

# Goodness-of-Fit Tests based on Series Estimators in Nonparametric Instrumental Regression\*

CHRISTOPH BREUNIG \*

*Humboldt-Universität zu Berlin*

March 2, 2015

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' asymptotic distributions under correct specification are derived and their consistency against any alternative model is shown. Under a sequence of local alternative hypotheses, the asymptotic distributions 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. A Monte Carlo study examines finite sample performance of the test statistics.

*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 Florens [2003] 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  $\varphi$  satisfies

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

---

\*This paper derives from my doctoral dissertation, completed under the guidance of Enno Mammen. I would like to thank two anonymous referees for comments and suggestions that greatly improved the paper. I also benefited from helpful comments by Jan Johannes, James Stock, Federico Crudu, and Petyo Bonev. This work was supported by the DFG-SNF research group FOR916.

\*Department of Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany, e-mail: christoph.breunig@hu-berlin.de

for an error term  $U$ . Here,  $Z$  contains potentially endogenous entries, that is,  $\mathbb{E}[U|Z]$  may not be zero. Model (1.1) does not involve the *a priori* assumption that the structural function is known up to finitely many parameters. By considering this 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], Newey and Powell [2003] or Johannes and Schwarz [2010] consider sieve minimum distance estimators of  $\varphi$ , while Darolles et al. [2011], Hall and Horowitz [2005], Gagliardini and Scaillet [2011] or Florens et al. [2011] study penalized least squares estimators. When the methods of analysis are widened to include nonparametric techniques, one must confront two major 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  $\varphi$  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, that is, the solution does not depend continuously on the data.

In light of the difficulties of estimating the nonparametric function  $\varphi$  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  $\varphi$  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  $\varphi$ . These model simplifications have important potential since they might increase the accuracy of estimators of  $\varphi$  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  $Z$ , that is, whether certain regressors can be omitted from the structural function  $\varphi$ . By a withdrawal of regressors which are independent of the instrument, identification in the restricted model might be possible although  $\varphi$  is not identified in the original model (1.1).

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  $\varphi$ . 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 testing procedure is entirely based on series estimation and hence is easy to implement. We use approximating functions to estimate the conditional moment restriction implied by

the model (1.1) where  $\varphi$  is replaced by an estimator under each conjectured hypothesis. 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  $(Z, W)$ . In addition, a Monte Carlo indicates that the finite sample power of our tests exceed that of existing tests.

The paper is organized as follows. In Section 2, we start with a simple hypothesis test, that is, whether  $\varphi$  coincides with a known function  $\varphi_0$ . We obtain the test's asymptotic distribution under the null hypothesis and its consistency against any fixed alternative model. Moreover, we judge its power by considering linear local alternatives and establish uniform consistency over a class of functions. In Sections 3–5 we consider a parametric specification test, a test of exogeneity, and a nonparametric specification test. The goodness-of-fit statistics are obtained by replacing  $\varphi_0$  in the statistic of Section 2 by an appropriate estimator. In each case, the asymptotic distribution under correct specification and power statements against alternative models are derived. In Section 6, we investigate the finite sample properties of our tests by Monte Carlo simulations. 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  $\varphi_0$  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  $\varphi_0$  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 notation.

**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  $\mathcal{L}_Z^2 = \{\phi : \mathbb{E}|\phi(Z)|^2 < \infty\}$  to  $\mathcal{L}_W^2 = \{\psi : \mathbb{E}|\psi(W)|^2 < \infty\}$ . Consequently, model (1.1) can be written as

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

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

**Assumptions.** Our test statistic based on a sequence of approximating functions  $\{f_l\}_{l \geq 1}$  in  $\mathcal{L}_W^2$ . Let  $\mathcal{W}$  denote the support of  $W$  and the marginal density of  $W$  by  $p_W$ . Let  $\nu$  be a probability density function that is strictly positive on  $\mathcal{W}$ . We assume throughout the paper that  $\{f_l\}_{l \geq 1}$  forms an orthonormal basis in  $\mathcal{L}_\nu^2(\mathbb{R}^{d_w}) := \{\phi : \int \phi^2(s)\nu(s)ds < \infty\}$  where  $d_w$  denotes the dimension of  $W$ . For instance, if  $\mathcal{W} \subset [a, b]$  then a natural choice of  $\nu$  would be  $\nu(w) = 1/(b - a)$  for  $w \in [a, b]$  and zero otherwise.

**ASSUMPTION 1.** *There exist constants  $\eta_f, \eta_p \geq 1$  such that (i)  $\sup_{l \geq 1} \int |f_l(s)|^4 \nu(s) ds \leq \eta_f$  and (ii)  $\sup_{w \in \mathcal{W}} \{p_W(w)/\nu(w)\} \leq \eta_p$  with  $\nu$  being strictly positive on  $\mathcal{W}$ .*

Assumption 1 (i) restricts the magnitude of the approximating functions  $\{f_j\}_{j \geq 1}$  which is necessary for our proof to determine the asymptotic behavior of our test statistic. This

assumption holds for sufficiently large  $\eta_f$  if the basis  $\{f_l\}_{l \geq 1}$  is uniformly bounded, such as trigonometric bases. Moreover, Assumption 1 (i) is satisfied by Hermite polynomials. Assumption 1 (ii) is satisfied if, for instance,  $p_W/\nu$  is continuous and  $\mathcal{W}$  is compact. 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 a constant  $\sigma > 0$  such that  $\mathbb{E}[U^4|W] \leq \sigma^4$ .*

The conditional moment condition on the error term  $U$  helps to establish the asymptotic distribution of our test statistics. The following assumption ensures identification of  $\varphi$  in the model (2.1).

**ASSUMPTION 3.** *The conditional expectation operator  $T$  is nonsingular.*

Under Assumption 3, the hypothesis  $H_0$  is equivalent to  $g = T\varphi_0$  which is used to construct our test statistic below. Note that the asymptotic results under null hypotheses considered in Sections 2–4 hold true even if  $T$  is singular. If Assumption 3 fails, however, our test has no power against alternative models whose structural function satisfies  $\varphi = \varphi_0 + \delta$  with  $\delta$  belonging to the null space of  $T$ .

We will see below that the power of our test can be increased by carrying out an additional smoothing step. Therefore, we introduce a smoothing operator  $L$  mapping  $\mathcal{L}_W^2$  to  $\mathcal{L}_W^2$ . In contrast to the unknown conditional expectation operator  $T$ , which has to be estimated, the operator  $L$  can be chosen by the econometrician. Let  $L$  have an eigenvalue decomposition given by  $\{\tau_j^{1/2}, f_j\}_{j \geq 1}$ . We allow in this paper for a wide range of smoothing operators. In particular,  $L$  may be the identity operator, that is, no smoothing step is carried out. We only require the following condition on the operator  $L$  determined by the sequence of eigenvalues  $\tau = (\tau_j)_{j \geq 1}$ .

**ASSUMPTION 4.** *The weighting sequence  $\tau$  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, that is,  $L\phi(s) = \int_0^1 \ell(s, t)\phi(t)dt$  for some function  $\phi \in \mathcal{L}^2[0, 1] = \{\phi : \int_0^1 \phi^2(s)ds < \infty\}$  and some kernel function  $\ell : [0, 1]^2 \rightarrow \mathbb{R}$ . In order to ensure  $L\phi \in \mathcal{L}^2[0, 1]$  it is sufficient to assume  $\int_0^1 \int_0^1 |\ell(s, t)|^2 ds dt < \infty$ . Let  $\{\tau_j^{1/2}, f_j\}_{j \geq 1}$  be the eigenvalue decomposition of  $L$ . By Parseval's identity

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

where the right hand side is only finite if the sequence  $\tau$  decays sufficiently fast. In our case, if we apply a smoothing operator  $L$  with  $\sum_{j=1}^{\infty} \tau_j < \infty$  then our test statistics converges also to a weighted series of chi-squared random variables. In addition, we allow for a milder degree of smoothing or no smoothing at all and show below that then asymptotic normality of our test statistics can be obtained.  $\square$

**Notation.** For a matrix  $A$  we denote its transposed by  $A^t$ , its inverse by  $A^{-1}$ , and its generalized inverse by  $A^-$ . The euclidean norm is denoted by  $\|\cdot\|$  which in case of a matrix denotes the spectral norm, that is  $\|A\| = (\text{trace}(A^t A))^{1/2}$ . The norms on  $L_Z^2$  and  $L_W^2$  are denoted by  $\|\phi\|_Z^2 := \mathbb{E}|\phi(Z)|^2$  for  $\phi \in L_Z^2$  and  $\|\psi\|_W^2 := \mathbb{E}|\psi(W)|^2$  for  $\psi \in L_W^2$ . The  $k \times k$  identity matrix is denoted by  $I_k$ . For a vector  $V$  we write  $\text{diag}(V)$  for the diagonal

matrix with diagonal elements being the values of  $V$ . Moreover,  $e_m(Z)$  and  $f_m(W)$  denote random vectors with entries  $e_j(Z)$  and  $f_j(W)$ ,  $1 \leq j \leq m$ , respectively. For any weighting sequence  $w$  we introduce vectors  $e_m^w(Z)$  and  $f_m^w(W)$  with entries  $e_j^w(Z) = \sqrt{w_j}e_j(Z)$  and  $f_j^w(W) = \sqrt{w_j}f_j(W)$ ,  $1 \leq j \leq m$ . We write  $a_n \sim b_n$  when there exist constants  $c, c' > 0$  such that  $cb_n \leq a_n \leq c'b_n$  for all sufficiently large  $n$ .

## 2.2. The test statistic and its asymptotic distribution

Nonsingularity of the conditional expectation operator  $T$  and the smoothing operator  $L$  implies that the null hypothesis  $H_0$  is equivalent to  $L(g - T\varphi_0) = 0$ . Note that  $\|L(g - T\varphi_0)\|_W = 0$  if and only if  $\int |L(g - T\varphi_0)(w)p_W(w)/\nu(w)|^2 \nu(w)dw = 0$  since the Lebesgue measure  $\nu$  is strictly positive on  $W$ . Moreover, since  $\{f_j\}_{j \geq 1}$  is an orthonormal basis with respect to  $\nu$  we obtain by Parseval's identity

$$\int |L(g - T\varphi_0)(w)p_W(w)/\nu(w)|^2 \nu(w)dw = \sum_{j=1}^{\infty} \mathbb{E}[(g - T\varphi_0)(W)f_j^\tau(W)]^2. \quad (2.2)$$

Now we truncate the infinite sum at some integer  $m_n$  which grows with the sample size  $n$ . This ensures consistency of our testing procedure. Further, replacing the expectation by sample mean we obtain our test statistic

$$S_n := \sum_{j=1}^{m_n} \tau_j |n^{-1} \sum_{i=1}^n (Y_i - \varphi_0(Z_i))f_j(W_i)|^2. \quad (2.3)$$

We reject the hypothesis  $H_0$  if  $nS_n$  becomes too large. When no additional smoothing is carried out, that is,  $L$  is the identity operator, then  $\tau_j = 1$  for all  $j \geq 1$ . 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  $m \geq 1$  let  $\Sigma_m$  be the covariance matrix of the random vector  $Uf_m^\tau(W)$  with entries  $s_{jl} = \mathbb{E}[U^2 f_j^\tau(W)f_l^\tau(W)]$ ,  $1 \leq j, l \leq m$ . Then the trace and the Frobenius norm of  $\Sigma_m$  are respectively denoted by

$$\mu_m := \sum_{j=1}^m s_{jj} \quad \text{and} \quad \varsigma_m := \left( \sum_{j,l=1}^m s_{jl}^2 \right)^{1/2}.$$

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

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

$$\varsigma_{m_n}^{-1} = o(1) \quad \text{and} \quad \left( \sum_{j=1}^{m_n} \tau_j \right)^3 = o(n) \quad (2.4)$$

then under  $H_0$

$$(\sqrt{2}\varsigma_{m_n})^{-1}(nS_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

**REMARK 2.2.** Since  $\varsigma_{m_n}^2 \leq \eta_p \sigma^4 \left( \sum_{j=1}^{m_n} \tau_j \right)^2$  (cf. proof of Theorem 2.2) condition  $\varsigma_{m_n}^{-1} = o(1)$  implies that  $\sum_{j=1}^{m_n} \tau_j$  tends to infinity as  $n$  increases. Moreover, from condition (2.4) we see that by choosing a stronger decaying sequence  $\tau$  the parameter  $m_n$  may be chosen larger. From the following theorem we see that if  $\sum_{j=1}^{m_n} \tau_j = O(1)$  only  $m_n^{-1} = o(1)$  is required.  $\square$

In the following result, we establish the asymptotic distribution of our test when the sequence of weights  $\tau$  may have a stronger decay than in Theorem 2.1, that is, we consider the case where  $\tau$  satisfies  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . This holds, for instance, if the sequence  $\tau$  satisfies  $\tau_j \sim j^{-(1+\varepsilon)}$  for any  $\varepsilon > 0$ . In this case, the asymptotic distribution changes and additional definitions have to be made. Let  $\Sigma$  be the covariance matrix of the infinite dimensional centered vector  $(U f_j^\tau(W))_{j \geq 1}$ . The ordered eigenvalues of  $\Sigma$  are denoted by  $(\lambda_j)_{j \geq 1}$ . Below, we introduce a sequence  $\{\chi_{1j}^2\}_{j \geq 1}$  of independent random variables that are distributed as chi-square with one degree of freedom.

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

$$\sum_{j=1}^{m_n} \tau_j = O(1) \quad \text{and} \quad m_n^{-1} = o(1) \quad (2.5)$$

then under  $H_0$

$$n S_n \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2.$$

**REMARK 2.3 (Estimation of Critical Values).** The asymptotic results of Theorem 2.1 and 2.2 depend on unknown population quantities. As we see in the following, the critical values can be easily estimated. Let  $\mathbf{W}_m(\tau)$  denote a  $n \times m$  matrix with entries  $f_j^\tau(W_i)$  for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ . Moreover,  $\mathbf{U}_n = (Y_1 - \varphi_0(Z_1), \dots, Y_n - \varphi_0(Z_n))^t$ . In the setting of Theorem 2.1, we replace  $\Sigma_m$  by

$$\widehat{\Sigma}_m := n^{-1} \mathbf{W}_m(\tau)^t \text{diag}(\mathbf{U}_n)^2 \mathbf{W}_m(\tau).$$

Now the asymptotic result of Theorem 2.1 continues to hold if we replace  $\varsigma_{m_n}$  by the Frobenius norm of  $\widehat{\Sigma}_{m_n}$  and  $\mu_{m_n}$  by the trace of  $\widehat{\Sigma}_{m_n}$ . In the setting of Theorem 2.2, the asymptotic distribution is not pivotal and has to be approximated. First, the difference of critical values between  $\sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$  and the truncated sum  $\sum_{j=1}^{M_n} \lambda_j \chi_{1j}^2$  converges to zero if the integer  $M_n > 0$  tends to infinity (depending on  $n$ ). Second, replace  $(\lambda_j)_{1 \leq j \leq M_n}$  by  $(\widehat{\lambda}_j)_{1 \leq j \leq M_n}$  which are the ordered eigenvalues of  $\widehat{\Sigma}_{M_n}$ . Observe  $\max_{1 \leq j \leq M_n} |\widehat{\lambda}_j - \lambda_j| = \|\widehat{\Sigma}_{M_n} - \Sigma_{M_n}\| = O(M_n n^{-1/2})$  almost surely. Hence, the critical values of  $\sum_{j=1}^{M_n} \widehat{\lambda}_j \chi_{1j}^2$  converge in probability to the ones of the limiting distribution of  $n S_n$  if  $M_n = o(\sqrt{n})$ .  $\square$

### 2.3. Limiting behavior under local alternatives.

Let us study the power of the test statistic  $S_n$ , that is, the probability to reject a false hypothesis, against a sequence of linear local alternatives that tends to zero as  $n \rightarrow \infty$ . It is shown that the power of our tests essentially relies on the choice of the weighting sequence  $\tau$ .

Let us start with the case  $\varsigma_{m_n}^{-1} = o(1)$ . We consider the following sequence of linear local alternatives

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

for some function  $\delta \in \mathcal{L}_Z^4 := \{\phi : \mathbb{E}|\phi(Z)|^4 < \infty\}$ . The next result establishes asymptotic normality for the standardized test statistic  $S_n$ . Let us denote  $\delta_j := \sqrt{\tau_j} \mathbb{E}[\delta(Z) f_j(W)]$ .

**PROPOSITION 2.3.** *Given the conditions of Theorem 2.1 it holds under (2.6)*

$$(\sqrt{2}\varsigma_{m_n})^{-1}(nS_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2} \sum_{j=1}^{\infty} \delta_j^2, 1\right).$$

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

$$Y = \varphi_0(Z) + n^{-1/2}\delta(Z) + U \tag{2.7}$$

for some function  $\delta \in \mathcal{L}_Z^4$ . For the next result, the sequence  $\{\chi_{1j}^2(\delta_j/\lambda_j)\}_{j \geq 1}$  denotes independent random variables that are distributed as non-central chi-square with one degree of freedom and non-centrality parameters  $\delta_j/\lambda_j$ .

**PROPOSITION 2.4.** *Given the conditions of Theorem 2.2 it holds under (2.7)*

$$nS_n \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2(\delta_j/\lambda_j).$$

**REMARK 2.4.** We see from Proposition 2.3 that our test can detect linear alternatives at a rate  $\varsigma_{m_n}^{1/2} n^{-1/2}$ . On the other hand, if  $\sum_{j=1}^{m_n} \tau_j = O(1)$  then  $S_n$  can detect local linear alternatives at the faster rate  $n^{-1/2}$ . But still our test with  $L = \text{Id}$  can have better power against certain smooth classes of alternatives as illustrated by Hong and White [1995] and Horowitz and Spokoiny [2001]. Indeed, the next subsection shows that additional smoothing changes the class of alternatives over which uniform consistency can be obtained.  $\square$

## 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, that is,  $\mathbb{P}(\varphi = \varphi_0) < 1$ . 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.

The consistency properties require the following additional assumption.

**ASSUMPTION 5.** (i) *The function  $p_W/\nu$  is uniformly bounded away from zero.* (ii) *There exists a constant  $\sigma_o > 0$  such that  $\mathbb{E}[U^2|W] \geq \sigma_o^2$ .*

Assumption 5 (i) implies that  $\|LT(\varphi - \varphi_0)\|_W > 0$  for any structural function  $\varphi$  in the alternative. Further, Assumption 5 implies that  $\sum_{j=1}^{m_n} \tau_j^2 = O(\varsigma_{m_n}^2)$ .

**PROPOSITION 2.5.** *Assume that  $H_0$  does not hold. Let  $\mathbb{E}|Y - \varphi_0(Z)|^4 < \infty$  and let Assumption 5 (i) hold true. Consider the sequence  $(\alpha_n)_{n \geq 1}$  satisfying  $\alpha_n = o(n\varsigma_{m_n}^{-1})$ . Under the conditions of Theorem 2.1 we have*

$$\mathbb{P}\left((\sqrt{2}\varsigma_{m_n})^{-1}(nS_n - \mu_{m_n}) > \alpha_n\right) = 1 + o(1).$$

*Under the conditions of Theorem 2.2 we have  $\alpha_n = o(n)$  and*

$$\mathbb{P}(nS_n > \alpha_n) = 1 + o(1).$$



In the following, we specify a class of functions over which our test  $S_n$  is uniformly consistent. This essentially implies that there are no alternative functions in this class over which our test has low power. We show that our test is consistent uniformly over the class

$$\mathcal{G}_n^\rho = \left\{ \varphi \in \mathcal{L}_Z^2 : \|LT(\varphi - \varphi_0)\|_W^2 \geq \rho n^{-1} \varsigma_{m_n} \text{ and } \sup_{z \in \mathcal{Z}} |(\varphi - \varphi_0)(z)|^2 \leq C \right\}$$

where  $C > 0$  is a finite constant. Clearly, if  $H_0$  is false then  $\|LT(\varphi - \varphi_0)\|_W^2 \geq \rho \varsigma_{m_n} n^{-1}$  for all sufficiently large  $n$  and some  $\rho > 0$ . By Assumption 4 the sequence  $\tau$  is nonincreasing sequence with  $\tau_1 = 1$  and hence,  $\|LT(\varphi - \varphi_0)\|_W^2 \leq \|T(\varphi - \varphi_0)\|_W^2 \leq \|\varphi - \varphi_0\|_Z^2$  by Jensen's inequality. We conclude that  $\mathcal{G}_n^\rho$  contains all alternative functions whose  $\mathcal{L}_Z^2$ -distance to the structural function  $\varphi_0$  is at least  $n^{-1} \varsigma_{m_n}$  within a constant. If the coefficients  $\mathbb{E}[(\varphi - \varphi_0)(Z)f_j(W)]$  fluctuate for large  $j$  then  $\varphi$  does not belong to  $\mathcal{G}_n^\rho$  if the decay of  $\tau$  is too strong. On the other hand, if  $\mathbb{E}[(\varphi - \varphi_0)(Z)f_j(W)]$  is sufficiently small for  $j$  up to a finite constant then  $\varphi$  does not necessarily belong to  $\mathcal{G}_n^\rho$  with  $\tau$  having a slow decay. For the next result let  $q_{1\alpha}$  and  $q_{2\alpha}$  denote the  $1 - \alpha$  quantile of  $\mathcal{N}(0, 1)$  and  $\sum_{j=1}^\infty \lambda_j \chi_{1j}^2$ , respectively.

**PROPOSITION 2.6.** *Let Assumption 5 be satisfied. For any  $\varepsilon > 0$ , any  $0 < \alpha < 1$ , and any sufficiently large constant  $\rho > 0$  we have under the conditions of Theorem 2.1 that*

$$\lim_{n \rightarrow \infty} \inf_{\varphi \in \mathcal{G}_n^\rho} \mathbb{P} \left( (\sqrt{2} \varsigma_{m_n})^{-1} (nS_n - \mu_{m_n}) > q_{1\alpha} \right) \geq 1 - \varepsilon,$$

while under the conditions of Theorem 2.2

$$\lim_{n \rightarrow \infty} \inf_{\varphi \in \mathcal{G}_n^\rho} \mathbb{P} \left( nS_n > q_{2\alpha} \right) \geq 1 - \varepsilon.$$

### 3. A parametric specification test

In this section, we present a test whether the structural function  $\varphi$  is known up to a finite dimensional parameter. Let  $\Theta$  be a compact subspace of  $\mathbb{R}^k$  then we consider the null hypothesis  $H_p$  : there exists some  $\vartheta \in \Theta$  such that  $\varphi(\cdot) = \phi(\cdot, \vartheta)$  for a known function  $\phi$ . The alternative hypothesis is that there exists no  $\vartheta \in \Theta$  such that  $\varphi(\cdot) = \phi(\cdot, \vartheta)$  holds true.

#### 3.1. The test statistic and its asymptotic distribution

Under Assumptions 3 and 4, the null hypothesis  $H_p$  is equivalent to  $L(g - T\phi(\cdot, \vartheta)) = 0$  for some  $\vartheta \in \Theta$ . Thereby, to verify  $H_p$  we make use of the test statistic  $S_n$  given in (2.3) where  $\varphi_0$  is replaced by  $\phi(\cdot, \hat{\vartheta}_n)$  with  $\hat{\vartheta}_n$  being an estimator of  $\vartheta$ . Hence, our test statistic for a parametric specification is given by

$$S_n^p := \sum_{j=1}^{m_n} \tau_j |n^{-1} \sum_{i=1}^n (Y_i - \phi(Z_i, \hat{\vartheta}_n)) f_j(W_i)|^2.$$

If the test statistic  $S_n^p$  becomes too large then  $H_p$  has to be rejected. To obtain asymptotic results for the statistic  $S_n^p$  we require smoothness conditions of the function  $\phi$  with respect to its second argument. Below we denote the vector of partial derivatives of  $\phi$  with respect to  $\vartheta = (\vartheta_1, \dots, \vartheta_k)^t$  by  $\phi_\vartheta = (\phi_{\vartheta_l})_{1 \leq l \leq k}$  and the matrix of second-order partial derivatives by  $\phi_{\vartheta\vartheta} = (\phi_{\vartheta_j \vartheta_l})_{1 \leq j, l \leq k}$ .



**ASSUMPTION 6.** (i) Let  $\widehat{\vartheta}_n$  be an estimator satisfying  $\|\widehat{\vartheta}_n - \vartheta_0\| = O_p(n^{-1/2})$  for some  $\vartheta_0 \in \text{int}(\Theta)$  with  $\varphi(\cdot) = \phi(\cdot, \vartheta_0)$  if  $H_p$  holds true. (ii) The function  $\phi$  is twice partial differentiable with respect to its second argument. There exists some constant  $\eta_\phi \geq 1$  such that

$$\sup_{1 \leq l \leq k} \mathbb{E} |\phi_{\vartheta_l}(Z, \vartheta_0)|^4 \leq \eta_\phi \quad \text{and} \quad \sup_{1 \leq j, l \leq k} \mathbb{E} \sup_{\theta \in \Theta} |\phi_{\vartheta_j \vartheta_l}(Z, \theta)|^4 \leq \eta_\phi.$$

The following proposition establishes asymptotic normality of  $S_n^p$  after standardization.

**THEOREM 3.1.** Let Assumptions 1–4 and 6 hold true. If  $m_n$  satisfies (2.4), then under  $H_p$

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

In the following theorem, we state the asymptotic distribution of  $nS_n^p$  when  $\sum_{j=1}^{m_n} \tau_j = O(1)$ .

In this case, we assume that  $\widehat{\vartheta}_n$  satisfies under  $H_p$

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) = n^{-1/2} \sum_{i=1}^n h_{\underline{k}}(V_i) + o_p(1) \quad (3.1)$$

where  $V_i := (Y_i, Z_i, W_i, \vartheta_0)$  and  $h_{\underline{k}}(V_i) = (h_1(V_i), \dots, h_k(V_i))^t$  where  $h_j$ ,  $1 \leq j \leq k$ , are real valued functions. It is well known that this representation holds if  $\widehat{\vartheta}_n$  is the generalized method of moments estimator. Let  $\Sigma^p$  be the covariance matrix of the infinite dimensional centered vector  $(U f_j^\tau(W) - \mathbb{E}[f_j^\tau(W) \phi_{\vartheta}(Z, \vartheta_0)^t] h_{\underline{k}}(V))_{j \geq 1}$ . The ordered eigenvalues of  $\Sigma^p$  are denoted by  $(\lambda_j^p)_{j \geq 1}$ .

**THEOREM 3.2.** Let Assumptions 1–4 and 6 hold true. Assume that  $H_p$  holds true and  $\widehat{\vartheta}_n$  satisfies condition (3.1) with  $\mathbb{E} h_j(V) = 0$  and  $\mathbb{E} |h_j(V)|^4 < \infty$ ,  $1 \leq j \leq k$ . If  $m_n$  satisfies (2.5), then

$$n S_n^p \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^p \chi_{1j}^2.$$

**REMARK 3.1.** [Estimation of Critical Values] For the estimation of critical values of Theorem 3.1 and 3.2, let us define  $\mathbf{U}_n^p = (Y_1 - \phi(Z_1, \widehat{\vartheta}_n), \dots, Y_n - \phi(Z_n, \widehat{\vartheta}_n))^t$ . We estimate the covariance matrix  $\Sigma_m$  by

$$\widehat{\Sigma}_m := n^{-1} \mathbf{W}_m(\tau)^t \text{diag}(\mathbf{U}_n^p)^2 \mathbf{W}_m(\tau).$$

Now the asymptotic result of Theorem 3.1 continues to hold if we replace  $\varsigma_{m_n}$  by the Frobenius norm of  $\widehat{\Sigma}_{m_n}$  and  $\mu_{m_n}$  by the trace of  $\widehat{\Sigma}_{m_n}$ . In the setting of Theorem 3.2, we replace  $\Sigma^p$  by a finite dimensional matrix. Let  $\mathbf{A}_k$  be a  $n \times k$  matrix with entries  $\phi_{\vartheta_l}(Z_i, \widehat{\vartheta}_n)$  for  $1 \leq i \leq n$ ,  $1 \leq l \leq k$  and  $\mathbf{h}_k(V) = (h_{\underline{k}}(V_1), \dots, h_{\underline{k}}(V_n))^t$ . Then define  $\mathbf{V}_k := n^{-1} \mathbf{h}_k(V) \mathbf{A}_k^t$ . Given a sufficiently large integer  $M > 0$  we estimate  $\Sigma^p$  by

$$\widehat{\Sigma}_M^p := n^{-1} \mathbf{W}_M(\tau)^t \left( \text{diag}(\mathbf{U}_n^p) - \mathbf{V}_k \right)^t \left( \text{diag}(\mathbf{U}_n^p) - \mathbf{V}_k \right) \mathbf{W}_M(\tau).$$

Hence, we approximate  $\sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{M_n} \widehat{\lambda}_j^p \chi_{1j}^2$  where  $(\widehat{\lambda}_j^p)_{1 \leq j \leq M_n}$  are the ordered eigenvalues of  $\widehat{\Sigma}_{M_n}^p$ . We have  $\max_{1 \leq j \leq M_n} |\widehat{\lambda}_j^p - \lambda_j^p| = o_p(1)$  if  $M_n = o(\sqrt{n})$ .  $\square$

### 3.2. Limiting behavior under local alternatives and consistency.

In the following, we study the power and consistency properties of the test statistic  $S_n^p$ . In the following, we consider a sequence of linear local alternatives (2.6) or (2.7) with  $\varphi_0 = \phi(\vartheta_0, \cdot)$ . Further, let  $\delta_\perp$  denote the projection of  $\delta$  onto the orthogonal complement of  $\phi(\cdot, \vartheta_0)$ ; that is,  $\mathbb{E}[\phi_\vartheta(Z, \vartheta_0)\delta_\perp(Z)] = 0$ . Let us denote  $\delta_{j\perp} := \sqrt{\tau_j} \mathbb{E}[\delta_\perp(Z)f_j(W)]$ .

**PROPOSITION 3.3.** *Let the conditions of Theorem 3.1 be satisfied. Then under (2.6) with  $\varphi_0 = \phi(\cdot, \vartheta_0)$  it holds*

$$(\sqrt{2}\varsigma_{m_n})^{-1}(n S_n^p - \mu_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2} \sum_{j=1}^{\infty} \delta_{j\perp}^2, 1\right).$$

Let the conditions of Theorem 3.2 be satisfied. Then under (2.7) with  $\varphi_0 = \phi(\cdot, \vartheta_0)$  it holds

$$n S_n^p \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^p \chi_{1j}^2(\delta_{j\perp}/\lambda_j^p).$$

**REMARK 3.2.** Under homoscedasticity, that is,  $\mathbb{E}[U^2|W] = \sigma_o^2$ ,  $W \sim \mathcal{U}[0, 1]$ , and  $L = \text{Id}$  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, if  $\sum_{j=1}^{m_n} \tau_j = O(1)$  then our test can detect local linear alternatives at a rate  $n^{-1/2}$  as in Horowitz [2006], which decreases more quickly than the rate obtained by Tripathi and Kitamura [2003].  $\square$

The next proposition establishes consistency of our test against a fixed alternative model. It is assumed that  $H_p$  is false, that is, there exists no  $\vartheta \in \Theta$  such that  $\varphi(\cdot) = \phi(\cdot, \vartheta)$ . In this situation,  $\vartheta_0$  denotes the probability limit of the estimator  $\widehat{\vartheta}_n$ .

**PROPOSITION 3.4.** *Assume that  $H_p$  does not hold. Let  $\mathbb{E}|Y - \phi(Z, \vartheta_0)|^4 < \infty$  and Assumption 5 (i) hold true. Let  $(\alpha_n)_{n \geq 1}$  as in Proposition 2.5. Under the conditions of Theorem 3.1 we have*

$$\mathbb{P}\left((\sqrt{2}\varsigma_{m_n})^{-1}(n S_n^p - \mu_{m_n}) > \alpha_n\right) = 1 + o(1).$$

Given the conditions of Theorem 3.2 it holds

$$\mathbb{P}(n S_n^p > \alpha_n) = 1 + o(1).$$

In the following, we show that  $S_n^p$  is consistent uniformly over the function class

$$\mathcal{H}_n^\rho = \left\{ \varphi \in \mathcal{L}_Z^2 : \|LT(\varphi - \phi(\cdot, \vartheta_0))\|_W^2 \geq \rho n^{-1}\varsigma_{m_n} \text{ and } \sup_{z \in \mathcal{Z}} |\varphi(z) - \phi(z, \vartheta_0)| \leq C \right\}$$

for some constant  $C > 0$  and  $\vartheta_0$  denotes the probability limit of  $\widehat{\vartheta}_n$ . Similarly as in the previous section, it can be seen that  $\mathcal{H}_n^\rho$  only contains functions whose  $\mathcal{L}_Z^2$  distance to  $\phi(\cdot, \vartheta_0)$  is at least  $n^{-1}\varsigma_{m_n}$  within a constant. For the next result let  $q_{1\alpha}$  and  $q_{2\alpha}$  denote the  $1 - \alpha$  quantile of  $\mathcal{N}(0, 1)$  and  $\sum_{j=1}^{\infty} \lambda_j^p \chi_{1j}^2$ , respectively.

**PROPOSITION 3.5.** *Let Assumption 5 be satisfied. 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*

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

whereas under the conditions of Theorem 3.2 it holds

$$\lim_{n \rightarrow \infty} \inf_{\varphi \in \mathcal{H}_n^\rho} \mathbb{P}(n S_n^p > q_{2\alpha}) \geq 1 - \varepsilon.$$

## 4. A nonparametric test of exogeneity

Endogeneity of regressors is a common problem in econometric applications. Falsely assuming exogeneity of the regressors leads to inconsistent estimators. On the other hand, 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, that is,  $\mathbb{E}[U|Z] = 0$  or equivalently  $\varphi(Z) = \mathbb{E}[Y|Z]$ . In this section, let  $\varphi_0(Z) = \mathbb{E}[Y|Z]$  then the hypothesis under consideration is given by  $H_e : \varphi = \varphi_0$ . The alternative hypothesis is that  $\varphi \neq \varphi_0$ .

### 4.1. The test statistic and its asymptotic distribution

To establish a test of exogeneity, let us first introduce an estimator of the conditional mean of  $Y$  given  $Z$ . This estimator is based on a sequence of approximating functions  $\{e_j\}_{j \geq 1}$  belonging to  $\mathcal{L}_Z^2$ . Further, let  $\mathbf{Z}_k$  denote a  $n \times k$  matrix with entries  $e_j(Z_i)$  for  $1 \leq i \leq n$  and  $1 \leq j \leq k$ . Moreover, let  $\mathbf{Y}_n = (Y_1, \dots, Y_n)^t$ . Then we define the estimator

$$\bar{\varphi}_k(\cdot) := e_{\underline{k}}(\cdot)^t \hat{\beta}_k \quad \text{where} \quad \hat{\beta}_k = (\mathbf{Z}_k^t \mathbf{Z}_k)^{-} \mathbf{Z}_k^t \mathbf{Y}_n. \quad (4.1)$$

In contrast to the parametric case we need to allow for  $k$  tending to infinity as  $n \rightarrow \infty$  in order to ensure consistency of the estimator  $\bar{\varphi}_k$ . Under conditions given below  $\mathbf{Z}_{k_n}^t \mathbf{Z}_{k_n}$  will be nonsingular with probability approaching one and hence its generalized inverse will be the standard inverse. Note that the asymptotic behavior of the estimator  $\bar{\varphi}_k$  was studied, for example, by Newey [1997].

Under Assumptions 3 and 4, the null hypothesis  $H_e$  is equivalent to  $L(g - T\varphi_0) = 0$ . Consequently, our test of exogeneity of  $Z$  is based on the goodness-of-fit statistic  $S_n$  introduced in (2.3) but where  $\varphi_0$  is replaced by the series estimator  $\bar{\varphi}_{k_n}$ . The proposed test statistic for  $H_e$  is now given by

$$S_n^e = \sum_{j=1}^{m_n} \tau_j |n^{-1} \sum_{i=1}^n (Y_i - \bar{\varphi}_{k_n}(Z_i)) f_j(W_i)|^2$$

where  $k_n$  and  $m_n$  tend to infinity as  $n \rightarrow \infty$ . The hypothesis of exogeneity of  $Z$  has to be rejected if  $S_n^e$  becomes too large.

For controlling the bias of the estimator  $\bar{\varphi}_{k_n}$  we specify in the following a rate of approximation (cf. Newey [1997]). Let  $\gamma = (\gamma_j)_{j \geq 1}$  be a nondecreasing sequence with  $\gamma_1 = 1$ . We assume that  $\varphi_0$  belongs to

$$\mathcal{F}_\gamma := \left\{ \phi \in \mathcal{L}_Z^2 : \sup_{z \in \mathcal{Z}} |\phi(z) - e_{k_n}(z)^t \beta_{k_n}|^2 = O(\gamma_{k_n}^{-1}) \text{ for some } \beta_{k_n} \in \mathbb{R}^{k_n} \right\}.$$

Here, the sequence of weights  $\gamma$  measures the approximation error of  $\varphi_0$  with respect to the functions  $\{e_j\}_{j \geq 1}$ .

**ASSUMPTION 7.** (i) Let  $\varphi_0 \in \mathcal{F}_\gamma$  with nondecreasing sequence  $\gamma$  satisfying  $j^2 = o(\gamma_j)$ . (ii) There exists some constant  $\eta_e \geq 1$  such that  $\sup_{z \in \mathcal{Z}} \|e_{k_n}(z)\|^2 \leq \eta_e k_n$ . (iii) The smallest eigenvalue of  $\mathbb{E}[e_{\underline{k}}(Z)e_{\underline{k}}(Z)^t]$  is bounded away from zero uniformly in  $k$ . (iv)  $\mathbb{E}[U^2|Z]$  is bounded.

Assumption 7 (i) determines the required asymptotic behavior of the rate  $\gamma$ . For splines and power series this assumption is satisfied if the number of continuous derivatives of  $\varphi_0$

divided by the dimension of  $Z$  equals two. Assumption 7 (ii) and (iii) restrict the magnitude of the approximating functions  $\{e_j\}_{j \geq 1}$  and impose nonsingularity of their second moment matrix.

We are now in the position to prove the following asymptotic result for the standardized test statistic  $S_n^e$ . Here, a key requirement is that  $k_n = o(\varsigma_{m_n})$  implying that  $k_n = o(\sum_{j=1}^{m_n} \tau_j)$  and, in particular,  $k_n = o(m_n)$  if the smoothing operator  $L$  is the identity.

**THEOREM 4.1.** *Let Assumptions 1–4 and 7 be satisfied. If*

$$n = o(\gamma_{k_n} \varsigma_{m_n}), \quad k_n = o(\varsigma_{m_n}), \quad \text{and} \quad \left( \sum_{j=1}^{m_n} \tau_j \right)^3 = o(n) \quad (4.2)$$

then under  $H_e$  it holds

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

**EXAMPLE 4.1.** Let  $Z$  be continuously distributed with  $\dim(Z) = r$  and set  $L = \text{Id}$ . Consider the polynomial case where  $\gamma_j \sim j^{2p/r}$  with  $p > 1$  and let  $m_n \sim n^\nu$  with  $0 < \nu < 1/3$ . Let Assumption 5 hold true then  $\sqrt{m_n} = O(\varsigma_{m_n})$ . Hence, condition (4.2) is satisfied if  $k_n \sim n^\kappa$  with

$$r(1 - \nu/2)/(2p) < \kappa < \nu/2. \quad (4.3)$$

This ensures that the bias of this estimator in the statistic  $S_n^e$  is asymptotically negligible. Note that condition (4.3) requires  $2p > r(2/\nu - 1)$ . Hence, with a larger dimension  $r$  also the smoothness of  $\varphi_0$  has to increase, reflecting the curse of dimensionality.  $\square$

The next result states an asymptotic distribution result for the statistic  $S_n^e$  if  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . Let  $\Sigma^e$  be the covariance matrix of the infinite dimensional centered vector  $(U(f_j^r(W) - \sum_{l \geq 1} \mathbb{E}[f_j^r(W)e_l(Z)]e_l(Z)))_{j \geq 1}$ . The ordered eigenvalues of  $\Sigma^e$  are denoted by  $(\lambda_j^e)_{j \geq 1}$ .

**THEOREM 4.2.** *Let Assumptions 1–4 and 7 be satisfied. If*

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad n = O(\gamma_{k_n}), \quad k_n^3 = o(n), \quad \text{and} \quad m_n^{-1} = o(1) \quad (4.4)$$

then under  $H_e$  it holds

$$n S_n^e \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^e \chi_{1j}^2.$$

**EXAMPLE 4.2.** Consider the setting of Example 4.1 but where the eigenvalues of  $L$  satisfy  $\tau_j \sim j^{-2}$ . Condition (4.4) is satisfied if  $m_n \sim n^\nu$  for some  $\nu > 0$  and  $k_n \sim n^\kappa$  with  $r/(2p) < \kappa < 1/3$ . Here, the required smoothness of  $\varphi_0$  is  $p > 3r/2$ . In contrast to the setting of Theorem 4.1, the estimator of  $\varphi_0$  needs to be undersmoothed. This ensures that the bias of this estimator in the statistic  $S_n^e$  is asymptotically negligible.  $\square$

**REMARK 4.1.** In contrast to Blundell and Horowitz [2007] no smoothness assumptions on the joint distribution of  $(Z, W)$  is required here. In addition, we do not need any assumption that links the smoothness of the regression function  $\varphi_0$  to the smoothness of the joint density of  $(Z, W)$ .  $\square$

**REMARK 4.2 (Estimation of Critical Values).** For the estimation of critical values of Theorem 4.1 and 4.2, let us define  $\mathbf{U}_n^e = (Y_1 - \bar{\varphi}_{k_n}(Z_1), \dots, Y_n - \bar{\varphi}_{k_n}(Z_n))^t$ . For any  $m \geq 1$  we estimate the covariance matrix  $\Sigma_m$  by

$$\widehat{\Sigma}_m := n^{-1} \mathbf{W}_m(\tau)^t \text{diag}(\mathbf{U}_n^e)^2 \mathbf{W}_m(\tau).$$

Now the asymptotic result of Theorem 4.1 continues to hold if we replace  $\varsigma_{m_n}$  by the Frobenius norm of  $\widehat{\Sigma}_{m_n}$  and  $\mu_{m_n}$  by the trace of  $\widehat{\Sigma}_{m_n}$ . This consistency is shown in Lemma 4.3. In the setting of Theorem 4.2, we replace  $\Sigma^e$  by a finite dimensional matrix

$$\widehat{\Sigma}_M^e := n^{-1} \mathbf{W}_M(\tau)^t \left( I_n - n^{-1} \mathbf{Z}_{k_n} \mathbf{Z}_{k_n}^t \right) \text{diag}(\mathbf{U}_n^e)^2 \left( I_n - n^{-1} \mathbf{Z}_{k_n} \mathbf{Z}_{k_n}^t \right) \mathbf{W}_M(\tau)$$

where  $M > 0$  is a sufficiently large integer. Let  $(\widehat{\lambda}_j^e)_{1 \leq j \leq M_n}$  denote the ordered eigenvalues of  $\widehat{\Sigma}_{M_n}^e$ . Hence, we approximate  $\sum_{j=1}^{\infty} \lambda_j^e \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{M_n} \widehat{\lambda}_j^e \chi_{1j}^2$  where  $\max_{1 \leq j \leq M_n} |\widehat{\lambda}_j^e - \lambda_j^e| = o_p(1)$  if  $M_n = o(\sqrt{n})$ .  $\square$

**LEMMA 4.3.** Consider  $\widehat{\Sigma}_{m_n}$  as defined in Remark 4.2. Under conditions of Theorem 4.1 or Theorem 4.2 the difference of its Frobenius norm to  $\varsigma_{m_n}$  and its trace to  $\mu_{m_n}$  converge in probability to zero.

## 4.2. Limiting behavior under local alternatives and consistency.

Similar to the previous sections we study the power and consistency properties of our test. Let us study the power of our test of exogeneity under linear local alternatives (2.6) or (2.7). In these cases, it holds  $\mathbb{E}[U|W] = 0$  but  $\mathbb{E}[U|Z] = -\varsigma_{m_n}^{1/2} n^{-1/2} \delta(Z)$  under (2.6) or  $\mathbb{E}[U|Z] = -n^{-1/2} \delta(Z)$  under (2.7).

**PROPOSITION 4.4.** Given the conditions of Theorem 4.1 and Assumption 5 (ii) it holds under (2.6)

$$(\sqrt{2}\varsigma_{m_n})^{-1} (n S_n^e - \mu_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2} \sum_{j=1}^{\infty} \delta_j^2, 1\right).$$

Given the conditions of Theorem 4.2 it holds under (2.7)

$$n S_n^e \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^e \chi_{1j}^2 (\delta_j / \lambda_j^e).$$

Let us now establish consistency of our tests when  $H_e$  does not hold, that is,  $\mathbb{P}(\varphi = \varphi_0) < 1$ .

**PROPOSITION 4.5.** Assume that  $H_e$  does not hold. Let  $\mathbb{E}|Y - \varphi_0(Z)|^4 < \infty$  and Assumption 5 (i) hold true. Let  $(\alpha_n)_{n \geq 1}$  as in Proposition 2.5. Under the conditions of Theorem 4.1 we have

$$\mathbb{P}\left((\sqrt{2}\varsigma_{m_n})^{-1} (n S_n^e - \mu_{m_n}) > \alpha_n\right) = 1 + o(1),$$

whereas in the setting of Theorem 4.2

$$\mathbb{P}(n S_n^e > \alpha_n) = 1 + o(1).$$

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

$$\mathcal{I}_n^\rho = \left\{ \varphi \in \mathcal{L}_Z^2 : \|LT(\varphi - \varphi_0)\|_W^2 \geq \rho n^{-1} \varsigma_{m_n} \text{ and } \sup_{z \in \mathcal{Z}} |(\varphi - \varphi_0)(z)| \leq C \right\}$$

form some constant  $C > 0$ . For the next result let  $q_{1\alpha}$  and  $q_{2\alpha}$  denote the  $1 - \alpha$  quantile of  $\mathcal{N}(0, 1)$  and  $\sum_{j=1}^{\infty} \lambda_j^e \chi_{1j}^2$ , respectively.

**PROPOSITION 4.6.** *Let Assumption 5 be satisfied. 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*

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

whereas under the conditions of Theorem 4.2 it holds

$$\lim_{n \rightarrow \infty} \inf_{\varphi \in \mathcal{I}_n^\rho} \mathbb{P} (n S_n^e > q_{2\alpha}) \geq 1 - \varepsilon.$$

## 5. A nonparametric specification test

A solution to the linear operator equation (2.1) only exists if  $g$  belongs to the range of  $T$ . This might be violated if, for instance, the instrument is not valid, that is,  $\mathbb{E}[U|W] \neq 0$ . In many economic applications *a priori* smoothness restriction on the unknown function can be justified which we capture by a set of functions  $\mathcal{F}$ . We consider the hypothesis  $H_{np}$ : there exists a solution  $\varphi_0 \in \mathcal{F}$  to (2.1). The alternative hypothesis is that there exists a solution (2.1) which does not belong to  $\mathcal{F}$ . Under the alternative only unsmooth functions solve the conditional moment restriction which can be interpreted as a failure of validity of the instrument  $W$ . We see in this section that our results allow also for a test of dimension reduction of the vector of regressors  $Z$ , that is, whether some regressors can be omitted from the structural function  $\varphi_0$ .

### 5.1. Nonparametric estimation method

**The nonparametric estimator.** In the following, we derive an estimator of  $\varphi_0$  under the null hypothesis  $H_{np}$ . For simplicity, assume that  $\mathcal{Z} = \mathcal{W}$  and consider a sequence  $\{e_j\}_{j \geq 1}$  of approximating functions which are orthonormal on  $\mathcal{Z}$  with respect to the Lebesgue measure  $\nu$ . Under conditions given below,  $\varphi_0$  has the expansion  $\varphi_0(\cdot) = \sum_{l=1}^{\infty} \int \varphi_0(z) e_l(z) \nu(z) dz e_l(\cdot)$ . Thereby, the conditional moment restriction under  $H_{np}$  leads to the following unconditional moment restrictions

$$\mathbb{E}[Y e_j(W)] = \sum_{l=1}^{\infty} \mathbb{E}[e_j(W) e_l(Z)] \int \varphi_0(z) e_l(z) \nu(z) dz \quad (5.1)$$

for  $j \geq 1$ . This motivates the following orthogonal series type estimator. Let  $\mathbf{Z}_k$  and  $\mathbf{Y}_n$  be as in the previous section and let  $\mathbf{X}_k$  denote a  $n \times k$  matrix with entries  $e_j(W_i)$  for  $1 \leq i \leq n$  and  $1 \leq j \leq k$ . Then for each  $k \geq 1$  we consider the estimator

$$\widehat{\varphi}_k(\cdot) := \underline{e}_k(\cdot)^t \widehat{\beta}_k \quad \text{where} \quad \widehat{\beta}_k = (\mathbf{X}_k^t \mathbf{Z}_k)^{-1} \mathbf{X}_k^t \mathbf{Y}_n. \quad (5.2)$$

Under conditions given below  $\mathbf{X}_{k_n}^t \mathbf{Z}_{k_n}$  will be nonsingular with probability approaching one and hence its generalized inverse will be the standard inverse. The nonparametric estimator  $\widehat{\varphi}_k$  given in (5.2) was studied by Johannes and Schwarz [2010], Horowitz [2011], and Horowitz [2012].

**Additional assumptions.** In the following, we specify *a priori* smoothness assumptions captured by the set  $\mathcal{F}$ . As noted by Horowitz [2012], uniformly consistent testing of  $H_{np}$  is only possible if the null is restricted that any solution to (2.1) is smooth. Here, we assume that under the null hypothesis  $\varphi_0$  belongs to the ellipsoid  $\mathcal{F} := \mathcal{F}_\gamma^\rho := \{\phi \in \mathcal{L}_Z^2 : \sum_{j=1}^\infty \gamma_j \mathbb{E}[\phi(Z)e_j(Z)]^2 \leq \rho\}$ . As in the previous section,  $\gamma = (\gamma_j)_{j \geq 1}$  measures the approximation error of  $\varphi_0$  with respect to the basis  $\{e_j\}_{j \geq 1}$ .

Further, as usual in the context of nonparametric instrumental regression, we specify some mapping properties of the conditional expectation operator  $T$ . Denote by  $\mathcal{T}$  the set of all nonsingular operators mapping  $\mathcal{L}_Z^2$  to  $\mathcal{L}_W^2$ . Given a sequence of weights  $v := (v_j)_{j \geq 1}$  and  $d \geq 1$  we define the subset  $\mathcal{T}_d^v$  of  $\mathcal{T}$  by

$$\mathcal{T}_d^v := \left\{ T \in \mathcal{T} : \int |(T\phi)(w)|^2 \nu(w) dw \leq d \sum_{j=1}^\infty v_j \left( \int \phi(z) e_j(z) \nu(z) dz \right)^2 \text{ for all } \phi \in \mathcal{L}_Z^2 \right\}.$$

If  $p_Z/\nu$  is bounded from above and  $p_W/\nu$  is uniformly bounded away from zero then the conditional expectation operator  $T$  belongs to  $\mathcal{T}_d^v$  with  $v_j = 1$ ,  $j \geq 1$ , due to Jensen's inequality. Notice that for all  $T \in \mathcal{T}_d^v$  it follows that  $\|Te_j\|_W^2 \leq d\eta_p v_j$  and thereby, the condition  $T \in \mathcal{T}_d^v$  links the operator  $T$  to the basis  $\{e_j\}_{j \geq 1}$ . In the following, we denote  $[T]_{\underline{k}} = \mathbb{E}[e_{\underline{k}}(W)e_{\underline{k}}(Z)^t]$  which is assumed to be a nonsingular matrix. In what follows, we introduce a stronger condition on the basis  $\{e_l\}_{l \geq 1}$ . We denote by  $\mathcal{T}_{d,D}^v$  for some  $D \geq d$  the subset of  $\mathcal{T}_d^v$  given by

$$\mathcal{T}_{d,D}^v := \left\{ T \in \mathcal{T}_d^v : [T]_{\underline{k}} \text{ is nonsingular and } \sup_{k \geq 1} \|\text{diag}(v_1, \dots, v_k)^{1/2} [T]_{\underline{k}}^{-1}\|^2 \leq D \right\}.$$

The class  $\mathcal{T}_{d,D}^v$  only contains operators  $T$  whose off-diagonal elements of  $[T]_{\underline{k}}^{-1}$  are sufficiently small for all  $k \geq 1$ . A similar diagonality restriction has been used by Hall and Horowitz [2005] or Breunig and Johannes [2011]. Besides the mapping properties for the operator  $T$  we need a stronger assumption for the basis under consideration. The following condition gathers conditions on the sequences  $\gamma$  and  $v$ .

**ASSUMPTION 8.** (i) Under  $H_{np}$ , let  $\varphi_0 \in \mathcal{F}_\gamma^\rho$  with nondecreasing sequence  $\gamma$  satisfying  $j^3 = o(\gamma_j)$ . (ii) The sequence  $\{e_j\}_{j \geq 1}$  is an orthogonal basis on  $\mathcal{Z} = \mathcal{W}$  with respect to  $\nu$ . (iii) There exists some constant  $\eta_e \geq 1$  such that  $\sup_{j \geq 1} \sup_{z \in \mathcal{Z}} |e_j(z)| \leq \eta_e$ . (iv) Let  $T \in \mathcal{T}_{d,D}^v$  with  $v$  being a strictly positive sequences such that  $v$  and  $(v_j/\tau_j)_{j \geq 1}$  are nonincreasing. (v)  $p_Z/\nu$  is bounded from above and  $p_W/\nu$  is uniformly bounded away from zero.

Due to Assumption 8 (iv) the degree of additional smoothing for our testing procedure must not be stronger than the degree of ill-posedness implied by the conditional expectation operator  $T$ . Under similar assumptions as above, Johannes and Schwarz [2010] show that mean integrated squared error loss of  $\hat{\varphi}_{k_n}$  attains the optimal rate of convergence  $\mathcal{R}_n := \max(\gamma_{k_n}^{-1}, \sum_{j=1}^{k_n} (nv_j)^{-1})$ . Due to Assumption 8 (v) we do not require orthonormal bases with respect to the unknown distribution  $(Z, W)$  (cf. Remark 3.2 of Breunig and Johannes [2011]).

## 5.2. The test statistic and its asymptotic distribution

As in the previous sections, our test is based on the observation that the null hypothesis  $H_{np}$  is equivalent to  $L(g - T\varphi_0) = 0$ . Our goodness-of-fit statistic for testing nonparametric



specifications is given by  $S_n$  where  $\varphi_0$  is replaced by the nonparametric estimator  $\widehat{\varphi}_{k_n}$  given in (5.2), that is,

$$S_n^{np} := \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_{i=1}^n (Y_i - \widehat{\varphi}_{k_n}(Z_i)) f_j(W_i) \right|^2.$$

If  $S_n^{np}$  becomes too large then there exists no function in  $\mathcal{F}_\gamma^\rho$  solving (2.1). The next result establishes asymptotic normality of  $S_n^{np}$  after standardization. Again, a key requirement to obtain this asymptotic distribution is that  $k_n = o(\varsigma_{m_n})$  implying that  $k_n = o(m_n)$  if the smoothing operator  $L$  is the identity. This corresponds to the test of overidentification in the parametric framework where more orthogonality restrictions than parameters are required.

**THEOREM 5.1.** *Let Assumptions 1–4 and 8 be satisfied. If*

$$nv_{k_n} = o(\gamma_{k_n} \varsigma_{m_n}), \quad k_n = o(\varsigma_{m_n}), \quad k_n \left( \sum_{j=1}^{m_n} \tau_j \right)^2 = O(nv_{k_n}), \quad \text{and} \quad \left( \sum_{j=1}^{m_n} \tau_j \right)^3 = o(n) \quad (5.3)$$

then it holds under  $H_{np}$

$$(\sqrt{2\varsigma_{m_n}})^{-1} \left( nS_n^{np} - \mu_{m_n} \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

**EXAMPLE 5.1.** Consider the setting of Example 4.1. In the *mildly ill posed case* where  $v_j \sim j^{-2a/r}$  for some  $a \geq 0$  condition (5.3) holds true 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*, that is,  $v_j \sim \exp(-j^{2a/r})$  for some  $a > 0$ , condition (5.3) is satisfied if, for example,  $m_n$  satisfies  $m_n = o(k_n^p)$  and  $k_n = o(\sqrt{m_n})$  where  $k_n \sim (\log n - \log(m_n^{3/2}))^{r/(2a)}$ .  $\square$

The next result states an asymptotic distribution of our test if  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . Let  $\Sigma^{np}$  be the covariance matrix of the infinite dimensional centered vector  $(U(f_j^\tau(W) - e_j^\tau(W)))_{j \geq 1}$ . The ordered eigenvalues of  $\Sigma^{np}$  are denoted by  $(\lambda_j^{np})_{j \geq 1}$ .

**THEOREM 5.2.** *Let Assumptions 1–4 and 8 be satisfied. If*

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad nv_{k_n} = o(\gamma_{k_n}), \quad k_n^3 = o(nv_{k_n}), \quad \text{and} \quad m_n^{-1} = o(1) \quad (5.4)$$

then it holds under  $H_{np}$

$$nS_n^{np} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2.$$

**EXAMPLE 5.2.** Consider the setting of Example 4.2. In the *mildly ill posed case*, that is,  $v_j \sim j^{-2a/r}$  for some  $a \geq 0$ , condition (5.4) is satisfied if  $m_n \sim n^\nu$  for some  $\nu > 0$  and  $k_n \sim n^\kappa$  with

$$r/(2a + 2p) < \kappa < r/(2a + 3r).$$

In the *severely ill posed case*, that is,  $v_j \sim \exp(-j^{2a/r})$  for some  $a > 0$ , condition (5.4) is satisfied if  $k_n \sim (\log(n^{1+\varepsilon}))^{r/(2a)}$  for any  $\varepsilon > 0$ . In contrast to Theorem 5.1, we require undersmoothing of the estimator  $\widehat{\varphi}_{k_n}$ .  $\square$

**REMARK 5.1.** If the basis  $\{e_j\}_{j \geq 1}$  coincides with  $\{f_j\}_{j \geq 1}$  then  $nS_n^{np}$  is asymptotically degenerate. To avoid this degeneracy problem we choose different bases functions and hence, sample splitting as used by Horowitz [2012] is not necessary here.  $\square$

**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  $\varphi_0 \in \mathcal{F}_\gamma^p$  to (2.1) only depending on  $Z'$ . To be more precise consider the test statistic

$$S_n'^{np} := \left\| n^{-1} \sum_{i=1}^n (Y_i - \widehat{\varphi}_{k_n}(Z'_i)) f_{m_n}^\tau(W_i) \right\|^2$$

where  $\widehat{\varphi}_{k_n}$  is the estimator (5.2) 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  $T' : \mathcal{L}_{Z'}^2 \rightarrow \mathcal{L}_W^2$  with  $T'\phi := \mathbb{E}[\phi(Z')|W]$ . It is interesting to note that if  $T$  is nonsingular then also  $T'$  is. 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  $Z'$ .  $\square$

In the mildly ill-posed case, 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  $\varphi_0$ . 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]). As the next example illustrates, a dimension reduction test can also weaken the required restrictions on the instrument to obtain identification of  $\varphi$  in the restricted model.

**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  $\phi \in \mathcal{L}_{Z^{(1)}}^2$  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  $(Z^{(2)}, W)$ . Thereby only the operator  $T^{(1)} : \mathcal{L}_{Z^{(1)}}^2 \rightarrow \mathcal{L}_W^2$  with  $T^{(1)}\phi := \mathbb{E}[\phi(Z^{(1)})|W]$  is nonsingular but  $T$  is singular. If our dimension reduction test of  $Z$  indicates that  $Z^{(2)}$  can be omitted from the structural function  $\varphi_0$  then we obtain identification in the restricted model.  $\square$

**REMARK 5.3.** [Estimation of Critical Values] For the estimation of critical values of Theorem 5.1 and 5.2, let us define  $\mathbf{U}_n^{np} = (Y_1 - \widehat{\varphi}_{k_n}(Z_1), \dots, Y_n - \widehat{\varphi}_{k_n}(Z_n))^t$ . For all  $m \geq 1$ , we estimate the covariance matrix  $\Sigma_m$  by

$$\widehat{\Sigma}_m := n^{-1} \mathbf{W}_m(\tau)^t \text{diag}(\mathbf{U}_n^{np})^2 \mathbf{W}_m(\tau).$$

Now the asymptotic result of Theorem 5.1 continues to hold if we replace  $\varsigma_{m_n}$  by the Frobenius norm of  $\widehat{\Sigma}_{m_n}$  and  $\mu_{m_n}$  by the trace of  $\widehat{\Sigma}_{m_n}$  (this is easily seen from the proof of Lemma 4.3 assuming that  $\{f_j\}_{j \geq 1}$  is uniformly bounded). In the setting of Theorem 5.2, we replace  $\Sigma^{np}$  by a finite dimensional matrix. Let  $\mathbf{V}_k := \mathbf{W}_k(\mathbf{Z}_k^t \mathbf{W}_k)^{-1} \mathbf{Z}_k^t$  for  $k \geq 1$ . Then for a sufficiently large integer  $M > 0$  we estimate  $\Sigma^{np}$  by

$$\widehat{\Sigma}_M^{np} := n^{-1} \mathbf{W}_M(\tau) (I_n - \mathbf{V}_{k_n})^t \text{diag}(\mathbf{U}_n^{np})^2 (I_n - \mathbf{V}_{k_n}) \mathbf{W}_M(\tau).$$

Hence, we approximate  $\sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{M_n} \widehat{\lambda}_j^{np} \chi_{1j}^2$  where  $(\widehat{\lambda}_j^{np})_{1 \leq j \leq M_n}$  are the ordered eigenvalues of  $\widehat{\Sigma}_{M_n}^{np}$  where  $\max_{1 \leq j \leq M_n} |\widehat{\lambda}_j^{np} - \lambda_j^{np}| = o_p(1)$  if  $M_n = o(\sqrt{n})$ .  $\square$

### 5.3. Limiting behavior under local alternatives and consistency.

Similar to the previous sections we study the power and consistency properties of our test. To study the power against local alternatives of the statistic  $S_n^{np}$  we consider alternative models with the function  $\varphi_{k_n}(\cdot) = e_{k_n}(\cdot)^t [T]_{k_n}^{-1} \mathbb{E}[Y f_{k_n}(W)]$ . We consider alternative models

$$Y = \varphi_{k_n}(Z) + \varsigma_{m_n}^{1/2} n^{-1/2} \delta(Z) + U \quad (5.5)$$

for some function  $\delta \in L_Z^4$  and where  $\mathbb{E}[U|W] = 0$ . Let  $\varphi$  be such that  $\mathbb{E}[Y - \varphi(Z)|W] = 0$ . Due to (5.5)  $\varphi$  does not belong  $\mathcal{F}_\gamma^\rho$  and hence  $H_{np}$  fails. Indeed, if  $\varphi \in \mathcal{F}_\gamma^\rho$  then we show in the appendix that  $\|T(\varphi - \varphi_{k_n})\|_W^2 = O(v_{k_n} \gamma_{k_n}^{-1}) = o(\varsigma_{m_n} n^{-1})$  due to condition (5.3) (or (5.4)), which is in contrast to (5.5).

**PROPOSITION 5.3.** *Let Assumption 5 (ii) be satisfied. Given the conditions of Theorem 5.1 it holds under (5.5)*

$$(\sqrt{2}\varsigma_{m_n})^{-1} (n S_n^{np} - \mu_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2} \sum_{j=1}^{\infty} \xi_j^2, 1\right).$$

Given the conditions of Theorem 5.2 it holds under (5.5) where  $\varsigma_{m_n}$  is replaced by 1 that

$$n S_n^{np} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2 (\delta_j / \lambda_j^{np}).$$

In the next proposition, we establish consistency of our test when  $H_{np}$  does not hold, that is, the solution to (2.1) does not belong to  $\mathcal{F}_\gamma^\rho$  for any sequence  $\gamma$  satisfying Assumption 8 and any sufficiently large constant  $0 < \rho < \infty$ .

**PROPOSITION 5.4.** *Assume that  $H_{np}$  does not hold. Let Assumption 5 (i) hold true. Let  $(\alpha_n)_{n \geq 1}$  be as in Proposition 2.5. Under the conditions of Theorem 5.1 and 5.2, respectively, we have*

$$\begin{aligned} \mathbb{P}\left((\sqrt{2}\varsigma_{m_n})^{-1} (n S_n^{np} - \mu_{m_n}) > \alpha_n\right) &= 1 + o(1), \\ \mathbb{P}(n S_n^{np} > \alpha_n) &= 1 + o(1). \end{aligned}$$

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

$$\mathcal{J}_n^\rho = \left\{ \varphi \in \mathcal{L}_Z^2 : \|LT(\varphi - \varphi_0)\|_W^2 \geq \rho n^{-1} \varsigma_{m_n} \text{ and } \sup_{z \in \mathcal{Z}} |(\varphi - \varphi_0)(z)| \leq C \right\}$$

where  $\varphi_0 \in \mathcal{F}_\gamma^\rho$  solves (2.1) and  $C > 0$  is a finite constant. For the next result let  $q_{1\alpha}$  and  $q_{2\alpha}$  denote the  $1 - \alpha$  quantile of  $\mathcal{N}(0, 1)$  and  $\sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2$ , respectively.

**PROPOSITION 5.5.** *Let Assumption 5 be satisfied. For any  $\varepsilon > 0$ , any  $0 < \alpha < 1$ , and any sufficiently large constant  $\rho > 0$  we have under the conditions of Theorem 5.1*

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

whereas under the conditions of Theorem 5.2 it holds

$$\lim_{n \rightarrow \infty} \inf_{\varphi \in \mathcal{J}_n^\rho} \mathbb{P}(n S_n^{np} > q_{2\alpha}) \geq 1 - \varepsilon.$$

## 6. Monte Carlo simulation

In this section, we study the finite-sample performance of our test by presenting the results of Monte Carlo experiments. There are 1000 Monte Carlo replications in each experiment. Results are presented for the nominal level 0.05. Realizations of  $Y$  were generated from

$$Y = \varphi(Z) + c_U U \quad (6.1)$$

for some constant  $c_U > 0$  specified below. The structural function  $\varphi$  and the joint distribution of  $(Z, W, U)$  varies in the experiments below. As basis  $\{f_j\}_{j \geq 1}$  we choose cosine basis functions given by  $f_j(t) = \sqrt{2} \cos(\pi j t)$  for  $j = 1, 2, \dots$  throughout this simulation study.

**Parametric Specification** Let us investigate the finite sample performance of our tests in the case of parametric specifications. Realizations  $(Z, W)$  were generated by  $W \sim \mathcal{U}[0, 1]$ ,  $Z = (\xi W + (1 - \xi) \varepsilon)^2$  where  $\xi = 0.8$  and  $\varepsilon \sim \mathcal{N}(0.5, 0.1)$ . Moreover, let  $U = \kappa \varepsilon + \sqrt{1 - \kappa^2} \epsilon$  with  $\kappa = 0.3$  and  $\epsilon \sim N(0, 1)$ . Then realizations of  $Y$  were generated by (6.1) with  $c_U = 0.2$  by an either linear function

$$\varphi(z) = z, \quad (6.2)$$

a polynomial of second degree

$$\varphi(z) = z - z^2, \quad (6.3)$$

or a polynomial of third degree

$$\varphi(z) = z - z^2 + \theta_3 z^3. \quad (6.4)$$

Given (6.4) is the correct model, then  $\theta_3 = 1.5$  if (6.2) is the null model and  $\theta_3 = 3$  if (6.2) is the null model. In Table 1 we depict the empirical rejection probabilities when using  $S_n^p$  with additional smoothing where either  $\tau_j = j^{-1}$  or  $\tau_j = j^{-2}$ ,  $j \geq 1$ , which we denote by  $S_n^{1p}$  or  $S_n^{2p}$ , respectively. When  $\tau_j = j^{-1}$  then the number of basis functions used is

Sample Size	Null Model	Alt. Model	Empirical Rejection probability		
			$S_n^{1p}$	$S_n^{2p}$	$H(2006)$ ' test
250	(6.2)	$H^p$ true	0.047	0.045	0.063
	(6.3)	$H^p$ true	0.049	0.050	0.059
	(6.2)	(6.3)	0.902	0.930	0.888
	(6.2)	(6.4)	0.730	0.732	0.653
	(6.3)	(6.4)	0.442	0.488	0.468
500	(6.2)	$H^p$ true	0.055	0.044	0.053
	(6.3)	$H^p$ true	0.051	0.053	0.059
	(6.2)	(6.3)	0.989	0.998	0.988
	(6.2)	(6.4)	0.899	0.894	0.780
	(6.3)	(6.4)	0.709	0.728	0.652

Table 1: Empirical Rejection probabilities for parametric specification

$m = 200$  while in the case of  $\tau_j = j^{-2}$  a choice of  $m = 100$  is sufficient. The critical values are estimated as described in Remark 3.1 where  $M = 150$  if  $\tau_j = j^{-1}$  and  $M = 100$  if  $\tau_j = j^{-2}$ . This choice of  $M$  ensures that the estimated eigenvalues  $\hat{\lambda}_j$  are sufficiently close to zero for all  $j \geq M$ . We compare our test statistic with the test of Horowitz [2006]. We follow his implementation using biweight kernels. The bandwidth used to estimate the joint density of  $(Z, W)$  was also selected by cross validation. As Table 1 illustrates, the results for  $S_n^{1p}$  and  $S_n^{2p}$  are quite similar. In both situations, our test is more powerful than the test of Horowitz [2006] when testing (6.2) against (6.4). In this simulation study, we observed that the estimated coefficients of  $T(\varphi - \phi(\vartheta_0, \cdot))$  have a fast decay. Consequently, the test statistic  $S_n$  with no weighting has less power, as we discussed in Subsection 2.4. In contrast, we will demonstrate by the end of this section that using weights can be inappropriate.

**Testing Exogeneity** We now turn to the test of exogeneity where the realizations  $(Z, W)$  are generated by  $W \sim \mathcal{U}[0, 1]$  and  $Z = \xi W + \sqrt{1 - \xi^2} \varepsilon$  with  $\xi = 0.7$ , and  $\varepsilon \sim \mathcal{U}[0, 1]$ . Moreover, let  $U = \kappa \varepsilon + \sqrt{1 - \kappa^2} \epsilon$  with  $\epsilon \sim \mathcal{U}[0, 1]$ . Here,  $\kappa$  measures the degree of endogeneity of  $Z$  and is varied among the experiments. The null hypothesis  $H_0$  holds true if  $\kappa = 0$  and is false otherwise. Now realizations of  $Y$  were generated by (6.1) with  $c_U = 1$  and the nonparametric structural function  $\varphi_1(z) = \sum_{j=1}^{\infty} (-1)^{j+1} j^{-1} \sin(j\pi z)$ . For computational reasons we truncate the infinite sum at  $K = 100$ . The resulting function is displayed in Figure 6. We estimate the structural relationship using Lagrange polynomials. Indeed, only a few basis functions are necessary to accurately approximate the true function. If we choose  $k_n$  too small or too large then the estimator will be a poor approximate of the true structural function and hence, the test statistic will reject  $H_{np}$ . In this experiment we set  $k_n = 4$  for  $n = 250$  and  $n = 500$ .

Sample Size	$\kappa$	Empirical Rejection probability using		
		$S_n^{1e}$	$S_n^{2e}$	BH(2007)' test
250	0.0	0.038	0.030	0.030
	0.15	0.209	0.314	0.153
	0.2	0.369	0.513	0.293
	0.25	0.591	0.716	0.504
500	0.0	0.043	0.043	0.052
	0.15	0.476	0.543	0.416
	0.2	0.749	0.809	0.693
	0.25	0.922	0.957	0.885

Table 2: Empirical Rejection probabilities for testing exogeneity

In Table 2 we depict the empirical rejection probabilities when using  $S_n^e$  with additional smoothing where either  $\tau_j = j^{-1}$  or  $\tau_j = j^{-2}$ ,  $j \geq 1$ , which we denote by  $S_n^{1e}$  or  $S_n^{2e}$ , respectively. The critical values of these statistics are estimated as described in Remark 4.2 with  $M = 50$  in case of  $\tau_j = j^{-1}$  and  $M = 40$  in case of  $\tau_j = j^{-2}$ . We compare our results with the test of Blundell and Horowitz [2007]. We follow their approach by choosing the bandwidth of the joint density of  $(Z, W)$  by cross validation. The bandwidth of the marginal of  $Z$  is  $n^{1/5-7/24}$  times the cross-validation bandwidth. As we see from Table 2,  $S_n^{1e}$  is slightly more powerful than the test of Blundell and Horowitz [2007]. If we choose a

stronger sequence, however, then our test statistic  $S_n^{2e}$  becomes considerably more powerful.

**Nonparametric Specification** Let us now study the finite sample of our test in the case of nonparametric specification. We generate the pair  $(Z, W)$  as in the parametric case described above. For the generation of the dependent variable  $Y$  we distinguish two cases. Besides the structural function  $\varphi_1(z) = \sum_{j=1}^{\infty} (-1)^{j+1} j^{-2} \sin(j\pi z)$  we also consider the function  $\varphi_2(z) = \sum_{j=1}^{\infty} ((-1)^{j+1} + 1)/4 j^{-2} \sin(j\pi z)$ . Again, for computational reasons we truncate the infinite sum at  $K = 100$ . The resulting functions are displayed in Figure 6. Further,  $Y$  is generated by (6.1) either with  $\varphi_1$  and  $c_U = 0.2$  or  $\varphi_2$  and  $c_U = 0.8$ . In both cases, we estimate the structural relationship using Lagrange polynomials with  $k_n = 4$  for  $n = 500$  and  $n = 1000$ .

If  $H_{np}$  is false then  $\mathbb{E}[U|W] \neq 0$  and we let  $\mathbb{E}[U|W] = \mathbb{E}[\rho(Z)|W]$  where  $\rho$  is defined below. Consequently, when  $H_{np}$  is false we generate realizations of  $Y$  from

$$Y = \varphi_l(Z) + \rho_j(Z) + c_U U$$

for  $l = 1, 2$  and  $j \geq 1$  where  $\rho_j(z) = c_j(\exp(2jz) \mathbb{1}_{\{z \leq 1/2\}} + \exp(2j(1-z)) \mathbb{1}_{\{z > 1/2\}} - 1)$  and  $c_j$  is a normalizing constant such that  $\int_0^1 \rho_j(z) dz = 0.5$ . The functions  $\rho_j$  are continuous but not differentiable at 0.5. Roughly speaking, the degree of roughness of  $\rho_j$  is larger for larger  $j$ . In Table 3, we depict the empirical rejection probabilities when using  $S_n^{np}$  with

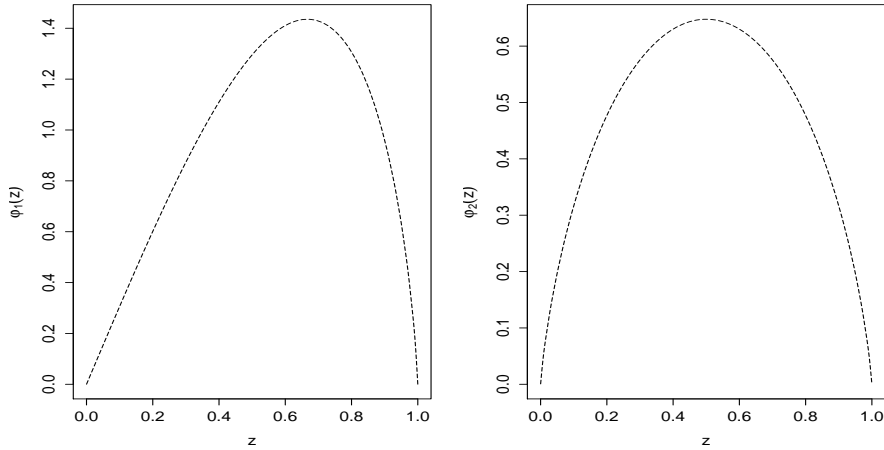


Figure 1: Graph of  $\varphi_1$  and  $\varphi_2$

either no smoothing or additional smoothing  $\tau_j = j^{-2}$ ,  $j \geq 1$ , which we denote by  $S_n^{0np}$  or  $S_n^{2np}$ , respectively. When no additional smoothing is applied then the number of basis functions  $f_j$  is given by  $m_n = 11$  if  $n = 500$  and  $m_n = 15$  if  $n = 1000$  and hence, the choice of  $m_n$  is slightly larger than  $n^{1/3}$  as suggested by the theoretical results. The critical values of these statistics are estimated as described in Remark 5.3 where in the case of  $S_n^{2np}$  we choose  $M = 100$ . We compare our results with the test of Horowitz [2012]. We observe that the statistic  $S_n^{0np}$  is less powerful than  $S_n^{2np}$  against the alternatives  $\rho_1$  and  $\rho_2$ .

In the following, we illustrate that using additional weighting can be inappropriate. Table 4 illustrates the power of our tests when the structural function  $\varphi_2$  is considered and realizations  $(Z, W)$  were generated by  $W \sim \mathcal{U}[0, 1]$ ,  $Z = (0.8W + 0.3\varepsilon)^2$  where  $\varepsilon \sim \mathcal{N}(0.5, 0.05)$ .

<i>Sample Size</i>	$\rho$	<i>Empirical Rejection probability using</i>		
		$S_n^{0np}$	$S_n^{2np}$	<i>H(2012)' test</i>
500	$H^{np}$ true	0.034	0.039	0.040
	$\rho_1$	0.099	0.382	0.258
	$\rho_2$	0.309	0.765	0.536
	$\rho_4$	0.498	0.884	0.712
1000	$H^{np}$ true	0.058	0.058	0.046
	$\rho_1$	0.405	0.672	0.427
	$\rho_2$	0.768	0.899	0.704
	$\rho_4$	0.920	0.943	0.808

Table 3: Empirical Rejection prob. for Nonparametric Specification for  $\varphi_1$  with  $c_U = 0.2$

In this case, we generate  $Y$  using (6.1) where  $c_U = 0.8$ . In this case, the estimates of the generalized coefficients of  $T(\varphi - \varphi_0)$  are more fluctuating and using weights is not appropriate here. Indeed, as we can see from Table 4, the test statistic  $S_n^{0np}$  with no smoothing is more powerful than  $S_n^{2np}$  were weighting  $\tau_j = j^{-2}$ ,  $j \geq 1$ , is used. In particular,  $S_n^{0np}$  is much more powerful than the test of Horowitz [2012].

<i>Sample Size</i>	$\rho$	<i>Empirical Rejection probability using</i>		
		$S_n^{0np}$	$S_n^{2np}$	<i>H(2012)' test</i>
500	$H^{np}$ true	0.022	0.044	0.044
	$\rho_3$	0.230	0.193	0.158
	$\rho_4$	0.400	0.319	0.245
	$\rho_5$	0.543	0.463	0.370
1000	$H^{np}$ true	0.044	0.049	0.052
	$\rho_3$	0.643	0.343	0.302
	$\rho_4$	0.836	0.579	0.518
	$\rho_5$	0.924	0.792	0.722

Table 4: Empirical Rejection prob. for Nonparametric Specification for  $\varphi_2$  with  $c_U = 0.8$

## 7. 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. The implementation of these statistics is straightforward. We have seen that the asymptotic results depend crucially on the choice of the smoothing operator  $L$ . By choosing a stronger decaying sequence  $\tau$ , our test becomes more powerful with respect to local alternatives but might lose desirable consistency properties. We gave heuristic arguments how to choose the weights in practice. In addition, in a Monte Carlo investigation our tests perform well in finite samples.



## A. Appendix

Throughout the Appendix, let  $C > 0$  denote a generic constant that may be different in different uses. For ease of notation let  $\sum_i = \sum_{i=1}^n$  and  $\sum_{i' < i} = \sum_{i=1}^n \sum_{i'=1}^{i-1}$ . Given  $m \geq 1$ ,  $\mathcal{E}_m$  and  $\mathcal{F}_m$  denote the subspace of  $\mathcal{L}_Z^2$  and  $\mathcal{L}_W^2$  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. Respectively, given functions  $\phi \in \mathcal{L}_Z^2$  and  $\psi \in \mathcal{L}_W^2$  we define by  $[\phi]_m$  and  $[\psi]_m$   $m$ -dimensional vectors with entries  $[\phi]_j = \mathbb{E}[\phi(Z)e_j(Z)]$  and  $[\psi]_l = \mathbb{E}[\psi(W)f_l(W)]$  for  $1 \leq j, l \leq m$ .

### A.1. Proofs of Section 2.

**PROOF OF THEOREM 2.1.** Under  $H_0$  we have  $(Y_i - \varphi_0(Z_i))f_m^\tau(W_i) = U_i f_m^\tau(W_i)$  for all  $m \geq 1$  and consequently we observe

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

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

$$\frac{1}{(\varsigma_m n)^2} \mathbb{E} \left| \sum_i \sum_{j=1}^m |U_i f_j^\tau(W_i)|^2 - s_{jj} \right|^2 = \frac{1}{n \varsigma_m^2} \mathbb{E} \left| \sum_{j=1}^m |U f_j^\tau(W)|^2 - s_{jj} \right|^2 \leq \frac{1}{n \varsigma_m^2} \mathbb{E} \|U f_m^\tau(W)\|^4.$$

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

$$\mathbb{E} \|U f_m^\tau(W)\|^4 \leq \max_{1 \leq j \leq m} \mathbb{E}|U f_j(W)|^4 \left( \sum_{j=1}^m \tau_j \right)^2 \leq \eta_f \eta_p \sigma^4 \left( \sum_{j=1}^m \tau_j \right)^2. \quad (\text{A.1})$$

Let  $m = m_n$  satisfy condition (2.4) then  $\mathbb{E} \|U f_{m_n}^\tau(W)\|^4 = o(n \varsigma_{m_n}^2)$ . Therefore, it is sufficient to prove

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

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 is sufficient to study the asymptotic behavior of  $n^{-1} \sum_{j=1}^{m_n} \sum_{i \neq i'} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'})$ . For any finite  $m \geq 1$  we obtain

$$\begin{aligned} & \mathbb{E} \left| \frac{1}{n} \sum_{j=1}^m \sum_{i \neq i'} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'}) - \frac{1}{n} \sum_{j=1}^{\infty} \sum_{i \neq i'} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'}) \right|^2 \\ & \leq \mathbb{E} \left[ \mathbb{E}[U_1^2 U_2^2 | W_1, W_2] \left( \sum_{j>m} f_j^\tau(W_1) f_j^\tau(W_2) \right)^2 \right] \leq \sigma^4 \eta_p \left( \sum_{j>m} \tau_j \right)^2 \end{aligned}$$

which, since  $\sum_{j \geq 1} \tau_j = O(1)$ , becomes sufficiently small (depending on  $m$ ). Note that  $(\frac{1}{\sqrt{n}} \sum_i U_i f_1^\tau(W_i), \dots, \frac{1}{\sqrt{n}} \sum_i U_i f_m^\tau(W_i)) \xrightarrow{d} \mathcal{N}(0, \Sigma_m)$ . Hence, for any finite  $m \geq 1$  we have

$$\sum_{j=1}^m \left| \frac{1}{\sqrt{n}} \sum_i U_i f_j^\tau(W_i) \right|^2 \xrightarrow{d} \sum_{j=1}^m \lambda_j \chi_{1j}^2$$

with  $\lambda_j$ ,  $1 \leq j \leq m$ , being eigenvalues of  $\Sigma_m$ . Moreover, we conclude for  $m \geq 1$

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^m \sum_{i \neq i'} U_i U_{i'} f_j^\tau(W_i) f_j^\tau(W_{i'})^t &= \sum_{j=1}^m \left( \left| \frac{1}{\sqrt{n}} \sum_i U_i f_j^\tau(W_i) \right|^2 - \frac{1}{n} \sum_i |U_i f_j^\tau(W_i)|^2 \right) \\ &\xrightarrow{d} \sum_{j=1}^m (\lambda_j \chi_{1j}^2 - s_{jj}). \end{aligned}$$

It is easily seen that  $\sum_{j=1}^m (\lambda_j \chi_{1j}^2 - s_{jj})$  has expectation zero. Hence, following the lines of page 198-199 of Serfling [1981] we obtain that  $\sum_{j > m} (\lambda_j \chi_{1j}^2 - s_{jj})$  becomes sufficiently small (depending on  $m$ ) and thus, completes the proof.  $\square$

**PROOF OF PROPOSITION 2.3.** For ease of notation let  $\delta_n(\cdot) := \zeta_{m_n}^{1/2} n^{-1/2} \delta(\cdot)$ . Under the sequence of alternatives (2.6) the following decomposition holds true

$$\begin{aligned} S_n &= \left\| n^{-1} \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\|^2 + 2 \left\langle n^{-1} \sum_i U_i f_{\underline{m}_n}^\tau(W_i), n^{-1} \sum_i \delta_n(Z_i) f_{\underline{m}_n}^\tau(W_i) \right\rangle \\ &\quad + \left\| n^{-1} \sum_i \delta_n(Z_i) f_{\underline{m}_n}^\tau(W_i) \right\|^2 =: I_n + 2II_n + III_n. \end{aligned}$$

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

$$\begin{aligned} n \mathbb{E} |II_n| &\leq \sum_{j=1}^{m_n} \tau_j (\mathbb{E} |U f_j(W)|^2 \mathbb{E} |\delta_n(Z) f_j(W)|^2)^{1/2} + \left( n \mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j [T \delta_n]_j U f_j(W) \right|^2 \right)^{1/2} \\ &\leq \sigma \sum_{j=1}^{m_n} \tau_j (\mathbb{E} |\delta_n(Z) f_j(W)|^2)^{1/2} + \sigma \eta_p \sqrt{n} \|T \delta_n\|_W. \end{aligned}$$

From the definition of  $\delta_n$  and condition (2.4) we infer that  $n \mathbb{E} |II_n| = o(\zeta_{m_n})$ . Consider  $III_n$ . Employing again the definition of  $\delta_n$  it is easily seen that  $n \zeta_{m_n}^{-1} III_n = \sum_{j=1}^{m_n} \tau_j [T \delta_n]_j^2 + o_p(1)$ . We conclude  $(\sqrt{2}\zeta_{m_n})^{-1} n III_n = 2^{-1/2} \sum_{j \geq 1} \delta_j^2 + o_p(1)$ , 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 2.2 it is straightforward to see that under the sequence of alternatives (2.7) it holds

$$\begin{aligned} &\frac{1}{n} \sum_{i \neq i'} \sum_{j=1}^{m_n} (U_i + \delta_n(Z_i))(U_{i'} + \delta_n(Z_{i'})) f_j^\tau(W_i) f_j^\tau(W_{i'}) \\ &= \sum_{j=1}^{\infty} \left( \left| \frac{1}{\sqrt{n}} \sum_i U_i f_j^\tau(W_i) + \frac{1}{n} \sum_i \delta(Z_i) f_j^\tau(W_i) \right|^2 - \frac{1}{n} \sum_i |U_i f_j^\tau(W_i)|^2 \right) \\ &\quad \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j} (\delta_j / \lambda_j) \end{aligned}$$

similar to the lines of page 198-199 of Serfling [1981] and hence the assertion follows.  $\square$

**PROOF OF PROPOSITION 2.5.** If  $H_0$  fails we observe that  $\sum_{j=1}^{\infty} \tau_j [T(\varphi - \varphi_0)]_j^2 = \int_{\mathcal{W}} |LT(\varphi - \varphi_0)(w) p_W(w) / \nu(w)|^2 \nu(w) dw \geq C \|LT(\varphi - \varphi_0)\|_W^2 > 0$  since  $p_W / \nu$  is uniformly bounded from zero and  $LT$  is nonsingular. Now since  $\varsigma_{m_n} \alpha_n + \mu_{m_n} = o(n)$  it is sufficient to show  $S_n = \sum_{j=1}^{\infty} \tau_j [T(\varphi - \varphi_0)]_j^2 + o_p(1)$ . We make use of the decomposition

$$\begin{aligned} S_n &= \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_i (Y_i - \varphi_0(Z_i)) f_j(W_i) - [T(\varphi - \varphi_0)]_j \right|^2 \\ &+ 2 \sum_{j=1}^{m_n} \tau_j \left( n^{-1} \sum_i (Y_i - \varphi_0(Z_i)) f_j(W_i) - [T(\varphi - \varphi_0)]_j \right) [T(\varphi - \varphi_0)]_j + \sum_{j=1}^{m_n} \tau_j [T(\varphi - \varphi_0)]_j^2 \\ &= I_n + II_n + III_n. \end{aligned}$$

Due to condition  $\mathbb{E} |Y - \varphi_0(Z)|^4 < \infty$  it is easily seen that  $I_n + II_n = o_p(1)$ , which proves the result.  $\square$

**PROOF OF PROPOSITION 2.6.** We make use of the decomposition

$$\begin{aligned} &\mathbb{P} \left( (\sqrt{2} \varsigma_{m_n})^{-1} (n S_n - \mu_{m_n}) > q_{1-\alpha} \right) \\ &\geq \mathbb{P} \left( \left\| n^{-1/2} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 + \left\| n^{-1/2} \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\|^2 - \mu_{m_n} \right. \\ &\quad \left. > \sqrt{2} \varsigma_{m_n} q_{1-\alpha} + 2 \left| \left\langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\rangle \right| \right). \end{aligned}$$

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

$$\left\langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\rangle = O_p \left( \max(\sqrt{n} \|LT(\varphi - \varphi_0)\|_W, \varsigma_{m_n}) \right).$$

Indeed, this is easily seen from

$$\mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j \mathbb{E} [(\varphi(Z) - \varphi_0(Z)) f_j(W)] \sum_i U_i f_j(W_i) \right|^2 \leq \sigma^2 \eta_p n \sum_{j=1}^{m_n} \mathbb{E} [(\varphi(Z) - \varphi_0(Z)) f_j^\tau(W)]^2$$

and further, denoting  $\psi_{ji} = (\varphi(Z_i) - \varphi_0(Z_i)) f_j(W) - \mathbb{E} [(\varphi(Z) - \varphi_0(Z)) f_j(W)]$ ,  $1 \leq j \leq m_n$ ,  $1 \leq i \leq n$ , from

$$\begin{aligned} \mathbb{E} \left| n^{-1} \sum_{i \neq i'} \sum_{j=1}^{m_n} \tau_j \psi_{ji} U_{i'} f_j(W_{i'}) \right|^2 &= \frac{n-1}{n} \sum_{j,j'=1}^{m_n} \tau_j \tau_{j'} \mathbb{E} [\psi_{j1} \psi_{j'1}] \mathbb{E} [U^2 f_j(W) f_{j'}(W)] \\ &\leq C \sum_{j,j'=1}^{m_n} \tau_j \tau_{j'} \mathbb{E} [U^2 f_j(W) f_{j'}(W)] \leq C \sigma^2 \mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j f_j(W) \right|^2 = O \left( \sum_{j=1}^{m_n} \tau_j^2 \right). \end{aligned}$$

Thereby, for all  $0 < \varepsilon' < 1$  there exists some constant  $C > 0$  such that

$$\begin{aligned} &\mathbb{P} \left( (\sqrt{2} \varsigma_{m_n})^{-1} (n S_n - \mu_{m_n}) > q_{1-\alpha} \right) \\ &\geq \mathbb{P} \left( \left\| n^{-1/2} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 + \left\| n^{-1/2} \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\|^2 - \mu_{m_n} \right. \\ &\quad \left. > \sqrt{2} \varsigma_{m_n} q_{1-\alpha} + C \max(\sqrt{n} \|LT(\varphi - \varphi_0)\|_W, \varsigma_{m_n}) \right) - \varepsilon'. \end{aligned}$$

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

$$\begin{aligned} & \left\| n^{-1/2} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 \geq n \sum_{j=1}^{m_n} \tau_j [T(\varphi - \varphi_0)]_j^2 \\ & - 2 \left\langle \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i) - n[LT(\varphi - \varphi_0)]_{\underline{m}_n}, [LT(\varphi - \varphi_0)]_{\underline{m}_n} \right\rangle = I_n + II_n. \end{aligned}$$

Consider  $II_n$ . For  $1 \leq j \leq m_n$  let  $s_j = \tau_j [T(\varphi - \varphi_0)]_j / (\sum_{j=1}^{m_n} \tau_j [T(\varphi - \varphi_0)]_j^2)^{1/2}$  then clearly  $\sum_{j=1}^{m_n} s_j^2 = 1$  and thus  $\mathbb{E} |\sum_{j=1}^{m_n} s_j f_j(W)|^2 \leq \eta_f \eta_p$ . Further, since  $\sup_{z \in \mathcal{Z}} |\varphi(z) - \varphi_0(z)|^2 \leq C$  we calculate

$$\begin{aligned} \mathbb{E} |II_n|^2 &= n \mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j ((\varphi(Z) - \varphi_0(Z)) f_j(W) - [T(\varphi - \varphi_0)]_j) [T(\varphi - \varphi_0)]_j \right|^2 \\ &\leq n \sum_{j=1}^{m_n} \tau_j [T(\varphi - \varphi_0)]_j^2 \mathbb{E} \left| \sum_{j=1}^{m_n} s_j (\varphi(Z) - \varphi_0(Z)) f_j(W) \right|^2 = O(n \|LT(\varphi - \varphi_0)\|_W^2) \end{aligned}$$

and hence  $II_n = O_p(\sqrt{n} \|LT(\varphi - \varphi_0)\|_W)$ . Consider  $I_n$ . Note that  $\|LT(\varphi - \varphi_0)\|_W^2 \leq C$  for all  $\varphi \in \mathcal{G}_n^\rho$  we have  $I_n \geq Cn \|LT(\varphi - \varphi_0)\|_W^2$  for  $n$  sufficiently large. Since on  $\mathcal{G}_n^\rho$  we have  $n \|LT(\varphi - \varphi_0)\|_W^2 \geq \rho \varsigma_{m_n}$  we obtain the result by choosing  $\rho$  sufficiently large.  $\square$

## A.2. Proofs of Section 3.

For ease of notation, we write in the following  $\phi(\cdot)$  for  $\phi(\cdot, \vartheta_0)$  and  $\phi_{\vartheta_l}(\cdot)$  for  $\phi_{\vartheta_l}(\cdot, \vartheta_0)$ .

**PROOF OF THEOREM 3.1.** The proof is based on the decomposition under  $H_p$

$$\begin{aligned} S_n^P &= \left\| n^{-1} \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\|^2 + 2 \left\langle n^{-1} \sum_i U_i f_{\underline{m}_n}^\tau(W_i), n^{-1} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\rangle \\ &\quad + \left\| n^{-1} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 = I_n + 2II_n + III_n. \quad (\text{A.3}) \end{aligned}$$

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  $III_n$ . It holds  $\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n) = \phi_{\vartheta}(Z_i)^t (\vartheta_0 - \hat{\vartheta}_n) + (\vartheta_0 - \hat{\vartheta}_n)^t \phi_{\vartheta\vartheta}(Z_i, \bar{\vartheta}_n)^t (\vartheta_0 - \hat{\vartheta}_n) / 2$  for some  $\bar{\vartheta}_n$  between  $\hat{\vartheta}_n$  and  $\vartheta_0$ . From the bounds imposed in Assumption 6 (ii) we infer

$$nIII_n \leq 2n \|\vartheta_0 - \hat{\vartheta}_n\|^2 \left( \sum_{l=1}^k \sum_{j=1}^{m_n} \tau_j [T\phi_{\vartheta_l}]_j^2 + \sum_{l=1}^k \sum_{j=1}^{m_n} \tau_j \left( \frac{1}{n} \sum_i \phi_{\vartheta_l}(Z_i) f_j(W_i) - [T\phi_{\vartheta_l}]_j \right)^2 \right) + o_p(1).$$

For each  $1 \leq l \leq k$  we have

$$\begin{aligned} \sum_{j=1}^{m_n} [T\phi_{\vartheta_l}]_j^2 &= \sum_{j=1}^{m_n} \left( \int_{\mathcal{W}} (T\phi_{\vartheta_l})(w) f_j(w) p_W(w) dw \right)^2 \leq \int_{\mathcal{W}} |(T\phi_{\vartheta_l})(w) p_W(w) / \nu(w)|^2 \nu(w) dw \\ &\leq \eta_p \|T\phi_{\vartheta_l}\|_W^2 \leq \eta_p \mathbb{E} |\phi_{\vartheta_l}(Z, \vartheta_0)|^2 \leq \eta_p \eta_\phi \quad (\text{A.4}) \end{aligned}$$

by applying Jensen's inequality. Moreover, we calculate

$$\sum_{l=1}^k \sum_{j=1}^{m_n} \mathbb{E} \left| \frac{1}{n} \sum_i \phi_{\vartheta_l}(Z_i) f_j(W_i) - [T\phi_{\vartheta_l}]_j \right|^2 \leq \frac{km_n}{n} \sup_{j, l \geq 1} \mathbb{E} |\phi_{\vartheta_l}(Z) f_j(W)|^2 \leq \eta^4 \frac{km_n}{n}. \quad (\text{A.5})$$

These estimates together with  $\|\vartheta_0 - \widehat{\vartheta}_n\| = O_p(n^{-1/2})$  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  $1 \leq l \leq k$

$$\begin{aligned} & \mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j \left( n^{-1/2} \sum_i U_i f_j(W_i) \left( n^{-1} \sum_i \phi_{\vartheta_l}(Z_i) f_j(W_i) - [T\phi_{\vartheta_l}]_j \right) \right) \right| \\ & \leq n^{-1/2} \sum_{j=1}^{m_n} \tau_j \left( \mathbb{E} |U f_j(W)|^2 \right)^{1/2} \left( \mathbb{E} |\phi_{\vartheta_l}(Z) f_j(W)|^2 \right)^{1/2} = O \left( n^{-1/2} \sum_{j=1}^{m_n} \tau_j \right) = o(\varsigma_{m_n}). \end{aligned}$$

Now since  $n^{1/2}(\vartheta_0 - \widehat{\vartheta}_n) = O_p(1)$  we infer

$$nII_n = n^{1/2}(\vartheta_0 - \widehat{\vartheta}_n)^t \sum_{j=1}^{m_n} \tau_j \left( \varsigma_{m_n}^{-1} n^{-1/2} \sum_i U_i f_j(W_i) \mathbb{E}[\phi_{\vartheta}(Z) f_j(W)] \right) + o_p(1).$$

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

$$\varsigma_{m_n}^{-2} n^{-1} \mathbb{E} \left| \sum_{j=1}^{m_n} \tau_j \sum_i U_i f_j(W_i) [T\phi_{\vartheta_l}]_j \right|^2 \leq \varsigma_{m_n}^{-2} \sigma^2 \eta_p \sum_{j=1}^{m_n} [T\phi_{\vartheta_l}]_j^2 \leq \varsigma_{m_n}^{-2} \sigma^2 \eta_p^2 \eta_f$$

which implies  $nII_n = o_p(\varsigma_{m_n})$  and thus, in light of decomposition (A.3), completes the proof.  $\square$

**PROOF OF THEOREM 3.2.** For  $1 \leq j \leq m_n$  we make use of the following decomposition

$$\begin{aligned} n^{-1/2} \sum_i f_j(W_i) \left( U_i + \phi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n) \right) &= n^{-1/2} \sum_i \left( f_j(W_i) U_i - \sum_{l=1}^k [T\phi_{\vartheta_l}]_j h_l(V_i) \right) \\ & \quad + \sum_{l=1}^k \left( n^{-1} \sum_i f_j(W_i) \phi_{\vartheta_l}(Z_i) - [T\phi_{\vartheta_l}]_j \right) \left( n^{-1/2} \sum_i h_l(V_i) \right) \\ & \quad + \sum_{l=1}^k n^{-1} \sum_i f_j(W_i) \phi_{\vartheta_l}(Z_i) r_l + o_p(1) = A_{nj} + B_{nj} + C_{nj} + o_p(1) \quad (\text{A.6}) \end{aligned}$$

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

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

Clearly, for all  $1 \leq i \leq n$  the random variables  $U_i f_j^t(W_i) + \mathbb{E} [f_j^t(W) \phi_{\vartheta}(Z)^t] h_{\underline{k}}(V_i)$ ,  $1 \leq j \leq m_n$ , are centered with bounded second moment. Due to the proof of Theorem 2.2 it is easily seen that  $\sum_{j=1}^{m_n} \tau_j A_{nj}^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^p \chi_{1j}^2$ . Inequality (A.5) yields  $\sum_{j=1}^{m_n} B_{nj}^2 = o_p(1)$ . Since  $\sum_{j=1}^{m_n} [T\phi_{\vartheta}]_j^2 \leq \eta_p \eta_{\phi}$  we have  $\|\mathbb{E} [f_{m_n}^t(W) \phi_{\vartheta}(Z)^t] r_{\underline{k}}\|^2 \leq k \eta_p \eta_{\phi} \|r_{\underline{k}}\|^2 = o_p(1)$  and hence  $\sum_{j=1}^{m_n} C_{nj}^2 = o_p(1)$ . Finally, the Cauchy-Schwarz inequality implies  $\sum_{j=1}^{m_n} \tau_j A_{nj} (B_{nj} + C_{nj}) = o_p(1)$ , which completes the proof.  $\square$

**PROOF OF PROPOSITION 3.3.** Without loss of generality we may assume  $\delta = \delta_\perp$  (otherwise replace  $\phi(Z_i)$  by  $\phi(Z_i) + \phi_\vartheta(Z_i)^t \mathbb{E}[\delta(Z)\phi_\vartheta(Z)]$ ). Consider the case  $\varsigma_{m_n}^{-1} = o(1)$ . Under the sequence of alternatives (2.6) the following decomposition holds true

$$\begin{aligned} S_n^p &= \left\| n^{-1} \sum_i (U_i + \varsigma_{m_n}^{1/2} n^{-1/2} \delta_\perp(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 \\ &\quad + 2 \left\langle n^{-1} \sum_i (U_i + \varsigma_{m_n}^{1/2} n^{-1/2} \delta_\perp(Z_i)) f_{\underline{m}_n}^\tau(W_i), n^{-1} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\rangle \\ &\quad + \left\| n^{-1} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\|^2. \end{aligned}$$

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

$$\left\langle n^{-1} \sum_i \delta_\perp(Z_i) f_{\underline{m}_n}^\tau(W_i), n^{-1/2} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\rangle = o_p(\sqrt{\varsigma_{m_n}}). \quad (\text{A.7})$$

Indeed, since  $\delta_{j\perp} = \sqrt{\tau_j} \mathbb{E}[\delta_\perp(Z) f_j(W)]$  we have

$$\begin{aligned} \sum_{j=1}^{m_n} \delta_{j\perp} n^{-1/2} \sum_i (\phi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_j(W_i) &= \sqrt{n} (\vartheta_0 - \hat{\vartheta}_n)^t \sum_{j=1}^{m_n} \delta_{j\perp} \mathbb{E}[\phi_\vartheta(Z) f_j(W)] + o_p(1) \\ &\leq \eta_p \eta_\phi \sqrt{n} \|\vartheta_0 - \hat{\vartheta}_n\| \sum_{j=1}^{\infty} \delta_{j\perp}^2 + o_p(1) = O_p(1) \end{aligned}$$

and hence (A.7) holds true.

Consider the case  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . We make use of decomposition (A.6) where  $U_i$  is replaced by  $U_i + n^{-1/2} \delta_\perp(Z_i)$ . Similarly to the proof of Proposition 2.4 it is easily seen that  $\sum_{j=1}^{m_n} \tau_j A_{nj}^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^p \chi_{1j}^2 (\delta_{j\perp} / \lambda_j^p)$ . Thereby, due to the proof of Theorem 3.2, the assertion follows.  $\square$

**PROOF OF PROPOSITION 3.4.** Consider first the case  $\varsigma_{m_n}^{-1} = o(1)$ . Similar to the proof of Theorem 3.1 we observe that  $\|n^{-1} \sum_i (\phi(Z_i, \vartheta_0) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i)\|^2 = o_p(1)$  and  $\|n^{-1} \sum_i (Y_i - \phi(Z_i, \vartheta_0)) f_{\underline{m}_n}^\tau(W_i)\|^2 = \sum_{j=1}^{\infty} \tau_j [T(\varphi - \phi(\cdot, \vartheta_0))]_j^2 + o_p(1)$ . Thus, the result follows as in the proof of Proposition 2.5. In case of  $\sum_{j=1}^{m_n} \tau_j = O(1)$ , we obtain similarly that  $S_n^p = \sum_{j=1}^{m_n} \tau_j |n^{-1} \sum_i ((Y_i - \phi(Z_i, \vartheta_0)) f_j(W_i))|^2 + o_p(1)$  and hence,  $S_n^p = \sum_{j=1}^{\infty} \tau_j [T(\varphi - \phi(\cdot, \vartheta_0))]_j^2 + o_p(1)$ .  $\square$

**PROOF OF PROPOSITION 3.5.** Consider the case  $\varsigma_{m_n}^{-1} = o(1)$ . The basic inequality  $(a - b)^2 \geq a^2/2 - b^2$ ,  $a, b \in \mathbb{R}$ , yields

$$\begin{aligned} &\mathbb{P}\left((\sqrt{2} \varsigma_{m_n})^{-1} (n S_n^p - \mu_{m_n}) > q_{1-\alpha}\right) \\ &\geq \mathbb{P}\left(1/2 \left\| n^{-1/2} \sum_i (\varphi(Z_i) - \phi(Z_i, \vartheta_0)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 + \left\| n^{-1/2} \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\|^2 - \mu_{m_n} \right. \\ &\quad > \sqrt{2} \varsigma_{m_n} q_{1-\alpha} + 2 \left| \left\langle n^{-1} \sum_i (\varphi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\rangle \right. \\ &\quad \left. + \left\| n^{-1/2} \sum_i (\phi(Z_i, \vartheta_0) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 \right). \quad (\text{A.8}) \end{aligned}$$

From the proof of Theorem 3.1 we infer  $\|n^{-1/2} \sum_i (\phi(Z_i, \hat{\vartheta}_n) - \phi(Z_i, \vartheta_0)) f_{\underline{m}_n}^\tau(W_i)\|^2 = o_p(\varsigma_{m_n})$  and

$$\begin{aligned} & \langle n^{-1} \sum_i (\varphi(Z_i) - \phi(Z_i, \hat{\vartheta}_n)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \rangle \\ &= \langle n^{-1} \sum_i (\varphi(Z_i) - \phi(Z_i, \vartheta_0)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \rangle + o_p(\varsigma_{m_n}). \end{aligned}$$

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

### A.3. Proofs of Section 4.

In the following, we denote  $[\hat{Q}]_{k_n} = n^{-1} \sum_i e_{k_n}(Z_i) e_{k_n}(Z_i)^t$ . By Assumption 7, the eigenvalues of  $\mathbb{E}[e_{k_n}(Z) e_{k_n}(Z)^t]$  are bounded away from zero and hence, it may be assumed that  $\mathbb{E}[e_{k_n}(Z) e_{k_n}(Z)^t] = \bar{I}_{k_n}$  (cf. Newey [1997], p. 161).

**PROOF OF THEOREM 4.1.** The proof is based on the decomposition (A.3) where the estimator  $\phi(\cdot, \hat{\vartheta}_n)$  is replaced by  $\bar{\varphi}_{k_n}(\cdot)$  given in (4.1). It holds  $nIII_n = o_p(\varsigma_{m_n})$ , which can be seen as follows. We make use of

$$\begin{aligned} III_n/2 &\leq \left\| \frac{1}{n} \sum_i (E_{k_n} \varphi_0(Z_i) - \bar{\varphi}_{k_n}(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\|^2 + \left\| \frac{1}{n} \sum_i (E_{k_n}^\perp \varphi_0)(Z_i) f_{\underline{m}_n}^\tau(W_i) \right\|^2 \\ &=: A_{n1} + A_{n2}. \end{aligned}$$

Consider  $A_{n1}$ . We observe

$$\begin{aligned} A_{n1} &\leq 2 \left\| \mathbb{E}[f_{\underline{m}_n}^\tau(W) e_{k_n}(Z)^t] [\hat{Q}]_{k_n}^{-1} ([\hat{Q}]_{k_n} [\varphi_0]_{k_n} - n^{-1} \sum_i Y_i e_{k_n}(Z_i)) \right\|^2 \\ &\quad + 2 \left\| E_{k_n} \varphi_0 - \bar{\varphi}_{k_n} \right\|_Z^2 \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \left| n^{-1} \sum_i e_l(Z_i) f_j(W_i) - [T]_{jl} \right|^2 \\ &=: 2B_{n1} + 2B_{n2}. \quad (\text{A.9}) \end{aligned}$$

For  $B_{n1}$  we evaluate due to the relation  $[\hat{Q}]_{k_n}^{-1} = I_{k_n} - [\hat{Q}]_{k_n}^{-1} ([\hat{Q}]_{k_n} - I_{k_n})$  that

$$\begin{aligned} B_{n1} &\leq 2 \left\| \mathbb{E}[f_{\underline{m}_n}^\tau(W) e_{k_n}(Z)^t] n^{-1} \sum_i (E_{k_n} \varphi_0(Z_i) - Y_i) e_{k_n}(Z_i) \right\|^2 \\ &\quad + 2 \left\| \mathbb{E}[f_{\underline{m}_n}^\tau(W) e_{k_n}(Z)^t] \right\|^2 \left\| [\hat{Q}]_{k_n} - I_{k_n} \right\|^2 \left\| [\hat{Q}]_{k_n}^{-1} \right\|^2 \left\| n^{-1} \sum_i (E_{k_n} \varphi_0(Z_i) - Y_i) e_{k_n}(Z_i) \right\|^2. \end{aligned}$$

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

$$\mathbb{E} \left\| [\hat{Q}]_{k_n} - I_{k_n} \right\|^2 \leq n^{-1} \sum_{l, l'=1}^{k_n} \mathbb{E} |e_l(Z) e_{l'}(Z)|^2 \leq \eta_e n^{-1} k_n^2.$$



Further, from  $\mathbb{E}[(E_{k_n}\varphi_0(Z) - Y)e_{k_n}(Z)] = 0$  we deduce

$$\begin{aligned} \mathbb{E} \left\| \mathbb{E}[f_{\underline{m}_n}^\tau(W)e_{k_n}(Z)^t]n^{-1} \sum_i (E_{k_n}\varphi_0(Z_i) - Y_i)e_{k_n}(Z_i) \right\|^2 \\ \leq n^{-1} \sum_{j=1}^{m_n} \mathbb{E} \left| \sum_{l=1}^{k_n} \mathbb{E}[f_j(W)e_l(Z)](E_{k_n}\varphi_0(Z) - Y)e_l(Z) \right|^2 \\ \leq C\eta_p n^{-1} \sum_{j=1}^{m_n} \sum_{l=1}^{k_n} \mathbb{E}[f_j(W)e_l(Z)]^2 = O(n^{-1}k_n) \end{aligned}$$

where we used the definition of  $\mathcal{F}_\gamma$  and that  $\mathbb{E}[U^2|Z]$  is bounded. Moreover, since the difference of eigenvalues of  $[\widehat{Q}]_{k_n}$  and  $I_{k_n}$  is bounded by  $\|[\widehat{Q}]_{k_n} - I_{k_n}\|$ , the smallest eigenvalue of  $[\widehat{Q}]_{k_n}$  converges in probability to one and hence,  $\|[\widehat{Q}]_{k_n}^{-1}\|^2 = 1 + o_p(1)$ . Further, note that  $\|\mathbb{E}[f_{\underline{m}_n}^\tau(W)e_{k_n}(Z)^t]\|^2 \leq \sum_{j=1}^{m_n} \sum_{l=1}^{k_n} \mathbb{E}[f_j(W)e_l(Z)]^2 = O(k_n)$ . Consequently,

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

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  $A_{n2}$ . Similar to the derivation of (A.4) we obtain

$$\mathbb{E} \left\| n^{-1} \sum_i (E_{k_n}^\perp \varphi_0)(Z_i) f_{\underline{m}_n}^\tau(W_i) \right\|^2 \leq 2\eta_p \|E_{k_n}^\perp \varphi_0\|_Z^2 + 2n^{-1} \sum_{j=1}^{m_n} \mathbb{E} |E_{k_n}^\perp \varphi_0(Z) f_j(W)|^2.$$

We have

$$\sum_{j=1}^{m_n} \tau_j \mathbb{E} |(E_{k_n}^\perp \varphi_0)(Z) f_j(W)|^2 = O\left(\gamma_{k_n}^{-1} \sum_{j=1}^{m_n} \tau_j\right) = o(\varsigma_{m_n}) \quad (\text{A.11})$$

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

$$\begin{aligned} nII_n &\leq \left| \sum_{j=1}^{m_n} \tau_j \sum_i U_i f_j(W_i) ([\varphi_0]_{k_n} - [\bar{\varphi}]_{k_n})^t \left( n^{-1} \sum_i e_{k_n}(Z_i) f_j(W_i) - \mathbb{E}[e_{k_n}(Z) f_j(W)] \right) \right| \\ &+ \left| \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} ([\varphi_0]_l - [\bar{\varphi}]_l) \left( \sum_i U_i f_j(W_i) [T]_{jl} \right) \right| \\ &+ \left| \sum_{j=1}^{m_n} \tau_j \left( \sum_i U_i f_j(W_i) \right) \left( n^{-1} \sum_i E_{k_n}^\perp \varphi_0(Z_i) f_j(W_i) - \mathbb{E}[E_{k_n}^\perp \varphi_0(Z) f_j(W)] \right) \right| \\ &+ \left| \sum_{j=1}^{m_n} \tau_j \left( \sum_i U_i f_j(W_i) \right) \mathbb{E}[E_{k_n}^\perp \varphi_0(Z) f_j(W)] \right| = C_{n1} + C_{n2} + C_{n3} + C_{n4}. \quad (\text{A.12}) \end{aligned}$$

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

$$C_{n1} \leq \|E_{k_n}\varphi_0 - \bar{\varphi}_{k_n}\|_Z \sum_{j=1}^{m_n} \tau_j \left| \sum_i U_i f_j(W_i) \right| \left( \sum_{l=1}^{k_n} \left| n^{-1} \sum_i e_l(Z_i) f_j(W_i) - \mathbb{E}[e_l(Z) f_j(W)] \right|^2 \right)^{1/2}.$$

From  $\mathbb{E} |\sum_i U_i f_j(W_i)|^2 \leq n \eta_f \sigma^2$ , 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

$$C_{n2} \leq \|E_{k_n} \varphi_0 - \bar{\varphi}_{k_n}\|_Z \left( \sum_{l=1}^{k_n} \left| \sum_{j=1}^{m_n} \sum_i U_i f_j(W_i) [T]_{jl} \right|^2 \right)^{1/2}.$$

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

$$\mathbb{E} C_{n3} \leq \sum_{j=1}^{m_n} \tau_j (\mathbb{E} |U f_j(W)|^2)^{1/2} (\mathbb{E} |E_{k_n}^\perp \varphi_0(Z) f_j(W)|^2)^{1/2} = O\left(\gamma_{k_n}^{-1/2} \sum_{j=1}^{m_n} \tau_j\right) = o(\varsigma_{m_n}).$$

Consider  $C_{n4}$ . We calculate

$$\mathbb{E} |C_{n4}|^2 \leq n \eta_p \sigma^2 \sum_{j=1}^{m_n} [T E_{k_n}^\perp \varphi_0]_j^2 \leq n \eta_p^2 \sigma^2 \|T E_{k_n}^\perp \varphi_0\|_W^2 = O(n \gamma_{k_n}^{-1}) = o(\varsigma_{m_n}).$$

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

**PROOF OF THEOREM 4.2.** Employing the equality  $[\widehat{Q}]_{k_n}^{-1} = I_{k_n} - [\widehat{Q}]_{k_n}^{-1}([\widehat{Q}]_{k_n} - I_{k_n})$  we obtain for all  $1 \leq j \leq m_n$

$$\begin{aligned} & n^{-1/2} \sum_i f_j(W_i) (U_i + \varphi_0(Z_i) - \bar{\varphi}_{k_n}(Z_i)) \\ &= n^{-1/2} \sum_i \left( f_j(W_i) U_i + \mathbb{E}[f_j(W) e_{k_n}(Z)^t] e_{k_n}(Z_i) (\varphi_0(Z_i) - Y_i) \right) \\ & \quad + n^{-1/2} \sum_i \mathbb{E}[f_j(W) e_{k_n}(Z)^t] [\widehat{Q}]_{k_n}^{-1} ([\widehat{Q}]_{k_n} - I_{k_n}) e_{k_n}(Z_i) (E_{k_n} \varphi_0(Z_i) - Y_i) \\ & \quad + \left( n^{-1} \sum_i f_j(W_i) e_{k_n}(Z_i) - \mathbb{E}[f_j(W) e_{k_n}(Z)^t] \right) \sqrt{n} ([\varphi_0]_{k_n} - [\bar{\varphi}_{k_n}]_{k_n}) \\ & \quad - n^{-1/2} \sum_i \mathbb{E}[f_j(W) e_{k_n}(Z)^t] e_{k_n}(Z_i) E_{k_n}^\perp \varphi_0(Z_i) = A_{nj} + B_{nj} + C_{nj} + D_{nj}. \quad (\text{A.13}) \end{aligned}$$

Due to Assumption 7 (ii) we may assume that  $\{e_1, \dots, e_k\}$  forms an orthonormal system in  $\mathcal{L}_Z^2$  and hence  $\sum_{l=1}^k \mathbb{E}[f_j(W) e_l(Z)]^2$  is bounded uniformly in  $k$ . Thereby, we conclude  $\sum_{j=1}^{m_n} \tau_j \sum_{l > k_n} \mathbb{E}[f_j(W) e_l(Z)] e_l(\cdot) = o(1)$ . Now following line by line the proof of Theorem 2.2 we deduce

$$\sum_{j=1}^{m_n} \tau_j A_{nj}^2 = \sum_{j=1}^m \tau_j \mathbb{E} \left| n^{-1/2} \sum_i U_i \left( f_j(W_i) + \sum_{l \geq 1} \mathbb{E}[f_j(W) e_l(Z)] e_l(Z_i) \right) \right|^2 + o_p(1) \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^e \chi_{1j}^2.$$

Moreover, we see similarly to the proof of Theorem 4.1 that  $\sum_{j=1}^{m_n} \tau_j (B_{nj}^2 + C_{nj}^2 + D_{nj}^2) = o_p(1)$ , which completes the proof.  $\square$

**PROOF OF LEMMA 4.3.** Note that the squared Frobenius norm of  $\widehat{\Sigma}_{m_n} - \Sigma_{m_n}$  is given by

$$\begin{aligned}
& \sum_{j,l=1}^{m_n} \left| n^{-1} \sum_i (Y_i - \bar{\varphi}_{k_n}(Z_i))^2 f_j^\tau(W_i) f_l^\tau(W_i) - s_{jl} \right|^2 \\
& \leq \| \bar{\varphi}_{k_n} - E_{k_n} \varphi \|_Z^4 \sum_{j,l=1}^{m_n} \mathbb{E} \left[ \| e_{k_n}(Z) \|^2 f_j^\tau(W) f_l^\tau(W) \right]^2 \\
& \quad + \sum_{j,l=1}^{m_n} \mathbb{E} \left[ (E_{k_n}^\perp \varphi_0(Z))^2 f_j^\tau(W) f_l^\tau(W) \right]^2 + o_p(1) \\
& \leq \| \bar{\varphi}_{k_n} - E_{k_n} \varphi \|_Z^4 O\left( \left( \sum_{j=1}^{m_n} \tau_j \right)^2 \right) + O\left( (\gamma_{k_n}^{-1} \sum_{j=1}^{m_n} \tau_j)^2 \right) + o_p(1) = o_p(1)
\end{aligned}$$

by using relation (A.10). Consequently, the Frobenius norm of  $\widehat{\Sigma}_{m_n}$  equals  $\varsigma_{m_n} + o_p(1)$ . Consistency of the trace of  $\widehat{\Sigma}_{m_n}$  is seen similarly.  $\square$

**PROOF OF PROPOSITION 4.4.** Similar to the proof of Proposition 3.3 it is sufficient to show

$$\left\langle n^{-1} \sum_i \delta(Z_i) f_{\underline{m}_n}^\tau(W_i), n^{-1/2} \sum_i (\varphi_0(Z_i) - \bar{\varphi}_{k_n}(Z_i)) f_{\underline{m}_n}^\tau(W_i) \right\rangle = o_p(\sqrt{\varsigma_{m_n}}). \quad (\text{A.14})$$

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

$$\begin{aligned}
& \sum_{j=1}^{m_n} \tau_j [T\delta]_j \frac{1}{\sqrt{n}} \sum_i (E_{k_n} \varphi_0(Z_i) - \bar{\varphi}_{k_n}(Z_i)) f_j(W_i) \\
& \leq \sqrt{n} \|T\delta\|_\tau \|T(E_{k_n} \varphi_0 - \bar{\varphi}_{k_n})\|_W + o_p(1) = o_p(\varsigma_{m_n}).
\end{aligned}$$

Similarly to the upper bounds of  $C_{n3}$  and  $C_{n4}$  in the proof of Theorem 4.1 it is straightforward to see that  $\sum_{j=1}^{m_n} \tau_j [T\delta]_j n^{-1/2} \sum_i E_{k_n}^\perp \varphi_0(Z_i) f_j(W_i) = o_p(\varsigma_{m_n})$  and, hence equation (A.14) holds true. Consider the case  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . 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  $\sum_{j=1}^{m_n} \tau_j A_{nj}^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2 (\delta_j / \lambda_j^e)$ . Thereby, due to the proof of Theorem 4.2, the assertion follows.  $\square$

**PROOF OF PROPOSITION 4.5.** Similar to the proof of Proposition 3.4.  $\square$

**PROOF OF PROPOSITION 4.6.** We make use of inequality (A.8) where  $\phi(\cdot, \widehat{\vartheta}_n)$  is replaced by  $\bar{\varphi}_{k_n}$ . From the proof of Theorem 4.1 we infer  $\|n^{-1/2} \sum_i (\bar{\varphi}_{k_n}(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i)\|^2 = o_p(\varsigma_{m_n})$  and

$$\begin{aligned}
& \left\langle n^{-1} \sum_i (\varphi(Z_i) - \bar{\varphi}_{k_n}(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\rangle \\
& = \left\langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \right\rangle + o_p(\varsigma_{m_n})
\end{aligned}$$

uniformly over all  $\varphi \in \mathcal{I}_n^\rho$ . Thus, following line by line the proof of Proposition 2.6, the assertion follows.  $\square$

#### A.4. Proofs of Section 5.

Recall that  $[T]_k = \mathbb{E}[e_k(W)e_k(Z)^t]$ . Further, we denote  $[\widehat{T}]_k = n^{-1} \sum_{i=1}^n e_k(W_i)e_k(Z_i)^t$  and  $[\widehat{g}]_k = n^{-1} \sum_{i=1}^n Y_i e_k(W_i)$ . In the following, we introduce the function  $\varphi_{k_n}(\cdot) := e_{k_n}(\cdot)^t [T]_{k_n}^{-1} [g]_{k_n}$  which belongs to  $\mathcal{L}_Z^2$ . For all  $k \geq 1$  let us denote  $\Omega_k := \{\|[\widehat{T}]_k^{-1}\| \leq \sqrt{n}\}$  and  $\mathcal{U}_k := \{\|R_k\| \| [T]_k^{-1} \| \leq 1/2\}$  where  $R_k = [\widehat{T}]_k - [T]_k$ . 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  $\mathbb{1}_{\Omega_{k_n}} = 1 + o_p(1)$ . For a sequence of weights  $\omega = (\omega_j)_{j \geq 1}$  we define the weighted norm  $\|\phi\|_\omega = (\sum_{j \geq 1} \omega_j (\int_{\mathcal{Z}} \phi(z) e_j(z) \nu(z) dz)^2)^{1/2}$ .

**PROOF OF THEOREM 5.1.** For the proof we make use of decomposition (A.3) where the estimator  $\phi(\cdot, \widehat{\vartheta}_n)$  is replaced by  $\widehat{\varphi}_{k_n}$  given in (5.2). Consider  $III_n$ . Observe

$$\begin{aligned} III_n &\leq 2 \|n^{-1} \sum_i (\varphi_{k_n}(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_{\underline{m}_n}^\tau(W_i)\|^2 \\ &\quad + 2 \|n^{-1} \sum_i (\varphi_{k_n}(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i)\|^2 = 2A_{n1} + 2A_{n2}. \end{aligned} \quad (\text{A.15})$$

Consider  $A_{n1}$ . Making use of the relation  $[\widehat{T}]_{k_n} [T]_{k_n}^{-1} [g]_{k_n} - [\widehat{g}]_{k_n} = n^{-1} \sum_i f_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)$  we obtain

$$\begin{aligned} A_{n1} &\leq 4 \| \mathbb{E}[f_{\underline{m}_n}(W) e_{k_n}(Z)^t] [T]_{k_n}^{-1} \|^2 \|n^{-1} \sum_i f_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 \\ &\quad + 4 \| \mathbb{E}[f_{\underline{m}_n}(W) e_{k_n}(Z)^t] [T]_{k_n}^{-1} R_{k_n} [\widehat{T}]_{k_n}^{-1} n^{-1} \sum_i f_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 \\ &\quad + 2 \| \varphi_{k_n} - \widehat{\varphi}_{k_n} \|_v^2 \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \nu_l^{-1} \|n^{-1} \sum_i e_l(Z_i) f_j(W_i) - [T]_{jl}\|^2 \\ &= 4B_{n1} + 4B_{n2} + 2B_{n3}. \end{aligned}$$

From Lemma A.1 of Breunig and Johannes [2011] we deduce  $\|n^{-1/2} \sum_i e_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 = O_p(k_n)$  and since  $\| \mathbb{E}[f_{\underline{m}_n}(W) e_{k_n}(Z)^t] [T]_{k_n}^{-1} \| = O(1)$  we have  $nB_{n1} = o(\varsigma_{m_n})$ . Further, consider  $B_{n2}$ . By employing  $\|[\widehat{T}]_k^{-1}\| \mathbb{1}_{\mathcal{U}_k} \leq 2 \| [T]_k^{-1} \|$  and  $\|[\widehat{T}]_k^{-1}\|^2 \mathbb{1}_{\Omega_k} \leq n$  for all  $k \geq 1$  it follows

$$\begin{aligned} B_{n2} \mathbb{1}_{\Omega_{k_n}} (\mathbb{1}_{\mathcal{U}_{k_n}} + \mathbb{1}_{\mathcal{U}_{k_n}^c}) &= O \left( 4 \| [T]_{k_n}^{-1} \|^2 \|R_{k_n}\|^2 \|n^{-1} \sum_i f_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 \right. \\ &\quad \left. + n \|R_{k_n}\|^2 \|n^{-1} \sum_i f_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 \mathbb{1}_{\mathcal{U}_{k_n}^c} \right). \end{aligned}$$

Further, since  $n \|R_{k_n}\|^2 = O_p(k_n^2)$  (cf. Lemma A.1 of Breunig and Johannes [2011]) and  $n \|R_{k_n}\|^2 \|n^{-1/2} \sum_i e_{k_n}(W_i) (\varphi_{k_n}(Z_i) - Y_i)\|^2 \mathbb{1}_{\mathcal{U}_{k_n}^c} = o_p(1)$  (cf. proof of Proposition 3.1 of Breunig and Johannes [2011]) it follows  $nB_{n2} \mathbb{1}_{\Omega_{k_n}} = o(\varsigma_{m_n})$ . This together with estimate

(A.5) implies  $nA_{n1} = o_p(\varsigma_{m_n})$ . Consider  $A_{n2}$ . We observe

$$\begin{aligned} \mathbb{E} A_{n2} &\leq 2\|T(\varphi_{k_n} - \varphi_0)\|_W^2 + 2n^{-1} \mathbb{E} \|(\varphi_{k_n}(Z) - \varphi_0(Z)) f_{\underline{m}_n}^\tau(W)\|^2 \\ &\leq 2\eta_p d \|\varphi_{k_n} - \varphi_0\|_v^2 + \frac{2}{n} \sum_{l \geq 1} l^2 \left( \int_{\mathcal{Z}} (\varphi_{k_n} - \varphi_0)(z) e_l(z) \nu(z) dz \right)^2 \sum_{j=1}^{m_n} \tau_j \sum_{l \geq 1} l^{-2} \mathbb{E} |e_l(Z) f_j(W)|^2 \\ &= O\left( \frac{\nu_{k_n}}{\gamma_{k_n}} \|\varphi_{k_n} - \varphi_0\|_\gamma^2 + \|\varphi_{k_n} - \varphi_0\|_\gamma^2 \frac{k_n^2}{n\gamma_{k_n}} \sum_{j=1}^{m_n} \tau_j \right). \quad (\text{A.16}) \end{aligned}$$

where we used Lemma A.2 of Johannes and Schwarz [2010], i.e.,  $\|\varphi_{k_n} - \varphi_0\|_w^2 = O(w_{k_n} \gamma_{k_n}^{-1})$  for a nondecreasing sequence  $w$ . Condition (5.3) together with the estimate  $k_n^2 \leq \sigma^4 \sum_{j=1}^{m_n} \tau_j$  for  $n$  sufficiently large implies  $nA_{n2} = o_p(\varsigma_{m_n})$ . Consequently, due to (A.15) 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  $\bar{\varphi}_{k_n}$  and  $E_{k_n}^\perp \varphi_0$  are replaced by  $\hat{\varphi}_{k_n}$  and  $\varphi_{k_n} - \varphi_0$ , respectively. Consider  $C_{n1}$ . We calculate

$$C_{n1} \leq \|\hat{\varphi}_{k_n} - \varphi_{k_n}\|_v \sum_{j=1}^{m_n} \tau_j \left| \sum_i U_i f_j(W_i) \right| \left( \sum_{l=1}^{k_n} \nu_l^{-1} |n^{-1} \sum_i e_l(Z_i) f_j(W_i) - [T]_{jl}|^2 \right)^{1/2}$$

Since  $\sqrt{n} \|\hat{\varphi}_{k_n} - \varphi_{k_n}\|_v = o_p(\varsigma_{m_n}^{1/2})$  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

$$\begin{aligned} C_{n2} &= \left| \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} [T]_{jl} \int_{\mathcal{Z}} (\hat{\varphi}_{k_n} - \varphi_{k_n})(z) e_l(z) \nu(z) dz \left( \sum_i U_i f_j(W_i) \right) \right| \\ &\leq (n \|\hat{\varphi}_{k_n} - \varphi_{k_n}\|_v^2)^{1/2} \left( \sigma^2 \sum_{l=1}^{k_n} \nu_l^{-1} \sum_{j=1}^{m_n} [T]_{jl}^2 \right)^{1/2} + o_p(1) = o(\varsigma_{m_n}) \end{aligned}$$

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

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

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

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

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.  $\square$

**PROOF OF THEOREM 5.2.** Observe  $[\widehat{T}]_{k_n}[T]_{k_n}^{-1}[g]_{k_n} - [\widehat{g}]_{k_n} = n^{-1} \sum_i e_{k_n}(W_i)(\varphi_{k_n}(Z_i) - Y_i)$  and hence, for all  $1 \leq j \leq m_n$

$$\begin{aligned}
& n^{-1/2} \sum_i f_j(W_i)(U_i + \varphi_0(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) \\
&= n^{-1/2} \sum_i \left( f_j(W_i)U_i + \mathbb{E} [f_j(W)e_{k_n}(Z)^t] [T]_{k_n}^{-1} e_{k_n}(W_i)(\varphi_{k_n}(Z_i) - Y_i) \right) \\
&\quad - n^{-1/2} \sum_i \mathbb{E} [f_j(W)e_{k_n}(Z)^t] [T]_{k_n}^{-1} R_{k_n} [\widehat{T}]_{k_n}^{-1} e_{k_n}(W_i)(\varphi_{k_n}(Z_i) - Y_i) \\
&+ \left( n^{-1} \sum_i f_j(W_i)e_{k_n}(Z_i)^t - \mathbb{E} [f_j(W)e_{k_n}(Z)^t] \right) [\widehat{T}]_{k_n}^{-1} \left( n^{-1/2} \sum_i e_{k_n}(W_i)(\varphi_{k_n}(Z_i) - Y_i) \right) \\
&\quad + n^{-1/2} \sum_i (\varphi_0(Z_i) - \varphi_{k_n}(Z_i)) f_j(W_i) = A_{nj} + B_{nj} + C_{nj} + D_{nj}. \quad (\text{A.17})
\end{aligned}$$

Consider  $A_{nj}$ . For each  $j \geq 1$ , note that  $\|\mathbb{E} [f_j(W)e_{k_n}(Z)^t] [T]_{k_n}^{-1}\|$  is bounded uniformly in  $k$  and further that  $\mathbb{E}[e_{k_n}(W)(\varphi_{k_n}(Z) - \varphi_0(Z))] = 0$ . Now similarly to the proof of Theorem 4.2 we conclude

$$\sum_{j=1}^{m_n} \tau_j A_{nj}^2 = \sum_{j=1}^{m_n} \tau_j \left| n^{-1/2} \sum_i U_i (f_j(W_i) - e_j(W_i)) \right|^2 + o_p(1) \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2.$$

Moreover, as in the proof of Theorem 5.1 it can be seen that  $\sum_{j=1}^{m_n} \tau_j (B_{nj}^2 + C_{nj}^2 + D_{nj}^2) = o_p(1)$ , which proves the result  $\square$

**PROOF OF PROPOSITION 5.3.** Consider the case  $\varsigma_{m_n}^{-1} = o(1)$ . Further, under (5.5) we observe by following the upper bound for  $A_{n1}$  in the proof of Theorem 5.1 that

$$\begin{aligned}
& \sum_{j=1}^{m_n} \left| n^{-1/2} \sum_i (\widehat{\varphi}_{k_n}(Z_i) - \varphi(Z_i)) f_j^\tau(W_i) \right|^2 \\
&= \sum_{j=1}^{m_n} \frac{1}{n} \left| \sum_i (\widehat{\varphi}_{k_n}(Z_i) - \varphi_{k_n}(Z_i)) f_j^\tau(W_i) \right|^2 + \sum_{j=1}^{m_n} \left| \frac{\varsigma_{m_n}}{n} \sum_i \delta(Z_i) f_j^\tau(W_i) \right|^2 + o_p(\varsigma_{m_n}) \\
&= \varsigma_{m_n} \sum_{j=1}^{m_n} \delta_j^2 + o_p(\varsigma_{m_n}).
\end{aligned}$$

Consequently, the result follows as in the proof of Theorem 5.1. For  $\sum_{j=1}^{m_n} \tau_j = O(1)$  we conclude similarly.  $\square$

**PROOF OF PROPOSITION 5.4.** Following the lines of the proof of Theorem 5.1 it can be seen that  $\|n^{-1} \sum_i (\widehat{\varphi}_{k_n}(Z_i) - \varphi_{k_n}(Z_i)) f_{m_n}^\tau(W_i)\|^2 = o_p(1)$  with  $\varphi \in \mathcal{F}_\gamma^\rho$ . On the other hand,  $\|n^{-1} \sum_i (\varphi_{k_n}(Z_i) - \varphi(Z_i)) f_{m_n}^\tau(W_i)\|^2 = C \|T(\varphi_{k_n} - \varphi)\|_W^2 + o_p(1)$ . Further, since  $\|T(\varphi_{k_n} - \varphi)\|_W^2 = \|g - T\varphi\|_W^2 + o(1)$  the result follows as in the proof of Proposition 3.4.  $\square$

**PROOF OF PROPOSITION 5.5.** We make use of inequality (A.8) where  $\phi(\cdot, \widehat{\vartheta}_n)$  is replaced by  $\widehat{\varphi}_{k_n}$ . From the proof of Theorem 5.1 we infer  $\|n^{-1/2} \sum_i (\widehat{\varphi}_{k_n}(Z_i) - \varphi_0(Z_i)) f_{m_n}^\tau(W_i)\|^2 =$

$o_p(\varsigma_{m_n})$  and

$$\begin{aligned} & \langle n^{-1} \sum_i (\varphi(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \rangle \\ &= \langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f_{\underline{m}_n}^\tau(W_i), \sum_i U_i f_{\underline{m}_n}^\tau(W_i) \rangle + o_p(\varsigma_{m_n}) \end{aligned}$$

uniformly over all  $\varphi \in \mathcal{J}_n^p$ . Consequently, following line by line the proof of Proposition 2.6, the assertion follows.  $\square$

### 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

$$Q_{ni} := \begin{cases} \sum_{l=1}^{i-1} X_{li}, & \text{for } i = 2, \dots, n, \\ 0, & \text{for } i = 1 \text{ and } i > n. \end{cases} \quad (\text{A.18})$$

Then clearly

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

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

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

$$\sum_{i=1}^{\infty} \mathbb{E} |Q_{ni}|^2 \leq 1 \quad \text{for all } n \geq 1, \quad (\text{A.19})$$

$$\sum_{i=1}^{\infty} Q_{ni}^2 = 1 + o_p(1), \quad (\text{A.20})$$

$$\sup_{i \geq 1} |Q_{ni}| = o_p(1) \quad (\text{A.21})$$

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

*Proof.* See Awad [1981].  $\square$

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.18). Let Assumptions 1–4 be satisfied and assume  $(\sum_{j=1}^{m_n} \tau_j)^3 = o(n)$ . Then conditions (A.19)–(A.21) hold true.*

*Proof.* Proof of (A.19). Observe that  $\mathbb{E}[X_{1i}X_{1i'}] = 0$  for  $i \neq i'$  and thus, for  $i = 2, \dots, n$  we have

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

by the definition of  $\varsigma_{m_n}$ . Thereby, we conclude

$$\sum_{i=1}^n \mathbb{E}|Q_{ni}|^2 = \frac{2}{n^2} \sum_{i=1}^{n-1} i = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n} \quad (\text{A.22})$$

which proves (A.19).

Proof of (A.20). Using relation (A.22) we observe

$$\mathbb{E}\left|\sum_{i=1}^n Q_{ni}^2 - 1\right|^2 = \sum_{i=1}^n \mathbb{E}Q_{ni}^4 + 2 \sum_{i < i'} \mathbb{E}Q_{ni}^2 Q_{ni'}^2 - 1 + o(1) =: I_n + II_n - 1 + o(1).$$

Consider  $I_n$ . Observe that

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

where we used that  $\mathbb{E}[U f_j(W)] = 0$ . Since  $\sum_{i=1}^n 3(i-1)(i-2) = n(n-1)(n-2)$  we conclude

$$I_n \leq \frac{4}{n^4\varsigma_{m_n}^4} \left(\sum_{j=1}^{m_n} \tau_j\right)^3 \left(\frac{n(n-1)}{2} \sum_{j=1}^{m_n} \tau_j (\mathbb{E}|U f_j(W)|^4)^2 + n(n-1)(n-2) \sum_{j=1}^{m_n} \varsigma_{jj}^2 \mathbb{E}|U f_j(W)|^4\right)$$

Therefore, applying  $\max_{j \geq 1} \mathbb{E}|U f_j(W)|^4 \leq \eta_f \eta_p \sigma^4$  and  $\sum_{j=1}^{m_n} \tau_j = o(n^{1/3})$  yields  $I_n = o(1)$ . Consider  $II_n$ . We calculate for  $i < i'$

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

Consider  $A_{ii'}$ . Exploiting relation (A.22) and using  $\sum_{i < i'} (i-1) = \sum_{i'=1}^n (i'-1)(i'-2)/2 = n(n-1)(n-2)/6$  and further  $\sum_{i < i'} (i-1)(i'-3) = \sum_{i'=1}^n (i'-3)(i'-2)(i'-1)/2 =$



$n(n-1)(n-2)(n-3)/8$  we obtain

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

Moreover, applying the Cauchy-Schwarz inequality twice gives

$$\begin{aligned} \sum_{j, j', l, l'=1}^{m_n} s_{jj' l l'} \mathbb{E} U^4 f_j^\tau(W) f_{j'}^\tau(W) f_l^\tau(W) f_{l'}^\tau(W) &\leq \max_{1 \leq j \leq m_n} \mathbb{E} |U f_j(W)|^4 \left( \sum_{j, j'=1}^{m_n} \sqrt{\tau_j \tau_{j'}} s_{jj'} \right)^2 \\ &\leq \eta_f \eta_p \sigma^4 \zeta_{m_n}^2 \left( \sum_{j=1}^{m_n} \tau_j \right)^2. \end{aligned}$$

Thereby, it holds  $2 \sum_{i < i'} \mathbb{E} A_{ii'} = 1 + o(1)$ . Now consider  $B_{ii'}$ . Since  $\{f_l\}_{l \geq 1}$  forms an orthonormal basis on the support of  $W$  we obtain

$$\begin{aligned} \mathbb{E} \left( \sum_{k=1}^{i-1} X_{ki}^2 \right) \left( \sum_{k \neq k'}^{i'-1} X_{ki'} X_{k'i'} \right) &= 2 \sum_{k=1}^{i-1} \mathbb{E} X_{ki}^2 X_{ki'} X_{i'i} \\ &\leq \frac{8(i-1)}{n^4 \zeta_{m_n}^4} \sum_{j, j'=1}^{m_n} \mathbb{E} \left| U_1^3 f_j^\tau(W_1) f_{j'}^\tau(W_1) U_2^3 f_j^\tau(W_2) f_{j'}^\tau(W_2) \sum_{l, l'=1}^{m_n} \varsigma_{ll'} f_l^\tau(W_1) f_{l'}^\tau(W_2) \right| \\ &\leq \frac{8(i-1) \sigma^2 \eta_p^2}{n^4 \zeta_{m_n}^3} \left( \sum_{j, j'=1}^{m_n} \mathbb{E} |U^2 f_j^\tau(W) f_{j'}^\tau(W)|^2 \right) \leq \frac{8(i-1) \sigma^6 \eta_f \eta_p^3}{n^4 \zeta_{m_n}^3} \left( \sum_{j=1}^{m_n} \tau_j \right)^2. \end{aligned}$$

This, together with relation (A.22), yields  $\sum_{i < i'} \mathbb{E} B_{ii'} = o(1)$ . Further, it is easily seen that  $\sum_{i < i'} \mathbb{E} C_{ii'} = o(1)$ . Consider  $D_{ii'}$ . Using twice the law of iterated expectation gives

$$\begin{aligned} \mathbb{E} D_{ii'} &= \mathbb{E} \left( \sum_{k \neq k'}^{i-1} X_{ki} X_{k'i} \right) \left( \sum_{k \neq k'}^{i'-1} X_{ki'} X_{k'i'} \right) = 4 \sum_{k < k'}^{i-1} \mathbb{E} X_{ki} X_{k'i} X_{ki'} X_{k'i'} \\ &= 4 \sum_{k < k'}^{i-1} \mathbb{E} [X_{ki} X_{k'i} \mathbb{E}[X_{ki'} X_{k'i'} | (Y_k, Z_k, W_k), (Y_{k'}, Z_{k'}, W_{k'}), (Y_i, Z_i, W_i)]] \\ &= \frac{8}{n^2 \zeta_{m_n}^2} \sum_{k < k'}^{i-1} \mathbb{E} \left[ \mathbb{E}[X_{ki} X_{k'i} | (Y_k, Z_k, W_k), (Y_{k'}, Z_{k'}, W_{k'})] \sum_{j, j'=1}^{m_n} s_{jj'} U_k f_j^\tau(W_k) U_{k'} f_{j'}^\tau(W_{k'}) \right] \\ &= \frac{8}{n^4 \zeta_{m_n}^4} \mathbb{E} \left| \sum_{j, j'=1}^{m_n} s_{jj'} U_1 f_j^\tau(W_1) U_2 f_{j'}^\tau(W_2) \right|^2 (i-1)(i-2) \leq \frac{8\sigma^4 \eta_p^2}{n^4 \zeta_{m_n}^2} (i-1)(i-2). \end{aligned}$$

Since  $\zeta_{m_n}^{-1} = o(1)$  we obtain

$$\sum_{i < i'} \mathbb{E} D_{ii'} \leq \frac{8\sigma^4 \eta_p^2}{n^4 \zeta_{m_n}^2} \sum_{i < i'} (i-1)(i-2) = \frac{2\sigma^4 \eta_p^2 n(n-1)(n-2)(n-3)}{3\zeta_{m_n}^2 n^4} = o(1)$$

and hence  $2 \sum_{i < i'} \mathbb{E} Q_{ni}^2 Q_{ni'}^2 = 1 + o(1)$ .

Proof of (A.21). Note that  $\mathbb{P}(\sup_{i \geq 1} |Q_{ni}| > \varepsilon) \leq \sum_{i=1}^n \mathbb{P}(Q_{ni}^2 > \varepsilon^2)$  and, hence the assertion follows from the Markov inequality.  $\square$

## 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. Inverse problems and structural econometrics: The example of instrumental variables. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, editors, *Advances in Economics and Econometrics: Theory and Applications – Eight World Congress*, volume 36 of *Econometric Society Monographs*. Cambridge University Press, 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. A specification test for nonparametric instrumental variable regression. Swiss Finance Institute Research Paper No. 07-13, 2007.
- P. Gagliardini and O. Scaillet. Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167:6175, 2011.
- 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. *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. Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394, 2011.
- J. L. Horowitz. Specification testing in nonparametric instrumental variables estimation. *Journal of Econometrics*, 167:383–396, 2012.
- 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. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147 – 168, 1997.
- 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.
- R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 1981.
- G. Tripathi and Y. Kitamura. Testing conditional moment restrictions. *Annals of Statistics*, 31(6):2059–2095, 2003.