the instrument identification in the restricted model might be obtained.

5.1. Nonparametric estimation method

The nonparametric estimator. Since $[T]_{\underline{k}} = \mathbb{E}\, f_{\underline{k}}(W) e_{\underline{k}}(Z)^t$ and $[g]_{\underline{k}} = \mathbb{E}\, Y f_{\underline{k}}(W)$ we construct estimators by using their empirical counterparts, i.e.,

$$[\widehat{T}]_{\underline{k}} := \frac{1}{n}\sum_{i=1}^n f_{\underline{k}}(W_i) e_{\underline{k}}(Z_i)^t \quad \text{ and } \quad [\widehat{g}]_{\underline{k}} := \frac{1}{n}\sum_{i=1}^n Y_i f_{\underline{k}}(W_i).$$

Throughout this section $[T]_{\underline{k}}$ is assumed to be nonsingular for k sufficiently large, so that its inverse $[T]_{\underline{k}}^{-1}$ exists. Then the orthogonal series type estimator of Johannes and Schwarz [2010] for the structural function φ is defined for all $k \geqslant 1$ by

$$\widehat{\varphi}_k(\cdot) := \begin{cases} e_{\underline{k}}(\cdot)^t [\widehat{T}]_{\underline{k}}^{-1} [\widehat{g}]_{\underline{k}}, & \text{if } [\widehat{T}]_{\underline{k}} \text{ is nonsingular and } \|[\widehat{T}]_{\underline{k}}^{-1}\| \leqslant \sqrt{n}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Additional assumptions. As usual in the context of ill-posed inverse problems, we specify some mapping properties of the operator under consideration. Denote by $\mathcal{T}$ the set of all nonsingular compact operators mapping $\mathcal{L}^2_Z$ into $\mathcal{L}^2_W$. Given a sequence of weights $\upsilon := (\upsilon_j)_{j\geqslant 1}$ and $d \geqslant 1$ we define the subset $\mathcal{T}_d^\upsilon$ of $\mathcal{T}$ by

$$\mathcal{T}_d^\upsilon := \Big\{T \in \mathcal{T} : \quad \|\phi\|_\upsilon^2/d \leqslant \|T\phi\|_W^2 \leqslant d\,\|\phi\|_\upsilon^2 \quad \text{for all } \phi \in \mathcal{L}^2_Z\Big\}. \tag{5.2}$$

Notice that for all $T \in \mathcal{T}_d^\upsilon$ it follows that $\|Te_j\|_W^2 \sim \upsilon_j$. In what follows, we introduce a stronger condition which involves the basis $\{f_l\}_{l\geqslant 1}$ in $\mathcal{L}^2_W$ and thus, extends the link condition $T \in \mathcal{T}$. We denote by $\mathcal{T}_{d,D}^\upsilon$ for some $D \geqslant d$ the subset of $\mathcal{T}_d^\upsilon$ given by

$$\mathcal{T}_{d,D}^\upsilon := \Big\{T \in \mathcal{T}_d^\upsilon : \quad \sup_{k\geqslant 1}\|[\nabla_\upsilon]_{\underline{k}}^{1/2}[T]_{\underline{k}}^{-1}\|^2 \leqslant D\Big\}. \tag{5.3}$$

If the operator T has a singular value decomposition $\{\upsilon_j^{1/2}, e_j, f_j\}_{j\geqslant 1}$ then $[T]_{\underline{k}}$ is equivalent to the diagonal matrix $[\nabla_\upsilon]_{\underline{k}}^{1/2}$ for all $k \geqslant 1$ and, hence condition $T \in \mathcal{T}_d^\upsilon$ is equivalent to $T \in \mathcal{T}_{d,D}^\upsilon$. Moreover, the class $\mathcal{T}_{d,D}^\upsilon$ only contains operators T whose off-diagonal elements of $[T]_{\underline{k}}^{-1}$ are sufficiently small for all $k \geqslant 1$. A similar diagonality restriction has been used by Hall and Horowitz [2005] and Horowitz [2012]. Besides the mapping properties for the operator T we need a stronger assumption for the basis under consideration.

Assumption 5. *There exists $\eta \geqslant 1$ such that the joint distribution of (Z, W) satisfies*

(i) $\sup_{j\geqslant 1}\mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$ *and* $\sup_{l\in\mathbb{N}}\mathbb{E}[f_l^4(W)] \leqslant \eta^4$*;*

(ii) $\sup_{j,l\geqslant 1}\mathbb{E}\,|e_j(Z)f_l(W) - \mathbb{E}[e_j(Z)f_l(W)]|^k \leqslant \eta^k k!$*,* $k = 3, 4, \ldots$*.*

The following condition gathers conditions on the sequences γ and υ.

Assumption 6. *Let γ and υ be strictly positive sequences of weights with $\gamma_1 = \upsilon_1 = 1$ such that γ is nondecreasing with $\sup_{k\geqslant 1} k^3/\gamma_k < \infty$ and both sequences υ and υ/τ are nonincreasing.*

Remark 5.1. Under Assumptions 2–6, Johannes and Schwarz [2010] establish minimax optimality of the estimator $\widehat{\varphi}_{k_n}$ given in (5.1). More precisely, it is shown that mean integrated squared error loss of $\widehat{\varphi}_{k_n}$ attains the lower rate of convergence $\mathcal{R}_n := \max\big(\gamma_{k_n}^{-1}, \sum_{j=1}^{k_n}(n\upsilon_j)^{-1}\big)$ within a constant if the parameter k_n is chosen appropriately. □

5.2. The test statistic and its asymptotic distribution

Our goodness-of-fit statistic for testing nonparametric specifications is given by S_n where φ_0 is replaced by the nonparametric estimator $\widehat{\varphi}_{k_n}$ given in (5.1), i.e.,

$$S_n^{\mathrm{np}} := \|n^{-1}\sum_{i=1}^n \big(Y_i - \widehat{\varphi}_{k_n}(Z_i)\big) f_{\underline{m_n}}^\tau(W_i)\|^2.$$

The next result establishes asymptotic normality of S_n^{np} after standardization.

Theorem 5.1. *Let Assumptions 2–6 be satisfied. Moreover, $T \in \mathcal{T}_{d,D}^\upsilon$ and $\varphi_0 \in \mathcal{F}_\gamma^\rho$. If*

$$n\upsilon_{k_n} = o(\gamma_{k_n}\varsigma_{m_n}),\ k_n = o(\varsigma_{m_n}),\ k_n\Big(\sum_{j=1}^{m_n}\tau_j\Big)^2 = o(n\upsilon_{k_n}),\ \textit{and}\ m_n\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2 = o(n) \tag{5.4}$$

then it holds under H_{np}

$$(\sqrt{2}\varsigma_{m_n})^{-1}\Big(nS_n^{np}-\mu_{m_n}\Big)\xrightarrow{d}\mathcal{N}(0,1)\qquad as\ \ n\to\infty.$$

Example 5.1. Consider the setting of Example 4.1 where additionally $\upsilon_j\sim j^{-2a/r}$ for some $a>0$. Then condition (5.4) holds if $k_n\sim n^\kappa$ with $\kappa<\nu/2$ and

$$r(1-\nu/2)/(2a+2p)<\kappa<r(1-2\nu)/(2a+r).$$

In the severely ill posed case, i.e., $\upsilon_j\sim\exp(-j^{2a/r})$, $a>0$, condition (5.4) is satisfied if, for example, m_n satisfies $m_n=o(k_n^p)$ and $k_n^2=o(m_n)$ where $k_n\sim\big(\log n-\log(m_n^{3/2})\big)^{r/(2a)}$. □

The next result states an asymptotic distribution result of our test if $\sum_{j=1}^{m_n}\tau_j=O(1)$. Let us denote $U_W:=Y-e_{\underline{k_n}}(Z)^t[T]_{\underline{k_n}}^{-1}[g]_{\underline{k_n}}$ then clearly $\mathbb{E}[f_l(W)U_W]=0$ for all $1\leqslant l\leqslant k_n$. Let $\Sigma_{m_n}^{\mathrm{np}}$ be the covariance matrix of $Uf_{\underline{m_n}}^\tau(W)+U_W\,\mathbb{E}[f_{\underline{m_n}}^\tau(W)e_{\underline{k_n}}(Z)^t][T]_{\underline{k_n}}^{-1}f_{\underline{k_n}}(W)$. Then we define $\varsigma_{m_n}^{\mathrm{np}}=\varsigma(\Sigma_{m_n}^{\mathrm{np}})$, $\mu_{m_n}^{\mathrm{np}}=\mu(\Sigma_{m_n}^{\mathrm{np}})$, and $\mathcal{V}^{\mathrm{np}}=\mathcal{V}(\Sigma_{m_n}^{\mathrm{np}})$ where $\varsigma(\cdot)$, $\mu(\cdot)$, and $\mathcal{V}(\cdot)$ are given in Definition 2.1. In the proof of the following theorem we show by employing the extended link condition $T\in\mathcal{T}_{d,D}^{\upsilon}$ that $\varsigma_{m_n}^{\mathrm{np}}=O(1)$.

Theorem 5.2. *Let Assumptions 2–6 be satisfied. Moreover, $T\in\mathcal{T}_{d,D}^{\upsilon}$ and $\varphi_0\in\mathcal{F}_\gamma^\rho$. If*

$$\sum_{j=1}^{m_n}\tau_j^2=O(1),\quad n\upsilon_{k_n}=o(\gamma_{k_n}),\ \ and\ m_nk_n^2=o(n\upsilon_{k_n})\tag{5.5}$$

then it holds under H_{np}

$$(\sqrt{2}\varsigma_{m_n}^{np})^{-1}\big(nS_n^{np}-\mu_{m_n}^{np}\big)\xrightarrow{d}\mathcal{N}(0,\mathcal{V}^{np})\qquad as\ \ n\to\infty.$$

Example 5.2. Consider the setting of Example 4.2 where additionally $\upsilon_j\sim j^{-2a/r}$ for some $a>0$. Condition (5.5) is satisfied if $m_n\sim n^\nu$ with $0<\nu<1/2$ and $k_n\sim n^\kappa$ with

$$r/(2a+2p)<\kappa<r(1-\nu)/(2a+2r).$$

In the severely ill posed case, i.e., $\upsilon_j\sim\exp(-j^{2a/r})$, $a>0$, condition (5.5) is satisfied if, for example, m_n satisfies $m_n=o(k_n^p)$ and $k_n^2=o(m_n)$ where $k_n\sim\big(\log n-\log(m_n^2)\big)^{r/(2a)}$. □

Remark 5.2. Let Z' be a vector containing only entries of Z with $\dim(Z')<\dim(Z)$. It is easy to generalize our previous result for a test of H'_{np}: there exists a solution $\varphi_0\in\mathcal{F}_\gamma^\rho$ to (2.1) only depending on Z'. To be more precise consider the test statistic

$$S_n'^{\mathrm{np}}:=\|n^{-1}\sum_{i=1}^n\big(Y_i-\widehat{\varphi}_{k_n}(Z_i')\big)f_{\underline{m_n}}^\tau(W_i)\|^2$$

where $\widehat{\varphi}_{k_n}$ is the estimator (5.1) based on an iid. sample $(Y_1,Z_1',W_1),\dots,(Y_n,Z_n',W_n)$ of (Y,Z',W). Under H'_{np} we consider the conditional expectation operator $T':\mathcal{L}_{Z'}^2\to\mathcal{L}_W^2$ with $(T'\phi)(W):=\mathbb{E}[\phi(Z')|W]$. It is interesting to note that if T is nonsingular then also T' is one to one. Hence, for a test of H'_{np} we may replace Assumption 3 by the weaker condition that T' is nonsingular. Moreover, under H'_{np} the results of Theorem 5.1 and 5.2 still hold true if we replace Z by Z'. □

In the mildly ill-posed case, i.e., the singular values of T have a polynomial decay, the estimation precision suffers from the curse of dimensionality. Hence, by the test of dimension reduction of Z we can increase the accuracy of estimation of φ. On the other hand, in the severely ill-posed case the rate of convergence is independent of the dimension of Z (cf. Chen and Reiß [2011]). But still our dimension reduction test has an important implication concerning identification of φ. As the next example illustrates identification in the restricted model can be possible even if the structural function is not identified in the original version.

Example 5.3. Let $Z = (Z^{(1)}, Z^{(2)})$ where both, $Z^{(1)}$ and $Z^{(2)}$ are endogenous vectors of regressors. But only $Z^{(1)}$ satisfies a sufficiently strong relationship with the instrument W in the sense that for all $\phi \in \mathcal{L}^2_{Z^{(1)}}$ condition $\mathbb{E}[\phi(Z^{(1)})|W] = 0$ implies $\phi = 0$. In this example, we do not assume that this completeness condition is fulfilled for the joint distribution of $(Z^{(2)}, W)$. This can be interpreted as an insufficiency of correlation between $Z^{(2)}$ and W. Thereby only the operator $T^{(1)} : \mathcal{L}^2_{Z^{(1)}} \to \mathcal{L}^2_W$ with $T^{(1)} = \mathbb{E}[\phi(Z^{(1)})|W]$ is nonsingular but T is singular. If our dimension reduction test of Z indicates that $Z^{(2)}$ can be omitted from the structural function φ then we obtain identification in the restricted model. □

5.3. Limiting behavior under local alternatives and consistency.

Similar to the previous sections we study the power and consistency properties of our test.

Proposition 5.3. *Given the conditions of Proposition 5.1 it holds under (2.5)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\, S_n^{np} - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}\big(2^{-1/2}\|T\delta\|_\tau^2, 1\big) \qquad \textit{as } n \to \infty.$$

Given the conditions of Proposition 5.2 it holds under (2.6)

$$(\sqrt{2}\varsigma_{m_n}^{np})^{-1}\big(n\, S_n^{np} - \mu_{m_n}^{np}\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma_\infty^{np})^{-1}\|T\delta\|_\tau^2, \mathcal{V}^{np}\big) \qquad \textit{as } n \to \infty.$$

In the next proposition we establish consistency of our test when H_{np} does not hold, i.e., there exists no function in $\mathcal{F}_\gamma^\rho$ that solves (2.1) for any sequence γ satisfying Assumption 6 and any sufficiently large constant $0 < \rho < \infty$.

Proposition 5.4. *Assume that H_{np} does not hold. Let $\mathbb{E}\,|Y - \varphi_0(Z)|^4 < \infty$. Consider a sequence λ such that $\lambda_n = o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 5.1 and 5.2, respectively, we have*

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(nS_n^{np} - \mu_{m_n}\big) > \lambda_n\Big) = 1 + o(1),$$
$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n}^{np})^{-1}\big(nS_n^{np} - \mu_{m_n}^{np}\big) > \lambda_n\Big) = 1 + o(1).$$

In the following we show that our tests are consistent uniformly over the function class

$$\mathcal{J}_n^\rho = \Big\{\varphi \in \mathcal{F}_{\gamma'}^\rho : \inf_{\varphi_0 \in \mathcal{F}_\gamma^\rho} \|F_{m_n} T(\varphi - \varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n} n^{-1}\Big\}$$

where the sequence $\gamma' := (\gamma'_j)_{j\geqslant 1}$ satisfies $\gamma'_j \sim j^2$. Hence, under Assumption 6, i.e., $\sup_{k\geqslant 1} k^3/\gamma_k < \infty$, it holds $\mathcal{F}_\gamma^\rho \subset \mathcal{F}_{\gamma'}^\rho$. Again, $q_{1-\alpha}$ denotes the $1-\alpha$ quantile of $\mathcal{N}(0,1)$.

Proposition 5.5. *For any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ we have under the conditions of Theorem 5.1*

$$\lim_{n\to\infty} \inf_{\varphi \in \mathcal{J}_n^\rho} \mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\, S_n^{np} - \mu_{m_n}\big) > q_{1-\alpha}\Big) \geqslant 1 - \varepsilon,$$

whereas under the conditions of Theorem 5.2 it holds

$$\lim_{n\to\infty} \inf_{\varphi\in\mathcal{J}_n^\rho} \mathbb{P}\Big(\big(\sqrt{2}\,\varsigma_{m_n}^{np}\,\mathcal{V}^{np}\big)^{-1}\big(n\,S_n^{np} - \mu_{m_n}^{np}\big) > q_{1-\alpha}\Big) \geqslant 1-\varepsilon.$$

Remark 5.3. The condition $\inf_{\varphi_0\in\mathcal{F}_\gamma^\rho} \|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2 \geqslant \rho\,\varsigma_{m_n}n^{-1}$ for $\varphi\in\mathcal{J}_n^\rho$ rules out alternatives with generalized Fourier coefficients $[T(\varphi-\varphi_0)]_j$ converging to zero too fast. Horowitz [2012] gave an example where uniform consistency over these alternatives can not be achieved. On the other hand, if $[T(\varphi-\varphi_0)]_j$ oscillates for large $j\geqslant 1$ then one might chose a weaker decaying sequence τ in order to ensure consistency uniformly over these alternatives. □

6. Conclusion

Based on the methodology of series estimation, we have developed in this paper a family of goodness-of-fit statistics and derived their asymptotic properties. We have seen that the asymptotic results depend crucially on the choice of the smoothing operator L. For the theory we had to distinguish two cases namely that $\varsigma_{m_n}^{-1} = O(1)$ and $\sum_{j=1}^{m_n}\tau_j = O(1)$. By choosing a stronger decaying sequence τ, our test becomes more powerful with respect to local alternatives but might lose desirable consistency properties. Although our results hold for any decaying sequence τ, it is of great interest how to choose this sequence in practice. Moreover, in the case of exogeneity or nonparametric specification, one may also use estimators for φ where the dimension parameter k_n adapts to the unknown smoothness of φ as well as to the unknown decay of the singular values of T.

A. Appendix

A.1. Proofs of Section 2.

Proof of Theorem 2.1. Under H_0 we have $(Y_i-\varphi_0(Z_i))f_{\underline{m}}^\tau(W_i) = U_i f_{\underline{m}}^\tau(W_i)$ for all $m\geqslant 1$ and consequently we observe

$$\varsigma_{m_n}^{-1}\big(nS_n - \mu_{m_n}\big) = \frac{1}{\varsigma_{m_n}n}\sum_{i=1}^n\sum_{j=1}^{m_n}\big(|U_i f_j^\tau(W_i)|^2 - \varsigma_{jj}\big) + \frac{1}{\varsigma_{m_n}n}\sum_{i\neq i'}\sum_{j=1}^{m_n} U_iU_{i'}f_j^\tau(W_i)f_j^\tau(W_{i'})$$

where the first summand tends in probability to zero as $n\to\infty$. Indeed, since $\mathbb{E}\,|Uf_j(W)|^2 - \varsigma_{jj} = 0$, $j\geqslant 1$, it holds for all $m\geqslant 1$

$$\frac{1}{(\varsigma_m n)^2}\,\mathbb{E}\,\Big|\sum_{i=1}^n\sum_{j=1}^m |U_i f_j^\tau(W_i)|^2 - \varsigma_{jj}\Big|^2 = \frac{1}{n\varsigma_m^2}\,\mathbb{E}\,\Big|\sum_{j=1}^m |Uf_j^\tau(W)|^2 - \varsigma_{jj}\Big|^2 \leqslant \frac{1}{n\varsigma_m^2}\,\mathbb{E}\,\|Uf_{\underline{m}}^\tau(W)\|^4.$$

By using Assumptions 1 and 2, i.e., $\sup_{j\in\mathbb{N}}\mathbb{E}\,|f_j(W)|^4 \leqslant \eta^4$ and $\mathbb{E}[U^2|W]\leqslant\sigma^2$, we conclude

$$\mathbb{E}\,\|Uf_{\underline{m}}^\tau(W)\|^4 \leqslant m\sum_{j=1}^m \tau_j^2\,\mathbb{E}\,|Uf_j(W)|^4 \leqslant \eta^4\sigma^4 m\sum_{j=1}^m\tau_j^2. \tag{A.1}$$

Let $m = m_n$ satisfy condition (2.3) then $\mathbb{E}\,\|U f^\tau_{\underline{m_n}}(W)\|^4 = o\big(n\varsigma^2_{m_n}\big)$. Therefore, it is sufficient to prove

$$\sqrt{2}(\varsigma_{m_n} n)^{-1} \sum_{i\neq i'} \sum_{j=1}^{m_n} U_i U_{i'} f^\tau_j(W_i) f^\tau_j(W_{i'}) \xrightarrow{d} \mathcal{N}(0,1). \tag{A.2}$$

Since $\varsigma_{m_n} = o(1)$ this follows from Lemma A.2 and thus, completes the proof. $\square$

Proof of Theorem 2.2. Similarly to the proof of Theorem 2.1 it can be seen that it is sufficient to show

$$n^{-1} \sum_{i\neq i'} \sum_{j=1}^{m_n} U_i U_{i'} f^\tau_j(W_i) f^\tau_j(W_{i'}) \xrightarrow{d} \mathcal{N}(0,\mathcal{V}).$$

This result is due to Lemma A.2 since $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$. $\square$

Proof of Proposition 2.3. For ease of notation let $\delta_n(\cdot) := \varsigma^{1/2}_{m_n} n^{-1/2}\delta(\cdot)$. Under the sequence of alternatives (2.5) the following decomposition holds true

$$S_n = \big\|n^{-1}\sum_{i=1}^n U_i f^\tau_{\underline{m_n}}(W_i)\big\|^2 + 2\langle n^{-1}\sum_{i=1}^n U_i f^\tau_{\underline{m_n}}(W_i), n^{-1}\sum_{i=1}^n \delta_n(Z_i) f^\tau_{\underline{m_n}}(W_i)\rangle$$
$$+ \big\|n^{-1}\sum_{i=1}^n \delta_n(Z_i) f^\tau_{\underline{m_n}}(W_i)\big\|^2 =: I_n + 2II_n + III_n.$$

Due to Theorem 2.1 we have $(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\, I_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0,1)$. Consider II_n. We observe

$$n\,\mathbb{E}\,|II_n| \leqslant \sum_{j=1}^{m_n} \tau_j \big(\mathbb{E}\,|U f_j(W)|^2\, \mathbb{E}\,|\delta_n(Z) f_j(W)|^2\big)^{1/2} + \Big(n\,\mathbb{E}\Big|\sum_{j=1}^{m_n} \tau_j [T\delta_n]_j U f_j(W)\Big|^2\Big)^{1/2}$$
$$\leqslant \sigma \sum_{j=1}^{m_n} \tau_j \big(\mathbb{E}\,|\delta_n(Z) f_j(W)|^2\big)^{1/2} + \sigma\sqrt{n}\|T\delta_n\|_\tau.$$

From the definition of δ_n and condition (2.3) we infer that $n\,\mathbb{E}\,|II_n| = o(\varsigma_{m_n})$. Consider III_n. Employing again the definition of δ_n yields

$$n\varsigma^{-1}_{m_n} III_n = \sum_{j=1}^{m_n} \tau_j \big|n^{-1}\sum_{i=1}^n \delta(Z_i) f_j(W_i) - [T\delta]_j\big|^2$$
$$+ 2\sum_{j=1}^{m_n} \tau_j [T\delta]_j \big(n^{-1}\sum_{i=1}^n \delta(Z_i) f_j(W_i) - [T\delta]_j\big) + \sum_{j=1}^{m_n} \tau_j [T\delta]_j^2 =: A_{n1} + 2A_{n2} + A_{n3}.$$

Clearly, $\mathbb{E}\,A_{n1} \leqslant n^{-1}\sum_{j=1}^{m_n} \tau_j\, \mathbb{E}\,|\delta(Z) f_j(W)|^2$ and $\mathbb{E}\,A^2_{n2} \leqslant n^{-1}\|T\delta\|^2_\tau \sum_{j=1}^{m_n} \mathbb{E}\,|\delta(Z) f^\tau_j(W)|^2$. Since $\mathbb{E}\,|\delta(Z)|^4 < \infty$ we conclude $A_{n1} = o_p(1)$ and $A_{n2} = o_p(1)$. On the other hand, it holds $A_{n3} = \|T\delta\|^2_\tau + o(1)$. We conclude $(\sqrt{2}\varsigma_{m_n})^{-1} n III_n = (\sqrt{2})^{-1}\|T\delta\|^2_\tau + o_p(1)$, which completes the proof. $\square$

Proof of Proposition 2.4. Let $\delta_n(\cdot) := n^{-1/2}\delta(\cdot)$. Similarly to the proof of Theorem 2.1 it is straightforward to see that under the sequence of alternatives (2.6) it holds

$$(\sqrt{2}\varsigma_{m_n})^{-1}\big(n\,S_n-\mu_{m_n}\big) = \frac{1}{\sqrt{2}\varsigma_{m_n}n}\sum_{i\neq i'}\sum_{j=1}^{m_n}(U_i+\delta_n(Z_i))(U_{i'}+\delta_n(Z_{i'}))f_j^\tau(W_i)f_j^\tau(W_{i'})+o_p(1).$$

In the following we make use of the decomposition

$$\begin{aligned}
&\sum_{i\neq i'}\sum_{j=1}^{m_n}(U_i+\delta_n(Z_i))(U_{i'}+\delta_n(Z_{i'}))f_j^\tau(W_i)f_j^\tau(W_{i'})\\
&=2\sum_{i'<i}\sum_{j=1}^{m_n}\tau_j(U_i+\delta_n(Z_i))f_j(W_i)\big((U_{i'}+\delta_n(Z_{i'}))f_j(W_{i'})-[T\delta_n]_j\big)\\
&+\sum_{j=1}^{m_n}\tau_j[T\delta_n]_j\sum_{i=1}^{n}(i-1)\big((U_i+\delta_n(Z_i))f_j(W_i)-[T\delta_n]_j\big)+\varsigma_{m_n}^{-1}\sum_{j=1}^{m_n}\tau_j[T\delta]_j^2\\
&\qquad\quad=I_n+II_n+III_n.
\end{aligned}$$

Due to Lemma A.3 it holds $(\sqrt{2}\varsigma_{m_n}n)^{-1}I_n\xrightarrow{d}\mathcal{N}(0,\mathcal{V})$. In addition, $(\sqrt{2}\varsigma_{m_n}n)^{-1}III_n = \varsigma_\infty^{-1}\|T\delta\|_\tau^2+o_p(1)$. Thereby, the result follows from $II_n=o_p(n)$ since

$$\begin{aligned}
n^{-2}\,\mathbb{E}\,II_n^2 &\leqslant n^{-2}\Big(\sum_{j=1}^{m_n}[T\delta_n]_j^2\Big)\sum_{j=1}^{m_n}\tau_j^2\sum_{i=1}^{n}(i-1)\,\mathbb{E}\,\big|(U_i+\delta_n(Z_i))f_j(W_i)-[T\delta_n]_j\big|^2\\
&\leqslant n^{-1}\|T\delta\|_W^2\sum_{j=1}^{m_n}\tau_j^2\big(\,\mathbb{E}\,|Uf_j(W)|^2+\mathbb{E}\,|\delta_n(Z)f_j(W)|^2\big)=o(1)
\end{aligned}$$

where we used $\sum_{j=1}^{m_n}\tau_j^2=O(1)$, $\mathbb{E}\,|Uf_j(W)|^2\leqslant\sigma^2$, and $\mathbb{E}\,|\delta(Z)|^4<\infty$. □

Proof of Proposition 2.5. Since $\varsigma_{m_n}\lambda_n+\mu_{m_n}=o(n)$ it is sufficient to show $S_n = \|T(\varphi-\varphi_0)\|_\tau^2+o_p(1)$. We make use of the decomposition

$$\begin{aligned}
S_n&=\sum_{j=1}^{m_n}\tau_j\Big|n^{-1}\sum_{i=1}^{n}(Y_i-\varphi_0(Z_i))f_j(W_i)-[T(\varphi-\varphi_0)]_j\Big|^2\\
&+2\sum_{j=1}^{m_n}\tau_j\Big(n^{-1}\sum_{i=1}^{n}(Y_i-\varphi_0(Z_i))f_j(W_i)-[T(\varphi-\varphi_0)]_j\Big)[T(\varphi-\varphi_0)]_j+\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2\\
&\qquad\quad=I_n+II_n+III_n.
\end{aligned}$$

Due to condition $\mathbb{E}\,|Y-\varphi_0(Z)|^4<\infty$ it is easily seen that $I_n+II_n=o_p(1)$. On the other hand $III_n=\|T(\varphi-\varphi_0)\|_\tau^2+o(1)$, which proves the result. □

Proof of Proposition 2.6. We make use of the decomposition

$$\begin{aligned}
&\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\,S_n-\mu_{m_n}\big)>q_{1-\alpha}\Big)\\
&\qquad\geqslant\mathbb{P}\Big(\Big\|n^{-1/2}\sum_{i=1}^{n}(\varphi(Z_i)-\varphi_0(Z_i))f^\tau_{\underline{m_n}}(W_i)\Big\|^2+\Big\|n^{-1/2}\sum_{i=1}^{n}U_if^\tau_{\underline{m_n}}(W_i)\Big\|^2-\mu_{m_n}\\
&\qquad\qquad>\sqrt{2}\,\varsigma_{m_n}q_{1-\alpha}+2|\langle n^{-1}\sum_{i=1}^{n}(\varphi(Z_i)-\varphi_0(Z_i))f^\tau_{\underline{m_n}}(W_i),\sum_{i=1}^{n}U_if^\tau_{\underline{m_n}}(W_i)\rangle|\Big).
\end{aligned}$$

Uniformly over all $\varphi \in \mathcal{G}_n^\rho$ it holds

$$\langle n^{-1}\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^\tau(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\rangle = O_p\big(\sqrt{n}\|F_{m_n}T(\varphi-\varphi_0)\|_\tau\big). \quad \text{(A.3)}$$

Indeed, we observe

$$\mathbb{E}\,\Big|\sum_{j=1}^{m_n}\tau_j\,\mathbb{E}[(\varphi(Z)-\varphi_0(Z))f_j(W)]\sum_{i=1}^n U_i f_j(W_i)\Big|^2 \leqslant \sigma^2 n\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2$$

which yields (A.3). Thereby, for all $0<\varepsilon'<1$ there exists some constant $C>0$ such that

$$\begin{aligned}&\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\,S_n-\mu_{m_n}\big)>q_{1-\alpha}\Big)\\ &\geqslant \mathbb{P}\Big(\big\|n^{-1/2}\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^\tau(W_i)\big\|^2+\big\|n^{-1/2}\sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\big\|^2-\mu_{m_n}\\ &\qquad\qquad > \sqrt{2}\,\varsigma_{m_n}q_{1-\alpha}+C\sqrt{n}\|F_{m_n}T(\varphi-\varphi_0)\|_\tau\Big)-\varepsilon'.\end{aligned}$$

Note that $\big\|n^{-1/2}\sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\big\|^2=\mu_{m_n}+O_p(\varsigma_{m_n})$ due to Theorem 2.1. Moreover,

$$\begin{aligned}&\big\|n^{-1/2}\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^\tau(W_i)\big\|^2 \geqslant n\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2\\ &-2\big|\langle\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^\tau(W_i)-n[LT(\varphi-\varphi_0)]_{\underline{m_n}},[LT(\varphi-\varphi_0)]_{\underline{m_n}}\rangle = I_n+II_n.\end{aligned}$$

Consider II_n. For $1\leqslant j\leqslant m_n$ let $s_j=\tau_j[T(\varphi-\varphi_0)]_j/\|F_{m_n}T(\varphi-\varphi_0)\|_\tau$ then clearly $\sum_{j=1}^{m_n}s_j^2=1$ and thus $\mathbb{E}\,|\sum_{j=1}^{m_n}s_jf_j(W)|^2=1$. Further, since $\sup_{z\in \mathrm{Supp}(Z)}|\varphi(z)-\varphi_0(z)|^2\leqslant \rho$ we calculate

$$\begin{aligned}\mathbb{E}\,|II_n|^2&=n\,\mathbb{E}\,\Big|\sum_{j=1}^{m_n}\tau_j\big((\varphi(Z)-\varphi_0(Z))f_j(W)-[T(\varphi-\varphi_0)]_j\big)[T(\varphi-\varphi_0)]_j\Big|^2\\ &\leqslant n\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2\,\mathbb{E}\,\Big|\sum_{j=1}^{m_n}s_j(\varphi(Z)-\varphi_0(Z))f_j(W)\Big|^2\leqslant \rho\,\|F_{m_n}T(\varphi-\varphi_0)\|_\tau^2\end{aligned}$$

and hence $II_n=O_p(1)$. Note that $I_n-C\sqrt{n}\|F_{m_n}T(\varphi-\varphi_0)\|_\tau\geqslant I_n/2$ for n sufficiently large. Since on $\mathcal{G}_n^\rho$ we have $I_n\geqslant\rho\,\varsigma_{m_n}$ we obtain the result by choosing ρ sufficiently large. □

A.2. Proofs of Section 3.

Proof of Theorem 3.1. The proof is based on the decomposition under H_p

$$\begin{aligned}S_n^\mathrm{p}=\big\|n^{-1}\sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\big\|^2&+2\langle n^{-1}\sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i), n^{-1}\sum_{i=1}^n(\varphi_0(Z_i)-\breve{\varphi}_k(Z_i))f_{\underline{m_n}}^\tau(W_i)\rangle\\ &+\|n^{-1}\sum_{i=1}^n((\varphi_0(Z_i)-\breve{\varphi}_k(Z_i))f_{\underline{m_n}}^\tau(W_i)\|^2=I_n+2II_n+III_n. \quad \text{(A.4)}\end{aligned}$$

Due to Theorem 2.1 it holds $(\sqrt{2}\varsigma_{m_n})^{-1}(nI_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0,1)$. Consider III_n. We observe

$$III_n \leqslant 2\|\varphi_0 - \breve{\varphi}_k\|_Z^2 \Big(\sum_{l=1}^{k}\sum_{j=1}^{m_n} \tau_j [T]_{jl}^2 + \sum_{l=1}^{k}\sum_{j=1}^{m_n} \tau_j \big(n^{-1}\sum_{i=1}^{n} e_l(Z_i) f_j(W_i) - [T]_{jl}\big)^2 \Big).$$

For each $1 \leqslant l \leqslant k$ we have $\sum_{j=1}^{m_n}[T]_{jl}^2 \leqslant \|Te_l\|_W^2 \leqslant 1$ by applying Jensen's inequality. Moreover, we calculate

$$\sum_{l=1}^{k}\sum_{j=1}^{m_n} \mathbb{E}\,\Big|n^{-1}\sum_{i=1}^{n} e_l(Z_i) f_j(W_i) - [T]_{jl}\Big|^2 \leqslant \frac{km_n}{n} \sup_{j,l\geqslant 1} \mathbb{E}\,|e_l(Z)f_j(W)|^2 \leqslant \eta^4 \frac{km_n}{n}. \quad \text{(A.5)}$$

These estimates together with $\|\varphi_0 - \breve{\varphi}_k\|_Z = O_p(n^{-1/2})$ imply $nIII_n = o_p(\varsigma_{m_n})$. We are left with the proof of $nII_n = o_p(\varsigma_{m_n})$. We observe for each $1 \leqslant l \leqslant k$

$$\mathbb{E}\Big|\sum_{j=1}^{m_n} \tau_j \Big(\varsigma_{m_n}^{-1} n^{-1/2} \sum_{i=1}^{n} U_i f_j(W_i)\big(n^{-1}\sum_{i=1}^{n} e_l(Z_i) f_j(W_i) - [T]_{jl}\big)\Big)\Big|$$
$$\leqslant \varsigma_{m_n}^{-1} n^{-1/2} \sum_{j=1}^{m_n} \tau_j \big(\mathbb{E}\,|Uf_j(W)|^2\big)^{1/2} \big(\mathbb{E}\,|e_l(Z)f_j(W)|^2\big)^{1/2} \leqslant \sigma\eta^4 \varsigma_{m_n}^{-1} n^{-1/2} \sum_{j=1}^{m_n} \tau_j = o(1).$$

Now since $n^{1/2}([\varphi_0]_{\underline{k}} - [\breve{\varphi}_k]_{\underline{k}}) = O_p(1)$ we infer

$$nII_n = n^{1/2} \sum_{l=1}^{k} ([\varphi_0]_l - [\breve{\varphi}_k]_l) \sum_{j=1}^{m_n} \tau_j \Big(\varsigma_{m_n}^{-1} n^{-1/2} \sum_{i=1}^{n} U_i f_j(W_i)[T]_{jl}\Big) + o_p(1).$$

We observe for each $1 \leqslant l \leqslant k$

$$\varsigma_{m_n}^{-2} n^{-1}\, \mathbb{E}\Big|\sum_{j=1}^{m_n} \tau_j \sum_{i=1}^{n} U_i f_j(W_i)[T]_{jl}\Big|^2 \leqslant \varsigma_{m_n}^{-2}\sigma^2 \sum_{j=1}^{m_n}[T]_{jl}^2 \leqslant \varsigma_{m_n}^{-2}\sigma^2$$

which implies $nII_n = o_p(\varsigma_{m_n})$ and thus, in light of decomposition (A.4), completes the proof. □

Proof of Theorem 3.2. For $1 \leqslant j \leqslant m_n$ we make use of the following decomposition

$$n^{-1/2}\sum_{i=1}^{n} f_j(W_i)\Big(U_i + \sum_{l=1}^{k} e_l(Z_i)([\varphi_0]_l - [\breve{\varphi}_k]_l)\Big) = n^{-1/2}\sum_{i=1}^{n}\Big(f_j(W_i)U_i + \sum_{l=1}^{k}[T]_{jl} h_l(V_i)\Big)$$
$$+ \sum_{l=1}^{k}\Big(n^{-1}\sum_{i=1}^{n} f_j(W_i)e_l(Z_i) - [T]_{jl}\Big)\Big(n^{-1/2}\sum_{i=1}^{n} h_l(V_i)\Big)$$
$$+ \sum_{l=1}^{k} n^{-1}\sum_{i=1}^{n} f_j(W_i)e_l(Z_i) r_l = A_{nj} + B_{nj} + C_{nj} \quad \text{(A.6)}$$

where $r_{\underline{k}} = (r_1, \dots, r_k)^t$ is a stochastic vector satisfying $r_{\underline{k}} = o_p(1)$. Consequently, under H_{p} we have

$$nS_n^{\mathrm{p}} = \sum_{j=1}^{m_n} \tau_j A_{nj}^2 + 2\sum_{j=1}^{m_n} \tau_j A_{nj}(B_{nj} + C_{nj}) + \sum_{j=1}^{m_n} \tau_j (B_{nj} + C_{nj})^2.$$

Clearly, for all $1 \leqslant i \leqslant n$ the random variables $U_i f_j^\tau(W_i) + \mathbb{E}\left[f_j^\tau(W) e_{\underline{k}}(Z)^t\right] h_{\underline{k}}(V_i)$, $1 \leqslant j \leqslant m_n$, are centered with bounded fourth moment. Following line by line the proof of Lemma A.2 it is easily seen that $(\sqrt{2}\varsigma_{m_n}^{\mathrm{p}})^{-1}\big(\sum_{j=1}^{m_n} \tau_j A_{nj}^2 - \mu_{m_n}^{\mathrm{p}}\big) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}^{\mathrm{p}})$. Exploiting inequality (A.5) it is easily seen that $\sum_{j=1}^{m_n} B_{nj}^2 = O_p(m_n n^{-1}) = o_p(1)$. Since $\sum_{j=1}^{m_n}[T]_{jl}^2 \leqslant 1$ we have $\|\mathbb{E}[f_{\underline{m_n}}(W) e_{\underline{k}}(Z)^t] r_{\underline{k}}\|^2 \leqslant \sum_{j=1}^{m_n}\sum_{l=1}^{k}[T]_{jl}^2 \|r_{\underline{k}}\|^2 \leqslant k\|r_{\underline{k}}\|^2 = o_p(1)$ and hence $\sum_{j=1}^{m_n} C_{nj}^2 = o_p(1)$. Finally, condition $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ implies $\mathbb{E}\,\big|\sum_{j=1}^{m_n}\tau_j^2 A_{nj}\big|^2 \leqslant \sup_{j\geqslant 1}\big\{\mathbb{E}\,A_{nj}^2\big\}\big(\sum_{j=1}^{m_n}\tau_j^2\big)^2 = O(1)$ and thereby, we have $\mathbb{E}\,\big|\sum_{j=1}^{m_n}\tau_j A_{nj}(B_{nj} + C_{nj})\big|^2 \leqslant \sum_{j=1}^{m_n}\tau_j^2 A_{nj}^2 \sum_{j=1}^{m_n}(B_{nj} + C_{nj})^2 = o_p(1)$, which completes the proof. □

Proof of Proposition 3.3. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Under the sequence of alternatives (2.5) the following decomposition holds true

$$S_n^{\mathrm{p}} = S_n + 2\langle n^{-1}\sum_{i=1}^n (U_i + \varsigma_{m_n}^{1/2} n^{-1/2}\delta(Z_i)) f_{\underline{m_n}}^\tau(W_i), n^{-1}\sum_{i=1}^n(\varphi_0(Z_i) - \breve{\varphi}_k(Z_i)) f_{\underline{m_n}}^\tau(W_i)\rangle + \big\|n^{-1}\sum_{i=1}^n(\varphi_0(Z_i) - \breve{\varphi}_k(Z_i)) f_{\underline{m_n}}^\tau(W_i)\big\|^2.$$

Due to Proposition 2.3 and the proof of Theorem 3.1 it is sufficient to show

$$\langle n^{-1}\sum_{i=1}^n \delta(Z_i) f_{\underline{m_n}}^\tau(W_i), n^{-1/2}\sum_{i=1}^n (\varphi_0(Z_i) - \breve{\varphi}_k(Z_i)) f_{\underline{m_n}}^\tau(W_i)\rangle = o_p(\sqrt{\varsigma_{m_n}}). \tag{A.7}$$

Since $\sum_{j=1}^{m_n}[T]_{jl}^2 \leqslant 1$ we conclude

$$\begin{aligned}\sum_{j=1}^{m_n}\tau_j[T\delta]_j n^{-1/2}\sum_{i=1}^n(\varphi_0(Z_i) - \breve{\varphi}_k(Z_i)) f_j(W_i) &= \sum_{l=1}^{k}\sqrt{n}([\varphi_0]_l - [\breve{\varphi}_k]_l)\sum_{j=1}^{m_n}\tau_j[T\delta]_j[T]_{jl} + o_p(1)\\ &\leqslant \sqrt{n}\|T\delta\|_\tau \|\varphi_0 - \breve{\varphi}_k\|_Z + o_p(1) = O_p(1)\end{aligned}$$

and hence (A.7) holds true.
Consider the case $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$. We make use of decomposition (A.6) where U_i is replaced by $U_i + n^{-1/2}\delta(Z_i)$. Similarly to the proof of Proposition 2.4 it is easily seen that $(\sqrt{2}\varsigma_{m_n}^{\mathrm{p}})^{-1}\big(\sum_{j=1}^{m_n}\tau_j A_{nj}^2 - \mu_{m_n}^{\mathrm{p}}\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma_\infty^{\mathrm{p}})^{-1}\|T\delta\|_\tau^2, \mathcal{V}^{\mathrm{p}}\big)$. Thereby, due to the proof of Theorem 3.2, the assertion follows. □

Proof of Proposition 3.4. It is sufficient to prove $S_n^{\mathrm{p}} = \|T(\varphi - \varphi_0)\|_\tau^2 + o_p(1)$. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Since $\|n^{-1}\sum_{i=1}^n((\varphi_0(Z_i) - \breve{\varphi}_k(Z_i)) f_{\underline{m_n}}^\tau(W_i)\|^2 = o_p(1)$ (cf. proof of Theorem 3.1) and $\|n^{-1}\sum_{i=1}^n(Y_i - \varphi_0(Z_i)) f_{\underline{m_n}}^\tau(W_i)\|^2 = \|T(\varphi - \varphi_0)\|_\tau^2 + o_p(1)$ (cf. proof of Proposition 2.5) the result follows. In case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ we infer from the proof of Theorem 3.2 that $S_n^{\mathrm{p}} = \sum_{j=1}^{m_n}\tau_j\big|n^{-1}\sum_{i=1}^n\big((Y_i - \varphi_0(Z_i)) f_j(W_i) + \sum_{l=1}^k [T]_{jl} h_l(V_i)\big)\big|^2 + o_p(1)$. Since $h_l(V_i)$, $1 \leqslant l \leqslant k$, are centered random variables we obtain, similarly to the proof of Proposition 2.5, that $S_n^{\mathrm{p}} = \|T(\varphi - \varphi_0)\|_\tau^2 + o_p(1)$. □

Proof of Proposition 3.5. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Let $\varphi_0 \in \mathcal{E}_k$ with estimator $\breve{\varphi}_k$. The basic inequality $(a-b)^2 \geqslant a^2/2 - b^2$, $a, b \in \mathbb{R}$, yields

$$\begin{aligned}
&\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\big(n\,S_n^{\mathrm{p}} - \mu_{m_n}\big) > q_{1-\alpha}\Big)\\
&\qquad \geqslant \mathbb{P}\Big(1/2\big\|n^{-1/2}\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^{\tau}(W_i)\big\|^2 + \big\|n^{-1/2}\sum_{i=1}^n U_i f_{\underline{m_n}}^{\tau}(W_i)\big\|^2 - \mu_{m_n}\\
&\qquad\qquad > \sqrt{2}\,\varsigma_{m_n} q_{1-\alpha} + 2|\langle n^{-1}\sum_{i=1}^n(\varphi(Z_i)-\breve{\varphi}_k(Z_i))f_{\underline{m_n}}^{\tau}(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^{\tau}(W_i)\rangle|\\
&\qquad\qquad\qquad + \big\|n^{-1/2}\sum_{i=1}^n(\breve{\varphi}_k(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^{\tau}(W_i)\big\|^2\Big). \qquad \text{(A.8)}
\end{aligned}$$

From the proof of Theorem 3.1 we infer $\big\|n^{-1/2}\sum_{i=1}^n(\breve{\varphi}_k(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^{\tau}(W_i)\big\|^2 = o_p(\varsigma_{m_n})$ and

$$\begin{aligned}
&\langle n^{-1}\sum_{i=1}^n(\varphi(Z_i)-\breve{\varphi}_k(Z_i))f_{\underline{m_n}}^{\tau}(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^{\tau}(W_i)\rangle\\
&\qquad = \langle n^{-1}\sum_{i=1}^n(\varphi(Z_i)-\varphi_0(Z_i))f_{\underline{m_n}}^{\tau}(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^{\tau}(W_i)\rangle + o_p(\varsigma_{m_n})
\end{aligned}$$

uniformly over all $\varphi \in \mathcal{H}_n^\rho$. In addition, let s_j be as in the proof of Proposition 2.6, then condition $\sup_{j\geqslant 1}\mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$ yields

$$\mathbb{E}\Big|\sum_{j=1}^{m_n} s_j(\varphi(Z)-\varphi_0(Z))f_j(W)\Big|^2 \leqslant 2\rho + 2\sum_{l=1}^k[\varphi_0]_l^2\sum_{l=1}^k\mathbb{E}\,\Big|e_l(Z)\sum_{j=1}^{m_n} s_j f_j(W)\Big|^2 = O(1).$$

Thus, following line by line the proof of Proposition 2.6, the assertion follows. In case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ the assertion follows similarly. □

A.3. Proofs of Section 4.

Let $\mathcal{A}_k = \{\|[\widehat{\mathrm{Id}}]_{\underline{k}}^{-1}\| \leqslant 2\}$ and $\mathcal{B}_k = \{\|[\widehat{\mathrm{Id}}]_{\underline{k}} - [\mathrm{Id}]_{\underline{k}}\| \leqslant 1/2\}$. Their complements are denoted by $\mathcal{A}_k^c$ and $\mathcal{B}_k^c$, respectively. By the usual Neumann series argument we observe on $\mathcal{B}_k$ that $\|[\widehat{\mathrm{Id}}]_{\underline{k}}^{-1}\| \leqslant (1 - \|[\widehat{\mathrm{Id}}]_{\underline{k}} - [\mathrm{Id}]_{\underline{k}}\|)^{-1} \leqslant 2$ and hence $\mathcal{B}_k \subset \mathcal{A}_k$.

Proof of Theorem 4.1. The proof is based on the decomposition (A.4) where the estimator $\breve{\varphi}_k$ is replaced by $\overline{\varphi}_{k_n}$ given in (4.1). It holds $nIII_n = o_p(\varsigma_{m_n})$, which can be seen as follows. We make use of

$$III_n/2 \leqslant \big\|\frac{1}{n}\sum_{i=1}^n(E_{k_n}\varphi_0(Z_i) - \overline{\varphi}_{k_n}(Z_i))f_{\underline{m_n}}^{\tau}(W_i)\big\|^2 + \big\|\frac{1}{n}\sum_{i=1}^n\big(E_{k_n}^{\perp}\varphi_0\big)(Z_i)f_{\underline{m_n}}^{\tau}(W_i)\big\|^2 =: A_{n1} + A_{n2}.$$

Consider A_{n1}. We observe

$$\begin{aligned}
A_{n1} \leqslant 2\|T(E_{k_n}\varphi_0 - \overline{\varphi}_{k_n})\|_W^2 + 2\|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}|n^{-1}\sum_{i=1}^n e_l(Z_i)f_j(W_i) - [T]_{jl}|^2&\\
=: 2B_{n1} + 2B_{n2}.& \qquad \text{(A.9)}
\end{aligned}$$

We have $\mathbb{E}\, e_{\underline{k_n}}(Z)e_{\underline{k_n}}(Z)^t = [\mathrm{Id}]_{\underline{k_n}}$. For B_{n1} we evaluate due to Jensen's inequality

$$B_{n1} \leqslant \|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2 \leqslant 2\big\|[\widehat{\mathrm{Id}}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i)\big\|^2\, \mathbb{1}_{\mathcal{A}_{k_n}}$$
$$+ 8\big\|[\widehat{\mathrm{Id}}]_{\underline{k_n}} - [\mathrm{Id}]_{\underline{k_n}}\big\|^2\big\|[\widehat{\mathrm{Id}}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i)\big\|^2 + \|E_{k_n}\varphi_0\|_Z^2\, \mathbb{1}_{\mathcal{A}^c_{k_n}}\,.$$

Since the spectral norm of a matrix is bounded by its Frobenius norm it holds

$$\mathbb{E}\,\big\|[\widehat{\mathrm{Id}}]_{\underline{k_n}} - [\mathrm{Id}]_{\underline{k_n}}\big\|^2 \leqslant n^{-1}\sum_{l,l'=1}^{k_n} \mathbb{E}\,|e_l(Z)e_{l'}(Z)|^2 \leqslant \eta^4 n^{-1}k_n^2$$

and condition $\varphi \in \mathcal{F}_\gamma^\rho$ together with $\sum_{l=1}^{k_n}\gamma_l^{-1} \leqslant \pi^2/6$ for n sufficiently large yields

$$\mathbb{E}\,\big\|[\widehat{\mathrm{Id}}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i)\big\|^2 \leqslant n^{-1}\sum_{j=1}^{k_n}\mathbb{E}\,\Big|e_j(Z)\sum_{l=1}^{k_n}[\varphi_0]_l e_l(Z) - Y e_j(Z)\Big|^2$$
$$\leqslant 2n^{-1}\|\varphi_0\|_\gamma^2\sum_{j=1}^{k_n}\mathbb{E}\left[e_j^2(Z)\sum_{l=1}^{k_n}\gamma_l^{-1}e_l^2(Z)\right] + 2n^{-1}k_n\eta^2\,\mathbb{E}[Y^4] \leqslant 2\eta^2\Big(\frac{\pi^2\eta^2\rho}{6} + \mathbb{E}[Y^4]\Big)n^{-1}k_n.$$

Moreover, $\mathbb{1}_{\mathcal{A}^c_{k_n}} = o_p(1)$ since $\mathbb{P}(\mathcal{B}^c_{k_n}) \leqslant 4\,\mathbb{E}\,\|[\widehat{\mathrm{Id}}]_{\underline{k_n}} - [\mathrm{Id}]_{\underline{k_n}}\|^2 = o(1)$ and $\mathcal{B}_{k_n} \subset \mathcal{A}_{k_n}$. Consequently,

$$n\|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2 = O_p(k_n) \tag{A.10}$$

and since $k_n = o(\varsigma_{m_n})$ we proved $nB_{n1} = o_p(\varsigma_{m_n})$. In addition, applying inequality (A.5) together with equation (A.10) yields $nB_{n2} = o_p(\varsigma_{m_n})$. Consequently, $nA_{n1} = o(\varsigma_{m_n})$. Consider A_{n2}. Jensen's inequality gives

$$\mathbb{E}\,\big\|n^{-1}\sum_{i=1}^n\big(E_{k_n}^\perp\varphi_0\big)(Z_i)f^\tau_{\underline{m_n}}(W_i)\big\|^2 \leqslant 2\|E_{k_n}^\perp\varphi_0\|_Z^2 + 2n^{-1}\sum_{j=1}^{m_n}\mathbb{E}\,|E_{k_n}^\perp\varphi_0(Z)f_j(W)|^2.$$

Note that condition (4.2) implies $k_n^2 < \sigma_{m_n}^2 \leqslant \sigma^4 m_n$ for n sufficiently large. Due to the Cauchy Schwarz inequality

$$\sum_{j=1}^{m_n}\mathbb{E}\,|E_{k_n}^\perp\varphi_0(Z)f_j(W)|^2 \leqslant \sum_{j=1}^{m_n}\sum_{l>k_n}l^2[\varphi_0]_l^2\sum_{l>k_n}l^{-2}\,\mathbb{E}\,|e_l(Z)f_j(W)|^2$$
$$\leqslant \eta^4\frac{\pi^2}{6}\frac{m_n k_n^2}{\gamma_{k_n}}\|E_{k_n}^\perp\varphi_0\|_\gamma^2 = o(\varsigma_{m_n}) \tag{A.11}$$

and $n\|E_{k_n}^\perp\varphi_0\|_Z^2 \leqslant n\gamma_{k_n}^{-1}\|E_{k_n}^\perp\varphi_0\|_\gamma^2 = o(\varsigma_{m_n})$. Hence, $nIII_n = o_p(\varsigma_{m_n})$. Consider II_n. We calculate

$$nII_n \leqslant \Big|\sum_{j=1}^{m_n}\tau_j\sum_{i=1}^{n}U_if_j(W_i)([\varphi_0]_{\underline{k_n}}-[\overline{\varphi}]_{\underline{k_n}})^t\Big(n^{-1}\sum_{i=1}^{n}e_{\underline{k_n}}(Z_i)f_j(W_i)-\mathbb{E}\left[e_{\underline{k_n}}(Z)f_j(W)\right]\Big)\Big|$$

$$+\Big|\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}([\varphi_0]_l-[\overline{\varphi}]_l)\Big(\sum_{i=1}^{n}U_if_j(W_i)[T]_{jl}\Big)\Big|$$

$$+\Big|\sum_{j=1}^{m_n}\tau_j\Big(\sum_{i=1}^{n}U_if_j(W_i)\Big)\Big(n^{-1}\sum_{i=1}^{n}E_{k_n}^{\perp}\varphi_0(Z_i)f_j(W_i)-\mathbb{E}[E_{k_n}^{\perp}\varphi_0(Z)f_j(W)]\Big)\Big|$$

$$+\Big|\sum_{j=1}^{m_n}\tau_j\Big(\sum_{i=1}^{n}U_if_j(W_i)\Big)\,\mathbb{E}[E_{k_n}^{\perp}\varphi_0(Z)f_j(W)]\Big| = C_{n1}+C_{n2}+C_{n3}+C_{n4}. \quad \text{(A.12)}$$

Consider C_{n1}. Applying twice the Cauchy Schwarz inequality gives

$$C_{n1} \leqslant \Big(\sum_{j=1}^{m_n}\tau_j^2\big|\sum_{i=1}^{n}U_if_j(W_i)\big|^2\Big)^{1/2}\|E_{k_n}\varphi_0-\overline{\varphi}_{k_n}\|_Z \times\Big(\sum_{j=1}^{m_n}\sum_{l=1}^{k_n}|n^{-1}\sum_{i=1}^{n}e_l(Z_i)f_j(W_i)-\mathbb{E}[e_l(Z)f_j(W)]|^2\Big)^{1/2}.$$

From $\mathbb{E}\,|\sum_{i=1}^n U_if_j(W_i)|^2 \leqslant n\sigma^2$, relation (A.10), and inequality (A.5) we infer $C_{n1} = o_p(\varsigma_{m_n})$ due to condition (4.2). For C_{n2} we evaluate

$$C_{n2} \leqslant \|E_{k_n}\varphi_0-\overline{\varphi}_{k_n}\|_Z\Big(\sum_{l=1}^{k_n}\big|\sum_{j=1}^{m_n}\sum_{i=1}^{n}U_if_j(W_i)[T]_{jl}\big|^2\Big)^{1/2}$$

Estimate $\sum_{j=1}^{m_n}\sum_{l=1}^{k_n}[T]_{jl}^2 \leqslant k_n$ together with (A.10) yields $C_{n2} = o_p(1)$. Consider C_{n3}. Since $\mathbb{E}[U^2|W] \leqslant \sigma^2$ we conclude similarly as in inequality (A.11) that

$$\mathbb{E}\,C_{n3} \leqslant \sum_{j=1}^{m_n}\tau_j\big(\mathbb{E}\,|Uf_j(W)|^2\big)^{1/2}\big(\mathbb{E}\,|E_{k_n}^{\perp}\varphi_0(Z)f_j(W)|^2\big)^{1/2} \leqslant \eta^2\frac{\pi\sigma}{\sqrt{6}}\frac{k_n}{\sqrt{\gamma_{k_n}}}\|E_{k_n}^{\perp}\varphi_0\|_\gamma\sum_{j=1}^{m_n}\tau_j = o(\varsigma_{m_n})$$

where we used $\sum_{j=1}^{m_n}\tau_j \leqslant \sqrt{m_n}\big(\sum_{j=1}^{m_n}\tau_j^2\big)^{1/2}$. Consider C_{n4}. We calculate

$$\mathbb{E}\,|C_{n4}|^2 \leqslant n\sigma^2\sum_{j=1}^{m_n}[TE_{k_n}^{\perp}\varphi_0]_j^2 \leqslant n\sigma^2\|TE_{k_n}^{\perp}\varphi_0\|_W^2 \leqslant n\gamma_{k_n}^{-1}\sigma^2\|E_{k_n}^{\perp}\varphi_0\|_\gamma^2 = o(\varsigma_{m_n}).$$

Consequently, in light of decomposition (A.12) we obtain $nII_n = o(\varsigma_{m_n})$, which completes the proof. □

Proof of Theorem 4.2. Employing $[\widehat{\mathrm{Id}}]_{\underline{k_n}}^{-1} = [\mathrm{Id}]_{\underline{k_n}} - [\widehat{\mathrm{Id}}]_{\underline{k_n}}^{-1}([\widehat{\mathrm{Id}}]_{\underline{k_n}} - [\mathrm{Id}]_{\underline{k_n}})$ and $[\widehat{\mathrm{Id}}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i) = n^{-1}\sum_{i=1}^n e_{\underline{k_n}}(Z_i)(E_{k_n}\varphi(Z_i) - Y_i)$ we obtain for all $1 \leqslant j \leqslant m_n$

$$n^{-1/2}\sum_{i=1}^n f_j(W_i)\Big(U_i + e_{\underline{k_n}}(Z_i)^t\big([\varphi_0]_{\underline{k_n}} - [\widehat{\mathrm{Id}}]_{\underline{k_n}}^{-1}\frac{1}{n}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i)\big) + E_{k_n}^\perp \varphi_0(Z_i)\Big)$$
$$= n^{-1/2}\sum_{i=1}^n \Big(f_j(W_i)U_i + \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right]e_{\underline{k_n}}(Z_i)\big(E_{k_n}\varphi_0(Z_i) - Y_i\big)\Big)$$
$$+ n^{-1/2}\sum_{i=1}^n \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][\widehat{\mathrm{Id}}]_{\underline{k_n}}^{-1}([\widehat{\mathrm{Id}}]_{\underline{k_n}} - [\mathrm{Id}]_{\underline{k_n}})e_{\underline{k_n}}(Z_i)\big(E_{k_n}\varphi_0(Z_i) - Y_i\big)$$
$$+ \Big(n^{-1}\sum_{i=1}^n f_j(W_i)e_l(Z_i) - \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right]\Big)\sqrt{n}\big([\varphi_0]_{\underline{k_n}} - [\widehat{\mathrm{Id}}]_{\underline{k_n}}^{-1}\frac{1}{n}\sum_{i=1}^n Y_i e_{\underline{k_n}}(Z_i)\big)$$
$$+ n^{-1/2}\sum_{i=1}^n E_{k_n}^\perp \varphi_0(Z_i) f_j(W_i) = A_{nj} + B_{nj} + C_{nj} + D_{nj}. \quad \text{(A.13)}$$

For all $1 \leqslant i \leqslant n$ the random variables $U_i f_j^\tau(W_i) + \big(E_{k_n}\varphi_0(Z_i) - Y_i\big)\,\mathbb{E}\left[f_j^\tau(W)e_{\underline{k_n}}(Z)^t\right]e_{\underline{k_n}}(Z_i)$, $1 \leqslant j \leqslant m_n$, are centered under H_{p} with bounded fourth moment. More precisely, due to condition $\mathbb{E}[U_Z^4|Z] \leqslant \sigma^4$ where $U_Z = E_{k_n}\varphi_0(Z) - Y$ we calculate for all $1 \leqslant j \leqslant m_n$

$$\mathbb{E}\left|U f_j(W) + \sum_{l=1}^{k_n}[T]_{jl}e_l(Z)U_Z\right|^4 \leqslant 8\,\mathbb{E}\left|U f_j(W)\right|^4 + 8\,\mathbb{E}\left|(E_{k_n}T^* f_j)(Z)U_Z\right|^4$$
$$\leqslant 8\sigma^4\eta^4 + 8\sigma^4\,\mathbb{E}\left|(E_{k_n}T^* f_j)(Z)\right|^4$$

which is bounded since $\mathbb{E}\left|(E_{k_n}T^* f_j)(Z)\right|^4 \leqslant \mathbb{E}\left|(T^* f_j)(Z)\right|^4 \leqslant \mathbb{E}\left|f_j(W)\right|^4 \leqslant \eta^4$ by using well known properties of projections on Banach spaces. Now following line by line the proof of Lemma A.2 it is easily seen that $(\sqrt{2}\varsigma_{m_n}^{\mathrm{e}})^{-1}\big(\sum_{j=1}^{m_n}\tau_j A_{nj}^2 - \mu_{m_n}^{\mathrm{e}}\big) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}^{\mathrm{e}})$. Moreover, similarly to the proof of Theorem 4.1 it is easily seen that $\sum_{j=1}^{m_n} B_{nj}^2 = O_p\big(n^{-1}m_n k_n^2\big) = o_p(1)$, $\sum_{j=1}^{m_n} C_{nj}^2 = o_p(1)$ and $\sum_{j=1}^{m_n} D_{nj}^2 = o_p(1)$. Since $\mathbb{1}_{\mathcal{A}_{k_n}} = 1 + o_p(1)$ the result follows similarly to the proof of Theorem 3.2. □

Proof of Proposition 4.3. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Similar to the proof of Proposition 3.3 it is sufficient to show

$$\langle n^{-1}\sum_{i=1}^n \delta(Z_i) f_{\underline{m_n}}^\tau(W_i), n^{-1/2}\sum_{i=1}^n(\varphi_0(Z_i) - \overline{\varphi}_{k_n}(Z_i)) f_{\underline{m_n}}^\tau(W_i)\rangle = o_p(\sqrt{\varsigma_{m_n}}). \quad \text{(A.14)}$$

By employing Jensen's inequality and estimate (A.10) we obtain

$$\sum_{j=1}^{m_n}\tau_j[T\delta]_j\frac{1}{\sqrt{n}}\sum_{i=1}^n(E_{k_n}\varphi_0(Z_i) - \overline{\varphi}_{k_n}(Z_i))f_j(W_i)$$
$$\leqslant \sqrt{n}\|T\delta\|_\tau\|T(E_{k_n}\varphi_0 - \overline{\varphi}_{k_n})\|_W + o_p(1) = o_p(\varsigma_{m_n}).$$

Similarly to the upper bounds of C_{n3} and C_{n4} in the proof of Theorem 4.1 it is straightforward to see that $\sum_{j=1}^{m_n}\tau_j[T\delta]_j n^{-1/2}\sum_{i=1}^n E_{k_n}^\perp\varphi_0(Z_i)f_j(W_i) = o_p(\varsigma_{m_n})$ and, hence equation (A.14) holds true. Consider the case $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$. We make use of decomposition

(A.13) where U_i is replaced by $U_i + n^{-1/2}\delta(Z_i)$. Similarly to the proof of Proposition 2.4 it is easily seen that $(\sqrt{2}\varsigma^{\mathrm{e}}_{m_n})^{-1}\big(\sum_{j=1}^{m_n}\tau_j A_{nj}^2 - \mu^{\mathrm{e}}_{m_n}\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma^{\mathrm{p}}_{\infty})^{-1}\|T\delta\|_{\tau}^2, \mathcal{V}^{\mathrm{e}}\big)$. Thereby, due to the proof of Theorem 4.2, the assertion follows. □

Proof of Proposition 4.4. Similar to the proof of Proposition 3.4. □

Proof of Proposition 4.5. We make use of inequality (A.8) where $\breve{\varphi}_k$ is replaced by $\overline{\varphi}_{k_n}$. From the proof of Proposition 4.1 we infer $\big\|n^{-1/2}\sum_{i=1}^n(\overline{\varphi}_{k_n}(Z_i)-\varphi_0(Z_i))f^{\tau}_{\underline{m_n}}(W_i)\big\|^2 = o_p(\varsigma_{m_n})$ and

$$\langle n^{-1}\sum_{i=1}^n(\varphi(Z_i) - \overline{\varphi}_{k_n}(Z_i))f^{\tau}_{\underline{m_n}}(W_i), \sum_{i=1}^n U_i f^{\tau}_{\underline{m_n}}(W_i)\rangle$$
$$= \langle n^{-1}\sum_{i=1}^n(\varphi(Z_i) - \varphi_0(Z_i))f^{\tau}_{\underline{m_n}}(W_i), \sum_{i=1}^n U_i f^{\tau}_{\underline{m_n}}(W_i)\rangle + o_p(\varsigma_{m_n})$$

uniformly over all $\varphi \in \mathcal{I}_n^{\rho}$. In addition, let s_j be as in the proof of Proposition 2.6 then condition $\sup_{j\geqslant 1}\mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$ yields

$$\mathbb{E}\Big|\sum_{j=1}^{m_n} s_j(\varphi(Z)-\varphi_0(Z))f_j(W)\Big|^2 \leqslant \|\varphi-\varphi_0\|_{\gamma}^2\sum_{l=1}^{\infty}\gamma_l^{-1}\,\mathbb{E}\,\Big|e_l(Z)\sum_{j=1}^{m_n}s_jf_j(W)\Big|^2 = O(1). \quad \text{(A.15)}$$

Thus, following line by line the proof of Proposition 2.6, the assertion follows. In case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ the assertion follows similarly. □

A.4. Proofs of Section 5.

Proof of Theorem 5.1. For the proof we make use of decomposition (A.4) where the estimator $\breve{\varphi}_k$ is replaced by $\widehat{\varphi}_{k_n}$ given in (5.1). Consider III_n. Observe

$$III_n \leqslant 2\|n^{-1}\sum_{i=1}^n(\varphi_{k_n}(Z_i) - \widehat{\varphi}_{k_n}(Z_i))f^{\tau}_{\underline{m_n}}(W_i)\|^2$$
$$+ 2\|n^{-1}\sum_{i=1}^n\big(\varphi_{k_n}(Z_i) - \varphi_0(Z_i)\big)f^{\tau}_{\underline{m_n}}(W_i)\|^2 = 2A_{n1} + 2A_{n2}. \quad \text{(A.16)}$$

Consider A_{n1}. We evaluate by applying Cauchy Schwarz inequality

$$A_{n1} \leqslant 2\|T(\varphi_{k_n} - \widehat{\varphi}_{k_n})\|_W^2 + 2\|\varphi_{k_n} - \widehat{\varphi}_{k_n}\|_{\upsilon}^2\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}\upsilon_l^{-1}|n^{-1}\sum_{i=1}^n e_l(Z_i)f_j(W_i) - [T]_{jl}|^2.$$

The link condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ yields $\|T(\widehat{\varphi}_{k_n} - \varphi_{k_n})\|_W^2 \leqslant d\|\widehat{\varphi}_{k_n} - \varphi_{k_n}\|_{\upsilon}^2$. From Theorem 2.6 of Johannes and Schwarz [2010] and condition (5.4) we infer $n\|\widehat{\varphi}_{k_n} - \varphi_{k_n}\|_{\upsilon}^2 = O_p\big(\max(n\upsilon_{k_n}\gamma_{k_n}^{-1}, k_n)\big) = o_p(\varsigma_{m_n})$. This together with estimate (A.5) implies $nA_{n1} =$

$o_p(\varsigma_{m_n})$. Consider A_{n2}. We observe

$$\mathbb{E}\,A_{n2} \leqslant 2\|T(\varphi_{k_n}-\varphi_0)\|_W^2 + 2n^{-1}\,\mathbb{E}\,\|\big(\varphi_{k_n}(Z)-\varphi_0(Z)\big)f^{\tau}_{\underline{m_n}}(W)\|^2$$
$$\leqslant 2d\|\varphi_{k_n}-\varphi_0\|_v^2 + 2n^{-1}\sum_{l\geqslant 1} l^2([\varphi_{k_n}]_l-[\varphi_0]_l)^2\sum_{j=1}^{m_n}\tau_j\sum_{l\geqslant 1} l^{-2}\,\mathbb{E}\,|e_l(Z)f_j(W)|^2$$
$$\leqslant 8Dd^2\rho\Big(\frac{\upsilon_{k_n}}{\gamma_{k_n}}\|\varphi_{k_n}-\varphi_0\|_\gamma^2 + \frac{\pi^2}{6}\eta^4\|\varphi_{k_n}-\varphi_0\|_\gamma^2\frac{k_n^2}{n\gamma_{k_n}}\sum_{j=1}^{m_n}\tau_j\Big). \quad \text{(A.17)}$$

where we used Lemma A.2 of Johannes and Schwarz [2010], i.e., $\|\varphi_{k_n}-\varphi_0\|_w^2 \leqslant 4Dd\rho w_{k_n}\gamma_{k_n}^{-1}$ for a nondecreasing sequence w. Condition (5.4) together with the estimate $k_n^2 \leqslant \sigma^4\sum_{j=1}^{m_n}\tau_j$ for n sufficiently large implies $nA_{n2} = o_p(\varsigma_{m_n})$. Consequently, due to (A.16) we have shown $nIII_n = o_p(\varsigma_{m_n})$. The proof of $nII_n = o_p(\varsigma_{m_n})$ is based on decomposition (A.12) where $\overline{\varphi}_{k_n}$ and $E_{k_n}^{\perp}\varphi_0$ are replaced by $\widehat{\varphi}_{k_n}$ and $\varphi_{k_n}-\varphi_0$, respectively. Consider C_{n1}. We calculate

$$C_{n1} \leqslant \|\widehat{\varphi}_{k_n}-\varphi_{k_n}\|_v\sum_{j=1}^{m_n}\tau_j\Big|\sum_{i=1}^{n}U_if_j(W_i)\Big|\Big(\sum_{l=1}^{k_n}\upsilon_l^{-1}\Big|n^{-1}\sum_{i=1}^{n}e_l(Z_i)f_j(W_i)-[T]_{jl}\Big|^2\Big)^{1/2}$$

Since $\sqrt{n}\|\widehat{\varphi}_{k_n}-\varphi_{k_n}\|_v = o_p(\varsigma_{m_n}^{1/2})$ we obtain, similarly as in the proof of Theorem 4.1, $C_{n1} = o_p(\varsigma_{m_n})$. Consider C_{n2}. Again similarly to the proof of Theorem 4.1 we observe

$$\mathbb{E}\,C_{n2} = \mathbb{E}\Big|\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}[T]_{jl}([\widehat{\varphi}_{k_n}]_l-[\varphi_{k_n}]_l)\Big(\sum_{i=1}^{n}U_if_j(W_i)\Big)\Big|$$
$$\leqslant \big(n\,\mathbb{E}\,\|\widehat{\varphi}_{k_n}-\varphi_{k_n}\|_v^2\big)^{1/2}\Big(\sigma^2\sum_{l=1}^{k_n}\upsilon_l^{-1}\sum_{j=1}^{m_n}[T]_{jl}^2\Big)^{1/2} = o(\varsigma_{m_n})$$

by exploiting $\sum_{j=1}^{m_n}[T]_{jl}^2 = \|Te_l\|_W^2 \leqslant d\upsilon_l$. Consider C_{n3}. Since $\mathbb{E}[U^2|W] \leqslant \sigma^2$ we conclude similarly as in inequality (A.11) using Lemma A.2 of Johannes and Schwarz [2010]

$$\mathbb{E}\,C_{n3} \leqslant \sigma\sum_{j=1}^{m_n}\tau_j\big(\,\mathbb{E}\,|(\varphi_{k_n}(Z)-\varphi_0(Z))f_j(W)|^2\big)^{1/2} \leqslant \eta^2\frac{\pi\sigma}{\sqrt{6}}\frac{k_n}{\sqrt{\gamma_{k_n}}}\|\varphi_{k_n}-\varphi_0\|_\gamma\sum_{j=1}^{m_n}\tau_j = o(\varsigma_{m_n}).$$

Consider C_{n4}. Again exploring the link condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ and Lemma A.2 of Johannes and Schwarz [2010] we calculate

$$\mathbb{E}\,|C_{n4}|^2 \leqslant n\sigma\sum_{j=1}^{m_n}[T(\varphi_{k_n}-\varphi_0)]_j^2 \leqslant n\sigma\|T(\varphi_{k_n}-\varphi_0)\|_W^2$$
$$\leqslant n\sigma d\|\varphi_{k_n}-\varphi_0\|_v^2 \leqslant 4Dd\rho\sigma\frac{n\upsilon_{k_n}}{\gamma_{k_n}}\|\varphi_{k_n}-\varphi_0\|_\gamma^2 = o(\varsigma_{m_n}).$$

Consequently, the estimates for C_{n1}, C_{n2}, C_{n3}, and C_{n4} imply $nII_n = o_p(\varsigma_{m_n})$, which completes the proof. □

Proof of Theorem 5.2. For all $k \geqslant 1$ let us denote $\Omega_k := \{\|[\widehat{T}]_{\underline{k}}^{-1}\| \leqslant \sqrt{n}\}$ and $\mho_k := \{\|Q_k\|\|[T]_{\underline{k}}^{-1}\| \leqslant 1/2\}$ where $Q_k = [\widehat{T}]_{\underline{k}} - [T]_{\underline{k}}$. Let $\varphi_{k_n}(\cdot) := e_{\underline{k_n}}(\cdot)^t[T]_{\underline{k_n}}^{-1}[g]_{\underline{k_n}}$. Observe

$[\widehat{T}]_{\underline{k_n}}[\varphi_{k_n}]_{\underline{k_n}} - [\widehat{g}]_{\underline{k_n}} = n^{-1}\sum_{i=1}^n f_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)$ and hence, for all $1 \leqslant j \leqslant m_n$

$$
\begin{aligned}
n^{-1/2}\sum_{i=1}^n f_j(W_i)\Big(U_i + e_{\underline{k_n}}(Z_i)^t\big([\varphi_{k_n}]_{\underline{k_n}} - [\widehat{T}]_{\underline{k_n}}^{-1}[\widehat{g}]_{\underline{k_n}}\big) + \varphi_0(Z_i) - \varphi_{k_n}(Z_i)\Big)\\
= n^{-1/2}\sum_{i=1}^n \Big(f_j(W_i)U_i + \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\Big)\\
- n^{-1/2}\sum_{i=1}^n \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} Q_{k_n}[\widehat{T}]_{\underline{k_n}}^{-1} f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\\
+\Big(n^{-1}\sum_{i=1}^n f_j(W_i)e_{\underline{k_n}}(Z_i)^t - \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right]\Big)[\widehat{T}]_{\underline{k_n}}^{-1}\Big(n^{-1/2}\sum_{i=1}^n f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\Big)\\
+ n^{-1/2}\sum_{i=1}^n \big(\varphi_0(Z_i) - \varphi_{k_n}(Z_i)\big)f_j(W_i) = A_{nj} + B_{nj} + C_{nj} + D_{nj}. \quad \text{(A.18)}
\end{aligned}
$$

Consider A_{nj}. The random variables $U_i f_j^\tau(W_i) + \mathbb{E}\left[f_j^\tau(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)$, $1 \leqslant j \leqslant m_n$, are centered with bounded second moment. More precisely, condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ together with Lemma A.1 of Breunig and Johannes [2011] yields

$$
\mathbb{E}\,\big|\,\mathbb{E}[f_j(W)e_{\underline{k_n}}(Z)^t][T]_{\underline{k_n}}^{-1} f_{\underline{k_n}}(W)\big(\varphi_{k_n}(Z) - Y\big)\big|^2 \leqslant 2D\sum_{l=1}^{k_n} \upsilon_l^{-1}[T]_{jl}^2(\sigma^2 + C(\gamma)\eta^2\|\varphi - \varphi_{k_n}\|_\gamma^2).
$$

Moreover, exploiting condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ yields $\sum_{l=1}^{k_n}\upsilon_l^{-1}[T]_{jl}^2 = \|F_{k_n}\sqrt{\nabla_{1/\upsilon}}T^* f_j\|_W^2 \leqslant \|T\sqrt{\nabla_{1/\upsilon}^*}\|^2 \leqslant d$. In addition we calculate for all $1 \leqslant j \leqslant m_n$

$$
\mathbb{E}\left|U f_j^\tau(W) + \mathbb{E}\left[f_j^\tau(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} f_{\underline{k_n}}(W)(\varphi_{k_n}(Z) - Y)\right|^4 = O(k_n)
$$

since $\|\,\mathbb{E}[f_j^\tau(W)e_{\underline{k_n}}(Z)^t][\nabla_\upsilon]_{\underline{k_n}}^{-1/2}\|^4 = \big(\sum_{l=1}^{k_n}\upsilon_l^{-1}[T]_{jl}^2\big)^2 \leqslant d^4$ and $\mathbb{E}\,|f_j(W)(Y - \varphi_{k_n}(Z))|^4 = O(1)$ (cf. Lemma A.1 of Breunig and Johannes [2011]). Thus, with the fourth moment growing only at rate k_n, by following line by line the proof of Lemma A.2 it is easily seen that $(\sqrt{2}\varsigma_{m_n}^{\mathrm{np}})\big(\sum_{j=1}^{m_n}\tau_j A_{nj}^2 - \mu_{m_n}^{\mathrm{np}}\big) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}^{\mathrm{np}})$. Note that $\mathbb{E}\,\mathbb{1}_{\Omega_{k_n}^c} = \mathbb{P}(\Omega_{k_n}^c) = o(1)$ (cf. proof of Proposition 3.1 of Breunig and Johannes [2011]) and, hence $\mathbb{1}_{\Omega_{k_n}} = 1 + o_p(1)$. Consider B_{nj}. By employing $\|[\widehat{T}]_{\underline{k}}^{-1}\|\,\mathbb{1}_{\mho_k} \leqslant 2\|[T]_{\underline{k}}^{-1}\|$ and $\|[\widehat{T}]_{\underline{k}}^{-1}\|^2\,\mathbb{1}_{\Omega_k} \leqslant n$ for all $k \geqslant 1$ it follows

$$
\begin{aligned}
\sum_{j=1}^{m_n} B_{nj}^2\,\mathbb{1}_{\Omega_{k_n}} &= \sum_{j=1}^{m_n} B_{nj}^2\,\mathbb{1}_{\Omega_{k_n}}(\mathbb{1}_{\mho_{k_n}} + \mathbb{1}_{\mho_{k_n}^c})\\
&\leqslant \big\|\,\mathbb{E}[f_{\underline{m_n}}(W)e_{\underline{k_n}}(Z)^t][T]_{\underline{k_n}}^{-1}\big\|^2\Big(4\|[T]_{\underline{k_n}}^{-1}\|^2\|Q_{k_n}\|^2\|n^{-1/2}\sum_{i=1}^n f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\|^2\\
&\qquad + n\|Q_{k_n}\|^2\|n^{-1/2}\sum_{i=1}^n f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\|^2\,\mathbb{1}_{\mho_{k_n}^c}\Big)
\end{aligned}
$$

Condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ implies $\|\,\mathbb{E}\,f_{\underline{m_n}}^\tau(W)e_{\underline{k_n}}(Z)^t[T]_{\underline{k_n}}^{-1}\|^2 \leqslant D\|\,\mathbb{E}\,f_{\underline{m_n}}^\tau(W)e_{\underline{k_n}}(Z)^t[\nabla_\upsilon]_{\underline{k_n}}^{-1/2}\|^2$. Moreover, let K be a linear operator on $\mathcal{L}_Z^2$ with eigenvalue decomposition $\{\upsilon_j^{1/2}, e_j\}_{j\geqslant 1}$

and let $\mathcal{S}_{k_n} = \{\phi \in \mathcal{E}_{k_n}, \|\phi\|_Z = 1\}$, then by the spectral theorem

$$\|\mathbb{E} f_{\underline{m_n}}(W) e_{\underline{k_n}}(Z)^t [\nabla_\upsilon]_{\underline{k_n}}^{-1/2}\|^2 = \sup_{\phi\in\mathcal{S}_{k_n}} \sum_{j=1}^{m_n} \Big|\sum_{l=1}^{k_n} [T]_{jl} \upsilon_l^{-1/2} [\phi]_l\Big|^2$$

$$= \sup_{\phi\in\mathcal{S}_{k_n}} \sum_{j=1}^{m_n} [TK^{-1}\phi]_j^2 \leqslant \sup_{\phi\in\mathcal{S}_{k_n}} \|TK^{-1}\phi\|_W^2 \leqslant d \sup_{\phi\in\mathcal{S}_{k_n}} \|K^{-1}\phi\|_\upsilon^2 = d.$$

Now $n\|Q_{k_n}\|^2 = O_p(k_n^2)$ and $\|n^{-1/2}\sum_{i=1}^n f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\|^2 = O_p(k_n)$ due to Lemma A.1 of Breunig and Johannes [2011]. In addition similarly to their proof of Proposition 3.1 it can be seen that $n\|Q_{k_n}\|^2\|n^{-1/2}\sum_{i=1}^n f_{\underline{k_n}}(W_i)\big(\varphi_{k_n}(Z_i) - Y_i\big)\|^2 \mathbb{1}_{\mho_{k_n}^c} = o_p(1)$. Consequently, $\sum_{j=1}^{m_n} B_{nj}^2 \mathbb{1}_{\Omega_{k_n}} = o_p(1)$. Similarly, it is easily seen that $\sum_{j=1}^{m_n} C_{nj}^2 \mathbb{1}_{\Omega_{k_n}} = o_p(1)$ and $\sum_{j=1}^{m_n} D_{nj}^2 = o_p(1)$. Hence, since $\mathbb{1}_{\Omega_{k_n}} = 1 + o_p(1)$ the result follows similarly to the proof of Theorem 3.2. □

Proof of Proposition 5.3. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Similar to the proof of Proposition 3.3 it is sufficient to show

$$\langle n^{-1}\sum_{i=1}^n \delta(Z_i) f_{\underline{m_n}}^\tau(W_i), n^{-1/2}\sum_{i=1}^n (\varphi_0(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_{\underline{m_n}}^\tau(W_i)\rangle = o_p(\sqrt{\varsigma_{m_n}}). \qquad (A.19)$$

Due to the link condition $T \in \mathcal{T}_{d,D}^\upsilon$ we obtain

$$\sum_{j=1}^{m_n} \tau_j [T\delta]_j \frac{1}{\sqrt{n}} \sum_{i=1}^n (\varphi_{k_n}(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_j(W_i) \leqslant \sqrt{dn}\|T\delta\|_\tau \|\varphi_{k_n} - \widehat{\varphi}_{k_n}\|_\upsilon + o_p(1) = o_p(\varsigma_{m_n}).$$

As in the proof of Theorem 5.1 it can be seen $\sum_{j=1}^{m_n} \tau_j [T\delta]_j \sum_{i=1}^n (\varphi_0(Z_i) - \varphi_{k_n}(Z_i)) f_j(W_i) = o_p(\sqrt{n}\varsigma_{m_n})$ and, hence equation (A.19) holds true. Consider the case $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$. We make use of decomposition (A.18) where U_i is replaced by $U_i + n^{-1/2}\delta(Z_i)$. Similarly to the proof of Proposition 2.4 it is seen that $(\sqrt{2}\varsigma_{m_n}^{\mathrm{np}})^{-1}\big(\sum_{j=1}^{m_n} \tau_j A_{nj}^2 - \mu_{m_n}^{\mathrm{np}}\big) \xrightarrow{d} \mathcal{N}\big((\sqrt{2}\varsigma_\infty^{\mathrm{np}})^{-1}\|T\delta\|_\tau^2, \mathcal{V}^{\mathrm{np}}\big)$. Thereby, due to the proof of Theorem 3.2, the assertion follows. □

Proof of Proposition 5.4. Similar to the proof of Proposition 3.4. □

Proof of Proposition 5.5. We make use of inequality (A.8) where $\breve{\varphi}_k$ is replaced by $\widehat{\varphi}_{k_n}$. From the proof of Proposition 5.1 we infer $\big\|n^{-1/2}\sum_{i=1}^n (\widehat{\varphi}_{k_n}(Z_i) - \varphi_0(Z_i)) f_{\underline{m_n}}^\tau(W_i)\big\|^2 = o_p(\varsigma_{m_n})$ and

$$\langle n^{-1}\sum_{i=1}^n (\varphi(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_{\underline{m_n}}^\tau(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\rangle$$

$$= \langle n^{-1}\sum_{i=1}^n (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m_n}}^\tau(W_i), \sum_{i=1}^n U_i f_{\underline{m_n}}^\tau(W_i)\rangle + o_p(\varsigma_{m_n})$$

uniformly over all $\varphi \in \mathcal{J}_n^\rho$. Consequently, by using inequality (A.15) and following line by line the proof of Proposition 2.6, the assertion follows. In case of $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$ the assertion follows similarly. □

A.5. Technical assertions.

Let us introduce $X_{ii'} := \sqrt{2}(\varsigma_{m_n} n)^{-1} \sum_{j=1}^{m_n} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'})$ and

$$Q_{ni} := \begin{cases} \sum_{l=1}^{i-1} X_{li}, & \text{for } i = 2, \dots, n, \\ 0, & \text{for } i = 1 \text{ and } i > n. \end{cases} \tag{A.20}$$

Then clearly

$$(\sqrt{2}\varsigma_{m_n} n)^{-1} \sum_{i \neq i'} \sum_{j=1}^{m_n} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'}) = \sqrt{2}(\varsigma_{m_n} n)^{-1} \sum_{i<i'} \sum_{j=1}^{m_n} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'})$$
$$= \sum_{i<i'} X_{ii'} = \sum_{i=1}^{n} Q_{ni}.$$

Let $\mathcal{B}_{ni} := \mathcal{B}((Z_1, Y_1, W_1), \dots, (Z_i, Y_i, W_i))$, $1 \leqslant i \leqslant n$, $n \geqslant 1$, be the σ-algebra generated by $(Z_1, Y_1, W_1), \dots, (Z_i, Y_i, W_i)$. Since $U_i f_j^\tau(W_i)$, $1 \leqslant i \leqslant n$, are centered random variables it follows that $\{(\sum_{i'=1}^{i} Q_{ni'}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale for each $n \geqslant 1$ and hence $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale difference array for each $n \geqslant 1$. Moreover, it satisfies the conditions of Proposition A.1 as shown in the following technical result.

Proposition A.1. *If $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale difference array for each $n \geqslant 1$ satisfying conditions*

$$\sum_{i=1}^{\infty} \mathbb{E}\,|Q_{ni}|^2 \leqslant 1 \quad \textit{for all } n \geqslant 1, \tag{A.21}$$

$$\sum_{i=1}^{\infty} Q_{ni}^2 = \nu + o_p(1) \quad \textit{for some constant } \nu > 0, \tag{A.22}$$

$$\sup_{i \geqslant 1} |Q_{ni}| = o_p(1) \tag{A.23}$$

then

$$\sum_{i=1}^{\infty} Q_{ni} \xrightarrow{d} N(0, \nu).$$

Proof. See Awad [1981]. □

Note that this result has been also applied by Ghorai [1980] to establish asymptotic normality of an orthogonal series type density estimator.

Lemma A.2. *Let Q_{ni} be defined as in (A.20). Let Assumptions 1–4 be satisfied and assume $m_n\big(\sum_{j=1}^{m_n} \tau_j^2\big)^2 = o(n)$. Then the conditions (A.21)–(A.23) hold true where $\nu = 1$ if $\varsigma_{m_n}^{-1} = o(1)$ and $\nu = 1 + \big(4/3\varsigma_\infty^4\big) \sum_{j,j',l,l'=1}^{\infty} \varsigma_{jj'} \varsigma_{ll'} \varsigma_{jl} \varsigma_{j'l'}$ if $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$.*

Proof. Proof of (A.21). Observe that $\mathbb{E}[X_{1i} X_{1i'}] = 0$ for $i \neq i'$ and thus, for $i = 2, \dots, n$ we have

$$\mathbb{E}\,|Q_{ni}|^2 = \mathbb{E}\,|X_{1i} + \dots + X_{i-1,i}|^2 = (i-1)\,\mathbb{E}\,|X_{12}|^2 = \frac{2(i-1)}{n^2 \varsigma_{m_n}^2}\,\mathbb{E}\,|\sum_{j=1}^{m_n} U_1 f_j^\tau(W_1) U_2 f_j^\tau(W_2)|^2$$
$$= \frac{2(i-1)}{n^2 \varsigma_{m_n}^2} \sum_{j,j'=1}^{m_n} (\mathbb{E}\, U^2 f_j^\tau(W) f_{j'}^\tau(W))^2 = \frac{2(i-1)}{n^2}$$

by the definition of ς_{m_n}. Thereby, we conclude

$$\sum_{i=1}^{n} \mathbb{E}\,|Q_{ni}|^2 = \frac{2}{n^2}\sum_{i=1}^{n-1} i = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n} \tag{A.24}$$

which proves (A.21).
Proof of (A.22). We make use of the decomposition

$$\Big|\sum_{i=1}^{n} Q_{ni}^2\Big|^2 = \sum_{i=1}^{n} Q_{ni}^4 + \sum_{i\neq i'} Q_{ni}^2 Q_{ni'}^2 =: I_n + II_n. \tag{A.25}$$

Consider I_n. Observe that

$$\begin{aligned}\mathbb{E}\,|Q_{ni}|^4 = \mathbb{E}\,\Big|\sum_{i'=1}^{i-1} X_{i'i}\Big|^4 &= \mathbb{E}\,\Big|\frac{\sqrt{2}}{n\varsigma_{m_n}}\sum_{j=1}^{m_n}\tau_j U_i f_j(W_i)\sum_{i'=1}^{i-1} U_{i'} f_j(W_{i'})\Big|^4 \\ &\leqslant \frac{4m_n}{n^4\varsigma_{m_n}^4}\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2\sum_{j=1}^{m_n}\mathbb{E}\,|Uf_j(W)|^4\,\mathbb{E}\,\Big|\sum_{i'=1}^{i-1} U_{i'} f_j(W_{i'})\Big|^4 \\ &= \frac{4m_n}{n^4\varsigma_{m_n}^4}\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2\sum_{j=1}^{m_n}\mathbb{E}\,|Uf_j(W)|^4\Big((i-1)\,\mathbb{E}\,|Uf_j(W)|^4 + 3(i-1)(i-2)\varsigma_{jj}^2\Big)\end{aligned}$$

where we used that $\mathbb{E}[Uf_j(W)] = 0$. Since $\sum_{i=1}^{n} 3(i-1)(i-2) = n(n-1)(n-2)$ (proof by induction) and τ is nonincreasing with $\tau_1 = 1$ we conclude

$$\sum_{i=1}^{n}\mathbb{E}\,Q_{ni}^4 \leqslant \frac{4m_n}{n^4\varsigma_{m_n}^4}\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2\Big(\frac{n(n-1)}{2}\sum_{j=1}^{m_n}(\mathbb{E}\,|Uf_j(W)|^4)^2 + n(n-1)(n-2)\sum_{j=1}^{m_n}\varsigma_{jj}^2\,\mathbb{E}\,|Uf_j(W)|^4\Big)$$

By using Assumptions (1) and (2), i.e., $\sup_{j\in\mathbb{N}}\mathbb{E}[f_j^4(W)] \leqslant \eta^4$ and $\mathbb{E}[U^4|W] \leqslant \sigma^4$, we get $\max_{1\leqslant j\leqslant m_n}\mathbb{E}\,|Uf_j(W)|^4 \leqslant \eta^4\sigma^4$ and thus, $\sum_{i=1}^{n}\mathbb{E}\,|Q_{ni}|^4 = o(1)$. Consider II_n. We calculate for $i < i'$

$$\begin{aligned}Q_{ni}^2 Q_{ni'}^2 = \Big(\sum_{k=1}^{i-1} X_{ki}^2\Big)\Big(\sum_{k=1}^{i'-1} X_{ki'}^2\Big) + \Big(\sum_{k=1}^{i-1} X_{ki}^2\Big)\Big(\sum_{k\neq k'}^{i'-1} X_{ki'}X_{k'i'}\Big) \\ + \Big(\sum_{k\neq k'}^{i-1} X_{ki}X_{k'i}\Big)\Big(\sum_{k=1}^{i'-1} X_{ki'}^2\Big) + \Big(\sum_{k\neq k'}^{i-1} X_{ki}X_{k'i}\Big)\Big(\sum_{k\neq k'}^{i'-1} X_{ki'}X_{k'i'}\Big) \\ =: A_{ii'} + B_{ii'} + C_{ii'} + D_{ii'}.\end{aligned}$$

Consider $A_{ii'}$. By exploiting relation (A.24) we have

$$\mathbb{E}\,\Big|\sum_{i=1}^{n}\sum_{k=1}^{i-1} X_{ki}^2 - 1\Big|^2 = \sum_{i,i'=1}^{n}\mathbb{E}\,A_{ii'} - 2\sum_{i=1}^{n}(i-1)\,\mathbb{E}\,|X_{12}|^2 + 1 = \sum_{i,i'=1}^{n}\mathbb{E}\,A_{ii'} - 1 + o(1). \tag{A.26}$$

It is easily seen that $\mathbb{E}\,A_{ii'} = 2(i-1)\,\mathbb{E}\,X_{12}^2 X_{23}^2 + (i-1)(i'-3)(\mathbb{E}\,X_{12}^2)^2$. Moreover, since $\sum_{i<i'}(i-1) = \sum_{i'=1}^{n}\sum_{i=1}^{i'-1}(i-1) = \sum_{i'=1}^{n}(i'-1)(i'-2)/2 = n(n-1)(n-2)/6$ and

$\sum_{i<i'}(i-1)(i'-3) = \sum_{i'=1}^{n}(i'-3)(i'-2)(i'-1)/2 = n(n-1)(n-2)(n-3)/8$ (proof by induction) we obtain

$$\begin{aligned}
2\sum_{i<i'} \mathbb{E}\, A_{ii'} &= 4\,\mathbb{E}\, X_{12}^2 X_{23}^2 \sum_{i<i'}(i-1) + 2(\mathbb{E}\, X_{12}^2)^2 \sum_{i<i'}(i-1)(i'-3) \\
&= \frac{8n(n-1)(n-2)}{3n^4 \varsigma_{m_n}^4}\Big(\sum_{j,j',l,l'=1}^{m_n} \varsigma_{jj'}\varsigma_{ll'}\,\mathbb{E}\, U^4 f_j^\tau(W) f_{j'}^\tau(W) f_l^\tau(W) f_{l'}^\tau(W)\Big) \\
&\qquad\qquad + \frac{n(n-1)(n-2)(n-3)}{n^4}.
\end{aligned}$$

Moreover, by applying Cauchy Schwarz's inequality twice

$$\begin{aligned}
\sum_{j,j',l,l'=1}^{m_n} \varsigma_{jj'}\varsigma_{ll'}\,\mathbb{E}\, U^4 f_j^\tau(W) f_{j'}^\tau(W) f_l^\tau(W) f_{l'}^\tau(W) &\leqslant \max_{1\leqslant j\leqslant m_n} \mathbb{E}\,|U f_j(W)|^4 \Big(\sum_{j,\,j'=1}^{m_n} \sqrt{\tau_j \tau_{j'}}\varsigma_{jj'}\Big)^2 \\
&\leqslant \sigma^4 \eta^4 \varsigma_{m_n}^2 m_n \sum_{j=1}^{m_n} \tau_j^2.
\end{aligned}$$

Thereby, since $m_n \sum_{j=1}^{m_n} \tau_j^2 = o(n)$ it holds $2\sum_{i<i'} \mathbb{E}\, A_{ii'} = 1 + o(1)$. Obviously $\sum_{i=1}^{n} \mathbb{E}\, A_{ii} = o(1)$. Hence decomposition (A.26) yields $2\sum_{i<i'} A_{ii'} = 1 + o_p(1)$. Now consider $B_{ii'}$. Since $\big(\sum_{l=1}^{m_n} \varsigma_{ll}\big)^2 \leqslant m_n \sum_{l=1}^{m_n} \varsigma_{ll}^2 \leqslant m_n \varsigma_{m_n}^2$ and $\varsigma_{jj'} = \mathbb{E}\, U^2 f_j^\tau(W) f_{j'}^\tau(W)$ we conclude

$$\begin{aligned}
|\,\mathbb{E}\, B_{ii'}| &= \big|2\sum_{k=1}^{i-1} \mathbb{E}\, X_{ki}^2 X_{ki'} X_{ii'}\big| \\
&\leqslant \frac{8(i-1)}{n^4 \varsigma_{m_n}^4} \sum_{j,j',l,l'=1}^{m_n} |\varsigma_{ll'}\,\mathbb{E}\, U^3 f_j^\tau(W) f_{j'}^\tau(W) f_l^\tau(W)\,\mathbb{E}\, U^3 f_j^\tau(W) f_{j'}^\tau(W) f_{l'}^\tau(W)| \\
&\leqslant \frac{8(i-1)}{n^4 \varsigma_{m_n}^4}\Big(\sum_{l,l'=1}^{m_n} \varsigma_{ll'}\sqrt{\varsigma_{ll}\varsigma_{l'l'}}\Big)\Big(\sum_{j,j'=1}^{m_n} \mathbb{E}\,|U^2 f_j^\tau(W) f_{j'}^\tau(W)|^2\Big) \\
&\leqslant \frac{8(i-1)}{n^4 \varsigma_{m_n}^4}\Big(\sum_{l,l'=1}^{m_n} \varsigma_{ll'}^2\Big)^{1/2}\Big(\sum_{l=1}^{m_n} \varsigma_{ll}\Big)\,\mathbb{E}\,\|U f_{\underline{m_n}}^\tau(W)\|^4 \leqslant \frac{8(i-1)\sqrt{m_n}}{n^4 \varsigma_{m_n}^2}\,\mathbb{E}\,\|U f_{\underline{m_n}}^\tau(W)\|^4.
\end{aligned}$$

Estimate (A.1), i.e., $\mathbb{E}\,\|U f_{\underline{m_n}}^\tau(W)\|^4 \leqslant \eta^4 \sigma^4 m_n \sum_{j=1}^{m_n} \tau_j^2$, yields $\sum_{i<i'} |\,\mathbb{E}\, B_{ii'}| = o(1)$ as $n \to \infty$. Also it is easily seen that $\mathbb{E}\, C_{ii'} = 0$ and hence $\sum_{i<i'} \mathbb{E}\, C_{ii'} = 0$. For $D_{ii'}$ we treat two cases seperately. First consider the case $\varsigma_{m_n}^{-1} = o(1)$. Using twice the law of iterated

expectation gives

$$\begin{aligned}
\mathbb{E}\, D_{ii'} &= \mathbb{E}\Big(\sum_{k\neq k'}^{i-1} X_{ki}X_{k'i}\Big)\Big(\sum_{k\neq k'}^{i'-1} X_{ki'}X_{k'i'}\Big) = 4\sum_{k<k'}^{i-1}\mathbb{E}\, X_{ki}X_{k'i}X_{ki'}X_{k'i'}\\
&= 4\sum_{k<k'}^{i-1}\mathbb{E}\left[X_{ki}X_{k'i}\,\mathbb{E}[X_{ki'}X_{k'i'}|(Y_k,Z_k,W_k),(Y_{k'},Z_{k'},W_{k'}),(Y_i,Z_i,W_i)]\right]\\
&= \frac{8}{n^2\varsigma_{m_n}^2}\sum_{k<k'}^{i-1}\mathbb{E}\left[\mathbb{E}[X_{ki}X_{k'i}|(Y_k,Z_k,W_k),(Y_{k'},Z_{k'},W_{k'})]\sum_{j,j'=1}^{m_n}\varsigma_{jj'}U_kf_j^\tau(W_k)U_{k'}f_{j'}^\tau(W_{k'})\right]\\
&= \frac{8}{n^4\varsigma_{m_n}^4}\,\mathbb{E}\Big|\sum_{j,j'=1}^{m_n}\varsigma_{jj'}U_1f_j^\tau(W_1)U_2f_{j'}^\tau(W_2)\Big|^2(i-1)(i-2)\\
&= \frac{8}{n^4\varsigma_{m_n}^4}\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}(i-1)(i-2).
\end{aligned}$$

Since $\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}\leqslant\varsigma_{m_n}^2$ and $\varsigma_{m_n}^{-1}=o(1)$ we obtain

$$\sum_{i<i'}\mathbb{E}\, D_{ii'}\leqslant\frac{8}{n^4\varsigma_{m_n}^2}\sum_{i<i'}(i-1)(i-2)=\frac{2n(n-1)(n-2)(n-3)}{3\varsigma_{m_n}^2n^4}=o(1)$$

and hence $2\sum_{i<i'}\mathbb{E}\, Q_{ni}^2Q_{ni'}^2=1+o(1)$. Second, consider the case $\sum_{j=1}^{m_n}\tau_j^2=O(1)$. It holds

$$\sum_{i,i'=1}^{n}D_{ii'}=\Big(2(n\varsigma_{m_n})^{-2}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\varsigma_{jj'}\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big)^2+o_p(1). \tag{A.27}$$

Indeed, condition $\sum_{j=1}^{m_n}\tau_j^2=O(1)$ yields

$$\begin{aligned}
&\mathbb{E}\Big|\frac{2}{(n\varsigma_{m_n})^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\big(U_i^2f_j^\tau(W_i)f_{j'}^\tau(W_i)-\varsigma_{jj'}\big)\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big|^2\\
&\leqslant\frac{4}{(n\varsigma_{m_n})^4}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\tau_j^2\tau_{j'}^2\,\mathbb{E}\,|U^2f_j(W)f_{j'}(W)|^2\,\mathbb{E}\Big|\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j(W_k)f_{j'}(W_{k'})\Big|^2\\
&\leqslant\frac{4\sigma^8\eta^8}{(n\varsigma_{m_n})^4}\Big(\sum_{j=1}^{m_n}\tau_j^2\Big)^2\sum_{i=1}^{n}(i-1)(i-2)=o(1)
\end{aligned}$$

which proves (A.27). Moreover, since

$$\begin{aligned}
&\mathbb{E}\Big|\frac{2}{(n\varsigma_{m_n})^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\varsigma_{jj'}\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big|^2\\
&=\frac{4n}{(n\varsigma_{m_n})^4}\sum_{i=1}^{n}\sum_{k\neq k'}^{i-1}\mathbb{E}\Big|\sum_{j,j'=1}^{m_n}\varsigma_{jj'}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big|^2=\frac{4}{3\,\varsigma_{m_n}^4}\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}+o(1)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\Big|\frac{2}{(n\varsigma_{m_n})^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\varsigma_{jj'}\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big|^4\\
&=\frac{16}{n^6\varsigma_{m_n}^8}\sum_{i,i'=1}^{n}\sum_{k\neq k'}^{i-1}\sum_{l\neq l'}^{i'-1}\mathbb{E}\Big|\sum_{j,j'=1}^{m_n}\varsigma_{jj'}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big|^2\Big|\sum_{j,j'=1}^{m_n}\varsigma_{jj'}U_lU_{l'}f_j^\tau(W_l)f_{j'}^\tau(W_{l'})\Big|^2\\
&=\frac{16}{n^6\varsigma_{m_n}^8}\sum_{i,i'=1}^{n}(i-1)(i-2)(i'-3)(i'-4)\Big(\mathbb{E}\,\Big|U_1U_2\sum_{j,j'=1}^{m_n}\varsigma_{jj'}f_j^\tau(W_1)f_{j'}^\tau(W_2)\Big|^2\Big)^2+o(1)\\
&=\frac{16}{9\varsigma_{m_n}^8}\Big(\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}\Big)^2+o(1)
\end{aligned}$$

it holds

$$\mathbb{E}\Big|\Big(\frac{2}{(n\varsigma_{m_n})^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{m_n}\varsigma_{jj'}\sum_{k\neq k'}^{i-1}U_kU_{k'}f_j^\tau(W_k)f_{j'}^\tau(W_{k'})\Big)^2-\frac{4}{3\,\varsigma_{m_n}^4}\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}\Big|^2=o(1).$$

In light of equality (A.27) we have shown

$$\sum_{i,i'=1}^{n}D_{ii'}=\frac{4}{3\,\varsigma_{m_n}^4}\sum_{j,j',l,l'=1}^{m_n}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}+o_p(1).$$

Consequently, $2\sum_{i<i'}\mathbb{E}\,Q_{ni}^2Q_{ni'}^2=1+(4/3\varsigma_\infty^4)\sum_{j,j',l,l'=1}^{\infty}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}+o(1)$ which, in light of decomposition (A.25), completes the proof (A.22).
Proof of (A.23). Note that $\mathbb{P}\big(\sup_{i\geqslant 1}|Q_{ni}|>\varepsilon\big)\leqslant\sum_{i=1}^{n}\mathbb{P}\big(Q_{ni}^2>\varepsilon^2\big)$ and, hence the assertion follows from the Markov inequality. □

Consider the function $\delta\in L_Z^4$ introduced in Subsection 2.3. Let us denote

$$\widetilde{Q}_{ni}=\sqrt{2}(\varsigma_{m_n}n)^{-1}\sum_{i'=1}^{i-1}\sum_{j=1}^{m_n}\tau_j(U_i+n^{-1/2}\delta(Z_i))f_j(W_i)\big((U_{i'}+n^{-1/2}\delta(Z_{i'}))f_j(W_{i'})-n^{-1/2}[T\delta]_j\big).$$

Lemma A.3. *Under the conditions of Theorem 2.2 the process $\{(\widetilde{Q}_{ni},\mathcal{B}_{ni}),\ i\geqslant 1\}$ satisfies the conditions of Proposition A.1.*

Proof. From the definition of $\widetilde{Q}_{ni}$ we infer that $\{(\sum_{i'=1}^{i}\widetilde{Q}_{ni'},\mathcal{B}_{ni}),\ i\geqslant 1\}$ is a Martingale and, in particular $\{(\widetilde{Q}_{ni},\mathcal{B}_{ni}),\ i\geqslant 1\}$ forms a Martingale difference array. Moreover, following line by line the proof of Lemma A.2 it is easy to see that $\widetilde{Q}_{ni}$ satisfies conditions (A.21)–(A.23). Now Proposition A.1 yields $I_n\stackrel{d}{\to}\mathcal{N}(0,\nu)$ where $\nu=1+\big(4/3\varsigma_\infty^4\big)\sum_{j,j',l,l'=1}^{\infty}\varsigma_{jj'}\varsigma_{ll'}\varsigma_{jl}\varsigma_{j'l'}$, which completes the proof. □

References

C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71:1795–1843, 2003.

A. M. Awad. Conditional central limit theorems for martingales and reversed martingales. *The Indian Journal of Statistics, Series A*, 43:10–106, 1981.

R. Blundell and J. Horowitz. A nonparametric test of exogeneity. *Review of Economic Studies*, 74(4):1035–1058, Oct 2007.

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

C. Breunig and J. Johannes. Adaptive estimation of functionals in nonparametric instrumental regression. Technical report, University of Mannheim (submitted.), 2011.

X. Chen and M. Reiß. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(03):497–521, 2011.

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

S. G. Donald, G. Imbens, and W. K. Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117(1):55–93, 2003.

J. P. Florens, J. Johannes, and S. Van Bellegem. Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(03):472–496, 2011.

P. Gagliardini and O. Scaillet. Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167:6175, 2011.

P. Gagliardini and O. Scaillet. A specification test for nonparametric instrumental variable regression. Swiss Finance Institute Research Paper No. 07-13, 2007.

J. Ghorai. Asymptotic normality of a quadratic measure of the orthogonal series type density estimate. *Annals of the Institute of Statistical Mathematics*, 32:341–350, 1980.

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

Y. Hong and H. White. Consistent specification testing via nonparametric series regression. *Econometrica*, 63:1133–1159, 1995.

J. L. Horowitz. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74(2):521–538, 2006.

J. L. Horowitz. Specification testing in nonparametric instrumental variables estimation. *Journal of Econometrics*, 167:383–396, 2012.

J. L. Horowitz. Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394, 2011.

J. L. Horowitz and V. G. Spokoiny. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001.

J. Johannes and M. Schwarz. Adaptive nonparametric instrumental regression by model selection. Technical report, Université catholique de Louvain, 2010.

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

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

G. Tripathi and Y. Kitamura. Testing conditional moment restrictions. *Annals of Statistics*, 31(6):2059–2095, 2003.