Estimation and Testing of Instrumental Mean and Quantile Regression Models

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von
Christoph Breunig

Juni 2013

Abteilungssprecher: Prof. Dr. Martin Peitz
Referent: Prof. Dr. Enno Mammen
Korreferent: Prof. Dr. Jan Johannes

Tag der Verteidigung: 29. Juli 2013

Acknowledgments

First and foremost, I would like to thank my advisor Enno Mammen for providing me continuous support during my doctoral studies. His amazing intuition and deep understanding of statistics have been very influential to me. I cannot think of any better supervision. I am very grateful for that.

I would also like to thank Jan Johannes. He introduced me to the fascinating field of nonparametric statistics during my studies at the University of Heidelberg. My visits of the Université catholique de Louvain will stay in good remembrance.

During my doctoral studies in Mannheim, Petyo Bonev became a close friend. We had so many deep discussions which certainly shaped my way of thinking. I benefited from his broad knowledge in econometrics. Besides, we had a great time in Mannheim.

I very much enjoyed working with my colleagues at the Chair of Statistics, in particular Andreas Dzemski, Carsten Jentsch, Eun Ryung Lee, Anne Leucht, Florian Sarnetzki, Michael Vogt, and Martin Wahl. I also want to thank Anna Simoni for her helpful comments. Further, I would like to thank my fellow students at the graduate school CDSE. Moreover, I was pleased to be part of the DFG Research Group FOR916.

I thank my parents and my brothers for their continuous support. My friends for the time we had together. Marilena for being the sunshine of my life.

Mannheim, June 2013 Christoph Breunig

Contents

Introduction

Endogeneity is one of the most challenging problems in applied economics. It may occur when important causal factors are omitted, when variables can only be observed within a measurement error, or when a regressor is determined simultaneously with the response variable. Frequently, economists overcome this obstacle by using instrumental variables. Examples of instruments are rainfall variations to estimate the impact of economic growth on the likelihood of civil war in Africa in Miguel et al. [2004], the Vietnam era draft lottery to identify the effects of military service on subsequent civilian mortality in Hearst et al. [1986], or quarter of birth for estimating returns to schooling Angrist and Krueger [1991].

Applications of instrumental variables methods are often widely debated in the literature. One important issue is the validity of the instruments, that is, exogeneity to the error term (cf. Bound et al. [1995] as an example). Another concern regards unreasonable strong functional assumptions on the structural relationship, such as linearity (cf. Horowitz [2011b]). In both situations, the model is potentially misspecified which can lead to seriously erroneous conclusions. Therefore, the aim of this doctoral thesis is twofold. First, we provide testing procedures to check whether instrumental variable models are correctly specified. Second, we develop estimation methods that do not rely on implausible functional restrictions.

To minimize the likelihood of misspecification the nonparametric version of instrumental variable models became increasingly popular in the literature. In this work, we consider nonparametric instrumental mean and quantile regressions models. In these models, given a scalar dependent variable Y and a vector of regressors Z, the structural function φ satisfies

$$Y = \varphi(Z) + U \tag{0.1}$$

where the error term U might be correlated with the vector of regressors Z. But there is another variable W available, called an instrumental variable, that satisfies in the mean regression case

$$\mathbb{E}[U|W] = 0. \tag{0.2}$$

In instrumental quantile regression, we assume that the instrumental variable W satisfies

$$\mathbb{P}(U \leqslant 0|W) = q \tag{0.3}$$

for some quantile $0 < q < 1$. The quantile regression model (0.1) & (0.3) subsumes a nonseparable model $Y = h(Z, V)$ with function h being strictly monotonic in its second argument and unobservable V being independent W. Thereby, the quantile regression model (0.1) & (0.3) allows for heterogeneity in the unobservables.

When the methods of analysis are widened to include nonparametric techniques, however, one must confront two mayor challenges. First, identification requires stronger assumptions about the instrumental variables than for the parametric case. Second, the accuracy of any estimator of the structural function can be low, even for large sample sizes. Consequently, the need for statistically justified model simplifications is paramount.

In this doctoral thesis, we contribute to the literature new estimation and testing procedures in the nonparametric instrumental regression models (0.1) & (0.2) and (0.1) & (0.3). The chapters are self-contained and can be read separately. Each chapter ends with an appendix that contains the proofs.

In *Chapter 1*, which is based on a joint paper with Prof. Dr. Jan Johannes, we consider the problem of estimating a linear functional $\ell(\varphi)$ of the structural function φ in the mean regression model (0.1) & (0.2). We propose a plug-in estimator which is based on a dimension reduction technique and additional thresholding. It is shown that this estimator is consistent and can attain the minimax optimal rate of convergence under additional regularity conditions. This, however, requires an optimal choice of the dimension parameter m depending on certain characteristics of the structural function φ and the joint distribution of the regressor and the instrument, which are unknown in practice. We propose a fully data driven choice of m which combines model selection and Lepski's method. We show that the adaptive estimator attains the optimal rate of convergence up to a logarithmic factor. The theory in this paper is illustrated by considering classical smoothness assumptions and we discuss examples such as pointwise estimation or estimation of averages of the structural function φ. A Monte Carlo investigation illustrates that the difference of the adaptive parameter choice to the optimal one is small in finite samples.

In *Chapter 2*, we propose several tests of restricted specification in the mean regression model (0.1) & (0.2). Based on series estimators, test statistics are established that allow for tests of the general model against a parametric or nonparametric specification as well as a test of exogeneity of the vector of regressors. The tests are asymptotically normally distributed under correct specification and consistent against any alternative model. Under a sequence of local alternative hypotheses, the asymptotic distribution of the tests is derived. In Monte Carlo simulations we give examples where our tests exceeds the power of existing tests in finite samples.

Chapter 3 studies the quantile regression model (0.1) & (0.3). There are many environments in econometrics which require nonseparable modeling of a structural disturbance. Under certain key conditions, these models lead to the conditional quantile restriction (0.1) & (0.3) which is used in the literature to obtain identification and estimation results. These conditions concern validity of the instruments and monotonicity of the model in the nonseparable, continuously distributed disturbance. If one of these assumptions is violated the true structural function may not solve the conditional quantile restriction. Erroneously assuming the misspecified conditional quantile representation might lead to inconsistent estimators. This paper develops a methodology for testing the hypothesis whether the instrumental quantile regression model is correctly specified. Our test statistic is asymptotically normally distributed under correct specification and consistent against any alternative model. A Monte Carlo study examines its finite sample properties. As an empirical illustration we consider a quantile regression model describing the effect of class size on scholastic achievement.

1 Adaptive Estimation of Functionals in Nonparametric Instrumental Regression

1.1 Introduction

We consider estimation of the value of a linear functional of the structural function φ in a nonparametric instrumental regression model. The structural function characterizes the dependency of a response Y on the variation of an explanatory random variable Z by

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

for some error term U. In other words, the structural function equals not the conditional mean function of Y given Z. In this model, however, a sample from (Y, Z, W) is available, where W is a random variable, an instrument, such that

$$\mathbb{E}[U|W] = 0. \tag{1.1b}$$

Given some a-priori knowledge on the unknown structural function φ, captured by a function class $\mathcal{F}$, its estimation has been intensively discussed in the literature. In contrast, in this paper we are interested in estimating the value $\ell(\varphi)$ of a continuous linear functional $\ell : \mathcal{F} \to \mathbb{R}$. Important examples discussed in this paper are weighted average derivatives or point evaluation functionals which are both continuous under appropriate conditions on $\mathcal{F}$. We establish a lower bound of the maximal mean squared error for estimating $\ell(\varphi)$ over a wide range of classes $\mathcal{F}$ and functionals ℓ. As a step towards adaptive estimation, we propose in this paper a plug-in estimator of $\ell(\varphi)$ which is consistent and minimax optimal. This estimator is based on a linear Galerkin approach which involves the choice of a dimension parameter. We present a method for choosing this parameter in a data driven way combining model selection and Lepski's method. Moreover, it is shown that the adaptive estimator can attain the minimax optimal rate of convergence within a logarithmic factor.

Model (1.1a–1.1b) has been introduced first by Florens [2003] and Newey and Powell [2003], while its identification has been studied e.g. in Carrasco et al. [2006], Darolles et al. [2002], and Florens et al. [2011]. It is interesting to note that recent applications and extensions of this approach include nonparametric tests of exogeneity (Blundell and Horowitz [2007]), quantile regression models (Horowitz and Lee [2007]), or semiparametric modeling (Florens [2002]) to name but a few. For example, Ai and Chen [2003], Blundell et al. [2007], Chen and Reiß [2011] or Newey and Powell [2003] consider sieve minimum distance estimators of φ, while Darolles et al. [2002], Hall and Horowitz [2005], Gagliardini and Scaillet [2012a] or Florens et al. [2011] study penalized least squares estimators. A linear Galerkin approach to construct an estimator of φ coming from the inverse problem community (c.f. Efromovich and Koltchinskii [2001] or Hoffmann and Reiß [2004]) has been proposed by Johannes and Schwarz [2010]. But estimating the structural function φ as a whole involves the inversion of the conditional expectation operator of Z given W and generally leads to an ill-posed inverse problem (c.f. Newey and Powell [2003] or Florens [2003]). This essentially implies that all proposed estimators have under reasonable assumptions very poor rates of convergence. In contrast, it might be possible to estimate certain local features of φ, such as the value of a linear functional at the usual parametric rate of convergence.

The nonparametric estimation of linear functionals from Gaussian white noise observations is a subject of considerable literature (c.f. Speckman [1979], Li [1982] or Ibragimov and Has'minskii [1984] in case of direct observations, while in case of indirect observations we refer to Donoho and Low [1992], Donoho [1994] or Goldenshluger and Pereverzev [2000]). However, nonparametric instrumental regression is in general not a Gaussian white noise model. Moreover, this model involves the additional difficulty of dealing with an unkown operator. On the other hand, in the former setting the parametric estimation of linear functionals has been addressed in recent years in the econometrics literature. To be more precise, under restrictive conditions on the linear functional ℓ and the joint distribution of (Z,W) it is shown in Ai and Chen [2007], Santos [2011], and Severini and Tripathi [2010] that it is possible to construct $n^{1/2}$-consistent estimators of $\ell(\varphi)$. In this situation, efficiency bounds are derived by Ai and Chen [2007] and, when φ is not necessarily identified, by Severini and Tripathi [2010]. We show below, however, that $n^{1/2}$-consistency is not possible for a wide range of linear functionals ℓ and joint distributions of (Z,W).

In this paper, we establish a minimax theory for the nonparametric estimation of the value of a linear functional $\ell(\varphi)$ of the structural function φ. For this purpose, we consider a plug-in estimator $\widehat{\ell}_m := \ell(\widehat{\varphi}_m)$ of $\ell(\varphi)$, where the estimator $\widehat{\varphi}_m$ was proposed by Johannes and Schwarz [2010] and the integer m denotes a dimension to be chosen appropriately. The accuracy of $\widehat{\ell}_m$ is measured by its maximal mean squared error uniformly over the classes $\mathcal{F}$ and $\mathcal{P}$, where $\mathcal{P}$ captures conditions on the unknown joint distribution P_{UZW} of the random vector (U, Z, W), i.e., $P_{UZW} \in \mathcal{P}$. The class $\mathcal{F}$ reflects prior information on the structural function φ, e.g., its level of smoothness, and will be constructed flexible enough to characterize, in particular, differentiable or analytic functions. On the other hand, the condition $P_{UZW} \in \mathcal{P}$ specifies amongst others some mapping properties of the conditional expectation operator of Z given W implying a certain decay of its singular values. The construction of $\mathcal{P}$ allows us to discuss both a polynomial and an exponential decay of those singular values. Considering the maximal mean squared error over $\mathcal{F}$ and $\mathcal{P}$ we derive a lower bound for estimating $\ell(\varphi)$. Given an optimal choice m_n^* of the dimension we show that the lower bound is attained by $\widehat{\ell}_{m_n^*}$ up to a constant $C > 0$, i.e.,

$$\sup_{P_{UZW}\in\mathcal{P}} \sup_{\varphi\in\mathcal{F}} \mathbb{E}\,|\widehat{\ell}_{m_n^*} - \ell(\varphi)|^2 \leqslant C \inf_{\breve{\ell}} \sup_{P_{UZW}\in\mathcal{P}} \sup_{\varphi\in\mathcal{F}} \mathbb{E}\,|\breve{\ell} - \ell(\varphi)|^2$$

where the infimum on the right hand side runs over all possible estimators $\breve{\ell}$. Thereby, the estimator $\widehat{\ell}_{m_n^*}$ is minimax optimal even though the optimal choice m_n^* depends on the classes $\mathcal{F}$ and $\mathcal{P}$, which are unknown in practice.

The main issue addressed in this paper is the construction of a data driven selection method for the dimension parameter which adapts to the unknown classes $\mathcal{F}$ and $\mathcal{P}$. When estimating the structural function φ as a whole, adaptive estimators have been proposed by Loubes and Marteau [2009], Johannes and Schwarz [2010], and Horowitz [2011a]. Johannes and Schwarz [2010] consider an adaptive estimator based on a model selection approach (cf. Barron et al. [1999] and its detailed discussion in Massart [2007]) which attains the minimax optimal rate. The estimator of Loubes and Marteau [2009] attains this rate within a logarithmic term. Both papers crucially rely on the a-priori knowledge of the eigenfunctions which yields an orthogonal series estimator involving the estimated singular values of the conditional expectation operator. In econometric applications, however, the eigenfunctions of this operator are unknown. Recently, Horowitz [2011a] proposed an adaptive estimation procedure which is based on minimizing the asymptotic integrated mean-square error and does not involve the knowledge of the eigenfunctions of the operator.

For estimating linear functionals of the structural function φ, adaptive estimation procedures are not yet available. We propose a new method that is different from the above, does not involve a-priori knowledge of the eigenfunctions of the operator, and allows for a polynomial or exponential decay of its singular values. The methodology combines a model selection approach and Lepski's method (cf. Lepskij [1990]). It is inspired by the recent work of Goldenshluger and Lepski [2011]. To be more precise, the adaptive choice $\widehat{m}$ is defined as the minimizer of a random penalized contrast criterion [1], i.e.,

$$\widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}_n} \left\{\widehat{\Psi}_m + \widehat{\mathrm{pen}}_m\right\} \tag{1.2a}$$

with random integer $\widehat{M}_n$ and random penalty sequence $\widehat{\mathrm{pen}} := (\widehat{\mathrm{pen}}_m)_{m\geqslant 1}$, to be defined below, and the sequence of contrast $\widehat{\Psi} := (\widehat{\Psi}_m)_{m\geqslant 1}$ given by

$$\widehat{\Psi}_m := \max_{m\leqslant m'\leqslant \widehat{M}_n} \left\{|\widehat{\ell}_{m'} - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_{m'}\right\}. \tag{1.2b}$$

With this adaptive choice $\widehat{m}$ at hand the estimator $\widehat{\ell}_{\widehat{m}}$ is shown to be minimax optimal within a logarithmic factor over a wide range of classes $\mathcal{F}$ and $\mathcal{P}$. The appearance of the logarithmic factor within the rate is a known fact in the context of local estimation. Brown and Low [1996] show that it is unavoidable in the context of non-parametric Gaussian regression and, hence it is widely considered as an acceptable price for adaptation. This factor is also present in the work of Goldenshluger and Pereverzev [2000] where Lepski's method is applied in the presence of indirect Gaussian observations.

1.2 Complexity of functional estimation: a lower bound.

1.2.1 Notations and basic model assumptions.

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

$$g = T\varphi \tag{1.3}$$

where the function $g := \mathbb{E}[Y|W]$ belongs to L^2_W. In what follows we always assume that there exists a unique solution $\varphi \in L^2_Z$ of equation (1.3), i.e., g belongs to the range of T,

[1] For a sequence $(a_m)_{m\geqslant 1}$ having a minimal value in $A \subset \mathbb{N}$ set $\arg\min_{m\in A}\{a_m\} := \min\{m : a_m \leqslant a_{m'} \forall m' \in A\}$.

and that the null space of T is trivial (c.f. Engl et al. [2000] or Carrasco et al. [2006] in the special case of nonparametric instrumental regression). Estimation of the structural function φ is thus linked with the inversion of the operator T. Moreover, we suppose throughout the paper that T is compact which is under fairly mild assumptions satisfied (c.f. Carrasco et al. [2006]). Note that the proof of minimax optimality of our estimator does not rely on this assumption but it is used for the illustrations and remarks below. If T is compact then a continuous generalized inverse of T does not exist as long as the range of the operator T is an infinite dimensional subspace of L^2_W. This corresponds to the setup of ill-posed inverse problems.

In this section, we show that the obtainable accuracy of any estimator of the value $\ell(\varphi)$ of a linear functional can be essentially determined by regularity conditions imposed on the structural function φ and the conditional expectation operator T. In this paper, these conditions are characterized by different weighted norms in L^2_Z with respect to a pre-specified orthonormal basis $\{e_j\}_{j\geqslant 1}$ in L^2_Z, which we formalize now. Given a positive sequence of weights $w := (w_j)_{j\geqslant 1}$ we define the weighted norm $\|\phi\|^2_w := \sum_{j\geqslant 1} w_j |\langle \phi, e_j\rangle_Z|^2$, $\phi \in L^2_Z$, the completion $\mathcal{F}_w$ of L^2_Z with respect to $\|\cdot\|_w$ and the ellipsoid $\mathcal{F}^r_w := \{\phi \in \mathcal{F}_w : \|\phi\|^2_w \leqslant r\}$ with radius $r > 0$. We shall stress that the basis $\{e_j\}_{j\geqslant 1}$ does not necessarily correspond to the eigenfunctions of T^*T where T^* denotes the adjoint operator of T. In the following we write $a_n \lesssim b_n$ when there exists a generic constant $C > 0$ such that $a_n \leqslant C\, b_n$ for sufficiently large $n \in \mathbb{N}$ and $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously.

Minimal regularity conditions. Given a nondecreasing sequence of weights $\gamma := (\gamma_j)_{j\geqslant 1}$, we suppose, here and subsequently, that the structural function φ belongs to the ellipsoid $\mathcal{F}^\rho_\gamma$ for some $\rho > 0$. The ellipsoid $\mathcal{F}^\rho_\gamma$ captures all the prior information (such as smoothness) about the unknown structural function φ. Observe that the dual space of $\mathcal{F}_\gamma$ can be identified with $\mathcal{F}_{1/\gamma}$ where $1/\gamma := (\gamma_j^{-1})_{j\geqslant 1}$ (cf. Krein and Petunin [1966]). To be more precise, for all $\phi \in \mathcal{F}_\gamma$ the value $\langle h, \phi\rangle_Z$ is well defined for all $h \in \mathcal{F}_{1/\gamma}$ and by Riesz's Theorem there exists a unique $h \in \mathcal{F}_{1/\gamma}$ such that $\ell(\phi) = \langle h, \phi\rangle_Z =: \ell_h(\phi)$. In certain applications one might not only be interested in the performance of an estimation procedure of $\ell_h(\varphi)$ for a given representer h, but also for h varying over the ellipsoid $\mathcal{F}^\tau_\omega$ with radius $\tau > 0$ for a nonnegative sequence $\omega := (\omega_j)_{j\geqslant 1}$ satisfying $\inf_{j\geqslant 1}\{\omega_j\gamma_j\} > 0$. Obviously, $\mathcal{F}_\omega$ is a subset of $\mathcal{F}_{1/\gamma}$.

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

$$\mathcal{T}_d^{\upsilon} := \Big\{ T \in \mathcal{T} : \quad \|\phi\|_{\upsilon}^2/d \leqslant \|T\phi\|_W^2 \leqslant d\,\|\phi\|_{\upsilon}^2, \quad \forall \phi \in L_Z^2 \Big\}. \tag{1.4}$$

Notice first that any operator $T \in \mathcal{T}_d^{\upsilon}$ is injective if the sequence υ is strictly positive. Furthermore, for all $T \in \mathcal{T}_d^{\upsilon}$ it follows that $\upsilon_j/d \leqslant \|Te_j\|_W^2 \leqslant d\upsilon_j$ for all $j \geqslant 1$. If $(s_j)_{j\geqslant 1}$ denotes the ordered sequence of singular values of T then it is easily seen that $\upsilon_j/d \leqslant s_j^2 \leqslant d\upsilon_j$. In other words, the sequence υ specifies the decay of the singular values of T. In what follows, all the results are derived under regularity conditions on the structural function φ and the conditional expectation operator T described through the sequence γ and υ, respectively. We provide illustrations of these conditions below by assuming a “regular decay” of these sequences. The next assumption summarizes our minimal regularity conditions on these sequences.

Assumption 1.1. *Let $\gamma := (\gamma_j)_{j\geqslant 1}$, $\omega := (\omega_j)_{j\geqslant 1}$ and $\upsilon := (\upsilon_j)_{j\geqslant 1}$ be strictly positive sequences of weights with $\gamma_1 = \omega_1 = \upsilon_1 = 1$ such that γ is nondecreasing with $|j|^3\gamma_j^{-1} = o(1)$ as $j \to \infty$, ω satisfies $\inf_{j\geqslant 1}\{\omega_j\gamma_j\} > 0$ and υ is a nonincreasing sequence.*

Remark 1.2.1. *We illustrate Assumption 1.1 for typical choices of γ and υ usually studied in the literature (c.f. Hall and Horowitz [2005], Chen and Reiß [2011] or Johannes et al. [2011]). Let $[h]_j$ be the j-th generalized Fourier coefficient, i.e., $[h]_j := \mathbb{E}[h(Z)e_j(Z)]$, then we consider the cases*

(pp) $\gamma_j \sim |j|^{2p}$ *with* $p > 3/2$, $\upsilon_j \sim |j|^{-2a}$, $a > 0$, *and*

- *(i)* $[h]_j^2 \sim |j|^{-2s}$, $s > 1/2 - p$ *or*
- *(ii)* $\omega_j \sim |j|^{2s}$, $s > -p$.

(pe) $\gamma_j \sim |j|^{2p}$, $p > 3/2$ *and* $\upsilon_j \sim \exp(-|j|^{2a})$, $a > 0$, *and*

- *(i)* $[h]_j^2 \sim |j|^{-2s}$, $s > 1/2 - p$ *or*
- *(ii)* $\omega_j \sim |j|^{2s}$, $s > -p$.

(ep) $\gamma_j \sim \exp(|j|^{2p})$, $p > 0$ *and* $\upsilon_j \sim |j|^{-2a}$, $a > 0$, *and*

- *(i)* $[h]_j^2 \sim |j|^{-2s}$, $s \in \mathbb{R}$ *or*
- *(ii)* $\omega_j \sim |j|^{2s}$, $s \in \mathbb{R}$.

Note that condition $|j|^3\gamma_j^{-1} = o(1)$ as $j \to \infty$ is automatically satisfied for all $p > 0$ in case of (ep). In the other two cases this condition states under classical smoothness assumptions that, roughly speaking, the structural function φ has to be differentiable. Note that Hall *and Horowitz [2005], who only consider the polynomial case, assume* $2p+1 > 2a > p$ *with* $p > 0$ *and* $a > 1/2$ *which is more restrictive than Assumption 1.1 for* $a \geqslant 2$. □

We shall see that the minimax optimal rate is determined by the sequence $\mathcal{R}^h := (\mathcal{R}^h_n)_{n\geqslant 1}$, in case of a fixed representer h, and $\mathcal{R}^\omega := (\mathcal{R}^\omega_n)_{n\geqslant 1}$ in case of a representer varying over the class $\mathcal{F}^\tau_\omega$. These sequences are given for all $x \geqslant 1$ by

$$\mathcal{R}^h_x := \max\left\{\alpha^*_x \sum_{j=1}^{m^*_x} \frac{[h]_j^2}{\upsilon_j}, \sum_{j>m^*_x} \frac{[h]_j^2}{\gamma_j}\right\} \quad \text{and} \quad \mathcal{R}^\omega_x := \alpha^*_x \max_{1\leqslant j\leqslant m^*_x}\left\{\frac{1}{\omega_j \upsilon_j}\right\} \tag{1.5}$$

where $\alpha^*_x := \max\{\upsilon_{m^*_x}\gamma^{-1}_{m^*_x}, x^{-1}\}$. This corresponds to the usual variance and bias decomposition of the mean square error. Here the dimension parameter m^*_x is chosen to trade off both, that is, we let for $x \geqslant 1$

$$m^*_x := \arg\min_{m\in\mathbb{N}} \left\{\Big|\frac{\upsilon_m}{\gamma_m} - x^{-1}\Big|\right\}. \tag{1.6}$$

In case of adaptive estimation the rate of convergence is given by $\mathcal{R}^h_{\text{adapt}} := (\mathcal{R}^h_{n(1+\log n)^{-1}})_{n\geqslant 1}$ and $\mathcal{R}^\omega_{\text{adapt}} := (\mathcal{R}^\omega_{n(1+\log n)^{-1}})_{n\geqslant 1}$, respectively. For ease of notation let $m^\circ_n := m^*_{n(1+\log n)^{-1}}$ and $\alpha^\circ_n := \alpha^*_{n(1+\log n)^{-1}}$. The bounds established below need the following additional assumption, which is satisfied in all cases considered in Remark 1.2.1.

Assumption 1.2. *There exists a constant* $0 < \kappa \leqslant 1$ *such that for all* $n \geqslant 1$

$$\kappa \leqslant \frac{n\,\upsilon_{m^*_n}}{\gamma_{m^*_n}} \leqslant \kappa^{-1}. \tag{1.7}$$

Assumption 1.2 implies that $n\upsilon_{m^*_n}\gamma^{-1}_{m^*_n}$ is uniformly bounded from above and away from zero. Thereby, we can write $n\upsilon_{m^*_n} \sim \gamma_{m^*_n}$.

1.2.2 Lower bounds.

The results derived below involve assumptions on the conditional moments of the random variables U given W, captured by $\mathcal{U}_\sigma$, which contains all conditional distributions of U given W, denoted by $P_{U|W}$, satisfying $\mathbb{E}[U|W] = 0$ and $\mathbb{E}[U^4|W] \leqslant \sigma^4$ for some $\sigma > 0$. The next assertion gives a lower bound for the mean squared error of any estimator when estimating the value $\ell_h(\varphi)$ of a linear functional with given representer h and structural function φ in the function class $\mathcal{F}^\rho_\gamma$.

Theorem 1.2.1. *Assume an iid. n-sample of (Y, Z, W) from the model (1.1a–1.1b). Let γ and υ be sequences satisfying Assumptions 1.1 and 1.2. Suppose that* $\sup_{j\geqslant 1}\mathbb{E}[e_j^4(Z)|W] \leqslant$

η^4, $\eta \geqslant 1$, and $\sigma^4 \geqslant \big(\sqrt{3} + 4\rho\,\eta^2 \sum_{j\geqslant 1} \gamma_j^{-1}\big)^2$. Then for all $n \geqslant 1$ we have

$$\inf_{\breve{\ell}} \sup_{T\in\mathcal{T}_d^{\upsilon}} \sup_{P_{U|W}\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\,|\breve{\ell} - \ell_h(\varphi)|^2 \geqslant \frac{\kappa}{4} \min\left(\frac{1}{2\,d}\,,\,\rho\right) \mathcal{R}_n^h$$

where the first infimum runs over all possible estimators $\breve{\ell}$.

Note that in Theorem 1.2.1 and in the following results the marginal distribution of Z and W is kept fixed while only the dependency structure of the joint distribution of (Z, W) and of (U, Z, W) is allowed to vary.

Remark 1.2.2. *In the proof of the lower bound we consider a test problem based on two hypothetical structural functions. For each test function the condition $\sigma^4 \geqslant \big(\sqrt{3} + 4\rho\,\eta^2 \sum_{j\geqslant 1} \gamma_j^{-1}\big)^2$ ensures a certain complexity of the hypothetical model in a sense that it allows for Gaussian residuals. This specific case is only needed to simplify the calculation of the distance between distributions corresponding to different structural functions. A similar assumption has been used by Chen and Reiß [2011] in order to derive a lower bound for the estimation of the structural function φ itself. In particular, the authors show that in opposite to the present work an one-dimensional subproblem is not sufficient to describe the full difficulty in estimating φ.*

On the other hand, below we derive an upper bound assuming that $P_{U|W}$ belongs to $\mathcal{U}_\sigma$ and that the joint distribution of (Z, W) fulfills in addition Assumption 1.3. Obviously in this situation Theorem 1.2.1 provides a lower bound for any estimator as long as σ is sufficiently large. □

Remark 1.2.3. *The regularity conditions imposed on the structural function φ and the conditional expectation operator T involve only the basis $\{e_j\}_{j\geqslant 1}$ in L_Z^2. Therefore, the lower bound derived in Theorem 1.2.1 does not capture the influence of the basis $\{f_l\}_{l\geqslant 1}$ in L_W^2 used below to construct an estimator of the value $\ell_h(\varphi)$. In other words, this estimator attains the lower bound only if $\{f_l\}_{l\geqslant 1}$ is chosen appropriately.* □

Remark 1.2.4. *The rate $\mathcal{R}^h$ of the lower bound is never faster than the $\sqrt{n}$-rate, that is, $\mathcal{R}_n^h \geqslant n^{-1}$. Moreover, it is easily seen that the lower bound rate is parametric if and only if $\sum_{j\geqslant 1} [h]_j^2 \upsilon_j^{-1} < \infty$. This condition does not involve the sequence γ and hence, attaining a $\sqrt{n}$-rate is independent of the regularity conditions imposed on the structural function. Moreover, due to the link condition $T \in \mathcal{T}_d^{\upsilon}$ we have that Piccard's condition $\sum_{j\geqslant 1} [h]_j^2 \upsilon_j^{-1} < \infty$ is equivalent to h belonging to the range $\mathcal{R}(T^*)$, where T^* denotes the adjoint of T. Note that Severini and Tripathi [2010] showed in their Lemma 4.1 that $h \in \mathcal{R}(T^*)$ is necessary to obtain $\sqrt{n}$-estimability. Under appropriate conditions on φ and* *the joint distribution of (Y, Z, W) we show in the next section that $h \in \mathcal{R}(T^*)$ is also sufficient for $\sqrt{n}$-estimability.* □

The following assertion is due to Breunig and Johannes [2009] who establish a lower bound uniformly over the ellipsoid $\mathcal{F}_\omega^\tau$ of representer. Note that this result is a direct consequence of the lower bound in Theorem 1.2.1 with a fixed representer h. Indeed, if we consider the function $h^* := \tau\omega_{j^*}^{-1/2} e_{j^*}$ with $j^* := \arg\max_{1 \leqslant j \leqslant m_n^*}\{(\omega_j \upsilon_j)^{-1}\}$ then it obviously belongs to $\mathcal{F}_\omega^\tau$. Further, Corollary 1.2.2 follows by calculating the value of the lower bound in Theorem 1.2.1 for the specific representer h^* and, hence we omit its proof.

Corollary 1.2.2. *Let the assumptions of Theorem 1.2.1 be satisfied. Then for all $n \geqslant 1$ we have*

$$\inf_{\breve{\ell}} \sup_{T \in \mathcal{T}_d^\upsilon} \sup_{P_{U|W} \in \mathcal{U}_\sigma} \sup_{\varphi \in \mathcal{F}_\gamma^\rho,\, h \in \mathcal{F}_\omega^\tau} \mathbb{E}\,|\breve{\ell} - \ell_h(\varphi)|^2 \geqslant \frac{\tau\kappa}{4} \min\left(\frac{1}{2\,d}\,,\, \rho\right) \mathcal{R}_n^\omega$$

where the first infimum runs over all possible estimators $\breve{\ell}$.

Remark 1.2.5. *If the lower bound given in Corollary 1.2.2 tends to zero then $(\omega_j\gamma_j)_{j \geqslant 1}$ is a divergent sequence. In other words, without any additional restriction on φ, that is, $\gamma \equiv 1$, consistency of an estimator of $\ell_h(\varphi)$ uniformly over all $\varphi \in \mathcal{F}_\gamma^\rho$ and all $h \in \mathcal{F}_\omega^\tau$ is only possible under restrictions on the representer h in the sense that ω has to be a divergent sequence.* □

1.3 Minimax optimal estimation.

1.3.1 Estimation by dimension reduction and thresholding.

In addition to the basis $\{e_j\}_{j \geqslant 1}$ in L_Z^2 used to establish the lower bound we consider now also a second basis $\{f_l\}_{l \geqslant 1}$ in L_W^2. We comment on the choice of these basis functions in Remark 1.3.2.

Matrix and operator notations. Given $m \geqslant 1$, $\mathcal{E}_m$ and $\mathcal{F}_m$ denote the subspace of L_Z^2 and 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. Given an operator K from L_Z^2 to L_W^2 we denote its inverse by K^{-1} and its adjoint by K^*. If we restrict $F_m K E_m$ to an operator from $\mathcal{E}_m$ to $\mathcal{F}_m$, then it can be represented by a matrix $[K]_{\underline{m}}$ with entries $[K]_{l,j} = \langle K e_j, f_l\rangle_W$

for $1 \leqslant j, l \leqslant m$. Its spectral norm is denoted by $\|[K]_{\underline{m}}\|$, its inverse by $[K]_{\underline{m}}^{-1}$ and its transposed by $[K]_{\underline{m}}^{t}$. We write I for the identity operator and ∇_{υ} for the diagonal operator with singular value decomposition $\{\upsilon_j, e_j, f_j\}_{j\geqslant 1}$. Respectively, given functions $\phi \in L_Z^2$ and $\psi \in L_W^2$ we define by $[\phi]_{\underline{m}}$ and $[\psi]_{\underline{m}}$ m-dimensional vectors with entries $[\phi]_j = \langle \phi, e_j \rangle_Z$ and $[\psi]_l = \langle \psi, f_l \rangle_W$ for $1 \leqslant j, l \leqslant m$.

Consider the conditional expectation operator T associated with (Z, W). If $[e(Z)]_{\underline{m}}$ and $[f(W)]_{\underline{m}}$ denote random vectors with entries $e_j(Z)$ and $f_j(W)$, $1 \leqslant j \leqslant m$, respectively, then it holds $[T]_{\underline{m}} = \mathbb{E}\left\{[f(W)]_{\underline{m}}[e(Z)]_{\underline{m}}^t\right\}$. Throughout the paper $[T]_{\underline{m}}$ is assumed to be nonsingular for all $m \geqslant 1$, so that $[T]_{\underline{m}}^{-1}$ always exists. Note that it is a nontrivial problem to determine when such an assumption holds (cf. Efromovich and Koltchinskii [2001] and references therein).

Definition of the estimator. Let $(Y_1, Z_1, W_1), \dots, (Y_n, Z_n, W_n)$ be an iid. sample of (Y, Z, W). Since $[T]_{\underline{m}} = \mathbb{E}\left\{[f(W)]_{\underline{m}}[e(Z)]_{\underline{m}}^t\right\}$ and $[g]_{\underline{m}} = \mathbb{E}\left\{Y[f(W)]_{\underline{m}}\right\}$ we construct estimators by using their empirical counterparts, that is,

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

Then the estimator of the linear functional $\ell_h(\varphi)$ is defined for all $m \geqslant 1$ by

$$\widehat{\ell}_m := \begin{cases} [h]_{\underline{m}}^t[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}, & \text{if } [\widehat{T}]_{\underline{m}} \text{ is nonsingular and } \|[\widehat{T}]_{\underline{m}}^{-1}\| \leqslant \sqrt{n}, \\ 0, & \text{otherwise.} \end{cases} \tag{1.8}$$

In fact, the estimator $\widehat{\ell}_m$ is obtained from the linear functional $\ell_h(\varphi)$ by replacing the unknown structural function φ by an estimator proposed by Johannes and Schwarz [2010].

Remark 1.3.1. *If Z is continuously distributed one might be also interested in estimating the value $\int_{\mathcal{Z}} \varphi(z)h(z)dz$ where $\mathcal{Z}$ is the support of Z. Assume that this integral and also $\int_{\mathcal{Z}} h(z)e_j(z)dz$ for $1 \leqslant j \leqslant m$ are well defined. Then we can cover the problem of estimating $\int_{\mathcal{Z}} \varphi(z)h(z)dz$ by simply replacing $[h]_{\underline{m}}$ in the definition of $\widehat{\ell}_m$ by a m-dimensional vector with entries $\int_{\mathcal{Z}} h(z)e_j(z)dz$ for $1 \leqslant j \leqslant m$. Hence for $\int_{\mathcal{Z}} \varphi(z)h(z)dz$ the results below follow mutatis mutandis.* $\square$

Note that the orthonormal bases $\{e_j\}_{j\geqslant 1}$ in L_Z^2 and $\{f_l\}_{l\geqslant 1}$ in L_W^2 depend on the marginal distributions of Z and W. As we illustrate in the following remark, these marginals are not needed to be completely known in advance as long as they satisfy additional regularity conditions.

Remark 1.3.2. *Without loss of generality assume that the support of Z and W is confined to $[0,1]$ and denote $L^2_{[0,1]} := \{\phi : \int_0^1 \phi^2(z)dz < \infty\}$. If one assumes in addition that $L^2_{[0,1]} \subset L^2_Z$ and $L^2_W \subset L^2_{[0,1]}$ then it is possible to consider the restriction of T onto $L^2_{[0,1]}$. Note that this condition is satisfied if the density of Z is bounded from above and the density of W is uniformly bounded away from zero. For a detailed discussion we refer to a preliminary version of Darolles et al. [2002] or Section 2.2. of Florens [2002]. Further, let $\{e_j\}_{j\geqslant 1}$ and $\{f_j\}_{j\geqslant 1}$ be orthonormal bases in $L^2_{[0,1]}$. In this case, $(\mathbb{E}[e_l(Z)f_j(W)])_{j,l\geqslant 1}$ is the matrix representation of the restricted operator $(\widetilde{T}\phi)(\cdot) := \int_0^1 \phi(z)p_{ZW}(z,\cdot)dz$ on $L^2_{[0,1]}$ where p_{ZW} denotes the joint density of (Z,W). Moreover, due to Remark 1.3.1 the estimator of $\ell(\varphi)$ in this situation coincides with the estimator $\widehat{\ell}_m$ given in* (1.8) *and hence, the results below follow similarly.* □

Moment assumptions. Besides the link condition (1.4) for the conditional expectation operator T we need moment conditions on the basis, more specific, on the random variables $e_j(Z)$ and $f_l(W)$ for $j,l \geqslant 1$, which we summarize in the next assumption.

Assumption 1.3. *There exists $\eta \geqslant 1$ such that the joint distribution of (Z,W) satisfies*

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

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

Note that condition (ii) is also known as Cramer's condition, which is sufficient to obtain an exponential bound for large deviations of the random variable $e_j(Z)f_l(W) - \mathbb{E}[e_j(Z)f_l(W)]$ (c.f. Bosq [1998]). Moreover, any joint distribution of (Z,W) satisfies Assumption 1.3 for sufficiently large η if the basis $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are uniformly bounded, which holds, for example, for the trigonometric basis considered in Subsection 1.3.4.

1.3.2 Consistency.

The next assertion summarizes sufficient conditions to ensure consistency of the estimator $\widehat{\ell}_m$ introduced in (1.8). Let us introduce the function $\varphi_m \in \mathcal{E}_m$ which is uniquely defined by the vector of coefficients $[\varphi_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ and $[\varphi]_j = 0$ for $j \geqslant m+1$. Obviously, up to the threshold, the estimator $\widehat{\ell}_m$ is the empirical counterpart of $\ell_h(\varphi_m)$. In Proposition 1.3.1 consistency of the estimator $\widehat{\ell}_m$ is only obtained under the condition

$$\|\varphi - \varphi_m\|_\gamma = o(1) \text{ as } m \to \infty \tag{1.9}$$

which does not hold true in general. Obviously (1.9) implies the convergence of $\ell_h(\varphi_m)$ to $\ell_h(\varphi)$ as m tends to infinity for all $h \in \mathcal{F}_{1/\gamma}$.

Proposition 1.3.1. *Assume an iid. n-sample of (Y, Z, W) from the model (1.1a–1.1b) with $P_{U|W} \in \mathcal{U}_\sigma$ and joint distribution of (Z, W) fulfilling Assumption 1.3. Let the dimension parameter m_n satisfy $m_n^{-1} = o(1)$, $m_n = o(n)$,*

$$\|[h]^t_{\underline{m_n}}[T]^{-1}_{\underline{m_n}}\|^2 = o(n), \text{ and } m_n^3 \|[T]^{-1}_{\underline{m_n}}\|^2 = O(n) \text{ as } n \to \infty. \tag{1.10}$$

If (1.9) holds true then $\mathbb{E}\,|\widehat{\ell}_{m_n} - \ell_h(\varphi)|^2 = o(1)$ as $n \to \infty$ for all $\varphi \in \mathcal{F}_\gamma$ and $h \in \mathcal{F}_{1/\gamma}$.

Notice that condition (1.9) also involves the basis $\{f_l\}_{l\geqslant 1}$ in L^2_W. In what follows, we introduce an alternative but stronger condition to guarantee (1.9) which extends the link condition (1.4). We denote by $\mathcal{T}^{\upsilon}_{d,D}$ for some $D \geqslant d$ the subset of $\mathcal{T}^{\upsilon}_d$ given by

$$\mathcal{T}^{\upsilon}_{d,D} := \Big\{ T \in \mathcal{T}^{\upsilon}_d : [T]_{\underline{m}} \text{is nonsingular and } \|[\nabla_\upsilon]^{1/2}_{\underline{m}}[T]^{-1}_{\underline{m}}\|^2 \leqslant D \text{ for all } m \geqslant 1 \Big\}. \tag{1.11}$$

Remark 1.3.3. *If $T \in \mathcal{T}^{\upsilon}_d$ and if in addition its singular value decomposition is given by $\{s_j, e_j, f_j\}_{j\geqslant 1}$ then for all $m \geqslant 1$ the matrix $[T]_{\underline{m}}$ is diagonalized with diagonal entries $[T]_{j,j} = s_j$, $1 \leqslant j \leqslant m$. In this situation it is easily seen that $\sup_{m\in\mathbb{N}} \|[\nabla_\upsilon]^{1/2}_{\underline{m}}[T]^{-1}_{\underline{m}}\|^2 \leqslant d$ and, hence T satisfies the extended link condition (1.11), that is, $T \in \mathcal{T}^{\upsilon}_{d,D}$. Furthermore, it holds $\mathcal{T}^{\upsilon}_d = \mathcal{T}^{\upsilon}_{d,D}$ for suitable $D > 0$, if T is a small perturbation of $\nabla_\upsilon^{1/2}$ or if T is strictly positive (c.f. Efromovich and Koltchinskii [2001] or Cardot and Johannes [2010], respectively).* □

Remark 1.3.4. *Once both basis $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are specified the extended link condition (1.11) restricts the class of joint distributions of (Z, W) such that (1.9) holds true. Moreover, under (1.11) the estimator $\widehat{\varphi}_m$ of φ proposed by Johannes and Schwarz [2010] can attain the minimax optimal rate. In this sense, given a joint distribution of (Z, W) a basis $\{f_l\}_{l\geqslant 1}$ satisfying condition (1.11) can be interpreted as optimal instruments (c.f. Newey [1990]).* □

Remark 1.3.5. *For each pre-specified basis $\{e_j\}_{j\geqslant 1}$ we can theoretically construct a basis $\{f_l\}_{l\geqslant 1}$ such that (1.11) is equivalent to the link condition (1.4). To be more precise, if $T \in \mathcal{T}^{\upsilon}_d$, which involves only the basis $\{e_j\}_{j\geqslant 1}$, then the fundamental inequality of Heinz [1951] implies $\|(T^*T)^{-1/2}e_j\|^2_Z \leqslant d\upsilon_j^{-1}$. Thereby, the function $(T^*T)^{-1/2}e_j$ is an element of L^2_Z and, hence $f_j := T(T^*T)^{-1/2}e_j$, $j \geqslant 1$, belongs to L^2_W. Then it is easily checked that $\{f_l\}_{l\geqslant 1}$ is a basis of the closure of the range of T which may be completed to a basis of L^2_W. Obviously $[T]_{\underline{m}}$ is symmetric and moreover, strictly positive since $\langle Te_j, f_l\rangle_W =$*

$\langle (T^*T)^{1/2} e_j, e_l\rangle_Z$ for all $j, l \geqslant 1$. Thereby, we can apply Lemma A.3 in Cardot and Johannes [2010] which gives $\mathcal{T}_d^\upsilon = \mathcal{T}_{d,D}^\upsilon$ for sufficiently large D. We are currently exploring the data driven choice of the basis $\{f_l\}_{l\geqslant 1}$. □

Under the extended link condition (1.11) the next assertion summarizes sufficient conditions to ensure consistency.

Corollary 1.3.2. *The conclusion of Proposition 1.3.1 still holds true without imposing condition* (1.9)*, if the sequence υ satisfies Assumption 1.1, the conditional expectation operator T belongs to $\mathcal{T}_{d,D}^\upsilon$, and* (1.10) *is substituted by*

$$\sum_{j=1}^{m_n} [h]_j^2 \upsilon_j^{-1} = o(n) \quad \textit{and} \quad m_n^3 = O(n\upsilon_{m_n}) \quad \textit{as } n \to \infty. \tag{1.12}$$

1.3.3 An upper bound.

The last assertions show that the estimator $\widehat{\ell}_m$ defined in (1.8) is consistent for all structural functions and representers belonging to $\mathcal{F}_\gamma$ and $\mathcal{F}_{1/\gamma}$, respectively. The following theorem provides now an upper bound if φ belongs to an ellipsoid $\mathcal{F}_\gamma^\rho$. In this situation the rate $\mathcal{R}^h$ of the lower bound given in Theorem 1.2.1 provides up to a constant also an upper bound of the estimator $\widehat{\ell}_{m_n^*}$. Thus we have proved that the rate $\mathcal{R}^h$ is optimal and, hence $\widehat{\ell}_{m_n^*}$ is minimax optimal.

Theorem 1.3.3. *Assume an iid. n-sample of (Y, Z, W) from the model (1.1a–1.1b) with joint distribution of (Z, W) fulfilling Assumption 1.3. Let Assumptions 1.1 and 1.2 be satisfied. Suppose that the dimension parameter m_n^* given by* (1.6) *satisfies*

$$(m_n^*)^3 \max\Big\{ |\log \mathcal{R}_n^h|, (\log m_n^*)\Big\} = o(\gamma_{m_n^*}), \quad \textit{as } n \to \infty, \tag{1.13}$$

then we have for all $n \geqslant 1$

$$\sup_{T\in\mathcal{T}_{d,D}^\upsilon} \sup_{P_{U|W}\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\, |\widehat{\ell}_{m_n^*} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\, \mathcal{R}_n^h$$

for a constant $\mathcal{C} > 0$ only depending on the classes $\mathcal{F}_\gamma^\rho$, $\mathcal{T}_{d,D}^\upsilon$, the constants σ, η and the representer h.

The next result gives sufficient conditions for $\sqrt{n}$-estimability of $\ell_h(\varphi)$. The next corollary is a direct consequence of Theorem 1.3.3 and Remark 1.2.4, hence its proof is omitted.

Corollary 1.3.4. *Let the assumptions of Theorem 1.3.3 be satisfied. If in addition $h \in \mathcal{R}(T^*)$ then we have for all $n \geqslant 1$*

$$\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W}\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\,|\widehat{\ell}_{m_n^*} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\, n^{-1}$$

where $\mathcal{C}$ is as in Theorem 1.3.3.

Remark 1.3.6. *The last result together with Remark 1.2.4 established equivalence between condition $h \in \mathcal{R}(T^*)$ and $\sqrt{n}$-estimability of $\ell_h(\varphi)$ under appropriate conditions on φ and the joint distribution of (Y, Z, W) (as conjectured in Chapter 4, Remark (ii) of Severini and Tripathi [2010]). As illustrated in the next subsection, depending on the severness of the ill-posedness $\sqrt{n}$-estimability could not be possible even for smooth functionals. In the polynomial case (pp), condition $h \in \mathcal{R}(T^*)$ holds true only if $s > a + 1/2$. In case of (ep), $h \in \mathcal{R}(T^*)$ only if the representer h is analytic. In contrast to our framework, the estimation procedure of Santos [2011] crucially relies on condition $h \in \mathcal{R}(T^*)$ which implies the existence of a function $\vartheta \in L^2_W$ such that $\ell_h(\varphi) = \mathbb{E}[Y\vartheta(W)]$.* □

Note that Breunig and Johannes [2009] considered a similar estimator as $\widehat{\ell}_m$ given in (1.8) which requires the choice an additional regularization parameter. Breunig and Johannes [2009] show that their estimator attains the lower bound rate within a constant where the loss is measured uniformly over the class $\mathcal{F}_\omega^\tau$ of representer. In the following, we see that a similar result can be shown for the estimator $\widehat{\ell}_m$. Observe that $\|h\|^2_{1/\gamma} \leqslant \tau$ and $\mathcal{R}_n^h \leqslant \tau\, \alpha_n^* \max_{1\leqslant j\leqslant m_n^*}\{(\omega_j \upsilon_j)^{-1}\} = \tau\, \mathcal{R}_n^\omega$ for all $h \in \mathcal{F}_\omega^\tau$. Employing these estimates, the proof of the next result is similar to the proof of Theorem 1.3.3 and is thus omitted.

Corollary 1.3.5. *Let the assumptions of Theorem 1.3.3 be satisfied where we substitute condition (1.13) by $(m_n^*)^3 \max\{|\log \mathcal{R}_n^\omega|, (\log m_n^*)\} = o(\gamma_{m_n^*})$ as $n \to \infty$. Then we have*

$$\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W}\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{F}_\gamma^\rho,\, h\in\mathcal{F}_\omega^\tau} \mathbb{E}\,|\widehat{\ell}_{m_n^*} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\, \mathcal{R}_n^\omega$$

for a constant $\mathcal{C} > 0$ only depending on the classes $\mathcal{F}_\gamma^\rho$, $\mathcal{F}_\omega^\tau$, $\mathcal{T}_{d,D}^{\upsilon}$ and the constants σ, η.

1.3.4 Illustration by classical smoothness assumptions.

Let us illustrate our general results by considering classical smoothness assumptions. To simplify the presentation we follow Hall and Horowitz [2005], and suppose that the marginal distribution of the scalar regressor Z and the scalar instrument W are uniformly distributed on the interval $[0,1]$. All the results below can be easily extended to the multivariate case. In the univariate case, however, both Hilbert spaces L^2_Z and L^2_W equal $L^2[0,1]$. Moreover, as a basis $\{e_j\}_{j\geqslant 1}$ in $L^2[0,1]$ we choose the trigonometric basis given by

$$e_1 :\equiv 1,\ e_{2j}(t) := \sqrt{2}\cos(2\pi jt),\ e_{2j+1}(t) := \sqrt{2}\sin(2\pi jt), t\in[0,1],\ j\in\mathbb{N}.$$

In this subsection also the second basis $\{f_l\}_{l\geqslant 1}$ is given by the trigonometric basis. In this situation, the moment conditions formalized in Assumption 1.3 are automatically fulfilled.

Recall the typical choices of the sequences γ, ω, and υ introduced in Remark 1.2.1. If $\gamma_j \sim |j|^{2p}$, $p>0$, as in case (pp) and (pe), then $\mathcal{F}_\gamma$ coincides with the Sobolev space of p-times differential periodic functions (c.f. Neubauer [1988]). In case of (ep) it is well known that $\mathcal{F}_\gamma$ contains only analytic functions if $p>1$(c.f. Kawata [1972]). Furthermore, we consider two special cases describing a "regular decay" of the unknown singular values of T. In case of (pp) and (ep) we consider a polynomial decay of the sequence υ. Easy calculus shows that any operator T satisfying the link condition (1.4) acts like integrating (a)-times and hence is called *finitely smoothing* (c.f. Natterer [1984]). In case of (pe) we consider an exponential decay of υ and it can easily be seen that $T\in\mathcal{T}_d^\upsilon$ implies $\mathcal{R}(T)\subset C^\infty[0,1]$, therefore the operator T is called *infinitely smoothing* (c.f. Mair [1994]). In the next assertion we present the order of sequences $\mathcal{R}^h$ and $\mathcal{R}^\omega$ which were shown to be minimax-optimal. Note that the minimax optimal rate $\mathcal{R}^\omega$ in the cases (pp) and (pe) were already derived in Breunig and Johannes [2009] but are stated and proved here for the sake of completeness.

Proposition 1.3.6. *Assume an iid. n-sample of (Y,Z,W) from the model (1.1a–1.1b) with $T\in\mathcal{T}_{d,D}^\upsilon$ and $P_{U|W}\in\mathcal{U}_\sigma$. Then for the example configurations of Remark 1.2.1 we obtain*

(pp) $m_n^* \sim n^{1/(2p+2a)}$ *and*

(i) $$\mathcal{R}_n^h \sim \begin{cases} n^{-(2p+2s-1)/(2p+2a)}, & \text{if } s-a<1/2,\\ n^{-1}\log n, & \text{if } s-a=1/2,\\ n^{-1}, & \text{otherwise,}\end{cases}$$

(ii) $\mathcal{R}_n^\omega \sim \max(n^{-(p+s)/(p+a)}, n^{-1})$.

(pe) $m_n^* \sim \log(n(\log n)^{-p/a})^{1/(2a)}$ *and*

(i) $\mathcal{R}_n^h \sim (\log n)^{-(2p+2s-1)/(2a)}$,

(ii) $\mathcal{R}_n^\omega \sim (\log n)^{-(p+s)/a}$.

(ep) $m_n^* \sim \log(n(\log n)^{-a/p})^{1/(2p)}$ *and*

$$\text{(i)}\quad \mathcal{R}_n^h \sim \begin{cases} n^{-1}(\log n)^{(2a-2s+1)/(2p)}, & \text{if } s-a<1/2,\\ n^{-1}\log(\log n), & \text{if } s-a=1/2,\\ n^{-1}, & \text{otherwise,}\end{cases}$$

$$\text{(ii)}\quad \mathcal{R}_n^\omega \sim \max(n^{-1}(\log n)^{(a-s)/p}, n^{-1}).$$

Remark 1.3.7. *As we see from Proposition 1.3.6, if the value of a increases the obtainable optimal rate of convergence decreases. Therefore, the parameter a is often called degree of ill-posedness (c.f. Natterer [1984]). On the other hand, an increasing of the value p or s leads to a faster optimal rate. Moreover, in the cases (pp) and (ep) the parametric rate n^{-1} is obtained independent of the smoothness assumption imposed on the structural function φ (however, $p \geqslant 3/2$ is needed) if the representer is smoother than the degree of ill-posedness of T, i.e., (i) $s \geqslant a - 1/2$ and (ii) $s \geqslant a$. Moreover, it is easily seen that if $[h]_j \sim \exp(-|j|^s)$ or $\omega_j \sim \exp(|j|^{2s})$, $s > 0$, then the minimax convergence rates are always parametric for any polynomial sequences γ and υ.* □

Remark 1.3.8. *It is of interest to compare our results with those of Hall and Horowitz [2005] or Chen and Reiss [2008] who consider the estimation of the structural function as a whole. In the notations of Hall and Horowitz [2005], who consider only the case (pp), the decay of the eigenvalues of T^*T is assumed to be of order $j^{-\alpha}$, that is, $\alpha = 2a$ with $\alpha > 1$. Furthermore, they suppose a decay of the coefficients of the structural function of order $j^{-\beta}$, that is, $\beta = p + 1/2$ with $\beta > 1/2$. By using this parametrization, Hall and Horowitz [2005] obtain in the case (pp) the minimax rate of convergence $n^{-2p/(2a+2p+1)}$ (see also Chen and Reiß [2011]). Let us compare this rate when estimating φ at a point $t_0 \in [0,1]$ (cf. Example 1.3.1). Here, we have $s = 0$ and hence, obtain the minimax rate of convergence $n^{-(2p-1)/(2a+2p)}$. Roughly speaking, one looses $1/2$ of smoothness, which corresponds to the loss of smoothness of Sobolev embeddings in Hölder spaces. For any representer h with $2s > (2a+1)/(2a+2p+1)$, however, the rate of convergence for estimating $\ell_h(\varphi)$ in the case (pp) is faster than estimating φ as a whole.* □

Example 1.3.1. *Suppose we are interested in estimating the value $\varphi(t_0)$ of the structural function φ evaluated at a point $t_0 \in [0,1]$. Consider the representer given by $h_{t_0} = \sum_{j=1}^\infty e_j(t_0)e_j$. Let $\varphi \in \mathcal{F}_\gamma$. Since $\sum_{j\geqslant 1}\gamma_j^{-1} < \infty$ (cf. Assumption 1.1) it holds $h \in \mathcal{F}_{1/\gamma}$ and hence the point evaluation functional in $t_0 \in [0,1]$, i.e., $\ell_{h_{t_0}}(\varphi) = \varphi(t_0)$, is well defined.*

In this case, the estimator $\widehat{\ell}_m$ *introduced in* (1.8) *writes for all* $m \geqslant 1$ *as*

$$\widehat{\varphi}_m(t_0) := \begin{cases} [e(t_0)]^t_{\underline{m}}[\widehat{T}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}, & \textit{if } [\widehat{T}]_{\underline{m}} \textit{ is nonsingular and } \|[\widehat{T}]^{-1}_{\underline{m}}\| \leqslant \sqrt{n}, \\ 0, & \textit{otherwise} \end{cases}$$

where $\widehat{\varphi}_m$ *is an estimator proposed by Johannes and Schwarz [2010]. Let* $p \geqslant 3/2$ *and* $a > 0$*. Then the estimator* $\widehat{\varphi}_{m_n^*}(t_0)$ *attains within a constant the minimax optimal rate of convergence* $\mathcal{R}^{h_{t_0}}$*. Applying Proposition 1.3.6 gives*

(pp) $\mathcal{R}_n^{h_{t_0}} \sim n^{-(2p-1)/(2p+2a)}$,

(ep) $\mathcal{R}_n^{h_{t_0}} \sim (\log n)^{-(2p-1)/(2a)}$,

(ep) $\mathcal{R}_n^{h_{t_0}} \sim n^{-1}(\log n)^{(2a+1)/(2p)}$. □

Example 1.3.2. *We want to estimate the average value of the structural function* φ *over a certain interval* $[0, b]$ *with* $0 < b < 1$*. The linear functional of interest is given by* $\ell_h(\varphi) = \int_0^b \varphi(t)dt$ *with representer* $h := \mathbb{1}_{[0,b]}$*. Its Fourier coefficients are given by* $[h]_1 = b$*,* $[h]_{2j} = (\sqrt{2}\pi j)^{-1}\sin(2\pi jb)$*,* $[h]_{2j+1} = -(\sqrt{2}\pi j)^{-1}\cos(2\pi jb)$ *for* $j \geqslant 1$ *and, hence* $[h]_j^2 \sim j^{-2}$*. Again we assume that* $p \geqslant 3/2$ *and* $a > 0$*. Then the mean squared error of the estimator* $\widehat{\ell}_{m_n^*} = \int_0^b \widehat{\varphi}_{m_n^*}(t)dt$ *is bounded up to a constant by the minimax rate of convergence* $\mathcal{R}^h$*. In the three cases the order of* $\mathcal{R}_n^h$ *is given by*

(pp) $\mathcal{R}_n^h \sim \begin{cases} n^{-(2p+1)/(2p+2a)}, & \textit{if } a > 1/2, \\ n^{-1}\log n, & \textit{if } a = 1/2, \\ n^{-1}, & \textit{otherwise,} \end{cases}$

(ep) $\mathcal{R}_n^h \sim (\log n)^{-(2p+1)/(2a)}$,

(ep) $\mathcal{R}_n^h \sim \begin{cases} n^{-1}(\log n)^{(2a-1)/(2p)}, & \textit{if } a > 1/2, \\ n^{-1}\log(\log n), & \textit{if } a = 1/2, \\ n^{-1}, & \textit{otherwise.} \end{cases}$

As in the direct regression model where the average value of the regression function can be estimated with rate n^{-1} *we obtain the parametric rate in the case of (pp) and (ep) if* $a < 1/2$*.* □

Example 1.3.3. *Consider estimation of the weighted average derivative of the structural function* φ *with weight function* H*, i.e.,* $\int_0^1 \varphi'(t)H(t)dt$*. This functional is useful not only for estimating scaled coefficients of an index model, but also to quantify the average slope of*

structural functions. Assume that the weight function H is continuously differentiable and vanishes at the boundary of the support of Z, i.e., $H(0) = H(1) = 0$. Integration by parts gives $\int_0^1 \varphi'(t)H(t)dt = -\int_0^1 \varphi(t)h(t)dt = -\ell_h(\varphi)$ with representer h given by the derivative of H. The weighted average derivative estimator $\widehat{\ell}_{m_n^*} = -\int_0^1 \widehat{\varphi}_{m_n^*}(t)h(t)dt$ is minimax optimal. As an illustration consider the specific weight function $H(t) = 1-(2t-1)^2$ with derivative $h(t) = 4(1-2t)$ for $0 \leqslant t \leqslant 1$. It is easily seen that the Fourier coefficients of the representer h are $[h]_1 = 0$, $[h]_{2j} = 0$, $[h]_{2j+1} = 4\sqrt{2}(\pi j)^{-1}$ for $j \geqslant 1$ and, thus $[h]_{2j+1}^2 \sim j^{-2}$. Thus, for the particular choice of the weight function H the estimator $\widehat{\ell}_{m_n^*}$ attains up to a constant the optimal rate $\mathcal{R}^h$, which was already specified in Example 1.3.2.
$\square$

1.4 Adaptive estimation

In this section, we derive an adaptive estimation procedure for the value of the linear function $\ell_h(\varphi)$. This procedure is based on the estimator $\widehat{\ell}_{\widehat{m}}$ given in (1.8) with dimension parameter $\widehat{m}$ selected as a minimizer of the data driven penalized contrast criterion (1.2a–1.2b). The selection criterion (1.2a–1.2b) involves the random upper bound $\widehat{M}_n$ and the random penalty sequence $\widehat{\mathrm{pen}}$ which we introduce below. We show that the estimator $\widehat{\ell}_{\widehat{m}}$ attains the minimax rate of convergence within a logarithmic term. Moreover, we illustrate the cost due to adaption by considering classical smoothness assumptions.

In an intermediate step we do not consider the estimation of unknown quantities in the penalty function. Let us therefore consider a deterministic upper bound M_n and a deterministic penalty sequence $\mathrm{pen} := (\mathrm{pen}_m)_{m\geqslant 1}$, which is nondecreasing. These quantities are constructed such that they can be easily estimated in a second step. As an adaptive choice $\widetilde{m}$ of the dimension parameter m we propose the minimizer of a penalized contrast criterion, that is,

$$\widetilde{m} := \underset{1\leqslant m\leqslant M_n}{\arg\min} \{\Psi_m + \mathrm{pen}_m\} \tag{1.14a}$$

where the random sequence of contrast $\Psi := (\Psi_m)_{m\geqslant 1}$ is defined by

$$\Psi_m := \max_{m\leqslant m'\leqslant M_n} \Big\{ |\widehat{\ell}_{m'} - \widehat{\ell}_m|^2 - \mathrm{pen}_{m'} \Big\}. \tag{1.14b}$$

The fundamental idea to establish an appropriate upper bound for the risk of $\widehat{\ell}_{\widetilde{m}}$ is given by the following reduction scheme. Let us denote $m \wedge m' := \min(m, m')$. Due to the definition of Ψ and $\widetilde{m}$ we deduce for all $1 \leqslant m \leqslant M_n$

$$\begin{aligned}|\widehat{\ell}_{\widetilde{m}} - \ell_h(\varphi)|^2 &\leqslant 3\Big\{|\widehat{\ell}_{\widetilde{m}} - \widehat{\ell}_{\widetilde{m}\wedge m}|^2 + |\widehat{\ell}_{\widetilde{m}\wedge m} - \widehat{\ell}_m|^2 + |\widehat{\ell}_m - \ell_h(\varphi)|^2\Big\}\\ &\leqslant 3\Big\{\Psi_m + \mathrm{pen}_{\widetilde{m}} + \Psi_{\widetilde{m}} + \mathrm{pen}_m + |\widehat{\ell}_m - \ell_h(\varphi)|^2\Big\}\\ &\leqslant 6\{\Psi_m + \mathrm{pen}_m\} + 3|\widehat{\ell}_m - \ell_h(\varphi)|^2\end{aligned}$$

where the right hand side does not depend on the adaptive choice $\widetilde{m}$. Since the penalty sequence pen is nondecreasing we obtain

$$\Psi_m \leqslant 6 \max_{m \leqslant m' \leqslant M}\Big(|\widehat{\ell}_{m'} - \ell_h(\varphi_{m'})|^2 - \frac{1}{6}\,\mathrm{pen}_{m'}\Big)_+ + 3 \max_{m \leqslant m' \leqslant M_n} |\ell_h(\varphi_m - \varphi_{m'})|^2$$

where $(\cdot)_+$ denotes the positive part of a function. Combining the last estimate with the previous reduction scheme yields for all $1 \leqslant m \leqslant M_n$

$$|\widehat{\ell}_{\widetilde{m}} - \ell_h(\varphi)|^2 \leqslant 7\,\mathrm{pen}_m + 78\,\mathrm{bias}_m + 42 \max_{m \leqslant m' \leqslant M}\Big(|\widehat{\ell}_{m'} - \ell_h(\varphi_{m'})|^2 - \frac{1}{6}\,\mathrm{pen}_{m'}\Big)_+ \tag{1.15}$$

where $\mathrm{bias}_m := \sup_{m' \geqslant m}|\ell_h(\varphi_{m'} - \varphi)|^2$. We will prove below that $\mathrm{pen}_m + \mathrm{bias}_m$ is of the order $\mathcal{R}^h_{n(1+\log n)^{-1}}$. Moreover, we will bound the right hand side term appropriately with the help of Bernstein's inequality.

Let us now introduce the upper bound M_n and sequence of penalty pen_m used in the penalized contrast criterion (1.14a–1.14b). In the following, assume without loss of generality that $[h]_1 \neq 0$.

Definition 1.4.1. *For all $n \geqslant 1$ let $a_n := n^{1-1/\log(2+\log n)}(1+\log n)^{-1}$ and $M_n^h := \max\{1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor : \max_{1 \leqslant j \leqslant m} [h]_j^2 \leqslant n[h]_1^2\}$ then we define*

$$M_n := \min\Big\{2 \leqslant m \leqslant M_n^h : m^3 \|[T]_{\underline{m}}^{-1}\|^2 \max_{1 \leqslant j \leqslant m}[h]_j^2 > a_n\Big\} - 1$$

where we set $M_n := M_n^h$ if the min runs over an empty set. Thus, M_n takes values between 1 and M_n^h. Let $\varsigma_m^2 = 74\big(\mathbb{E}[Y^2] + \max_{1 \leqslant m' \leqslant m}\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2\big)$, then we define

$$\mathrm{pen}_m := 24\,\varsigma_m^2\,(1+\log n)\,n^{-1} \max_{1 \leqslant m' \leqslant m}\|[h]_{\underline{m'}}^t [T]_{\underline{m'}}^{-1}\|^2. \tag{1.16}$$

To apply Bernstein's inequality we need another assumption regarding the error term U. This is captured by the set $\mathcal{U}_\sigma^\infty$ for some $\sigma > 0$, which contains all conditional distributions $P_{U|W}$ such that $\mathbb{E}[U|W] = 0$, $\mathbb{E}[U^2|W] \leqslant \sigma^2$, and Cramer's condition hold, i.e.,

$$\mathbb{E}[|U|^k|W] \leqslant \sigma^k\, k!, \quad k = 3, 4, \ldots.$$

Moreover, besides Assumption 1.3 we need the following Cramer condition which is in particular satisfied if the basis $\{f_l\}_{l\geqslant 1}$ are uniformly bounded.

Assumption 1.4. *There exists $\eta \geqslant 1$ such that the distribution of W satisfies*

$$\sup_{j,l\in\mathbb{N}} \mathbb{E}\,|f_j(W)f_l(W) - \mathbb{E}[f_j(W)f_l(W)]|^k \leqslant \eta^k\, k!,\ k = 3, 4, \ldots.$$

We now present an upper bound for $\widehat{\ell}_{\widetilde{m}}$. As Goldenshluger and Pereverzev [2000] we face a logarithmic loss due to the adaptation.

Theorem 1.4.1. *Assume an iid. n-sample of (Y, Z, W) from the model (1.1a–1.1b) with $\mathbb{E}[Y^2] > 0$. Let Assumptions 1.1–1.4 be satisfied. Suppose that $(m_n^\circ)^3 \max_{1\leqslant j\leqslant m_n^\circ}[h]_j^2 = o(a_n \upsilon_{m_n^\circ})$ as $n \to \infty$. Then we have for all $n \geqslant 1$*

$$\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W}\in\mathcal{U}_\sigma^\infty} \sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}|\widehat{\ell}_{\widetilde{m}} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\,\mathcal{R}^h_{n(1+\log n)^{-1}}$$

for a constant $\mathcal{C} > 0$ only depending on the classes $\mathcal{F}_\gamma^\rho$, $\mathcal{T}_{d,D}^\upsilon$, the constants σ, η and the representer h.

Remark 1.4.1. *In all examples studied below the condition $(m_n^\circ)^3 \max_{1\leqslant j\leqslant m_n^\circ}[h]_j^2 = o(a_n\upsilon_{m_n^\circ})$ as n tends to infinity is satisfied if the structural function φ is sufficiently smooth. More precisely, in case of (pp) it suffices to assume $3 < 2p + 2\min(0, s)$. On the other hand, in case of (pe) or (ep) this condition is automatically fulfilled.* $\square$

In the following definition we introduce empirical versions of the integer M_n and the penalty sequence pen. Thereby, we complete the data driven penalized contrast criterion (1.2a–1.2b). This allows for a completely data driven selection method. For this purpose, we construct an estimator for ς_m^2 by replacing the unknown quantities by their empirical analog, that is,

$$\widehat{\varsigma}_m^2 := 74\Big(n^{-1}\sum_{i=1}^n Y_i^2 + \max_{1\leqslant m'\leqslant m} \|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2\Big).$$

With the nondecreasing sequence $(\widehat{\varsigma}_m^2)_{m\geqslant 1}$ at hand we only need to replace the matrix $[T]_{\underline{m}}$ by its empirical counterpart (cf. Subsection 1.3.1).

Definition 1.4.2. *Let a_n and M_n^h be as in Definition 1.4.1 then for all $n \geqslant 1$ define*

$$\widehat{M}_n := \min\Big\{2 \leqslant m \leqslant M_n^h : m^3 \|[\widehat{T}]_{\underline{m}}^{-1}\|^2 \max_{1 \leqslant j \leqslant m} [h]_j^2 > a_n\Big\} - 1$$

where we set $\widehat{M}_n := M_n^h$ if the min runs over an empty set. Thus, $\widehat{M}_n$ takes values between 1 and M_n^h. Then we introduce for all $m \geqslant 1$ an empirical analog of pen_m by

$$\widehat{\mathrm{pen}}_m := 204\, \widehat{\varsigma}_m^2 (1 + \log n) n^{-1} \max_{1 \leqslant m' \leqslant m} \|[h]_{\underline{m}'}^t [\widehat{T}]_{\underline{m}'}^{-1}\|^2. \tag{1.17}$$

Before we establish the next upper bound we introduce

$$M_n^+ := \min\Big\{2 \leqslant m \leqslant M_n^h : \upsilon_m^{-1} m^3 \max_{1 \leqslant j \leqslant m} [h]_j^2 > 4 D a_n\Big\} - 1 \tag{1.18}$$

where $M_n^+ := M_n^h$ if the min runs over an empty set. Thus, M_n^+ takes values between 1 and M_n^h. As in the partial adaptive case we attain the minimax rate of convergence $\mathcal{R}^h$ within a logarithmic term.

Theorem 1.4.2. *Let the assumptions of Theorem 1.4.1 be satisfied. Additionally suppose that $(M_n^+ + 1)^2 \log n = o(n \upsilon_{M_n^+ + 1})$ as $n \to \infty$ and $\sup_{j \geqslant 1} \mathbb{E}\, |e_j(Z)|^{20} \leqslant \eta^{20}$. Then for all $n \geqslant 1$ we have*

$$\sup_{T \in \mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W} \in \mathcal{U}_\sigma^\infty} \sup_{\varphi \in \mathcal{F}_\gamma^\rho} \mathbb{E}\, |\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\, \mathcal{R}^h_{n(1+\log n)^{-1}}$$

for a constant $\mathcal{C} > 0$ only depending on the classes $\mathcal{F}_\gamma^\rho$, $\mathcal{T}_{d,D}^{\upsilon}$, the constants σ, η and the representer h.

Remark 1.4.2. *Note that below in all examples illustrating Theorem 1.4.2 the condition $(M_n^+ + 1)^2 \log n = o(n \upsilon_{M_n^+ + 1})$ as n tends to infinity is automatically satisfied.* $\square$

As in the case of minimax optimal estimation we now present an upper bound uniformly over the class $\mathcal{F}_\omega^\tau$ of representer. For this purpose define $M_n^\omega := \max\{1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor : \max_{1 \leqslant j \leqslant m}(\omega_j^{-1}) \leqslant n\}$. In the definition of the bounds $\widehat{M}_n$, M_n^+, and M_n^- (cf. Appendix 1.4) we replace M_n^h and $\max_{1 \leqslant j \leqslant m} [h]_j^2$ by M_n^ω and $\max_{1 \leqslant j \leqslant m} \omega_j^{-1}$, respectively. Consequently, by employing $\|h\|_{1/\gamma}^2 \leqslant \tau$ and $\mathcal{R}_n^h \leqslant \tau\, \mathcal{R}_n^\omega$ for all $h \in \mathcal{F}_\omega^\tau$ the next result follows line by line the proof of Theorem 1.4.2 and hence its proof is omitted.

Corollary 1.4.3. *Under the conditions of Theorem 1.4.2 we have for all $n \geqslant 1$*

$$\sup_{T \in \mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W} \in \mathcal{U}_\sigma^\infty} \sup_{\varphi \in \mathcal{F}_\gamma^\rho,\, h \in \mathcal{F}_\omega^\tau} \mathbb{E}\, |\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\, \mathcal{R}^\omega_{n(1+\log n)^{-1}}$$

where the constant $\mathcal{C} > 0$ depends on the parameter spaces $\mathcal{F}_\gamma^\rho$, $\mathcal{F}_\omega^\tau$, $\mathcal{T}_{d,D}^\upsilon$, and the constants σ, η.

Illustration by classical smoothness assumptions. Let us illustrate the cost due to adaption by considering classical smoothness assumptions as discussed in Subsection 1.3.4. In Theorem 1.4.2 and Corollary 1.4.3, respectively, we have seen that the adaptive estimator $\widehat{\ell}_{\widehat{m}}$ attains within a constant the rates $\mathcal{R}^h_{\mathrm{adapt}}$ and $\mathcal{R}^\omega_{\mathrm{adapt}}$. Let us now present the orders of these rates by considering the example configurations of Remark 1.2.1. The proof of the following result is omitted because of the analogy with the proof of Proposition 1.3.6.

Proposition 1.4.4. *Assume an iid. n-sample of (Y, Z, W) from the model (1.1a–1.1b) with conditional expectation operator $T \in \mathcal{T}_{d,D}^\upsilon$, error term U such that $P_{U|W} \in \mathcal{U}_\sigma^\infty$, and $\mathbb{E}[Y^2] > 0$. Then for the example configurations of Remark 1.2.1 we obtain*

(pp) if in addition $3 < 2p + 2\min(s, 0)$ that $m_n^\circ \sim \big(n(1+\log n)^{-1}\big)^{1/(2p+2a)}$ and

(i) $$\mathcal{R}^h_{n(1+\log n)^{-1}} \sim \begin{cases} (n^{-1}(1+\log n))^{(2p+2s-1)/(2p+2a)}, & \text{if } s-a < 1/2\\ n^{-1}(1+\log n)^2, & \text{if } s-a = 1/2\\ n^{-1}(1+\log n), & \text{if } s-a > 1/2,\end{cases}$$

(ii) $$\mathcal{R}^\omega_{n(1+\log n)^{-1}} \sim \max\big((n^{-1}(1+\log n))^{(p+s)/(p+a)}, n^{-1}(1+\log n)\big).$$

(pe) $m_n^\circ \sim \log\big(n(1+\log n)^{-(a+p)/a}\big)^{1/2a}$ and

(i) $\mathcal{R}^h_{n(1+\log n)^{-1}} \sim (1+\log n)^{-(2p+2s-1)/(2a)}$,

(ii) $\mathcal{R}^\omega_{n(1+\log n)^{-1}} \sim (1+\log n)^{-(p+s)/a}$.

(ep) $m_n^\circ \sim \log\big(n(1+\log n)^{-(a+p)/p}\big)^{1/2p}$ and

(i) $$\mathcal{R}^h_{n(1+\log n)^{-1}} \sim \begin{cases} n^{-1}(1+\log n)^{(2a+2p-2s+1)/(2p)}, & \text{if } s-a < 1/2\\ n^{-1}(1+\log n)(\log\log n), & \text{if } s-a = 1/2\\ n^{-1}(1+\log n), & \text{if } s-a > 1/2,\end{cases}$$

(ii) $$\mathcal{R}^\omega_{n(1+\log n)^{-1}} \sim \max\big(n^{-1}(\log n)^{(a+p-s)/p}, n^{-1}(1+\log n)\big).$$

Let us revisit Examples 1.3.1 and 1.3.2. In the following, we apply the general theory to adaptive pointwise estimation and adaptive estimation of averages of the structural function φ.

Example 1.4.1. *Consider the point evaluation functional* $\ell_{h_{t_0}}(\varphi) = \varphi(t_0)$*,* $t_0 \in [0,1]$*, introduced in Example 1.3.1. In this case, the estimator* $\widehat{\ell}_{\widehat{m}}$ *with dimension parameter* $\widehat{m}$ *selected as a minimizer of criterion (1.2a–1.2b) writes as*

$$\widehat{\varphi}_{\widehat{m}}(t_0) := \begin{cases} [e(t_0)]^t_{\underline{\widehat{m}}}[\widehat{T}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}, & \textit{if } [\widehat{T}]_{\underline{\widehat{m}}} \textit{ is nonsingular and } \|[\widehat{T}]^{-1}_{\underline{\widehat{m}}}\| \leqslant \sqrt{n}, \\ 0, & \textit{otherwise} \end{cases}$$

where $\widehat{\varphi}_m$ *is an estimator proposed by Johannes and Schwarz [2010]. Then* $\widehat{\varphi}_{\widehat{m}}(t_0)$ *attains within a constant the rate of convergence* $\mathcal{R}^{h_{t_0}}_{adapt}$*. Applying Proposition 1.4.4 gives*

(pp) $\mathcal{R}^{h_{t_0}}_{n(1+\log n)^{-1}} \sim \big(n^{-1}(1+\log n)\big)^{(2p-1)/(2p+2a)}$,

(ep) $\mathcal{R}^{h_{t_0}}_{n(1+\log n)^{-1}} \sim (1+\log n)^{-(2p-1)/(2a)}$,

(ep) $\mathcal{R}^{h_{t_0}}_{n(1+\log n)^{-1}} \sim n^{-1}(1+\log n)^{(2a+2p+1)/(2p)}$. □

Example 1.4.2. *Consider the linear functional* $\ell_h(\varphi) = \int_0^b \varphi(t)dt$ *with representer* $h := \mathbb{1}_{[0,b]}$ *introduced in Example 1.3.2. The mean squared error of the estimator* $\widehat{\ell}_{\widehat{m}} = \int_0^b \widehat{\varphi}_{\widehat{m}}(t)dt$ *is bounded up to a constant by* $\mathcal{R}^h_{\mathrm{adapt}}$*. Applying Proposition 1.4.4 gives*

(pp) $$\mathcal{R}^{h}_{n(1+\log n)^{-1}} \sim \begin{cases} (n^{-1}(1+\log n))^{(2p+1)/(2p+2a)}, & \textit{if } a > 1/2, \\ n^{-1}(1+\log n)^2, & \textit{if } a = 1/2, \\ n^{-1}(1+\log n), & \textit{otherwise,} \end{cases}$$

(ep) $\mathcal{R}^{h}_{n(1+\log n)^{-1}} \sim (1+\log n)^{-(2p+1)/(2a)}$,

(ep) $$\mathcal{R}^{h}_{n(1+\log n)^{-1}} \sim \begin{cases} n^{-1}(1+\log n)^{(2a+2p-1)/(2p)}, & \textit{if } a > 1/2, \\ n^{-1}(1+\log n)(\log\log n), & \textit{if } a = 1/2, \\ n^{-1}(1+\log n), & \textit{otherwise.} \end{cases}$$

□

1.5 Monte Carlo experiments.

In this section, we examine the finite sample properties of our estimation procedure. We study first the point evaluation functional and thereafter, an average of the structural function. As in Subsection 1.3.4, we consider the case where Z and W are both scalar and $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ coincide with the trigonometric basis. Moreover, we generate the joint density of (Z, W) by the multivariate function $p_{ZW}(z,w) = C_\upsilon[e(z)]^t_{\underline{k}}([I]_{\underline{k}} + A_k)[\nabla_\upsilon]^{1/2}_{\underline{k}}[f(w)]_{\underline{k}}$ where C_υ is a normalization constant, $(\upsilon_j)_{j\geqslant 1}$ is a nondecreasing sequence which is specified below, and $k = 100$. Here, A_k is a randomly generated $k \times k$ matrix with spectral norm $1/2$. Due to the construction of the joint density p_{ZW} the link condition $T \in \mathcal{T}^\upsilon_d$ is satisfied for all $\phi \in \mathcal{E}_k$. Note that if A_k equals the zero matrix then this would correspond to the situation where the eigenfunctions of T coincide with the bases $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$. We generate samples of size $n = 1000$ using the relationship $Y = \mathbb{E}[\varphi(Z)|W] + V$ where $V \sim N(0, 0.01)$. The number of Monte Carlo replications is 1000.

In particular, we want to study the performance of our estimators in finite samples when the dimension parameter m is chosen by our adaptive procedure given in (1.2a–1.2b). The constants in the definition of the adaptive procedure, though suitable for the theory, may be chosen much smaller in practice. Here, we replace in definition of $\widehat{\mathrm{pen}}$ (given in (1.17)) and $\widehat{\varsigma}^2_m$ the constants 204 and 74 by 5 and 1, respectively. In addition, we adjust the upper bound $\widehat{M}$ in the following way. We replace a_n (given in Definition 1.4.1) by $40\, n(1 + \log n)^{-1}$.

Point wise estimation Let us consider the problem of pointwise estimation of $\varphi(z) = 10\, z^2 \sin(\pi\, z)$ for $z \in [0, 1]$ over an equidistantly spaced grid of length 50. We truncate its infinite dimensional vector of coefficients at a sufficiently large integer, say 100. In Figure 1.1, we compare the performance of the estimators with optimal parameter m^*_n (in the first column) and data driven parameter $\widehat{m}$ (in the second column). More precisely, at each point t_0 of the grid we choose m^*_n as the minimizer of the empirical mean of $|\widehat{\ell}_m - \ell_{h_{t_0}}(\varphi)|^2$. The first row represents (pp) with $\upsilon_j = j^{-1}$ while the second depicts (pe) with $\upsilon_j = \exp(-j)$. In case of (pp), the pointwise 95%–confidence bands are sufficiently tight to make significant statements about the curvature of φ. Not surprisingly, in case of (pe) the pointwise confidence bands are much wider. But also in this case the pointwise median of the adaptive estimators is very close to φ.

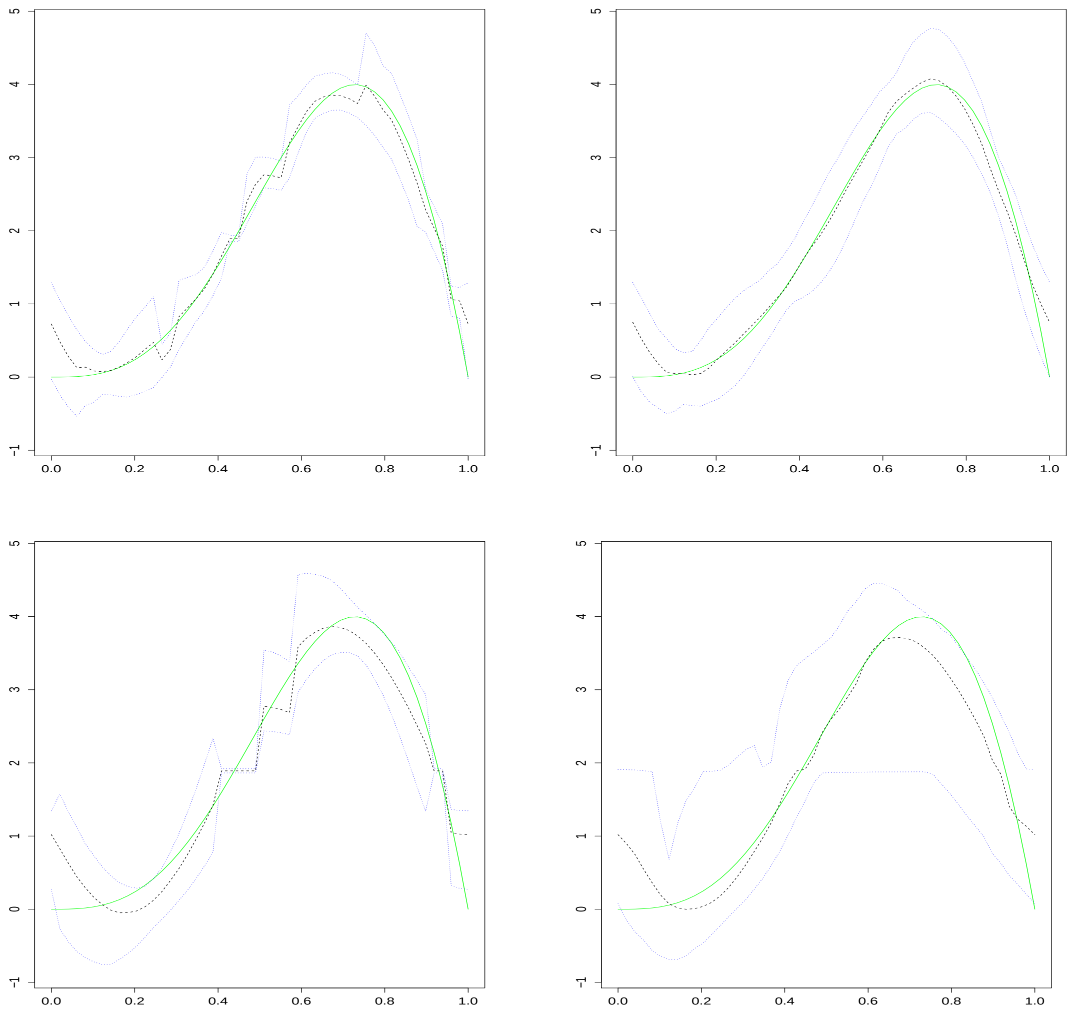

Figure 1.1: The green solid, black dashed, and blue dotted lines show φ, point-wise median of the estimators, and their 95% estimation band.

Estimation of averages We now consider the estimation of averages of the structural function. In the following, let $\varphi(z) = \sum_{j=1}^{100}(-1)^{j+1}j^{-2}e_j(z)$. We consider the problem of estimating the value of the linear functional $\int_0^{0.75}\varphi(z)dz \approx 0.898$. The empirical means from a Monte Carlo simulation are displayed in Table 1.1. Here, we choose m_n^* as the minimizer if the empirical mean of $|\widehat{\ell}_m - \int_0^{0.75}\varphi(z)dz|^2$. From Table 1.1 we see that the difference of the empirical means of $|\widehat{\ell}_{m_n^*} - \ell_h(\varphi)|^2$ and $|\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2$ are small.

Model υ_j	*Sample Size*	*Empirical mean of* $\lvert\widehat{\ell}_{m_n^*} - \ell_h(\varphi)\rvert^2$	$\lvert\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)\rvert^2$
j^{-1}	200	0.0218	0.0202
	1000	0.0058	0.0070
j^{-2}	200	0.0784	0.0770
	1000	0.0317	0.0300
j^{-3}	200	0.1295	0.1404
	1000	0.0931	0.1058
j^{-4}	200	0.1462	0.1533
	1000	0.1288	0.1462
$\exp(-j)$	200	0.0627	0.0619
	1000	0.0214	0.0313
$\exp(-2\,j)$	200	0.1275	0.1479
	1000	0.1080	0.1362
$\exp(-3\,j)$	200	0.1521	0.1555
	1000	0.1341	0.1538

Table 1.1: Results of Monte Carlo Simulations

Appendix

Proof of the lower bound given in Section 1.2.

Proof of Theorem 1.2.1. Let us define $\varphi_* := \Big(\frac{\zeta\,\kappa\,\alpha_n^*}{\sum_{l=1}^{m_n^*}[h]_l^2\upsilon_l^{-1}}\Big)^{1/2}\sum_{j=1}^{m_n^*}[h]_j\upsilon_j^{-1}e_j$ with $\zeta := \min(1/(2d),\rho)$. Since $(\gamma_j^{-1}\upsilon_j)_{j\geqslant 1}$ is nonincreasing and by using the definition of κ given in (1.7) it follows that φ_* and in particular $\varphi_\theta := \theta\varphi_*$ for $\theta\in\{-1,1\}$ belong to $\mathcal{F}_\gamma^\rho$. Let V be a Gaussian random variable with mean zero and variance one ($V\sim\mathcal{N}(0,1)$) which is independent of (Z,W). Consider $U_\theta := [T\varphi_\theta](W)-\varphi_\theta(Z)+V$, then $P_{U_\theta|W}$ belongs to $\mathcal{U}_\sigma$ for all $\sigma^4\geqslant(\sqrt{3}+4\rho\sum_{j\geqslant1}\gamma_j^{-1}\eta^2)^2$, which can be realized as follows. Obviously, we have $\mathbb{E}[U_\theta|W]=0$. Moreover, we have $\sup_j\mathbb{E}[e_j^4(Z)|W]\leqslant\eta^4$ implies $\mathbb{E}[\varphi_\theta^4(Z)|W]\leqslant\rho^2\big(\sum_{j\geqslant1}\gamma_j^{-1}\big)^2\,\mathbb{E}[e_j^4(Z)|W]\leqslant\rho^2\eta^4(\sum_{j\geqslant1}\gamma_j^{-1})^2$ and thus, $|[T\varphi_\theta](W)|^4\leqslant\mathbb{E}[\varphi_\theta^4(Z)|W]\leqslant\rho^2\eta^4(\sum_{j\geqslant1}\gamma_j^{-1})^2$. From the last two bounds we deduce $\mathbb{E}[U_\theta^4|W]\leqslant 16\,\mathbb{E}[\varphi_\theta^4(Z)|W]+6\,\mathbb{V}\mathrm{ar}(\varphi_\theta(Z)|W)+3\leqslant(\sqrt{3}+4\rho\,\eta^2\sum_{j\geqslant1}\gamma_j^{-1})^2$. Consequently, for each θ iid. copies (Y_i,Z_i,W_i), $1\leqslant i\leqslant n$, of (Y,Z,W) with $Y:=\varphi_\theta(Z)+U_\theta$ form an n-sample of the model (1.1a–1.1b) and we denote their joint distribution by P_θ and by $\mathbb{E}_\theta$ the expectation with respect to P_θ. In case of P_θ the conditional distribution of Y given W is Gaussian with mean $[T\varphi_\theta](W)$ and variance 1. The log-likelihood of P_1 with respect to P_{-1} is given by

$$\log\Big(\frac{dP_1}{dP_{-1}}\Big) = \sum_{i=1}^{n} 2(Y_i - [T\varphi_*](W_i))[T\varphi_*](W_i) + \sum_{i=1}^{n} 2|[T\varphi_*](W_i)|^2.$$

Since $T \in \mathcal{T}_d^{\upsilon}$ the Kullback-Leibler divergence satisfies the inequality $KL(P_1, P_{-1}) \leqslant \mathbb{E}_1[\log(dP_1/dP_{-1})] = 2n\|T\varphi_*\|_W^2 \leqslant 2nd\|\varphi_*\|_\upsilon^2$. It is well known that the Hellinger distance $H(P_1, P_{-1})$ satisfies $H^2(P_1, P_{-1}) \leqslant KL(P_1, P_{-1})$ and thus, employing again the definition of κ we have

$$H^2(P_1, P_{-1}) \leqslant 2nd\sum_{j=1}^{m_n^*}[\varphi_*]_j^2\upsilon_j = 2nd\frac{\zeta\,\kappa\,\alpha_n^*}{\sum_{l=1}^{m_n^*}[h]_l^2\upsilon_l^{-1}}\sum_{j=1}^{m_n^*}\frac{[h]_j^2}{\upsilon_j} = 2d\zeta\frac{\kappa\;\alpha_n^*}{n^{-1}} \leqslant 2\,d\,\zeta \leqslant 1. \tag{1.19}$$

Consider the Hellinger affinity $\rho(P_1, P_{-1}) = \int\sqrt{dP_1 dP_{-1}}$ then for any estimator $\breve{\ell}$ it holds

$$\begin{aligned}\rho(P_1, P_{-1}) &\leqslant \int\frac{|\breve{\ell} - \ell_h(\varphi_1)|}{2|\ell_h(\varphi_*)|}\sqrt{dP_1 dP_{-1}} + \int\frac{|\breve{\ell} - \ell_h(\varphi_{-1})|}{2|\ell_h(\varphi_*)|}\sqrt{dP_1 dP_{-1}}\\ &\leqslant \Big(\int\frac{|\breve{\ell} - \ell_h(\varphi_1)|^2}{4|\ell_h(\varphi_*)|^2}dP_1\Big)^{1/2} + \Big(\int\frac{|\breve{\ell} - \ell_h(\varphi_{-1})|^2}{4|\ell_h(\varphi_*)|^2}dP_{-1}\Big)^{1/2}.\end{aligned} \tag{1.20}$$

Due to the identity $\rho(P_1, P_{-1}) = 1 - \frac{1}{2}H^2(P_1, P_{-1})$ combining (1.19) with (1.20) yields

$$\mathbb{E}_1\,|\breve{\ell} - \ell_h(\varphi_1)|^2 + \mathbb{E}_{-1}\,|\breve{\ell} - \ell_h(\varphi_{-1})|^2 \geqslant \frac{1}{2}|\ell_h(\varphi_*)|^2. \tag{1.21}$$

Obviously, $|\ell_h(\varphi_*)|^2 = \zeta\kappa\alpha_n^*\sum_{j=1}^{m_n^*}[h]_j^2\upsilon_j^{-1}$. From (1.21) together with the last identity we conclude for any possible estimator $\breve{\ell}$

$$\begin{aligned}\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}}\ \sup_{P_{U|W}\in\mathcal{U}_\sigma}\ \sup_{\varphi\in\mathcal{F}_\gamma^\rho}\mathbb{E}\,|\breve{\ell} - \ell_h(\varphi)|^2 &\geqslant \sup_{\theta\in\{-1,1\}}\ \mathbb{E}_\theta\,|\breve{\ell} - \ell_h(\varphi_*^{(\theta)})|^2\\ &\geqslant \frac{1}{2}\Big\{\mathbb{E}_1\,|\breve{\ell} - \ell_h(\varphi_1)|^2 + \mathbb{E}_{-1}\,|\breve{\ell} - \ell_h(\varphi_{-1})|^2\Big\}\\ &\geqslant \frac{\kappa}{4}\,\min\Big(\frac{1}{2d}, \rho\Big)\,\alpha_n^*\sum_{j=1}^{m_n^*}[h]_j^2\upsilon_j^{-1}.\end{aligned} \tag{1.22}$$

Consider now $\widetilde{\varphi}_* := \left(\frac{\zeta\,\kappa}{\sum_{l>m_n^*}[h]_l^2\gamma_l^{-1}}\right)^{1/2}\sum_{j>m_n^*}[h]_j\gamma_j^{-1}e_j$, which belongs to $\mathcal{F}_\gamma^\rho$ since $\kappa \leqslant 1$ and $\zeta \leqslant \rho$. Moreover, since $(\gamma_j^{-1}\upsilon_j)_{j\geqslant 1}$ is nonincreasing and by using the definition of κ given in (1.7) we have

$$2nd\sum_{j>m_n^*}[\widetilde{\varphi}_*]_j^2\upsilon_j = 2nd\frac{\zeta\,\kappa}{\sum_{l>m_n^*}[h]_l^2\gamma_l^{-1}}\sum_{j>m_n^*}\frac{[h]_j^2\upsilon_j}{\gamma_j^2} \leqslant 2nd\zeta\frac{\kappa}{\gamma_{m_n^*}\upsilon_{m_n^*}^{-1}} \leqslant 2\,d\,\zeta \leqslant 1.$$

Thereby, following line by line the proof of (1.22) we obtain for any possible estimator $\breve{\ell}$

$$\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}}\ \sup_{P_{U|W}\in\mathcal{U}_\sigma}\ \sup_{\varphi\in\mathcal{F}_\gamma^\rho}\mathbb{E}\,|\breve{\ell}-\ell_h(\varphi)|^2 \geqslant \frac{1}{4}|\ell_h(\widetilde{\varphi}_*)|^2 = \frac{\kappa}{4}\min\left(\frac{1}{2d},\rho\right)\sum_{j>m_n^*}[h]_j^2\gamma_j^{-1}.$$

Combining, the last estimate and (1.22) implies the result of the theorem, which completes the proof. □

Proofs of Section 1.3.

We begin by defining and recalling notations to be used in the proofs of this section. For $m \geqslant 1$ recall $\varphi_m = \sum_{j=1}^m[\varphi_m]_je_j$ with $[\varphi_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ keeping in mind that $[T]_{\underline{m}}$ is nonsingular. Then the identities $[T(\varphi-\varphi_m)]_{\underline{m}} = 0$ and $[\varphi_m - E_m\varphi]_{\underline{m}} = [T]_{\underline{m}}^{-1}[TE_m^\perp\varphi]_{\underline{m}}$ hold true. We denote $Q_m := [\widehat{T}]_{\underline{m}} - [T]_{\underline{m}}$ and $V_m := [\widehat{g}]_{\underline{m}} - [\widehat{T}]_{\underline{m}}[\varphi_m]_{\underline{m}} = n^{-1}\sum_{i=1}^n(U_i + \varphi(Z_i)-\varphi_m(Z_i))[f(W_i)]_{\underline{m}}$, where obviously $\mathbb{E}\,V_m = 0$. Moreover, let us introduce the events

$$\Omega_m := \{\|[\widehat{T}]_{\underline{m}}^{-1}\| \leqslant \sqrt{n}\},\quad \mho_m := \{\sqrt{m}\|Q_m\|\|[T]_{\underline{m}}^{-1}\| \leqslant 1/2\}$$
$$\Omega_m^c := \{\|[\widehat{T}]_{\underline{m}}^{-1}\| > \sqrt{n}\}\quad\text{and}\quad \mho_m^c = \{\sqrt{m}\|Q_m\|\|[T]_{\underline{m}}^{-1}\| > 1/2\}.$$

Observe that if $\sqrt{m}\|Q_m\|\|[T]_{\underline{m}}^{-1}\| \leqslant 1/2$ then the identity $[\widehat{T}]_{\underline{m}} = [T]_{\underline{m}}\{I + [T]_{\underline{m}}^{-1}Q_m\}$ implies by the usual Neumann series argument that $\|[\widehat{T}]_{\underline{m}}^{-1}\| \leqslant 2\|[T]_{\underline{m}}^{-1}\|$. Thereby, if $\sqrt{n} \geqslant 2\|[T]_{\underline{m}}^{-1}\|$ we have $\mho_m \subset \Omega_m$. These results will be used below without further reference. We shall prove at the end of this section four technical Lemmata (1.5.1 – 1.5.4) which are used in the following proofs. Furthermore, we will denote by C universal numerical constants and by $C(\cdot)$ constants depending only on the arguments. In both cases, the values of the constants may change from line to line.

Proof of the consistency.

PROOF OF PROPOSITION 1.3.1. Consider for all $m \geqslant 1$ the decomposition

$$\begin{aligned}\mathbb{E}\,|\widehat{\ell}_m - \ell_h(\varphi)|^2 &= \mathbb{E}\,|\widehat{\ell}_m - \ell_h(\varphi)|^2\,\mathbb{1}_{\Omega_m} + |\ell_h(\varphi)|^2 P(\Omega_m^c)\\ &\leqslant 2\,\mathbb{E}\,|\widehat{\ell}_m - \ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m} + 2|\ell_h(\varphi_m-\varphi)|^2 + |\ell_h(\varphi)|^2 P(\Omega_m^c) \end{aligned}\tag{1.23}$$

where we bound each term separately. Let $\mho_m := \{\|Q_m\|\|[T]_{\underline{m}}^{-1}\| \leqslant 1/2\}$ and let $\mho_m^c$ denote its complement. By employing $\|[\widehat{T}]_{\underline{m}}^{-1}\|\,\mathbb{1}_{\mho_m} \leqslant 2\|[T]_{\underline{m}}^{-1}\|$ and $\|[\widehat{T}]_{\underline{m}}^{-1}\|^2\,\mathbb{1}_{\Omega_m} \leqslant n$ it follows that

$$\begin{aligned}|\widehat{\ell}_m-\ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m} &\leqslant 2\left|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}V_m\right|^2 + 2|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}Q_m[\widehat{T}]_{\underline{m}}^{-1}V_m|^2\,\mathbb{1}_{\Omega_m}(\mathbb{1}_{\mho_m}+\mathbb{1}_{\mho_m^c})\\ &\leqslant 2|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}V_m|^2 + 2\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2\Big\{4\|[T]_{\underline{m}}^{-1}\|^2\|Q_m\|^2\|V_m\|^2 + n\|Q_m\|^2\|V_m\|^2\,\mathbb{1}_{\mho_m^c}\Big\}.\end{aligned}$$

Thus, from estimate (1.27), (1.28), and (1.29) in Lemma 1.5.1 we infer

$$\begin{aligned}\mathbb{E}\,|\widehat{\ell}_m - \ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m} &\leqslant C(\gamma)n^{-1}\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2\eta^4(\sigma^2+\|\varphi-\varphi_m\|_\gamma^2)\\ &\qquad\times\Big\{1+\frac{m^3}{n}\|[T]_{\underline{m}}^{-1}\|^2 + m^3P^{1/4}(\mho_m^c)\Big\}.\end{aligned}\tag{1.24}$$

Let $m = m_n$ satisfying $m_n^{-1} = o(1)$, $m_n = o(n)$, and condition (1.10). We have $\sqrt{n} \geqslant 2\|[T]_{\underline{m_n}}^{-1}\|$ and thus, $\Omega_{m_n}^c \subset \mho_{m_n}^c$ for n sufficiently large. From Lemma 1.5.3 it follows that $m_n^{12}P(\mho_{m_n}^c) \leqslant 2\exp\{-m_n\,(32\eta^2n^{-1}{m_n}^3\|[T]_{\underline{m_n}}^{-1}\|^2)^{-1} + 14\log m_n\} = O(1)$ as $n \to \infty$ since $m_n(4n^{-1}m_n^3\|[T]_{\underline{m_n}}^{-1}\|^2)^{-1} \leqslant 4\eta^2 n$ for n sufficiently large. Thus, in particular $P(\Omega_{m_n}^c) = o(1)$. Consequently, as $n\to\infty$ we obtain $\mathbb{E}\,|\widehat{\ell}_{m_n} - \ell_h(\varphi_{m_n})|^2\,\mathbb{1}_{\Omega_{m_n}} = o(1)$ since $\|[h]_{\underline{m_n}}^t[T]_{\underline{m_n}}^{-1}\|^2 = o(n)$. Moreover, as $n\to\infty$ it holds $|\ell_h(\varphi_{m_n}) - \ell_h(\varphi)|^2 \leqslant \|h\|_{1/\gamma}\|\varphi - \varphi_{m_n}\|_\gamma = o(1)$ due to condition (1.9), and $|\ell_h(\varphi)|^2P(\Omega_{m_n}^c) \leqslant \|h\|_{1/\gamma}\|\varphi\|_\gamma P(\Omega_{m_n}^c) = o(1)$. This together with decomposition (1.23) proves the result. □

PROOF OF COROLLARY 1.3.2. The assertion follows directly from Proposition 1.3.1, it only remains to check conditions (1.9) and (1.10). We make use of decomposition $\|\varphi - \varphi_m\|_\gamma \leqslant \|E_m^\perp\varphi\|_\gamma + \|E_m\varphi - \varphi_m\|_\gamma$. As in the proof of Lemma 1.5.2 we conclude $\|E_m\varphi - \varphi_m\|_\gamma^2 \leqslant \|E_m^\perp\varphi\|_\gamma \sup_m \sup_{\|\phi\|_\gamma=1}\|T_m^{-1}F_mTE_m^\perp\phi\|_\gamma \leqslant Dd\|E_m^\perp\varphi\|_\gamma$. By using Lebesgue's dominated convergence theorem we observe $\|E_m^\perp\varphi\|_\gamma = o(1)$ as $m\to\infty$ and hence (1.9) holds. Condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ implies $\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2 \leqslant D\sum_{j=1}^m [h]_j^2\upsilon_j{}^{-1}$ and $\|[T]_{\underline{m}}^{-1}\|^2 \leqslant D\upsilon_m^{-1}$ for all $m \geqslant 1$ since υ is nonincreasing. Thereby, condition (1.12) implies condition (1.10), which completes the proof. □

Proof of the upper bound.

PROOF OF THEOREM 1.3.3. The proof is based on inequality (1.23). Applying estimate (1.32) in Lemma 1.5.2 gives $|\ell_h(\varphi_m-\varphi)|^2 \leqslant 2\rho\,\{\sum_{j>m}[h]_j^2\gamma_j^{-1}+Dd\,\upsilon_m\gamma_m^{-1}\sum_{j=1}^m[h]_j^2\upsilon_j^{-1}\}$ for all $\varphi\in\mathcal{F}_\gamma^\rho$ and $h\in\mathcal{F}_{1/\gamma}$. Since $|\ell_h(\varphi)|^2\leqslant\|\varphi\|_\gamma^2\|h\|_{1/\gamma}^2$ and $\|\varphi\|_\gamma^2\leqslant\rho$ we conclude

$$\mathbb{E}\,|\widehat{\ell}_m-\ell_h(\varphi)|^2\leqslant 2\,\mathbb{E}\,|\widehat{\ell}_m-\ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m}\\ +4\rho\Big\{\sum_{j>m}[h]_j^2\gamma_j^{-1}+dD\frac{\upsilon_m}{\gamma_m}\sum_{j=1}^m[h]_j^2\upsilon_j^{-1}\Big\}+\rho\|h\|_{1/\gamma}^2P(\Omega_m^c). \quad (1.25)$$

By employing $\|Q_m[\widehat{T}]_{\underline{m}}^{-1}\|^2\,\mathbb{1}_{\mho_m}\leqslant m^{-1}$ and $\|[\widehat{T}]_{\underline{m}}^{-1}\|^2\,\mathbb{1}_{\Omega_m}\leqslant n$ it follows that

$$|\widehat{\ell}_m-\ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m}\leqslant 2|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}V_m|^2+2m^{-1}\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2\|V_m\|^2\\ +2n\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2\|Q_m\|^2\|V_m\|^2\,\mathbb{1}_{\mho_m^c}\,.$$

Due to $T\in\mathcal{T}_{d,D}^\upsilon$ and $\varphi\in\mathcal{F}_\gamma^\rho$ we have $\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2\leqslant D\sum_{j=1}^m[h]_j^2/\upsilon_j$ and $\|\varphi-\varphi_m\|_\gamma^2\leqslant 2\,\rho\,(1+D\,d)$ (cf. (1.31) in Lemma 1.5.2), respectively. Thereby, similarly to the proof of Proposition 1.3.1 we get

$$\mathbb{E}\,|\widehat{\ell}_m-\ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m}\leqslant C(\gamma)D(\sigma^2+\eta^2dD\rho)n^{-1}\sum_{j=1}^m[h]_j^2\upsilon_j^{-1}\Big\{1+m^3P(\mho_m^c)^{1/4}\Big\}.$$

Combining the last estimate with (1.25) yields

$$\mathbb{E}\,|\widehat{\ell}_m-\ell_h(\varphi)|^2\leqslant C(\gamma)D(\sigma^2+\eta^2dD\rho)\max\Big\{\sum_{j>m}[h]_j^2\gamma_j^{-1},\max\Big(\frac{\upsilon_m}{\gamma_m},n^{-1}\Big)\sum_{j=1}^m[h]_j^2\upsilon_j^{-1}\Big\}\\ \times\Big\{1+m^3P(\mho_m^c)^{1/4}\Big\}+\rho\|h\|_{1/\gamma}^2P(\Omega_m^c). \quad (1.26)$$

Consider now the optimal choice $m=m_n^*$ defined in (1.6), then we have

$$\mathbb{E}\,|\widehat{\ell}_{m_n^*}-\ell_h(\varphi)|^2\leqslant C(\gamma)D\Big\{\sigma^2+\rho(\eta^2dD+\|h\|_{1/\gamma}^2)\Big\}\mathcal{R}_n^h\\ \times\Big\{1+(m_n^*)^3P(\mho_{m_n^*}^c)^{1/4}+(\mathcal{R}_n^h)^{-1}P(\Omega_{m_n^*}^c)\Big\}$$

and hence, the assertion follows by making use of Lemma 1.5.4. □

Technical assertions.

The following paragraph gathers technical results used in the proofs of Section 1.3. Below we consider the set $\mathbb{S}^m:=\{s\in\mathbb{R}^m:\|s\|=1\}$.

Lemma 1.5.1. *Suppose that $P_{U|W} \in \mathcal{U}_\sigma$ and that the joint distribution of (Z, W) satisfies Assumption 1.3. If in addition $\varphi \in \mathcal{F}_\gamma^\rho$ with γ satisfying Assumption 1.1, then for all $m \geqslant 1$ we have*

$$\sup_{s\in\mathbb{S}^m} \mathbb{E}\,|s^t\, V_m|^2 \leqslant 2n^{-1}(\sigma^2 + C(\gamma)\,\eta^2\|\varphi - \varphi_m\|_\gamma^2), \tag{1.27}$$

$$\mathbb{E}\|V_m\|^4 \leqslant C(\gamma)\,\big(n^{-1}m\,\eta^2(\sigma^2 + \|\varphi - \varphi_m\|_\gamma^2)\big)^2, \tag{1.28}$$

$$\mathbb{E}\|Q_m\|^8 \leqslant C\,\big(n^{-1}m^2\,\eta^2\big)^4. \tag{1.29}$$

PROOF. Proof of (1.27). Since $(\{U_i + \varphi(Z_i) - \varphi_m(Z_i)\}\sum_{j=1}^m s_j f_j(W_i))$, $1 \leqslant i \leqslant n$, are iid. with mean zero we have $\mathbb{E}\,|s^t\,V_m|^2 = n^{-1}\,\mathbb{E}\,|\{U + \varphi(Z) - \varphi_m(Z)\}\sum_{j=1}^m s_j f_j(W)|^2$. Then (1.27) follows from $\mathbb{E}[U^2|W] \leqslant (\mathbb{E}[U^4|W])^{1/2} \leqslant \sigma^2$ and from Assumption 1.3 (i), i.e., $\sup_{j\in\mathbb{N}}\mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$. Indeed, applying condition $|j|^3\gamma_j^{-1} = o(1)$ (cf. Assumption 1.1) gives $\sum_{j\geqslant 1}\gamma_j^{-1} \leqslant C(\gamma)$ and thus,

$$\begin{aligned}\mathbb{E}\,|\{\varphi(Z) - \varphi_m(Z)\}\sum_{j=1}^m s_j f_j(W)|^2 &\leqslant \|\varphi - \varphi_m\|_\gamma^2 \sum_{l=1}^\infty \gamma_l^{-1}\,\mathbb{E}\,|e_l(Z)\sum_{j=1}^m s_j f_j(W)|^2\\ &\leqslant C(\gamma)\,\eta^2\|\varphi - \varphi_m\|_\gamma^2 \sum_{j=1}^m s_j^2 = C(\gamma)\,\eta^2\|\varphi - \varphi_m\|_\gamma^2.\end{aligned}$$

Proof of (1.28). Observe that for each $1 \leqslant j \leqslant m$, $(\{U_i + \varphi(Z_i) - \varphi_m(Z_i)\}f_j(W_i))$, $1 \leqslant i \leqslant n$, are iid. with mean zero. It follows from Theorem 2.10 in Petrov [1995] that $\mathbb{E}\|V_m\|^4 \leqslant Cn^{-2}m^2\sup_{j\in\mathbb{N}}\mathbb{E}\,|\{U + \varphi(Z) - \varphi_m(Z)\}f_j(W)|^4$. Thereby, (1.28) follows from $\mathbb{E}[U^4|W] \leqslant \sigma^4$ and $\sup_{j\in\mathbb{N}}\mathbb{E}[f_j^4(W)] \leqslant \eta^4$ together with $\mathbb{E}\,|\{\varphi(Z) - \varphi_m(Z)\}f_j(W)|^4 \leqslant C(\gamma)\,\eta^4\|\varphi - \varphi_m\|_\gamma^4$, which can be realized as follows. Since $[T(\varphi - \varphi_m)]_j = 0$ we have $\{\varphi(Z) - \varphi_m(Z)\}f_j(W) = \sum_{l\geqslant 1}[\varphi - \varphi_m]_l\{e_l(Z)f_j(W) - [T]_{j,l}\}$. Furthermore, Assumption 1.3 (ii), i.e., $\sup_{j,l\in\mathbb{N}}\mathbb{E}\,|e_l(Z)f_j(W) - [T]_{j,l}|^4 \leqslant 4!\eta^4$, implies

$$\begin{aligned}\mathbb{E}\,|\{\varphi(Z) - \varphi_m(Z)\}f_j(W)|^4 &\leqslant \|\varphi - \varphi_m\|_\gamma^4\,\mathbb{E}\Big|\sum_{l\geqslant 1}\gamma_l^{-1}|e_l(Z)f_j(W) - [T]_{j,l}|^2\Big|^2\\ &\leqslant C(\gamma)\,\eta^4\|\varphi - \varphi_m\|_\gamma^4.\end{aligned}$$

Proof of (1.29). The random variables $(e_l(Z_i)f_j(W_i) - [T]_{j,l})$, $1 \leqslant i \leqslant n$, are iid. with mean zero for each $1 \leqslant j, l \leqslant m$. Hence, Theorem 2.10 in Petrov [1995] implies $\mathbb{E}\|Q_m\|^8 \leqslant Cn^{-4}m^8\sup_{j,l\in\mathbb{N}}\mathbb{E}\,|e_l(Z)f_j(W) - [T]_{j,l}|^8$ and thus, the assertion follows from Assumption 1.3 (ii), which completes the proof. □

Lemma 1.5.2. *If $T \in \mathcal{T}_{d,D}^{\upsilon}$ and $\varphi \in \mathcal{F}_{\gamma}^{\rho}$, then for all $m \geqslant 1$ we have*

$$\|E_m\varphi - \varphi_m\|_{\gamma}^2 \leqslant D\,d\,\rho, \tag{1.30}$$

$$\|\varphi - \varphi_m\|_{\gamma}^2 \leqslant 2\,(1 + D\,d)\,\rho, \tag{1.31}$$

$$|\langle h, \varphi - \varphi_m\rangle_Z|^2 \leqslant 2\,\rho \sum_{j>m} \frac{[h]_j^2}{\gamma_j} + 2\,D\,d\,\rho\,\frac{\upsilon_m}{\gamma_m}\sum_{j=1}^{m}\frac{[h]_j^2}{\upsilon_j}. \tag{1.32}$$

Proof. Consider (1.30). Since $T \in \mathcal{T}_{d,D}^{\upsilon}$ the identity $[E_m\varphi - \varphi_m]_{\underline{m}} = -[T]_{\underline{m}}^{-1}[TE_m^{\perp}\varphi]_{\underline{m}}$ implies $\|E_m\varphi - \varphi_m\|_{\upsilon}^2 \leqslant D\|TE_m^{\perp}\varphi\|_W^2 \leqslant Dd\|E_m^{\perp}\varphi\|_{\upsilon}^2$. Consequently,

$$\|E_m\varphi - \varphi_m\|_{\upsilon}^2 \leqslant D\,d\,\gamma_m^{-1}\upsilon_m\|\varphi\|_{\gamma}^2 \tag{1.33}$$

because $(\gamma_j^{-1}\upsilon_j)_{j\geqslant 1}$ is nonincreasing and thus, $\|E_m\varphi - \varphi_m\|_{\gamma}^2 \leqslant \gamma_m\upsilon_m^{-1}\,\|E_m\varphi - \varphi_m\|_{\upsilon}^2$. By combination of the last estimate and (1.33) we obtain the assertion (1.30). By employing the decomposition $\|\varphi - \varphi_m\|_{\gamma}^2 \leqslant 2\|\varphi - E_m\varphi\|_{\gamma}^2 + 2\|E_m\varphi - \varphi_m\|_{\gamma}^2$ the bound (1.31) follows from (1.30) and $\|\varphi - E_m\varphi\|_{\gamma}^2 \leqslant \|\varphi\|_{\gamma}^2$. It remains to show (1.32). Applying the Cauchy-Schwarz inequality gives $|\langle h, \varphi - E_m\varphi\rangle_Z|^2 \leqslant \|\varphi\|_{\gamma}^2\sum_{j>m}[h]_j^2\gamma_j^{-1}$ and $|\langle h, E_m\varphi - \varphi_m\rangle_Z|^2 \leqslant D\,d\,\|\varphi\|_{\gamma}^2\,\upsilon_m\gamma_m^{-1}\sum_{j=1}^m[h]_j^2\upsilon_j^{-1}$ by (1.33). Thereby (1.32) follows from the inequality $|\langle h, \varphi - \varphi_m\rangle_Z|^2 \leqslant 2|\langle h, \varphi - E_m\varphi\rangle_Z|^2 + 2|\langle h, E_m\varphi - \varphi_m\rangle_Z|^2$, which completes the proof. $\square$

Lemma 1.5.3. *Suppose that the joint distribution of (Z, W) satisfies Assumption 1.3. Then for all $n \geqslant 1$ and $m \geqslant 1$ we have*

$$P(m^{-2}n\|Q_m\|^2 \geqslant t) \leqslant 2\exp\Big(-\frac{t}{8\eta^2} + 2\log m\Big) \quad \textit{for all } 0 < t \leqslant 4\,\eta^2 n. \tag{1.34}$$

Proof. Our proof starts with the observation that for all $j, l \in \mathbb{N}$ the condition (ii) in Assumption 1.3 implies for all $t > 0$

$$P\Big(\Big|\sum_{i=1}^{n}\{e_j(Z_i)f_l(W_i) - \mathbb{E}[e_j(Z)f_l(W)]\}\Big| \geqslant t\Big) \leqslant 2\exp\left(\frac{-t^2}{4n\eta^2 + 2\eta t}\right),$$

which is just Bernstein's inequality (cf. Bosq [1998]). This implies for all $0 < t \leqslant 2\eta n$

$$\sup_{j,l\in\mathbb{N}} P\Big(\Big|\sum_{i=1}^{n}\{e_j(Z_i)f_l(W_i) - \mathbb{E}[e_j(Z)f_l(W)]\}\Big| \geqslant t\Big) \leqslant 2\exp\Big(-\frac{t^2}{8\eta^2 n}\Big). \tag{1.35}$$

It is well-known that $m^{-1}\|[A]_{\underline{m}}\| \leqslant \max_{1\leqslant j,l\leqslant m}|[A]_{j,l}|$ for any $m \times m$ matrix $[A]_{\underline{m}}$. Combining the last estimate and (1.35) we obtain for all $0 < t \leqslant 2\eta n^{1/2}$

$$P(m^{-1}n^{1/2}\|Q_m\| \geqslant t) \leqslant \sum_{j,l=1}^{m} P\Big(|\sum_{i=1}^{n}(e_j(Z_i)f_l(W_i) - \mathbb{E}[e_j(Z)f_l(W)])| \geqslant n^{1/2}t\Big)$$
$$\leqslant 2\exp\big(-\frac{t^2}{8\eta^2} + 2\log m\big).$$

□

Lemma 1.5.4. *Under the conditions of Theorem 1.3.3 we have for all $n \geqslant 1$*

$$(m_n^*)^{12}P(\mho_{m_n^*}^c) \leqslant C(\gamma, \upsilon, \eta, D) \tag{1.36}$$
$$(\mathcal{R}_n^h)^{-1}P(\Omega_{m_n^*}^c) \leqslant C(\gamma, \upsilon, \eta, h, D). \tag{1.37}$$

Proof. Proof of (1.36). Since $\|[T]_{\underline{m}}^{-1}\|^2 \leqslant D\upsilon_m^{-1}$ due to $T \in \mathcal{T}_{d,D}^{\upsilon}$ it follows from Lemma 1.5.3 for all $m, n \geqslant 1$ that

$$P(\mho_m^c) \leqslant P\Big(m^{-2}n\|Q_m\|^2 > \frac{n\upsilon_m}{4Dm^3}\Big) \leqslant 2\exp\Big(-\frac{n\upsilon_m}{32D\eta^2m^3} + 2\log m\Big)$$

since $(4Dm^3\upsilon_m^{-1})^{-1} \leqslant 1 \leqslant 4\eta^2$ for all $m \geqslant 1$. Due to condition (1.13) there exists $n_0 \geqslant 1$ such that $n\upsilon_{m_n^*} \geqslant 448D\eta^2(m_n^*)^3\log m_n^*$ for all $n \geqslant n_0$. Consequently, $(m_n^*)^{12}P(\mho_{m_n^*}^c) \leqslant 2$ for all $n \geqslant n_0$, while trivially $(m_n^*)^{12}P(\mho_{m_n^*}^c) \leqslant (m_{n_0}^*)^{12}$ for all $n \leqslant n_0$, which gives (1.36) since n_0 and $m_{n_0}^*$ depend on γ, υ, η and D only.

Consider (1.37). Let $n_0 \in \mathbb{N}$ such that $\max\{|\log \mathcal{R}_n^h|, (\log m_n^*)\}(m_n^*)^3 \leqslant n\upsilon_{m_n^*}(96D\eta^2)^{-1}$ for all $n \geqslant n_0$. Observe that $\mho_m \subset \Omega_m$ if $n \geqslant 4D\upsilon_m^{-1}$. Since $(m_n^*)^{-3}n\upsilon_{m_n^*} \geqslant 96D\eta^2$ for all $n \geqslant n_0$ it follows $n\upsilon_{m_n^*} \geqslant 4D$ for all $n \geqslant n_0$ and hence $(\mathcal{R}_n^h)^{-1}P(\Omega_{m_n^*}^c) \leqslant (\mathcal{R}_n^h)^{-1}P(\mho_{m_n^*}^c) \leqslant 2$ for all $n \geqslant n_0$ as in the proof of (1.36). Combining the last estimate and the elementary inequality $(\mathcal{R}_n^h)^{-1}P(\Omega_{m_n^*}^c) \leqslant (\mathcal{R}_{n_0}^h)^{-1}$ for all $n \leqslant n_0$ shows (1.37) since n_0 depends on γ, υ, η, h and D only, which completes the proof. □

Proofs of Section 1.3.4

Proof of Proposition 1.3.6. Proof of (pp). From the definition of m_n^* in (1.6) it follows $m_n^* \sim n^{1/(2p+2a)}$. Consider case (i). The condition $s - a < 1/2$ implies that $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}(m_n^*)^{2a-2s+1} \sim n^{-(2p+2s-1)/(2p+2a)}$ and moreover, we calculate $\sum_{j>m_n^*}|j|^{-2p-2s} \sim n^{-(2p+2s-1)/(2p+2a)}$ since $p + s > 1/2$. If $s - a = 1/2$ then $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}\log(n^{1/(2p+2a)})$ and $\sum_{j>m_n^*}|j|^{-2p-2s} \sim n^{-1}$. In the case of $s - a > 1/2$ it follows that $\sum_{j=1}^{m_n^*}|j|^{2a-2s}$ is bounded whereas $\sum_{j>m_n^*}|j|^{-2p-2s} \lesssim n^{-1}$

and hence, $\mathcal{R}_n^h \sim n^{-1}$. To prove (ii) we make use of Corollary 1.2.2. We observe that if $s-a \geqslant 0$ the sequence ωv is bounded from below, and hence $\mathcal{R}_n^\omega \sim n^{-1}$. Otherwise, the condition $s-a<0$ implies $\mathcal{R}_n^\omega \sim n^{-(p+s)/(p+a)}$.

Proof of (pe). Note that m_n^* satisfies $m_n^* \sim \log(n(\log n)^{-p/a})^{1/(2a)}$. In order to prove (i), we calculate that $\sum_{j>m_n^*}|j|^{-2p-2s} \sim (\log n)^{(-2p-2s+1)/(2a)}$ and moreover, $n^{-1}\sum_{j=1}^{m_n^*}\exp(|j|^{2a})|j|^{-2s} \lesssim (\log n)^{(-2p-2s+1)/(2a)}$. In case (ii) we immediately obtain $\mathcal{R}_n^\omega \sim (\log n)^{-(p+s)/a}$.

Proof of (ep). It holds true $m_n^* \sim \log(n(\log n)^{-a/p})^{1/(2p)}$. Consider case (i). If $s-a<1/2$ then $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}(\log n)^{(2a-2s+1)/(2p)}$. If $s-a=1/2$ we conclude $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}\log(\log(n))$. On the other hand, the condition $s-a>1/2$ implies that $\sum_{j=1}^{m_n^*}|j|^{2a-2s}$ is bounded and thus, we obtain the parametric rate n^{-1}. Moreover, it is easily seen that $\sum_{j>m_n^*}|j|^{-2s}\exp(-|j|^{2p}) \lesssim n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s}$. In case (ii) if $s-a\geqslant 0$ then the sequence ωv is bounded from below as mentioned above and thus, $\mathcal{R}_n^\omega \sim n^{-1}$. If $s-a<0$ then $\mathcal{R}_n^\omega \sim n^{-1}(\log n)^{(a-s)/p}$, which completes the proof. $\square$

Proofs of Section 1.4

At the end of this section we shall prove six technical Lemmata (1.5.7 – 1.5.12) which are used in the following proofs. Let us introduce a nondecreasing sequence $\Delta := (\Delta_m)_{m\geqslant 1}$ and its empirical analog $\widehat{\Delta} := (\widehat{\Delta}_m)_{m\geqslant 1}$ by $\Delta_m := \max_{1\leqslant m'\leqslant m}\|[h]_{\underline{m'}}^t[T]_{\underline{m'}}^{-1}\|^2$ and $\widehat{\Delta}_m := \max_{1\leqslant m'\leqslant m}\|[h]_{\underline{m'}}^t[\widehat{T}]_{\underline{m'}}^{-1}\|^2$, respectively. Similarly to M_n^+ introduced in (1.18) we define

$$M_n^- := \min\Big\{2\leqslant m\leqslant M_n^h : 4Dv_m^{-1}m^3\max_{1\leqslant j\leqslant m}[h]_j^2 > a_n\Big\}-1 \tag{1.38}$$

where we set $M_n^- := M_n^h$ if the set is empty. Thus, M_n^- takes values between 1 and M_n^h. In the following $\mathcal{C}>0$ denotes a constant only depending on the classes $\mathcal{F}_\gamma^\rho$, $\mathcal{T}_{d,D}^v$, the constants σ, η and the representer h. For ease of notation, the value of $\mathcal{C}>0$ may change from line to line.

Proof of Theorem 1.4.1. The proof of the theorem is based on inequality (1.15). Observe that by Lemma 1.5.10 we have $M_n^- \leqslant M_n \leqslant M_n^+$. Further, due to condition $(m_n^\circ)^3\max_{1\leqslant j\leqslant m_n^\circ}[h]_j^2 = o(a_n v_{m_n^\circ})$ there exists $n_0\geqslant 1$ only depending on h, γ, and v such that for all $n\geqslant n_0$ it holds $m_n^\circ \leqslant M_n^-$. We distinguish in the following the cases $n\geqslant n_0$ and $n<n_0$. First, consider $n\geqslant n_0$. Applying Corollary 1.5.6 together with estimate (1.15) implies

$$\mathbb{E}|\widehat{\ell}_{\widetilde{m}}-\ell_h(\varphi)|^2 \leqslant \mathcal{C}\Big\{\operatorname{pen}_{m_n^\circ}+\operatorname{bias}_{m_n^\circ}+n^{-1}\Big\}.$$

From the definition of pen_m we infer $\mathrm{pen}_m \leqslant 24(3\rho + 2\sigma^2)(1 + \log n)n^{-1}D \sum_{j=1}^{m}[h]_j^2 \upsilon_j^{-1}$ since $T \in \mathcal{T}_{d,D}^{\upsilon}$, $U \in \mathcal{U}_{\sigma}^{\infty}$, and $\varphi \in \mathcal{F}_{\gamma}^{\rho}$. Moreover, since $\varphi \in \mathcal{F}_{\gamma}^{\rho}$ and $h \in \mathcal{F}_{1/\gamma}$ estimate (1.32) in Lemma 1.5.2 implies for all $1 \leqslant m \leqslant M_n^-$ the inquality $\mathrm{bias}_m \leqslant \min_{1 \leqslant m' \leqslant M_n^-} 2\,\rho \left\{ \sum_{j>m'} [h]_j^2 \gamma_j^{-1} + dD\upsilon_{m'}\gamma_{m'}^{-1} \sum_{j=1}^{m'} [h]_j^2 \upsilon_j^{-1} \right\}$.

Consequently,

$$\mathbb{E}|\widehat{\ell}_{\widetilde{m}} - \ell_h(\varphi)|^2 \leqslant \mathcal{C}\Big\{ \max\Big(\sum_{j>m_n^\circ} [h]_j^2 \gamma_j^{-1}, \alpha_n^\circ \sum_{j=1}^{m_n^\circ} [h]_j^2 \upsilon_j^{-1} \Big) + n^{-1} \Big\}.$$

Consider now $n < n_0$. Observe that for all $1 \leqslant m \leqslant M_n^h$ it holds

$$\begin{aligned} |\widehat{\ell}_m - \ell_h(\varphi)|^2 &\leqslant 2|[h]_{\underline{m}}^t [\widehat{T}]_{\underline{m}}^{-1} V_m|^2 \,\mathbb{1}_{\Omega_m} + 2(|\ell_h(\varphi_m - \varphi)|^2 + |\ell_h(\varphi)|^2 \,\mathbb{1}_{\Omega_m^c}) \\ &\leqslant 2n\|[h]_{\underline{M_n^h}}\|^2 \|V_{M_n^h}\|^2 + 2(|\ell_h(\varphi_m - \varphi)|^2 + |\ell_h(\varphi)|^2 \,\mathbb{1}_{\Omega_m^c}). \end{aligned} \quad (1.39)$$

From the definition of M_n^h we infer $\|[h]_{\underline{M_n^h}}\|^2 \leqslant [h]_1^2 n^{5/4}$. Hence inequality (1.28) in Lemma 1.5.1, inequality (1.31) in Lemma 1.5.2 and Lemma 1.5.12 yield for all $\varphi \in \mathcal{F}_{\gamma}^{\rho}$ and $h \in \mathcal{F}_{1/\gamma}$

$$n\,\mathbb{E}\,|\widehat{\ell}_{\widetilde{m}} - \ell_h(\varphi)|^2 \leqslant 2\,[h]_1^2 n^{9/5} \|V_{M_n^h}\|^2 + 6\rho\|h\|_{1/\gamma}^2 (1 + Dd)n \leqslant \mathcal{C},$$

which proves the result. □

Lemma 1.5.5. *Consider* $(\widetilde{\mathrm{pen}}_m)_{m\geqslant 1}$ *with* $\widetilde{\mathrm{pen}}_m := 24(24\,\mathbb{E}[U^2] + 96\eta^2\rho\, m^3\gamma_m^{-1})(1+\log n)n^{-1}$. *Then under the conditions of Theorem 1.4.1 we have for all* $n \geqslant 1$

$$\sup_{T \in \mathcal{T}_{d,D}^{\upsilon}} \sup_{P_{U|W} \in \mathcal{U}_{\sigma}^{\infty}} \mathbb{E} \max_{m_n^\circ \leqslant m \leqslant M_n^+} \Big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\,\widetilde{\mathrm{pen}}_m \Big)_+ \leqslant \mathcal{C}\, n^{-1}.$$

PROOF. Similarly to the proof of Theorem 1.3.3 we obtain the decomposition

$$\begin{aligned} |\widehat{\ell}_m - \ell_h(\varphi_m)|^2 \leqslant 2|[h]_{\underline{m}}^t [T]_{\underline{m}}^{-1} V_m|^2 + 2m^{-1}\|[h]_{\underline{m}}^t [T]_{\underline{m}}^{-1}\|^2 \|V_m\|^2 + \\ 2n\|[h]_{\underline{m}}^t [T]_{\underline{m}}^{-1} Q_m\|^2 \|V_m\|^2 \,\mathbb{1}_{\mho_m^c} + |\ell_h(\varphi_m)|^2 \,\mathbb{1}_{\Omega_m^c}. \end{aligned}$$

Observe that $\|[h]_{\underline{m}}^t [T]_{\underline{m}}^{-1}\|^2 \leqslant \Delta_m$ for all $m \geqslant 1$ and hence, we have for all $m_n^\circ \leqslant m \leqslant$

M_n^+

$$\Big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\,\widetilde{\mathrm{pen}}_m\Big)_+ \leqslant 2\Delta_m\Big(\frac{|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}V_m|^2}{\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|^2} - \frac{\widetilde{\mathrm{pen}}_m}{24\Delta_m}\Big)_+$$
$$+\,2\Delta_m\Big(\frac{\|V_m\|^2}{m} - \frac{\widetilde{\mathrm{pen}}_m}{24\Delta_m}\Big)_+ + 2n\Delta_m\|Q_m\|^2\|V_m\|^2\,\mathbb{1}_{\mho_m^c} + |\ell_h(\varphi_m)|^2\,\mathbb{1}_{\Omega_m^c}$$
$$=: I_m + II_m + III_m + IV_m.$$

Consider the first two right hand side terms. We calculate

$$\mathbb{E}\max_{m_n^\circ\leqslant m\leqslant M_n^+}(I_m + II_m) \leqslant 4\max_{m_n^\circ\leqslant m\leqslant M_n^+}\sup_{s\in\mathbb{S}^m}\mathbb{E}\left(|s^tV_m|^2 - \frac{\widetilde{\mathrm{pen}}_m}{24\Delta_m}\right)_+\sum_{m=1}^{M_n^+}\Delta_m.$$

From the definition of $\widetilde{\mathrm{pen}}$ we infer for all $s\in\mathbb{S}^m$ and $m_n^\circ\leqslant m\leqslant M_n^+$

$$n\,\mathbb{E}\left(|s^tV_m|^2 - \frac{\widetilde{\mathrm{pen}}_m}{24\Delta_m}\right)_+ \leqslant 2\,\mathbb{E}\left((n^{-1/2}\sum_{i=1}^n U_i s^t[f(W_i)]_{\underline{m}})^2 - 12\,\mathbb{E}[U^2](1+\log n)\right)_+$$
$$+\,2\,\mathbb{E}\left((n^{-1/2}\sum_{i=1}^n(\varphi(Z_i)-\varphi_m(Z_i))s^t[f(W_i)]_{\underline{m}})^2 - 48\eta^2\rho\,m^3\gamma_m^{-1}(1+\log n)\right)_+$$
$$\leqslant C(\sigma,\eta,\gamma,\rho,D)\,n^{-1}$$

where the last inequality follows from Lemma 1.5.7 and 1.5.8. Due to the definition of M_n^+ and since Δ is nondecreasing we have

$$n^{-1}\sum_{m=1}^{M_n^+}\Delta_m \leqslant D(n\upsilon_{M_n^+})^{-1}(M_n^+)^2\max_{1\leqslant j\leqslant M_n^+}[h]_j^2 \leqslant 4D^2.$$

Consequently, $\mathbb{E}\max_{m_n^\circ\leqslant m\leqslant M_n^+}(I_m+II_m)\leqslant \mathcal{C}n^{-1}$. Further, we obtain for $\varphi\in\mathcal{F}_\gamma^\rho$ and $h\in\mathcal{F}_{1/\gamma}$

$$\mathbb{E}\max_{m_n^\circ\leqslant m\leqslant M_n^+}(III_m) \leqslant n\Delta_{M_n^+}(\mathbb{E}\,\|Q_{M_n^+}\|^8)^{1/4}(\mathbb{E}\,\|V_{M_n^+}\|^4)^{1/2}P^{1/4}\Big(\bigcup_{m=1}^{M_n^+}\mho_m^c\Big)$$
$$\leqslant C(\gamma)\,\eta^4(\sigma^2+(1+Dd)\rho)n^{-1}\Delta_{M_n^+}(M_n^+)^3P^{1/4}\Big(\bigcup_{m=1}^{M_n^+}\mho_m^c\Big)$$

where the last inequality is due to Lemma 1.5.1 and

$$\mathbb{E} \max_{m_n^\circ \leqslant m \leqslant M_n^+} (IV_m) \leqslant \rho \|h\|^2_{1/\gamma} P\Big(\bigcup_{m=1}^{M_n^+} \Omega_m^c \Big).$$

Now applying inequality $n^{-1}\Delta_{M_n^+}(M_n^+)^3 \leqslant 4D^2$ and Lemma 1.5.9 gives the upper bound $\mathbb{E}\max_{m_n^\circ \leqslant m \leqslant M_n^+}(III_m + IV_m) \leqslant \mathcal{C}n^{-1}$, which completes the proof. □

Corollary 1.5.6. *Under the conditions of Theorem 1.4.1 we have for all $n \geqslant 1$*

$$\sup_{T \in \mathcal{T}_{d,D}^v} \sup_{P_{U|W} \in \mathcal{U}_\sigma^\infty} \mathbb{E} \max_{m_n^\circ \leqslant m \leqslant M_n^+} \Big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\operatorname{pen}_m \Big)_+ \leqslant \mathcal{C}\, n^{-1}.$$

Proof. Observe that $m^3\gamma_m^{-1} = o(1)$ and $\|\varphi - \varphi_m\|_Z^2 = o(1)$ as $m \to \infty$ due to Assumption 1.1 and $T \in \mathcal{T}_{d,D}^v$ (cf. proof of Corollary 1.3.2), respectively. Thereby, there exists a constant n_0 only depending on γ, ρ, and η such that for all $n \geqslant n_0$ and $m \geqslant m_n^\circ$ we have

$$24\,\mathbb{E}[U^2] + 96\eta^2\rho\, m^3\gamma_m^{-1} \leqslant 72\big(\mathbb{E}[Y^2] + \|\varphi_m\|_Z^2 + \|\varphi - \varphi_m\|_Z^2\big) + 96\eta^2\rho\, m^3\gamma_m^{-1} \leqslant \varsigma_m^2. \quad (1.40)$$

We distinguish in the following the cases $n < n_0$ and $n \geqslant n_0$. First, consider $n < n_0$. Due to $n^{-1}\sum_{m=1}^{M_n^+}\Delta_m \leqslant 4D^2$ and inequality (1.27) in Lemma 1.5.1 we calculate for all $s \in \mathbb{S}^m$

$$\sum_{m=1}^{M_n^+} \Delta_m\, \mathbb{E}\Big(|s^t V_m|^2 - \frac{\operatorname{pen}_m}{24\Delta_m}\Big)_+ \leqslant \sum_{m=1}^{M_n^+} \Delta_m\, \mathbb{E}\,|s^t V_m|^2 \leqslant 8 n_0 D^2 \big(\sigma^2 + C(\gamma)\,\eta^2\, \|\varphi - \varphi_m\|_\gamma^2\big)\, n^{-1}.$$

Therefore, following line by line the proof of Lemma 1.5.5 it is easily seen that it holds $n\,\mathbb{E}\max_{m_n^\circ \leqslant m \leqslant M_n^+}\big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\operatorname{pen}_m\big)_+ \leqslant \mathcal{C}$. Consider now $n \geqslant n_0$. Inequality (1.40) implies $\widetilde{\operatorname{pen}}_m \leqslant \operatorname{pen}_m$ and thus, $\big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\operatorname{pen}_m\big)_+ \leqslant \big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\widetilde{\operatorname{pen}}_m\big)_+$ for all $m_n^\circ \leqslant m \leqslant M_n^+$. Thus, from Lemma 1.5.5 we infer $n\,\mathbb{E}\max_{m_n^\circ \leqslant m \leqslant M_n^+}\big(|\widehat{\ell}_m - \ell_h(\varphi_m)|^2 - \frac{1}{6}\operatorname{pen}_m\big)_+ \leqslant \mathcal{C}$, which completes the proof of the corollary.

□

PROOF OF THEOREM 1.4.2. Similarly to the proof of Theorem 1.4.1 and since $\widehat{\operatorname{pen}}$ is a nondecreasing sequence we have for all $1 \leqslant m \leqslant \widehat{M}_n$

$$|\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2 \lesssim \widehat{\operatorname{pen}}_m + \operatorname{bias}_m + \max_{m \leqslant m' \leqslant \widehat{M}_n} \Big(|\widehat{\ell}_{m'} - \ell_h(\varphi_{m'})|^2 - \frac{1}{6}\widehat{\operatorname{pen}}_{m'}\Big)_+ .$$

Let us introduce the set

$$\mathcal{A} := \Big\{\mathrm{pen}_m \leqslant \widehat{\mathrm{pen}}_m \leqslant 8\,\mathrm{pen}_m, \quad 1 \leqslant m \leqslant M_n^+\Big\} \cap \{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\},$$

then we conclude for all $1 \leqslant m \leqslant M_n^-$

$$|\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2\, \mathbb{1}_{\mathcal{A}} \lesssim \mathrm{pen}_m + \mathrm{bias}_m + \max_{m \leqslant m' \leqslant M_n^+} \Big(|\widehat{\ell}_{m'} - \ell_h(\varphi_{m'})|^2 - \frac{1}{6}\,\mathrm{pen}_{m'}\Big)_+.$$

Thereby, similarly as in the proof of Theorem 1.4.1 we obtain for all $\varphi \in \mathcal{F}_\gamma^\rho$ and $h \in \mathcal{F}_{1/\gamma}$ the upper bound for all $n \geqslant 1$

$$\mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2\, \mathbb{1}_{\mathcal{A}} \leqslant \mathcal{C}\,\mathcal{R}^h_{n(1+\log n)^{-1}}. \tag{1.41}$$

Let us now evaluate the risk of the adaptive estimator $\widehat{\ell}_{\widehat{m}}$ on $\mathcal{A}^c$. From the definition of M_n^h we infer $\|[h]_{\underline{M_n^h}}\|^2 \leqslant [h]_1^2\, n\, M_n^h$. Consequently, inequality (1.39) together with (1.28) in Lemma 1.5.1, (1.31) in Lemma 1.5.2 and Lemma 1.5.12 yields for all $\varphi \in \mathcal{F}_\gamma^\rho$ and $h \in \mathcal{F}_{1/\gamma}$

$$\begin{aligned}&\mathbb{E}\,|\widehat{\ell}_{\widehat{m}} - \ell_h(\varphi)|^2\, \mathbb{1}_{\mathcal{A}^c}\\ &\qquad \leqslant 2\,[h]_1^2\, n^2 M_n^h (\mathbb{E}\,\|V_{M_n^h}\|^4)^{1/2} P(\mathcal{A}^c)^{1/2} + 6\rho\|h\|^2_{1/\gamma}(1+Dd)P(\mathcal{A}^c) \leqslant \mathcal{C}\, n^{-1}.\end{aligned}$$

The result follows by combining the last inequality with (1.41). $\square$

Technical assertions.

The following paragraph gathers technical results used in the proofs of Section 1.4. In the following we denote $\xi_s(w) := \sum_{j=1}^m s_j f_j(w)$ where $s \in \mathbb{S}^m = \{s \in \mathbb{R}^m : \|s\| = 1\}$.

Lemma 1.5.7. *Let Assumptions 1.3 and 1.4 hold. Then for all $n \geqslant 1$ and $1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor$ we have*

$$\sup_{P_{U|W} \in \mathcal{U}_\sigma^\infty} \sup_{s \in \mathbb{S}^m} \mathbb{E}\Big[\Big(\frac{1}{n}\Big|\sum_{i=1}^n U_i \xi_s(W_i)\Big|^2 - 12\,\mathbb{E}[U^2](1+\log n)\Big)_+\Big] \leqslant C(\sigma,\eta)\, n^{-1}.$$

Proof. Let us denote $\delta = 12\,\mathbb{E}[U^2](1+\log n)$. Since the error term U satisfies Cramer's condition we may apply Bernstein's inequality and since $\mathbb{E}[U^2|W] \leqslant \sigma^2$ we have

$$\mathbb{E}\Big[\Big(\frac{1}{n}\Big|\sum_{i=1}^n U_i\xi_s(W_i)\Big|^2 - \delta\Big)_+ |W_1,\dots,W_n\Big]$$

$$= \int_0^\infty P\Big(\sum_{i=1}^n U_i\xi_s(W_i) \geqslant \sqrt{n(t+\delta)}|W_1,\dots,W_n\Big)dt$$

$$\leqslant \int_0^\infty \exp\Big(\frac{-n(t+\delta)}{8\sigma^2\sum_{i=1}^n|\xi_s(W_i)|^2}\Big)dt + \int_0^\infty \exp\Big(\frac{-\sqrt{n(t+\delta)}}{4\sigma\max_{1\leqslant i\leqslant n}|\xi_s(W_i)|}\Big)dt. \quad (1.42)$$

Consider the first summand of (1.42). Let us introduce the set

$$\mathcal{B} := \Big\{\forall 1 \leqslant j,l \leqslant m:\ |n^{-1}\sum_{i=1}^n f_j(W_i)f_l(W_i) - \delta_{jl}| \leqslant \frac{\log n}{3\sqrt{n}}\Big\}$$

where $\delta_{jl} = 1$ if $j = l$ and zero otherwise. Applying Cauchy-Schwarz's inequality twice we observe on $\mathcal{B}$ for all $n \geqslant 1$ and $1 \leqslant m \leqslant M_n^+$

$$|n^{-1}\sum_{i=1}^n |\xi_s(W_i)|^2 - 1|\,\mathbb{1}_{\mathcal{B}} \leqslant \sum_{j,l=1}^m |z_j||z_l||n^{-1}\sum_{i=1}^n f_j(W_i)f_l(W_i) - \delta_{jl}|\,\mathbb{1}_{\mathcal{B}} \leqslant \frac{1}{2}$$

since $n^{-1/4}\log n \leqslant 3/2$ for all $n \geqslant 1$. Thereby, it holds $n^{-1}\sum_{i=1}^n |\xi_s(W_i)|^2\,\mathbb{1}_{\mathcal{B}} \leqslant 3/2$ and thus,

$$n\,\mathbb{E}\Big[\int_0^\infty \exp\Big(\frac{-n(t+\delta)}{8\sigma^2\sum_{i=1}^n|\xi_s(W_i)|^2}\Big)dt\,\mathbb{1}_{\mathcal{B}}\Big] \leqslant 12\sigma^2\exp\Big(\log n - \frac{\delta}{12\sigma^2}\Big) \leqslant 6\sigma^2. \quad (1.43)$$

On the complement of $\mathcal{B}$ observe that $\sup_{j,l}\mathbb{V}\mathrm{ar}(f_j(W)f_l(W)) < \eta^2$ due that Assumption 1.3 (i) and thus, Assumption 1.4 together with Bernstein's inequality yields

$$P(\mathcal{B}^c) \leqslant \sum_{j,l=1}^m P\Big(3|\sum_{i=1}^n f_j(W_i)f_l(W_i) - \delta_{jl}| > \sqrt{n}\log n\Big)$$

$$\leqslant 2m^2\exp\Big(-\frac{n(\log n)^2}{36n\eta^4 + 6\eta\sqrt{n}\log n}\Big) \leqslant 2\exp\Big(2\log m - \frac{(\log n)^2}{42\eta^4}\Big).$$

By Assumption 1.3 (i) it holds $\mathbb{E}\,|\xi_s(W)|^4 \leqslant \mathbb{E}\,|\sum_{j=1}^m f_j^2(W)|^2 \leqslant m^2\eta^4$. Thereby

$$n\,\mathbb{E}\Big[\int_0^\infty \exp\Big(\frac{-n(t+\delta)}{8\sigma^2\sum_{i=1}^n|\xi_s(W_i)|^2}\Big)dt\,\mathbb{1}_{\mathcal{B}^c}\Big] \leqslant 8\sigma^2 n\big(\mathbb{E}\,|\xi_s(W_1)|^4 P(\mathcal{B}^c)\big)^{1/2} \leqslant 12\sigma^2\eta^2 \quad (1.44)$$

for all $n \geqslant \exp(126\eta^4)$ and $1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor$. Moreover, for $n < \exp(126\eta^4)$ it holds $n\,\mathbb{E}[|\xi_s(W_1)|^2\, \mathbb{1}_{\mathcal{B}^c}] < \exp(126\eta^4)$. Consider the second summand of (1.42). Since $x \mapsto \exp(-1/x)$, $x > 0$, is a concave function and $\mathbb{E}\,|\xi_s(W)|^4 \leqslant m^2\eta^4$ we deduce for all $1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor$

$$\begin{aligned}
\mathbb{E}\Big[\int_0^\infty \exp\Big(\frac{-\sqrt{n(t+\delta)}}{4\sigma \max_{1\leqslant i\leqslant n}|\xi_s(W_i)|}\Big)dt\Big] &\leqslant \int_0^\infty \exp\Big(\frac{-\sqrt{n(t+\delta)}}{4\sigma\,\mathbb{E}\max_{1\leqslant i\leqslant n}|\xi_s(W_i)|}\Big)dt \\
\leqslant \int_0^\infty \exp\Big(\frac{-\sqrt{n(t+\delta)}}{4\sigma(n\,\mathbb{E}\,|\xi_s(W)|^4)^{1/4}}\Big)dt &\leqslant \int_0^\infty \exp\Big(\frac{-n^{1/4}\sqrt{(t+\delta)}}{4\sigma\,\eta\sqrt{m}}\Big)dt \\
\leqslant 8\sigma\,\eta\sqrt{m/n}\exp\Big(\frac{-n^{1/4}\sqrt{\delta}}{4\sigma\,\eta\sqrt{m}}\Big)&\Big(n^{1/4}\sqrt{\delta}+4\sigma\,\eta\sqrt{m}\Big) \leqslant C(\sigma,\eta)n^{-1}. \qquad (1.45)
\end{aligned}$$

The assertion follows now by combining inequality (1.42) with (1.43), (1.44), and (1.45). $\square$

Lemma 1.5.8. *Let Assumptions 1.1 and 1.3 hold. Then for all $n \geqslant 1$ and $m \geqslant 1$ we have*

$$\sup_{T\in\mathcal{T}_{d,D}^{\upsilon}}\sup_{s\in\mathbb{S}^m}\mathbb{E}\Big[\Big(\frac{1}{n}\Big|\sum_{i=1}^n(\varphi-\varphi_m)(Z_i)\xi_s(W_i)\Big|^2-48\eta^2\rho\frac{m^3}{\gamma_m}(1+\log n)\Big)_+\Big] \leqslant C(\eta,\gamma,\rho,D)n^{-1}.$$

Proof. Let us consider a sequence $w := (w_j)_{j\geqslant 1}$ with $w_j := j^2$. Since $[T(\varphi-\varphi_m)]_{\underline{m}} = 0$ we conclude for $m \geqslant 1$, $s \in \mathbb{S}^m$, and $k = 2, 3, \dots$ that

$$\begin{aligned}
\mathbb{E}\,|(\varphi(Z)-\varphi_m(Z))\xi_s(W)|^k &= \mathbb{E}\,|\sum_{l=1}^\infty[\varphi-\varphi_m]_l\sum_{j=1}^m s_j(e_l(Z)f_j(W)-[T]_{jl})|^k \\
&\leqslant \|\varphi-\varphi_m\|_w^k\,\mathbb{E}\,|\sum_{l=1}^\infty w_l^{-1}\sum_{j=1}^m(e_l(Z)f_j(W)-[T]_{jl})^2|^{k/2} \\
&\leqslant \|\varphi-\varphi_m\|_w^k m^{k/2}(\pi/\sqrt{6})^k\sup_{j,l\in\mathbb{N}}\mathbb{E}\,|e_l(Z)f_j(W)-[T]_{jl}|^k
\end{aligned}$$

where due to Assumption 1.3 (i) $\sup_{j,l\in\mathbb{N}}\mathbb{V}\mathrm{ar}(e_l(Z)f_j(W)) \leqslant \eta^2$ and due to Assumption 1.3 (ii) it holds $\sup_{j,l\in\mathbb{N}}\mathbb{E}\,|e_l(Z)f_j(W)-[T]_{jl}|^k \leqslant k!\eta^k$ for $k \geqslant 3$. Moreover, similarly to the proof of (1.31) in Lemma 1.5.2 we conclude $m^{k/2}\|\varphi-\varphi_m\|_w^k \leqslant (m^3\gamma_m^{-1})^{k/2}(2+2Dd)^{k/2}\rho^{k/2}$. Let us denote $\mu_m := \eta\,(1+Dd)\sqrt{6\rho\,m^3\gamma_m^{-1}}$. Consequently, for all $m \geqslant 1$ we have $\mathbb{E}\,|(\varphi(Z)-\varphi_m(Z))\xi_s(W)|^2 \leqslant \mu_m^2$ and

$$\sup_{s\in\mathbb{S}^m}\mathbb{E}\,|(\varphi(Z)-\varphi_m(Z))\xi_s(W)|^k \leqslant \mu_m^k k! \text{ for } k = 3, 4, \dots. \qquad (1.46)$$

Now Bernstein's inequality gives for all $m \geqslant 1$

$$\sup_{s\in\mathbb{S}^m} \mathbb{E}\Big[\Big(\frac{1}{n}\Big|\sum_{i=1}^{n}(\varphi(Z_i)-\varphi_m(Z_i))\xi_s(W_i)\Big|^2 - 8\mu_m^2(1+\log n)\Big)_+\Big]$$
$$\leqslant 2\int_0^\infty \exp\Big(\frac{-(t+\delta)}{8\mu_m^2}\Big)dt + 2\int_0^\infty \exp\Big(\frac{-\sqrt{n(t+\delta)}}{4\mu_m}\Big)dt$$
$$\leqslant 16\mu_m^2\exp(-\log n)+16\mu_m n^{-1/2}\exp\Big(\frac{-\sqrt{n(1+\log n)}}{2}\Big)(4\mu_m+\sqrt{8n\mu_m^2(1+\log n)})$$
$$\leqslant C(\eta,\gamma,\rho,D)n^{-1}$$

and thus, the assertion follows.

$\square$

Lemma 1.5.9. *Let $T \in \mathcal{T}_{d,D}^{\upsilon}$. Then for all $n \geqslant 1$ it holds*

$$P\Big(\bigcup_{m=1}^{M_n^+} \mho_m^c\Big) \leqslant C(h,\upsilon,\eta,D)n^{-4}, \tag{1.47}$$
$$P\Big(\bigcup_{m=1}^{M_n^+} \Omega_m^c\Big) \leqslant C(h,\upsilon,\eta,D)\, n^{-1}. \tag{1.48}$$

Proof. Proof of (1.47). Since $T \in \mathcal{T}_{d,D}^{\upsilon}$ we have $\|[T]_{\underline{m}}^{-1}\|^2 \leqslant D\upsilon_m^{-1}$ and thus, exploiting Lemma 1.5.3 together with the definition of M_n^+ gives

$$n^4 P\Big(\bigcup_{m=1}^{M_n^+}\mho_m^c\Big) \leqslant 2\exp\Big(-\frac{1}{48\eta D}\frac{n\upsilon_{M_n^+}}{(M_n^+)^3} + 3\log M_n^+ + 4\log n\Big) \leqslant C(h,\upsilon,\eta,D).$$

Proof of (1.48). Due to the definition of M_n^+ there exists some $n_0 \geqslant 1$ such that $n \geqslant 4D\upsilon_{M_n^+}^{-1}$ for all $n \geqslant n_0$. Thereby, condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ implies $\max_{1\leqslant m\leqslant M_n^+}\|[T]_{\underline{m}}^{-1}\|^2 \leqslant D\upsilon_{M_n^+}^{-1} \leqslant n/4$ for all $n \geqslant n_0$. This gives $\bigcup_{m=1}^{M_n^+}\Omega_m^c \subset \bigcup_{m=1}^{M_n^+}\mho_m^c$ and inequality (1.48) follows by making use of (1.47). If $n < n_0$ then $nP(\bigcup_{m=1}^{M_n^+}\Omega_m^c) \leqslant n_0$ and the assertion follows since n_0 only depends on h, υ and D. $\square$

Lemma 1.5.10. *Let $T \in \mathcal{T}_{d,D}^{\upsilon}$. Then it holds $M_n^- \leqslant M_n \leqslant M_n^+$ for all $n \geqslant 1$.*

Proof. Consider $M_n^- \leqslant M_n$. If $M_n^- = 1$ or $M_n = M_n^h$ the result is trivial. If $M_n = 1$, then clearly $M_n^- = 1$. It remains to consider $M_n^- > 1$ and $M_n^h > M_n > 1$. Due to $T \in \mathcal{T}_{d,D}^{\upsilon}$ it holds $\|[T]_{\underline{M_n+1}}^{-1}\|^{-2} \geqslant D^{-1}\upsilon_{M_n+1}$ and thus, by the definition of M_n and M_n^-

it is easily seen that

$$\frac{\upsilon_{M_n^-}}{\max\limits_{1\leqslant j\leqslant M_n^-}[h]_j^2(M_n^-)^3} > \frac{4\upsilon_{M_n+1}}{\max\limits_{1\leqslant j\leqslant M_n+1}[h]_j^2(M_n+1)^3},$$

and thus, $M_n+1 > M_n^-$, i.e. $M_n \geqslant M_n^-$. Consider $M_n \leqslant M_n^+$. If $M_n = 1$ or $M_n^+ = M_n^h$ the result is trivial, while otherwise since $\upsilon_m^{-1} \leqslant \|[T]_{\underline{m}}^{-1}\|^2 \sup_{\|E_m\phi\|_\upsilon=1} \|F_m T E_m \phi\|^2 \leqslant D\|[T]_{\underline{m}}^{-1}\|^2$ due to condition $T \in \mathcal{T}_d^\upsilon$ with $d \leqslant D$ and by the definition of M_n and M_n^+ it follows

$$\frac{\upsilon_{M_n}}{\max\limits_{1\leqslant j\leqslant M_n}[h]_j^2 M_n^3} > \frac{4\upsilon_{M_n^++1}}{\max\limits_{1\leqslant j\leqslant M_n^++1}[h]_j^2(M_n^++1)^3}.$$

Thus, $M_n^+ + 1 > M_n$, i.e. $M_n^+ \geqslant M_n$, which completes the proof. $\square$

In the following, we make use of the notation $\sigma_Y^2 := \mathbb{E}[Y^2]$ and $\widehat{\sigma}_Y^2 := n^{-1}\sum_{i=1}^n Y_i^2$. Further, let us introduce the events

$$\mathcal{H} := \Big\{\|Q_m\|\|[T]_{\underline{m}}^{-1}\| \leqslant 1/4 \quad \forall 1 \leqslant m \leqslant (M_n^+ + 1)\Big\}, \tag{1.49}$$

$$\mathcal{G} := \Big\{\sigma_Y^2 \leqslant 2\,\widehat{\sigma}_Y^2 \leqslant 3\,\sigma_Y^2\Big\}, \tag{1.50}$$

$$\mathcal{J} := \Big\{\|[T]_{\underline{m}}^{-1} V_m\|^2 \leqslant \frac{1}{8}(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2) \quad \forall 1 \leqslant m \leqslant M_n^+\Big\}. \tag{1.51}$$

Lemma 1.5.11. *Let $T \in \mathcal{T}_{d,D}^\upsilon$. Then it holds $\mathcal{H} \cap \mathcal{G} \cap \mathcal{J} \subset \mathcal{A}$.*

Proof. For all $1 \leqslant m \leqslant M_n^+$ observe that condition $\|Q_m\|\|[T]_{\underline{m}}^{-1}\| \leqslant 1/4$ yields by the usual Neumann series argument that $\|([I]_{\underline{m}} + Q_m[T]_{\underline{m}}^{-1})^{-1} - [I]_{\underline{m}}\| \leqslant 1/3$. Thus, using the identity $[\widehat{T}]_{\underline{m}}^{-1} = [T]_{\underline{m}}^{-1} - [T]_{\underline{m}}^{-1}(([I]_{\underline{m}} + Q_m[T]_{\underline{m}}^{-1})^{-1} - [I]_{\underline{m}})$ we conclude

$$2\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\| \leqslant 3\|[h]_{\underline{m}}^t[\widehat{T}]_{\underline{m}}^{-1}\| \leqslant 4\|[h]_{\underline{m}}^t[T]_{\underline{m}}^{-1}\|.$$

Similarly, we have $2\|[T]_{\underline{m}}^{-1}v_m\| \leqslant 3\|[\widehat{T}]_{\underline{m}}^{-1}v_m\| \leqslant 4\|[T]_{\underline{m}}^{-1}v_m\|$ for all $v_m \in \mathbb{R}^m$. Thereby, since $[\widehat{T}]_{\underline{m}}^{-1}V_m = [\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}} - [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ we conclude

$$\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 \leqslant (32/9)\|[T]_{\underline{m}}^{-1}V_m\|^2 + 2\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2,$$
$$\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2 \leqslant (32/9)\|[T]_{\underline{m}}^{-1}V_m\|^2 + 2\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2.$$

On $\mathcal{J}$ it holds $\|[T]_{\underline{m}}^{-1}V_m\|^2 \leqslant \frac{1}{8}(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2)$. Thereby, the last two inequalities imply

$$(5/9)(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2) \leqslant \sigma_Y^2 + 2\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2,$$
$$\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2 \leqslant (22/9)\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + (4/9)\sigma_Y^2.$$

On $\mathcal{G}$ it holds $\sigma_Y^2 \leqslant 2\widehat{\sigma}_Y^2 \leqslant 3\sigma_Y^2$ which gives

$$(5/9)(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2) \leqslant (3/2)\widehat{\sigma}_Y^2 + 2\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2,$$
$$\|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2 + \widehat{\sigma}_Y^2 \leqslant (22/9)\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + (10/9)\sigma_Y^2.$$

Combing the last two inequalities we conclude for all $1 \leqslant m \leqslant M_n^+$

$$(5/18)\big(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2\big) \leqslant \|[\widehat{T}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\|^2 + \widehat{\sigma}_Y^2 \leqslant (22/9)\big(\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2\big).$$

Consequently, we have

$$\mathcal{H} \cap \mathcal{G} \cap \mathcal{J} \subset \Big\{4\,\Delta_m \leqslant 9\,\widehat{\Delta}_m \leqslant 16\,\Delta_m \text{ and } 5\,\varsigma_m^2 \leqslant 18\,\widehat{\varsigma}_m^2 \leqslant 44\,\varsigma_m^2 \quad \forall 1 \leqslant m \leqslant M_n^+\Big\}$$

and thus, $\mathcal{H} \cap \mathcal{G} \cap \mathcal{J} \subset \{\mathrm{pen}_m \leqslant \widehat{\mathrm{pen}}_m \leqslant 18\,\mathrm{pen}_m \ \forall 1 \leqslant m \leqslant M_n^+\}$. Moreover, it holds $\mathcal{H} \subset \{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\}$, which can be seen as follows. Consider $\{\widehat{M}_n < M_n^-\}$. In case of $\widehat{M}_n = M_n^h$ or $M_n^- = 1$ clearly $\{\widehat{M}_n < M_n^-\} = \emptyset$. Otherwise by the definition of $\widehat{M}_n$ it holds

$$\{\widehat{M}_n < M_n^-\} = \bigcup_{m=1}^{M_n^- -1} \Big\{\widehat{M}_n = m\Big\} \subset \Big\{\exists 2 \leqslant m \leqslant M_n^- : m^3\|[\widehat{T}]_{\underline{m}}^{-1}\|^2 \max_{1\leqslant j \leqslant m} [h]_j^2 > a_n\Big\}.$$

By the definition of M_n^- and the property $\|[T]_{\underline{m}}^{-1}\|^2 \leqslant D\upsilon_m^{-1}$ there exists $2 \leqslant m \leqslant M_n^-$ such that on $\{\widehat{M}_n < M_n^-\}$ it holds $\|[\widehat{T}]_{\underline{m}}^{-1}\|^2 > 4D\upsilon_m^{-1} \geqslant 4\,\|[T]_{\underline{m}}^{-1}\|^2$ and thereby,

$$\Big\{\widehat{M}_n < M_n^-\Big\} \subset \Big\{\exists 2 \leqslant m \leqslant M_n^- : \|[\widehat{T}]_{\underline{m}}^{-1}\|^2 \geqslant 4\,\|[T]_{\underline{m}}^{-1}\|^2\Big\}. \tag{1.52}$$

Consider $\{\widehat{M}_n > M_n^+\}$. In case of and $\widehat{M}_n = M_n^h$ or $M_n^- = 1$ clearly $\{\widehat{M}_n < M_n^-\} = \emptyset$. Otherwise, condition $T \in \mathcal{T}_d^\upsilon$ with $d \leqslant D$ implies $\upsilon_m^{-1} \leqslant D\|[T]_{\underline{m}}^{-1}\|^2$ as seen in the proof of Lemma 1.5.9. Thereby, we conclude similarly as above

$$\Big\{\widehat{M}_n > M_n^+\Big\} \subset \Big\{\|[T]_{\underline{M_n^+ +1}}^{-1}\|^2 \geqslant 4\|[\widehat{T}]_{\underline{M_n^+ +1}}^{-1}\|^2\Big\}. \tag{1.53}$$

Again applying the Neumann series argument we observe

$$\mathcal{H} \subset \Big\{\forall\, 1 \leqslant m \leqslant (M_n^+ + 1) : 2\|[T]_{\underline{m}}^{-1}\| \leqslant 3\|[\widehat{T}]_{\underline{m}}^{-1}\| \leqslant 4\|[T]_{\underline{m}}^{-1}\|\Big\},$$

which combined with (1.52) and (1.53) yields $\{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\}^c \subset \mathcal{H}^c$ and thus, completes the proof. □

Lemma 1.5.12. *Under the conditions of Theorem 1.4.2 we have for all* $n \geqslant 1$

$$n^4 (M_n^h)^4 P(\mathcal{A}^c) \leqslant \mathcal{C}.$$

Proof. Lemma 1.5.11 implies that $n^4(M_n^h)^4 P(\mathcal{A}^c) \leqslant n^4(M_n^h)^4\{P(\mathcal{H}^c) + P(\mathcal{J}^c) + P(\mathcal{G}^c)\}$. Therefore, the assertion follows if the right hand side is bounded by a constant $\mathcal{C}$, which we prove in the following. Consider $\mathcal{H}$. From condition $T \in \mathcal{T}_{d,D}^{\upsilon}$ and Lemma 1.5.3 we infer

$$n^4(M_n^h)^4 P(\mathcal{H}^c) \leqslant 2\exp\Big(-\frac{1}{128D\eta}\frac{n\upsilon_{M_n^+ +1}}{(M_n^+ + 1)^2} + 3\log(M_n^+ + 1) + 5\log n\Big) \leqslant C(h, \upsilon, \eta, D) \tag{1.54}$$

where the last inequality is due to condition $(M_n^+ + 1)^2 \log n = o(n\upsilon_{M_n^+ + 1})$. Consider $\mathcal{G}$. Due to condition $m^3\gamma_m^{-1} = o(1)$ as $m \to \infty$ and $U \in \mathcal{U}_\sigma^\infty$ we observe $\mathbb{E}[Y^k] \leqslant 2^k(\mathbb{E}[\varphi^k(Z)] + \mathbb{E}[U^k]) \leqslant C(\gamma, \rho, \sigma)\sup_{j\geqslant 1}\mathbb{E}[e_j^k(Z)]$. Thus, assumption $\sup_{j\geqslant 1}\mathbb{E}[e_j^{20}(Z)] \leqslant \eta^{20}$ together with Theorem 2.10 in Petrov [1995] imply

$$\begin{aligned} n^4(M_n^h)^4 P(\mathcal{G}^c) \leqslant n^5 P(|\widehat{\sigma}_Y^2 - \sigma_Y^2| > \sigma_Y^2/2) &\leqslant 1024\,\sigma_Y^{-20} n^5\, \mathbb{E}\,|n^{-1}\sum_{i=1}^n Y_i^2 - \sigma_Y^2|^{10} \\ &\leqslant 1024\,\sigma_Y^{-20}\,\mathbb{E}\,|Y^2 - \sigma_Y^2|^{10} \leqslant C(\gamma, \rho, \sigma, \eta). \end{aligned} \tag{1.55}$$

Consider $\mathcal{J}$. For all $m \geqslant 1$ observe that the centered random variables $(Y_i - \varphi(Z_i))f_j(W_i)$, $1 \leqslant i \leqslant n$, satisfy Cramer's condition (1.46) with $\mu_m = \eta\,(1 + Dd)\sqrt{6\rho\, m^3\gamma_m^{-1}} \leqslant C(\eta, \gamma, \rho, D)$. From (1.31) in Lemma 1.5.2, $\varphi \in \mathcal{F}_\gamma^\rho$, and $P_{U|W} \in \mathcal{U}_\sigma^\infty$ we infer $\|\varphi_m\|_Z^2 + \sigma_Y^2 \leqslant 4(2 + Dd)\rho + 2\sigma^2$. Moreover, it holds $\|[T]_{\underline{m}}^{-1}V_m\|^2 \leqslant D\upsilon_m^{-1}\|V_m\|^2$ by employing condition $T \in \mathcal{T}_{d,D}^{\upsilon}$. Now Bernstein's inequality yields for all $1 \leqslant m \leqslant M_n^+$

$$
\begin{aligned}
n^6 P\Big(\|[T]_{\underline{m}}^{-1} V_m\|^2 > (\|[T]_{\underline{m}}^{-1}[g]_{\underline{m}}\|^2 + \sigma_Y^2)/8\Big)& \\
\leqslant n^6 \sum_{j=1}^{m} P\Big(\Big|\sum_{i=1}^{n}(Y_i - \varphi(Z_i)) f_j(W_i)\Big|^2 > \frac{n^2 \upsilon_m}{8Dm}(\|\varphi_m\|_Z^2 + \sigma_Y^2)\Big)& \\
\leqslant 2\, n^6 m \, \exp\Big(-\frac{n^2 \upsilon_m m^{-1}(\|\varphi_m\|_Z^2 + \sigma_Y^2)}{32 D n \mu_m^2 + 16 \mu_m n \upsilon_m^{1/2} m^{-1/2}(\|\varphi_m\|_Z^2 + \sigma_Y^2)^{1/2}}\Big)& \\
\leqslant 2 \exp\Big(7 \log n - \frac{n\, \upsilon_{M_n^+}\, \sigma_Y^2}{M_n^+\, C(\sigma, \eta, \gamma, \rho, D)}\Big).&
\end{aligned}
$$

Due to the definition of M_n^+ the last estimate implies $n^4 (M_n^h)^4 P(\mathcal{J}^c) \leqslant \mathcal{C}$, which completes the proof. $\square$

2 Goodness-of-Fit Tests based on Series Estimators in Nonparametric Instrumental Regression

2.1 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 φ satisfies

$$Y = \varphi(Z) + U \quad \text{with} \quad \mathbb{E}[U|W] = 0 \tag{2.1}$$

for an error term U. Here, Z contains potentially endogenous entries, that is, $\mathbb{E}[U|Z]$ may not be zero. Model (2.1) 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 φ, while Darolles et al. [2002], Hall and Horowitz [2005], Gagliardini and Scaillet [2012a] 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 mayor challenges. First, identification in model (2.1) requires far stronger assumptions about the instrumental variables than for the parametric case (cf. Newey and Powell [2003]). Second, the accuracy of any estimator of φ can be low, even for large sample sizes. More precisely, Chen and Reiß

[2011] showed that for a large class of joint distributions of (Z, W) only logarithmic rates of convergence can be obtained. The reason for this slow convergence is that model (2.1) 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 φ in model (2.1), the need for statistically justified model simplifications is paramount. We do not face an ill posed inverse problem if a parametric structure of φ or exogeneity of Z can be justified. If these model simplifications are not supported by the data, one might still be interested in whether a smooth solution to model (2.1) exists and if some regressors could be omitted from the structural function φ. These model simplifications have important potential since they might increase the accuracy of estimators of φ or lower the required conditions imposed on the instrumental variables to ensure identification.

In this work we present a new family of goodness-of-fit statistics which allows for several restricted specification tests of the model (2.1). 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 φ. By a withdrawal of regressors which are independent of the instrument, identification in the restricted model might be possible although φ is not identified in the original model (2.1).

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 [2011b] 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 [2011b] tests whether the endogenous part of Z can be omitted from φ. Gagliardini and Scaillet [2007] and Horowitz [2012] develop nonparametric specification tests in an instrumental regression model. We like to emphasize that their test cannot be applied to model (2.1) 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 (2.1) where φ 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.

Our method is also applicable when an additional smoothing step is carried out. It is shown that the asymptotic behavior of our test relies crucially on the behavior of the smoothing operator. In particular, by carrying out additional smoothing the power with respect to local alternatives increases. On the other hand, the class of alternative models over which uniform consistency can be obtained might shrink. In this paper, we give heuristic arguments how to choose the smoothing operator.

2.2 A simple hypothesis test

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

2.2.1 Assumptions and notation.

The model revisited The nonparametric instrumental regression model (2.1) leads to a linear operator equation. To be more precise, let us introduce the conditional expectation operator $T\phi := \mathbb{E}[\phi(Z)|W]$ mapping $\mathcal{L}^2_Z = \{\phi : \mathbb{E}\,|\phi(Z)|^2 < \infty\}$ to $\mathcal{L}^2_W = \{\psi : \mathbb{E}\,|\psi(W)|^2 < \infty\}$. Consequently, model (2.1) can be written as

$$g = T\varphi \tag{2.2}$$

where the function $g := \mathbb{E}[Y|W]$ belongs to $\mathcal{L}^2_W$. Throughout the paper we assume that an iid. n-sample of (Y, Z, W) from the model (2.1) is available.

Assumptions. Our test statistic based on a sequence of approximating functions $\{f_l\}_{l\geqslant 1}$ in $\mathcal{L}^2_W$. Let $\mathcal{W}$ denote the support of W and the marginal density of W by p_W. We assume throughout the paper that $\{f_l\}_{l\geqslant 1}$ are orthonormal on the support of W with respect to the Lebesgue measure ν, that is, $\int_{\mathcal{W}} f_j(w) f_l(w)\nu(dw) = 1$ if $j = l$ and zero otherwise.

Assumption 2.1. *There exist constants $\eta_f, \eta_p \geqslant 1$ such that (i) $\sup_{l\geqslant 1}\Big(\int_{\mathcal{W}} |f_l(s)|^4\nu(ds)\Big) \leqslant \eta_f$ and (ii) $\sup_{w\in\mathcal{W}}\{p_W(w)/\nu(w)\} \leqslant \eta_p$ with ν being strictly positive on $\mathcal{W}$.*

Assumption 2.1 (i) restricts the magnitude of the approximating functions $\{f_j\}_{j\geqslant 1}$ which is necessary for our proof to determine the asymptotic behavior of our test statistic. This assumption holds for sufficiently large η_f if the basis $\{f_l\}_{l\geqslant 1}$ is uniformly bounded, such as trigonometric bases or B-splines that have been orthogonalized. Moreover, Assumption 2.1 (i) is satisfied by Hermite polynomials. Assumption 2.1 (ii) is satisfied if, for instance, p_W/ν 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 2.2. *There exists a constant $\sigma > 0$ such that $\mathbb{E}[U^4|W] \leqslant \sigma^4$.*

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 φ in the model (2.2).

Assumption 2.3. *The conditional expectation operator T is nonsingular.*

Under Assumption 2.3, 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 each null hypothesis considered below hold true even if T is singular. If Assumption 2.3 fails, however, our test has no power against alternative models whose structural function satisfies $\varphi = \varphi_0 + \delta$ with δ 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 the smoothing operator L on $\mathcal{L}^2_W$. In contrast to the unknown conditional expectation operator T, which has to be estimated, the operator L can be chosen by the econometrician. Let L have an eigenvalue decomposition given by $\{\tau_j^{1/2}, f_j\}_{j\geqslant 1}$. We allow in this paper for a wide range of smoothing operators. 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 $\tau = (\tau_j)_{j\geqslant 1}$.

Assumption 2.4. *The weighting sequence τ is positive, nonincreasing, and satisfies $\tau_1 = 1$.*

Assumption 2.4 ensures that the operator L is nonsingular.

Remark 2.2.1. *Horowitz [2006], Blundell and Horowitz [2007], and Horowitz [2011b] 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 $\ell : [0,1]^2 \to \mathbb{R}$. In order to ensure $L\phi \in \mathcal{L}^2[0,1]$ it is sufficient to assume $\int_0^1\int_0^1 |\ell(s,t)|^2 dsdt <$*

∞. Let $\{\tau_j^{1/2}, f_j\}_{j\geqslant 1}$ be the eigenvalue decomposition of L. By Parseval's identity

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

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

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 $\|A\| = (\text{trace}(A^tA))^{1/2}$. The norms on L^2_Z and L^2_W are denoted by $\|\phi\|_Z^2 := \mathbb{E}\,|\phi(Z)|^2$ for $\phi \in L^2_Z$ and $\|\psi\|_W^2 := \mathbb{E}\,|\psi(W)|^2$ for $\psi \in L^2_W$. The $k \times k$ identity matrix is denoted by I_k. For a vector V we write diag(V) for the diagonal matrix with diagonal elements being the values of V. Moreover, $e_{\underline{m}}(Z)$ and $f_{\underline{m}}(W)$ denote random vectors with entries $e_j(Z)$ and $f_j(W)$, $1 \leqslant j \leqslant m$, respectively. For any weighting sequence w we introduce vectors $e^w_{\underline{m}}(Z)$ and $f^w_{\underline{m}}(W)$ with entries $e^w_j(Z) = \sqrt{w_j} e_j(Z)$ and $f^w_j(W) = \sqrt{w_j} f_j(W)$, $1 \leqslant j \leqslant m$. We write $a_n \sim b_n$ when there exist constants $c, c' > 0$ such that $cb_n \leqslant a_n \leqslant c'b_n$ for all sufficiently large n.

2.2.2 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 $\int_{\mathcal{W}} |L(g-T\varphi_0)(w)p_W(w)/\nu(w)|^2 \nu(dw) = 0$ since the Lebesgue measure ν is strictly positive on $\mathcal{W}$. Moreover, since $\{f_j\}_{j\geqslant 1}$ is an orthonormal basis with respect to ν we obtain by Parseval's identity

$$\int_{\mathcal{W}} |L(g - T\varphi_0)(w)p_W(w)/\nu(w)|^2 \nu(dw) = \sum_{j=1}^\infty \mathbb{E}[(g - T\varphi_0)(W) f_j^\tau(W)]^2. \tag{2.3}$$

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

$$S_n := \sum_{j=1}^{m_n} \tau_j \big| n^{-1} \sum_{i=1}^{n} (Y_i - \varphi_0(Z_i)) f_j(W_i) \big|^2. \tag{2.4}$$

We reject the hypothesis H_0 if $n\,S_n$ becomes too large. When no additional smoothing is carried out, that is, L is the identity operator, then $\tau_j = 1$ for all $j \geqslant 1$. To achieve asymptotic normality we need to standardize our test statistic S_n by appropriate mean and variance, which we introduce in the following definition.

Definition 2.2.1. *For all $m \geqslant 1$ let Σ_m be the covariance matrix of the random vector $U f^{\tau}_{\underline{m}}(W)$ with entries $s_{jl} = \mathbb{E}\,[U^2 f^{\tau}_j(W) f^{\tau}_l(W)]$, $1 \leqslant j, l \leqslant m$. Then the trace and the Frobenius norm of Σ_m are respectively denoted by*

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

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 2.2.1. *Let Assumptions 2.1–2.4 hold true. If m_n satisfies*

$$\varsigma_{m_n}^{-1} = o(1) \quad \textit{and} \quad \Big(\sum_{j=1}^{m_n} \tau_j \Big)^3 = o(n) \tag{2.5}$$

then under H_0

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

Remark 2.2.2. *Since $\varsigma_{m_n}^2 \leqslant \eta_p\, \sigma^4 \sum_{j=1}^{m_n} \tau_j$ (cf. proof of Theorem 2.2.2) 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* (2.5) *we see that by choosing a stronger decaying sequence τ the parameter m_n may be chosen larger. From the following theorem we see that if $\sum_{j=1}^{m_n} \tau_j = O(1)$ only $m_n = o(1)$ is required.* $\square$

In the following result, we establish the asymptotic distribution of our test when the sequence of weights τ may have a stronger decay than in Theorem 2.2.1, that is, we consider the case where τ satisfies $\sum_{j=1}^{m_n} \tau_j = O(1)$. This holds, for instance, if the sequence τ 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 Σ be the covariance matrix of the infinite dimensional centered vector $\big(U f^{\tau}_j(W)\big)_{j \geqslant 1}$. The ordered eigenvalues of

Σ are denoted by $(\lambda_j)_{j\geqslant 1}$. Below, we introduce a sequence $\{\chi^2_{1j}\}_{j\geqslant 1}$ of independent random variables that are distributed as chi-square with one degree of freedom.

Theorem 2.2.2. *Let Assumptions 2.1–2.4 hold true. If m_n satisfies*

$$\sum_{j=1}^{m_n} \tau_j = O(1) \quad \textit{and} \quad m_n = o(1) \tag{2.6}$$

then under H_0

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

Remark 2.2.3 (Estimation of Critical Values)**.** *The asymptotic results of Theorem 2.2.1 and 2.2.2 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 $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$. Moreover, $\mathbf{U}_n = (Y_1 - \varphi_0(Z_1), \dots, Y_n - \varphi_0(Z_n))^t$. In the setting of Theorem 2.2.1, we replace Σ_m by*

$$\widehat{\Sigma}_m := \mathbf{W}_m(\tau)^t \, diag(\mathbf{U}_n)^2 \, \mathbf{W}_m(\tau)$$

Now the asymptotic result of Theorem 2.2.1 continues to hold if we replace ς_{m_n} by the Frobenius norm of $\widehat{\Sigma}_{m_n}$ and μ_{m_n} by the trace of $\widehat{\Sigma}_{m_n}$. In the setting of Theorem 2.2.2, the asymptotic distribution is not pivotal and has to approximated. First, note that the difference of critical values between $\sum_{j=1}^\infty \lambda_j \chi^2_{1j}$ and the truncated sum $\sum_{j=1}^M \lambda_j\, \chi^2_{1j}$ can be made arbitrarily small by choosing the integer $M > 0$ sufficiently large (cf. Horowitz [2006]). Second, replace $(\lambda_j)_{1\leqslant j\leqslant M}$ by $(\widehat{\lambda}_j)_{1\leqslant j\leqslant M}$ which are the ordered eigenvalues of $\widehat{\Sigma}_M$. Observe that $\max_{1\leqslant j\leqslant M} |\widehat{\lambda}_j - \lambda_j| = \|\widehat{\Sigma}_M - \Sigma_M\| = O(n^{-1/2})$ almost surely and hence the critical values of $\sum_{j=1}^M \widehat{\lambda}_j\, \chi^2_{1j}$ converge in probability to the ones of the limiting distribution of $n\, S_n$. □

2.2.3 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 $n \to \infty$. It is shown that the power of our tests essentially relies on the choice of the weighting sequence τ.

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

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

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 2.2.3. *Given the conditions of Theorem 2.2.1 it holds under* (2.7)

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

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

$$Y = \varphi_0(Z) + n^{-1/2}\delta(Z) + U \tag{2.8}$$

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

Proposition 2.2.4. *Given the conditions of Theorem 2.2.2 it holds under* (2.8)

$$n\,S_n \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j\,\chi_{1j}^2(\delta_j/\lambda_j) \quad \text{as } n \to \infty.$$

Remark 2.2.4. *We see from Proposition 2.2.3 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 $L = Id$ 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.* □

2.2.4 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 2.5. *(i) The function p_W/ν is uniformly bounded away from zero. (ii) There exists a constant $\sigma_o > 0$ such that $\mathbb{E}[U^2|W] \geqslant \sigma_o^2$.*

Assumption 2.5 (i) implies that $\|LT(\varphi - \varphi_0)\|_W > 0$ for any structural function φ in the alternative. Further, Assumption 2.5 implies that $\sum_{j=1}^{m_n} \tau_j^2 = O(\varsigma_{m_n}^2)$.

Proposition 2.2.5. *Assume that H_0 does not hold. Let $\mathbb{E}\,|Y - \varphi_0(Z)|^4 < \infty$ and let Assumption 2.5 (i) hold true. Consider the sequence $(\alpha_n)_{n\geqslant 1}$ satisfying $\alpha_n = o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 2.2.1 we have*

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

Under the conditions of Theorem 2.2.2 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

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

where $C > 0$ is a finite constant. Clearly, if H_0 is false then $\|LT(\varphi - \varphi_0)\|_W^2 \geqslant \rho\,\varsigma_{m_n} n^{-1}$ for all sufficiently large n and some $\rho > 0$. By Assumption 2.4 the sequence τ is nonincreasing sequence with $\tau_1 = 1$ and hence, $\|LT(\varphi - \varphi_0)\|_W^2 \leqslant \|T(\varphi - \varphi_0)\|_W^2 \leqslant \|\varphi - \varphi_0\|_Z^2$ by Jensen's inequality. We conclude that $\mathcal{G}_n^\rho$ contains all alternative functions whose $\mathcal{L}_Z^2$-distance to the structural function φ_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 φ does not belong to $\mathcal{G}_n^\rho$ if the decay of τ 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 φ does not necessarily belong to $\mathcal{G}_n^\rho$ with τ 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 2.2.6. *Let Assumption 2.5 be satisfied. For any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ we have under the conditions of Theorem 2.2.1 that*

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

while under the conditions of Theorem 2.2.2

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

2.3 A parametric specification test

In this section, we present a test whether the structural function φ is known up to a finite dimensional parameter. Let Θ 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 ϕ. The alternative hypothesis is that there exists no $\vartheta \in \Theta$ such that $\varphi(\cdot) = \phi(\cdot, \vartheta)$ holds true.

2.3.1 The test statistic and its asymptotic distribution

Under Assumptions 2.3 and 2.4, 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 (2.4) where φ_0 is replaced by $\phi(\cdot, \widehat{\vartheta}_n)$ with $\widehat{\vartheta}_n$ being an estimator of ϑ. Hence, our test statistic for a parametric specification is given by

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

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 ϕ with respect to its second argument. Below we denote the vector of partial derivatives of ϕ with respect to $\vartheta = (\vartheta_1, \dots, \vartheta_k)^t$ by $\phi_\vartheta = (\phi_{\vartheta_l})_{1\leqslant l\leqslant k}$ and the matrix of second-order partial derivatives by $\phi_{\vartheta\vartheta} = (\phi_{\vartheta_j\vartheta_l})_{1\leqslant j,l\leqslant k}$.

Assumption 2.6. *(i) Let $\widehat{\vartheta}_n$ be an estimator satisfying $\|\widehat{\vartheta}_n - \vartheta_0\| = O_p(n^{-1/2})$ for some $\vartheta_0 \in int(\Theta)$ with $\varphi(\cdot) = \phi(\cdot, \vartheta_0)$ if H_p holds true. (ii) The function ϕ is twice partial differentiable with respect to its second argument. There exists some constant $\eta_\phi \geqslant 1$ such* *that*

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

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

Theorem 2.3.1. *Let Assumptions 2.1–2.4 and 2.6 hold true. If m_n satisfies* (2.5)*, 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

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) = n^{-1/2}\sum_{i=1}^{n} h_{\underline{k}}(V_i) + o_p(1) \tag{2.9}$$

where $V_i := (Y_i, Z_i, W_i, \varphi_0)$ and $h_{\underline{k}}(V_i) = (h_1(V_i),\dots,h_k(V_i))^t$ where h_j, $1\leqslant j\leqslant k$, are real valued functions. It is well known that this representation holds if $\widehat{\vartheta}_n$ is the generalized method of moments estimator. In case of $\sum_{j=1}^{m_n}\tau_j = O(1)$ we have to modify the standardization of the statistic S_n^p as follows. Let Σ^p be the covariance matrix of the infinite dimensional centered vector $\big(Uf_j^\tau(W) - \mathbb{E}[f_j^\tau(W)\phi_\vartheta(Z,\vartheta_0)^t]h_{\underline{k}}(V)\big)_{j\geqslant1}$. The ordered eigenvalues of Σ^p are denoted by $(\lambda_j^p)_{j\geqslant1}$.

Theorem 2.3.2. *Let Assumptions2.1–2.4 and 2.6 hold true. Assume that H_p holds true and $\widehat{\vartheta}_n$ satisfies condition* (2.9) *with $\mathbb{E}\,h_j(V)=0$ and $\mathbb{E}\,|h_j(V)|^4<\infty$, $1\leqslant j\leqslant k$. If m_n satisfies* (2.6)*, then*

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

Remark 2.3.1. *[Estimation of Critical Values] For the estimation of critical values of Theorem 2.3.1 and 2.3.2, let us define $\mathbf{U}_n^p = \big(Y_1-\phi(Z_1,\widehat{\vartheta}_n),\dots,Y_n-\phi(Z_n,\widehat{\vartheta}_n)\big)^t$. We estimate the covariance matrix Σ_m by*

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

Now the asymptotic result of Theorem 2.3.1 continues to hold if we replace ς_{m_n} by the Frobenius norm of $\widehat{\Sigma}_{m_n}$ and μ_{m_n} by the trace of $\widehat{\Sigma}_{m_n}$. In the setting of Theorem 2.3.2, *we replace Σ^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 $1 \leqslant i \leqslant n$, $1 \leqslant l \leqslant k$ and $\mathbf{h}_k(V) = \big(h_{\underline{k}}(V_1), \dots, h_{\underline{k}}(V_n)\big)^t$. Then define $\mathbf{V}_k := n^{-1}\mathbf{h}_k(V)\mathbf{A}_k^t$. Given a sufficiently large integer $M > 0$ we estimate Σ^p by*

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

Hence, we approximate $\sum_{j=1}^\infty \lambda_j \chi_{1j}^2$ by the finite sum $\sum_{j=1}^M \widehat{\lambda}_j^p\, \chi_{1j}^2$ where $(\widehat{\lambda}_j^p)_{1\leqslant j\leqslant M}$ are the ordered eigenvalues of $\widehat{\Sigma}_M^p$. We have $\max_{1\leqslant j\leqslant M} |\widehat{\lambda}_j^p - \lambda_j^p| = O_p(n^{-1/2})$. □

2.3.2 Limiting behavior under local alternatives and consistency.

In the following, we study the power and consistency properties of the test statistic S_n^{p}. For the next result, we follow Härdle and Mammen [1993] and consider a sequence of linear local alternatives (2.7) or (2.8) with $\varphi_0 = \phi(\vartheta_0, \cdot)$ where δ is orthogonal to the class of parametric functions $\{\phi(\cdot, \vartheta_0) : \vartheta_0 \in \Theta\}$.

Proposition 2.3.3. *Let the conditions of Theorem 2.3.1 be satisfied. Then under* (2.7) *with $\varphi_0 = \phi(\vartheta_0, \cdot)$ and $\mathbb{E}[\phi_\vartheta(\vartheta_0, Z)\delta(Z)] = 0$ it holds*

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

Let the conditions of Theorem 2.3.2 be satisfied. Then under (2.8) *with $\varphi_0 = \phi(\vartheta_0, \cdot)$ and $\mathbb{E}[\phi_\vartheta(\vartheta_0, Z)\delta(Z)] = 0$ it holds*

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

Remark 2.3.2. *Under homoscedasticity, that is, $\mathbb{E}[U^2|W] = \sigma_o^2$, $W \sim \mathcal{U}[0,1]$, and $L = Id$ we see from Proposition 2.3.3 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].* □

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, ϑ_0 denotes the probability limit of the estimator $\widehat{\vartheta}_n$.

Proposition 2.3.4. *Assume that H_p does not hold. Let $\mathbb{E}\,|Y - \phi(Z, \vartheta_0)|^4 < \infty$ and Assumption 2.5 (i) hold true. Let $(\alpha_n)_{n\geqslant 1}$ as in Proposition 2.2.5. Under the conditions of*

Theorem 2.3.1 we have

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

Given the conditions of Theorem 2.3.2 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

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

for some constant $C > 0$ and ϑ_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 2.3.5. *Let Assumption 2.5 be satisfied. For any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ we have under the conditions of Theorem 2.3.1 that*

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

whereas under the conditions of Theorem 2.3.2 it holds

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

2.4 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 $\varphi \neq \varphi_0$.

2.4.1 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 $\{e_j\}_{j\geqslant 1}$ belonging to $\mathcal{L}^2_Z$. Further, let $\mathbf{Z}_k$ denote a $n \times k$ matrix with entries $e_j(Z_i)$ for $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant k$. Moreover, let $\mathbf{Y}_n = (Y_1, \dots, Y_n)^t$. Then we define the estimator

$$\overline{\varphi}_k(\cdot) := e_{\underline{k}}(\cdot)^t \widehat{\beta}_k \quad \text{where} \quad \widehat{\beta}_k = (\mathbf{Z}_k^t \mathbf{Z}_k)^- \mathbf{Z}_k^t \mathbf{Y}_n. \tag{2.10}$$

In contrast to the parametric case we need to allow for k tending to infinity as $n \to \infty$ in order to ensure consistency of the estimator $\overline{\varphi}_k$. 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 $\overline{\varphi}_k$ was studied, for example, by Newey [1997].

Under Assumptions 2.3 and 2.4, 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 (2.4) but where φ_0 is replaced by the series estimator $\overline{\varphi}_{k_n}$. The proposed test statistic for H_e is now given by

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

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

For controlling the bias of the estimator $\overline{\varphi}_{k_n}$ we specify in the following a rate of approximation (cf. Newey [1997]). Let $\gamma = (\gamma_j)_{j\geqslant 1}$ be a nondecreasing sequence with $\gamma_1 = 1$. We assume that φ_0 belongs to

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

Here, the sequence of weights γ measures the approximation error of φ_0 with respect to the functions $\{e_j\}_{j\geqslant 1}$.

Assumption 2.7. *(i) Let $\varphi_0 \in \mathcal{F}_\gamma$ with nondecreasing sequence γ satisfying $j^2 = o(\gamma_j)$. (ii) There exists some constant $\eta_e \geqslant 1$ such that $\sup_{z\in\mathcal{Z}} \|e_{\underline{k_n}}(z)\|^2 \leqslant \eta_e k_n$. (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 2.7 (i) determines the required asymptotic behavior of the rate γ. For splines and power series this assumption is satisfied if the number of continuous derivatives of φ_0 divided by the dimension of Z equals two. Assumption 2.7 (ii) and (iii) restrict the magnitude of the approximating functions $\{e_j\}_{j\geqslant 1}$ and impose nonsingularity of their second moment matrix.

We are now in the position to proof 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 2.4.1. *Let Assumptions 2.1–2.4 and 2.7 be satisfied. If*

$$n = o(\gamma_{k_n}\varsigma_{m_n}),\ k_n = o(\varsigma_{m_n}), \quad \textit{and} \quad \Big(\sum_{j=1}^{m_n} \tau_j\Big)^3 = o(n) \tag{2.11}$$

then under H_{e} it holds

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

Example 2.4.1. *Let Z be continuously distributed with $\dim(Z) = r$ and set $L = Id$. 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 2.5 hold true then $\sqrt{m_n} = O(\varsigma_{m_n})$. Hence, condition (2.11) is satisfied if $k_n \sim n^\kappa$ with*

$$r(1-\nu/2)/(2p) < \kappa < \nu/2. \tag{2.12}$$

Note that condition (2.12) requires $2p > r\,(2/\nu - 1)$. Hence, with a larger dimension r also the smoothness of φ_0 has to increase, reflecting the curse of dimensionality. □

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 Σ^{e} be the covariance matrix of the infinite dimensional centered vector $\big(U(f_j^\tau(W) - \sum_{l\geqslant 1}\mathbb{E}[f_j^\tau(W)e_l(Z)]e_l(Z))\big)_{j\geqslant 1}$. The ordered eigenvalues of Σ^{e} are denoted by $(\lambda_j^{\mathrm{e}})_{j\geqslant 1}$.

Theorem 2.4.2. *Let Assumptions 2.1–2.4 and 2.7 be satisfied. If*

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad n = O(\gamma_{k_n}), \quad k_n^3 = o(n),\ \textit{and} \quad m_n = o(1) \tag{2.13}$$

then under H_{e} it holds

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

Example 2.4.2. *Consider the setting of Example 2.4.1 but where the eigenvalues of L satisfy $\tau_j \sim j^{-2}$. Condition* (2.13) *is satisfied if $m_n \sim n^{\nu}$ for some $\nu > 0$ and $k_n \sim n^{\kappa}$ with $r/(2p) < \kappa < 1/3$. The estimator of φ_0 is undersmoothed. This ensures that the bias of this estimator in the statistic S_n^{e} is asymptotically negligible. Here, the required smoothness of φ_0 is $p > 3r/2$.* □

Remark 2.4.1. *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 φ_0 to the smoothness of the joint density of (Z, W).* □

Remark 2.4.2 (Estimation of Critical Values)**.** *For the estimation of critical values of Theorem 2.4.1 and 2.4.2, let us define $\mathbf{U}_n^{\mathrm{e}} = \big(Y_1 - \overline{\varphi}_{k_n}(Z_1), \dots, Y_n - \overline{\varphi}_{k_n}(Z_n)\big)^t$. For any $m \geqslant 1$ we estimate the covariance matrix $\varSigma_m$ by*

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

Now the asymptotic result of Theorem 2.4.1 continues to hold if we replace ς_{m_n} by the Frobenius norm of $\widehat{\varSigma}_{m_n}$ and μ_{m_n} by the trace of $\widehat{\varSigma}_{m_n}$. This consistency is shown in Lemma 2.4.3. In the setting of Theorem 2.4.2, we replace $\varSigma^{\mathrm{e}}$ by a finite dimensional matrix

$$\widehat{\varSigma}_M^{\mathrm{e}} := n^{-1}\mathbf{W}_M(\tau)^t \Big(I_n - n^{-1}\mathbf{Z}_{k_n}\mathbf{Z}_{k_n}^t\Big)\, diag(\mathbf{U}_n^{\mathrm{e}})^2 \, \Big(I_n - n^{-1}\mathbf{Z}_{k_n}\mathbf{Z}_{k_n}^t\Big)\mathbf{W}_M(\tau)$$

where $M > 0$ is a sufficiently large integer. Let $(\widehat{\lambda}_j^{\mathrm{e}})_{1\leqslant j \leqslant M}$ denote the ordered eigenvalues of $\widehat{\varSigma}_M^{\mathrm{e}}$. Hence, we approximate $\sum_{j=1}^{\infty} \lambda_j^{\mathrm{e}} \chi_{1j}^2$ by the finite sum $\sum_{j=1}^{M} \widehat{\lambda}_j^{\mathrm{e}}\, \chi_{1j}^2$ where $\max_{1\leqslant j\leqslant M} |\widehat{\lambda}_j^{\mathrm{e}} - \lambda_j^{\mathrm{e}}| = o_p(1)$. □

Lemma 2.4.3. *Consider $\widehat{\varSigma}_{m_n}$ as defined in Remark 2.4.2. Under conditions of Theorem 2.4.1 or Theorem 2.4.2 the difference of its Frobenius norm to ς_{m_n} and its trace to μ_{m_n} converge in probability to zero.*

2.4.2 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 of S_n^{e} against a sequence of local alternatives we proceed similarly as Ait-Sahalia et al. [2001]. More precisely, given a sequence of functions $(\varphi^{[n]})_{n\geqslant 1}$ where $\varphi^{[n]}$ satisfies $\|\varphi^{[n]}-\varphi_0\|_Z^2 = O(\gamma_{k_n}^{-1}, n^{-1}k_n)$. Then we consider alternative models $Y = \varphi(Z) + U$ with $\mathbb{E}[U|W] = 0$ where the structural function φ satisfies

$$\sqrt{n}\sup_{z\in\mathcal{Z}} |\varphi(z) - \varphi^{[n]}(z) - \sqrt{n}\varsigma_{m_n}^{-1}\delta(z)| = o(\varsigma_{m_n}) \tag{2.14}$$

with $\delta \in \mathcal{L}_Z^4$.

Proposition 2.4.4. *Given the conditions of Theorem 2.4.1 and Assumption 2.5 (ii) it holds under* (2.14)

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

Given the conditions of Theorem 2.4.2 it holds under (2.14) *with* ς_{m_n} *replaced by* 1 *we have*

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

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

Proposition 2.4.5. *Assume that* H_{e} *does not hold. Let* $\mathbb{E}\,|Y - \varphi_0(Z)|^4 < \infty$ *and Assumption 2.5 (i) hold true. Let* $(\alpha_n)_{n\geqslant 1}$ *as in Proposition 2.2.5. Under the conditions of Theorem 2.4.1 we have*

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

whereas in the setting of Theorem 2.4.2

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

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

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

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 $\sum_{j=1}^{\infty} \lambda_j^{\mathrm{e}}\, \chi_{1j}^2$, respectively.

Proposition 2.4.6. *Let Assumption 2.5 be satisfied. Under the conditions of Theorem 2.4.1 we have for any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ that*

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

whereas under the conditions of Theorem 2.4.2 it holds

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

2.5 A nonparametric specification test

A solution to the linear operator equation (2.2) only exists if g belongs to the range of T. This might be violated if, for instance, the instrument is not valid, that is, $\mathbb{E}[U|W] \neq 0$. We consider the hypothesis H_{np}: there exists a solution φ_0 to (2.2). The alternative hypothesis is that there exists no solution to (2.2). In addition, we see in this section that our results allow for a test of dimension reduction of the vector of regressors Z, that is, whether some regressors can be omitted from the structural function φ_0.

2.5.1 Nonparametric estimation method

The nonparametric estimator. In the following, we derive an estimator of φ_0 under the null hypothesis H_{np}. For simplicity, assume that $\mathcal{Z} = \mathcal{W}$ and consider a sequence $\{e_j\}_{j\geqslant 1}$ of approximating functions which are orthonormal on $\mathcal{Z}$ with respect to the Lebesque measure ν. Under conditions given below, φ_0 has the expansion $\varphi_0(\cdot) = \sum_{l=1}^\infty \int_{\mathcal{Z}} \varphi_0(z)e_l(z)\nu(dz)\,e_l(\cdot)$. Thereby, the conditional moment restriction under H_{np} leads to the following unconditional moment restrictions

$$\mathbb{E}[Ye_j(W)] = \sum_{l=1}^{\infty} \mathbb{E}[e_j(W)e_l(Z)] \int_{\mathcal{Z}} \varphi_0(z)e_l(z)\nu(dz) \tag{2.15}$$

for $j \geqslant 1$. 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 $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant k$. Then for each $k \geqslant 1$ we consider the estimator

$$\widehat{\varphi}_k(\cdot) := e_{\underline{k}}(\cdot)^t\widehat{\beta}_k \quad \text{where} \quad \widehat{\beta}_k = (\mathbf{X}_k^t\mathbf{Z}_k)^-\,\mathbf{X}_k^t\mathbf{Y}_n. \tag{2.16}$$

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 (2.16) was studied by Johannes and Schwarz [2010], Horowitz [2011b], and Horowitz [2012].

Additional assumptions. As noted by Horowitz [2012], uniformly consistent testing of H_{np} is only possible if the null is restricted that any solution to (2.2) is smooth. Hence, in the following we assume that φ_0 belongs to the ellipsoid $\mathcal{F}_\gamma^\rho := \{\phi \in \mathcal{L}_Z^2 : \sum_{j=1}^\infty \gamma_j \mathbb{E}[\phi(Z)e_j(Z)]^2 \leqslant \rho\}$. As in the previous section, $\gamma = (\gamma_j)_{j\geqslant 1}$ measures the approximation error of φ_0 with respect to the basis $\{e_j\}_{j\geqslant 1}$.

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 on $\mathcal{L}_W^2$. Given a sequence of weights $\upsilon := (\upsilon_j)_{j\geqslant 1}$ and $d \geqslant 1$ we define the subset $\mathcal{T}_d^\upsilon$ of $\mathcal{T}$ by

$$\mathcal{T}_d^\upsilon := \Big\{T \in \mathcal{T} : \quad \int_{\mathcal{W}} |(T\phi)(w)|^2 \nu(dw) \leqslant d \sum_{j=1}^\infty \upsilon_j \Big(\int_{\mathcal{Z}} \phi(z) e_j(z) \nu(dz)\Big)^2 \quad \text{for all } \phi \in \mathcal{L}_Z^2\Big\}.$$

If p_Z/ν is bounded from above and p_W/ν is uniformly bounded away from zero then the conditional expectation operator T belongs to $\mathcal{T}_d^\upsilon$ with $\upsilon_j = 1$, $j \geqslant 1$, due to Jensen's inequality. Notice that for all $T \in \mathcal{T}_d^\upsilon$ it follows that $\|Te_j\|_W^2 \leqslant d\,\eta_p \upsilon_j$ and thereby, the condition $T \in \mathcal{T}_d^\upsilon$ links the operator T to the basis $\{e_j\}_{j\geqslant 1}$. 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 $\{e_l\}_{l\geqslant 1}$. We denote by $\mathcal{T}_{d,D}^\upsilon$ for some $D \geqslant d$ the subset of $\mathcal{T}_d^\upsilon$ given by

$$\mathcal{T}_{d,D}^\upsilon := \Big\{T \in \mathcal{T}_d^\upsilon : [T]_{\underline{k}} \text{ is nonsingular and } \quad \sup_{k\geqslant 1} \|\operatorname{diag}(\upsilon_1, \dots, \upsilon_k)^{1/2} [T]_{\underline{k}}^{-1}\|^2 \leqslant D\Big\}.$$

The class $\mathcal{T}_{d,D}^\upsilon$ only contains operators T whose off-diagonal elements of $[T]_{\underline{k}}^{-1}$ are sufficiently small for all $k \geqslant 1$. A similar diagonality restriction has been used by Hall and Horowitz [2005]. 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 γ and υ.

Assumption 2.8. *(i) Let $\varphi_0 \in \mathcal{F}_\gamma^\rho$ with nondecreasing sequence γ satisfying $j^3 = o(\gamma_j)$. (ii) The sequence $\{e_j\}_{j\geqslant 1}$ is an orthogonal basis on $\mathcal{Z} = \mathcal{W}$ with respect to ν. (iii) There exists some constant $\eta_e \geqslant 1$ such that $\sup_{j\geqslant 1} \sup_{z\in\mathcal{Z}} |e_j(z)| \leqslant \eta_e$. (iv) Let $T \in \mathcal{T}_{d,D}^\upsilon$ with υ being a strictly positive sequences such that υ and $(\upsilon_j/\tau_j)_{j\geqslant 1}$ are nonincreasing. (v) The density p_Z/ν is bounded from above and p_W/ν is uniformly bounded away from zero.*

Note that by Assumption 2.8 (i) the alternative hypothesis is that there exists no function in $\mathcal{F}_\gamma^\rho$ solving (2.2). Due to Assumption 2.8 (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 $\widehat{\varphi}_{k_n}$ attains the optimal rate of convergence $\mathcal{R}_n := \max\big(\gamma_{k_n}^{-1}, \sum_{j=1}^{k_n}(n\upsilon_j)^{-1}\big)$. Due to Assumption 2.8 (v) we do not require orthonormal bases with respect to the unknown distribution (Z, W) (cf. Remark 1.3.2).

2.5.2 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 φ_0 is replaced by the nonparametric estimator $\widehat{\varphi}_{k_n}$ given in (2.16), that is,

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

If S_n^{np} becomes too large then there exists no function in $\mathcal{F}_\gamma^\rho$ solving (2.2). 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 2.5.1. *Let Assumptions 2.1–2.4 and 2.8 be satisfied. If*

$$n\upsilon_{k_n} = o(\gamma_{k_n}\varsigma_{m_n}),\ k_n = o(\varsigma_{m_n}),\ k_n\Big(\sum_{j=1}^{m_n}\tau_j\Big)^2 = O(n\upsilon_{k_n}),\ \textit{and}\ \Big(\sum_{j=1}^{m_n}\tau_j\Big)^3 = o(n) \quad (2.17)$$

then it holds under H_{np}

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

Example 2.5.1. *Consider the setting of Example 2.4.1. In the mildly ill posed case where $\upsilon_j \sim j^{-2a/r}$ for some $a \geqslant 0$ condition* (2.17) *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, $\upsilon_j \sim \exp(-j^{2a/r})$ for some $a > 0$, condition (2.17) *is satisfied if, for example, m_n satisfies $m_n = o(k_n^p)$ and $k_n = o(\sqrt{m_n})$ where $k_n \sim \big(\log n - \log(m_n^{3/2})\big)^{r/(2a)}$.* □

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

Theorem 2.5.2. *Let Assumptions 2.1–2.4 and 2.8 be satisfied. If*

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad n\upsilon_{k_n} = o(\gamma_{k_n}), \quad k_n^3 = o(n\upsilon_{k_n}), \text{ and } m_n = o(1) \tag{2.18}$$

then it holds under H_{np}

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

Example 2.5.2. *Consider the setting of Example 2.4.2. In the mildly ill posed case, that is, $\upsilon_j \sim j^{-2a/r}$ for some $a \geqslant 0$, condition* (2.18) *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, $\upsilon_j \sim \exp(-j^{2a/r})$ for some $a > 0$, condition (2.18) *is satisfied if $k_n \sim \big(\log(n^{1+\varepsilon})\big)^{r/(2a)}$ for any $\varepsilon > 0$. In both cases, we observe that the estimator $\widehat{\varphi}_{k_n}$ is undersmoothed.* □

Remark 2.5.1. *If the basis $\{e_j\}_{j\geqslant 1}$ coincides with $\{f_j\}_{j\geqslant 1}$ then $n\, S_n^{np}$ 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.* □

Remark 2.5.2. *Let Z' be a vector containing only entries of Z with $\dim(Z') < \dim(Z)$. It is easy to generalize our previous result for a test of H'_{np}: there exists a solution $\varphi_0 \in \mathcal{F}_\gamma^\rho$ to* (2.2) *only depending on Z'. To be more precise consider the test statistic*

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

where $\widehat{\varphi}_{k_n}$ is the estimator (2.16) *based on an iid. sample $(Y_1, Z_1', W_1), \ldots, (Y_n, Z_n', W_n)$ of (Y, Z', W). Under H'_{np} we consider the conditional expectation operator $T' : \mathcal{L}_{Z'}^2 \to \mathcal{L}_W^2$*

with $(T'\phi)(W) := \mathbb{E}[\phi(Z')|W]$*. It is interesting to note that if* T *is nonsingular then also* T' *is. Hence, for a test of* H'_{np} *we may replace Assumption 2.3 by the weaker condition that* T' *is nonsingular. Moreover, under* H'_{np} *the results of Theorem 2.5.1 and 2.5.2 still hold true if we replace* Z *by* Z'*.* □

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 φ_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 φ in the restricted model

Example 2.5.3. *Let* $Z = (Z^{(1)}, Z^{(2)})$ *where both,* $Z^{(1)}$ *and* $Z^{(2)}$ *are endogenous vectors of regressors. But only* $Z^{(1)}$ *satisfies a sufficiently strong relationship with the instrument* W *in the sense that for all* $\phi \in \mathcal{L}^2_{Z^{(1)}}$ *condition* $\mathbb{E}[\phi(Z^{(1)})|W] = 0$ *implies* $\phi = 0$*. In this example, we do not assume that this completeness condition is fulfilled for the joint distribution of* $(Z^{(2)}, W)$*. Thereby only the operator* $T^{(1)} : \mathcal{L}^2_{Z^{(1)}} \to \mathcal{L}^2_W$ *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* φ_0 *then we obtain identification in the restricted model.* □

Remark 2.5.3. *[Estimation of Critical Values] For the estimation of critical values of Theorem 2.5.1 and 2.5.2, let us define* $\mathbf{U}_n^{np} = \big(Y_1 - \widehat{\varphi}_{k_n}(Z_1), \dots, Y_n - \widehat{\varphi}_{k_n}(Z_n)\big)^t$*. For all* $m \geqslant 1$*, we estimate the covariance matrix* Σ_m *by*

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

Now the asymptotic result of Theorem 2.5.1 continues to hold if we replace ς_{m_n} *by the Frobenius norm of* $\widehat{\Sigma}_{m_n}$ *and* μ_{m_n} *by the trace of* $\widehat{\Sigma}_{m_n}$ *(this is easily seen from the proof of Lemma 2.4.3 assuming that* $\{f_j\}_{j\geqslant 1}$ *is uniformly bounded). In the setting of Theorem 2.5.2, we replace* Σ^{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* $k \geqslant 1$*. Then for a sufficiently large integer* $M > 0$ *we estimate* Σ^{np} *by*

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

Hence, we approximate $\sum_{j=1}^{\infty} \lambda_j^{np}\chi_{1j}^2$ *by the finite sum* $\sum_{j=1}^{M} \widehat{\lambda}_j^{np}\,\chi_{1j}^2$ *where* $(\widehat{\lambda}_j^{np})_{1\leqslant j\leqslant M}$ *are the ordered eigenvalues of* $\widehat{\Sigma}_M^{np}$ *where* $\max_{1\leqslant j\leqslant M} |\widehat{\lambda}_j^{np} - \lambda_j^{np}| = o_p(1)$*.* □

2.5.3 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 functions which solve the operator equation (2.2) but do belong to $\mathcal{F}_\gamma^\rho$. More precisely, given a sequence of functions $(\varphi^{[n]})_{n\geqslant 1}$ where $\varphi^{[n]}$ satisfies $\|\varphi^{[n]}-\varphi_0\|_Z^2 = \max\big(\gamma_{k_n}^{-1}, \sum_{j=1}^{k_n}(nv_j)^{-1}\big)$. Then we consider alternative models $Y=\varphi(Z)+U$ with $\mathbb{E}[U|W]=0$ where the structural function φ satisfies

$$\sqrt{n}\sup_{z\in\mathcal{Z}}\big|\varphi(z)-\varphi^{[n]}(z)-\sqrt{n}\varsigma_{m_n}^{-1}\delta(z)\big| = o_p(\varsigma_{m_n}) \tag{2.19}$$

for some $\delta\in\mathcal{L}_Z^4$.

Proposition 2.5.3. *Let Assumption 2.5 (ii) hold true. Given the conditions of Proposition 2.5.1 it holds under* (2.7)

$$(\sqrt{2}\varsigma_{m_n})^{-1}(n\,S_n^{np}-\mu_{m_n})\xrightarrow{d}\mathcal{N}\Big(2^{-1/2}\sum_{j=1}^\infty\delta_j^2,1\Big)\quad \textit{as } n\to\infty.$$

Given the conditions of Proposition 2.5.2 it holds under (2.8)

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

In the next proposition, we establish consistency of our test when H_{np} does not hold, that is, there exists no function in $\mathcal{F}_\gamma^\rho$ that solves (2.2) for any sequence γ satisfying Assumption 2.8 and any sufficiently large constant $0<\rho<\infty$.

Proposition 2.5.4. *Assume that H_{np} does not hold. Let $\mathbb{E}\,|Y-\varphi_0(Z)|^4<\infty$ and Assumption 2.5 (i) hold true. Let $(\alpha_n)_{n\geqslant 1}$ as in Proposition 2.2.5. Under the conditions of Theorem 2.5.1 and 2.5.2, respectively, we have*

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

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

$$\mathcal{J}_n^\rho=\Big\{\varphi\in\mathcal{L}_Z^2:\inf_{\varphi_0\in\mathcal{F}_\gamma^\rho}\|LT(\varphi-\varphi_0)\|_W^2\geqslant\rho n^{-1}\varsigma_{m_n}\text{ and }\sup_{z\in\mathcal{Z}}|(\varphi-\varphi_0)(z)|\leqslant C\Big\}$$

where $C > 0$ 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 2.5.5. *Let Assumption 2.5 be satisfied. For any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ we have under the conditions of Theorem 2.5.1*

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

whereas under the conditions of Theorem 2.5.2 it holds

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

2.6 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

$$Y = \varphi(Z) + c_U U \tag{2.20}$$

for some constant $c_U > 0$ specified below. The structural function φ and the joint distribution of (Z, W, U) varies in the experiments below. As basis $\{f_j\}_{j\geqslant 1}$ 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 $U = \kappa\,\varepsilon + \sqrt{1-\kappa^2}\,\varepsilon$ with $\kappa = 0.3$ and $\varepsilon \sim N(0,1)$. Then realizations of Y where generated by (2.20) with $c_U = 0.2$ by an either linear function

$$\varphi(z) = z, \tag{2.21}$$

a polynomial of second degree

$$\varphi(z) = z - z^2, \tag{2.22}$$

or a polynomial of third degree

$$\varphi(z) = z - z^2 + \theta_3 \, z^3. \tag{2.23}$$

Given (2.23) is the correct model, then $\theta_3 = 1.5$ if (2.21) is the null model and $\theta_3 = 3$ if (2.21) is the null model. In Table 2.1 we depict the empirical rejection probabilities when using S_n^p with additional smoothing where either $\tau_j = j^{-1}$ or $\tau_j = j^{-2}$, $j \geqslant 1$, which we denote by S_n^{1p} or S_n^{2p}, respectively. When $\tau_j = j^{-1}$ then the number of basis

Sample Size	*Null Model*	*Alt. Model*	*Empirical Rejection probability* S_n^{1p}	S_n^{2p}	*H(2006)' test*
250	(2.21)	H^p true	0.047	0.045	0.063
	(2.22)	H^p true	0.049	0.050	0.059
	(2.21)	(2.22)	0.902	0.930	0.888
	(2.21)	(2.23)	0.730	0.732	0.653
	(2.22)	(2.23)	0.442	0.488	0.468
500	(2.21)	H^p true	0.055	0.044	0.053
	(2.22)	H^p true	0.051	0.053	0.059
	(2.21)	(2.22)	0.989	0.998	0.988
	(2.21)	(2.23)	0.899	0.894	0.780
	(2.22)	(2.23)	0.709	0.728	0.652

Table 2.1: Empirical Rejection probabilities for parametric specification

functions used is $m = 200$ 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 2.3.1 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 $\widehat{\lambda}_j$ are sufficiently close to zero for all $j \geqslant M$. 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 2.1 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 (2.21) against (2.23). 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 $U = \kappa\,\varepsilon + \sqrt{1-\kappa^2}\,\varepsilon$ with $\varepsilon \sim \mathcal{U}[0,1]$. Here, κ 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 where generated by (2.20) 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 2.1. 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	κ	*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 2.2: Empirical Rejection probabilities for testing exogeneity

In Table 2.2 we depict the empirical rejection probabilities when using S_n^{e} with additional smoothing where either $\tau_j = j^{-1}$ or $\tau_j = j^{-2}$, $j \geqslant 1$, which we denote by S_n^{1e} or S_n^{2e}, respectively. The critical values of these statistics are estimated as described in Remark 2.4.2 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 2.2, 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 2.1. Further, Y is generated by (2.20) either with φ_1 and $c_U = 0.2$ or φ_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 ρ is defined below. Consequently, when H_{np} is false we generate realizations of Y from

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

for $l = 1, 2$ and $j \geqslant 1$ where $\rho_j(z) = c_j(\exp(2jz)\,\mathbb{1}_{\{z\leqslant 1/2\}} + \exp(2j(1-z))\,\mathbb{1}_{\{z>1/2\}} - 1)$ and c_j is a normalizing constant such that $\int_0^1 \rho_j(z)dz = 0.5$. The functions ρ_j are continuous but not differentiable at 0.5. Roughly speaking, the degree of roughness of ρ_j is larger for larger j. In Table 2.3, we depict the empirical rejection probabilities when

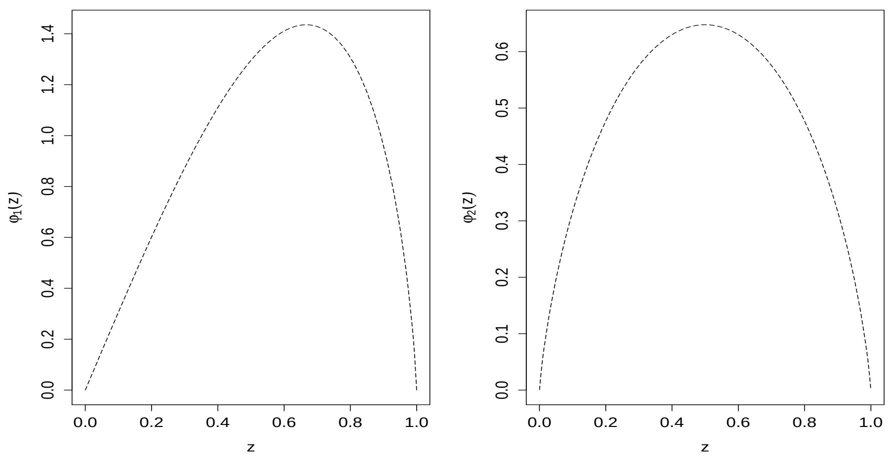

Figure 2.1: Graph of φ_1 and φ_2

using S_n^{np} with either no smoothing or additional smoothing $\tau_j = j^{-2}$, $j \geqslant 1$, 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

Sample Size	ρ	*Empirical Rejection probability using*		
		S_n^{0np}	S_n^{2np}	*H(2012)' test*
500	H^{np} true	0.034	0.039	0.040
	ρ_1	0.099	0.382	0.258
	ρ_2	0.309	0.765	0.536
	ρ_4	0.498	0.884	0.712
1000	H^{np} true	0.058	0.058	0.046
	ρ_1	0.405	0.672	0.427
	ρ_2	0.768	0.899	0.704
	ρ_4	0.920	0.943	0.808

Table 2.3: Empirical Rejection prob. for Nonparametric Spec. for φ_1 with $c_U = 0.2$

results. The critical values of these statistics are estimated as described in Remark 2.5.3 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 ρ_1 and ρ_2.

In the following, we illustrate that using additional weighting can be inappropriate. Table 2.4 illustrates the power of our tests when the structural function φ_2 is considered and realizations (Z, W) were generated by $W \sim \mathcal{U}[0,1]$, $Z = (0.8\,W + 0.3\,\varepsilon)^2$ where $\varepsilon \sim \mathcal{N}(0.5, 0.05)$. In this case, we generate Y using (2.20) where $c_U = 0.8$. In this case,

Sample Size	ρ	*Empirical Rejection probability using*		
		S_n^{0np}	S_n^{2np}	*H(2012)' test*
500	H^{np} true	0.022	0.044	0.044
	ρ_3	0.230	0.193	0.158
	ρ_4	0.400	0.319	0.245
	ρ_5	0.543	0.463	0.370
1000	H^{np} true	0.044	0.049	0.052
	ρ_3	0.643	0.343	0.302
	ρ_4	0.836	0.579	0.518
	ρ_5	0.924	0.792	0.722

Table 2.4: Empirical Rejection prob. for Nonparametric Spec. for φ_2 with $c_U = 0.8$

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 2.4, the test statistic S_n^{0np} with no smoothing is more powerful than S_n^{2np} were weighting $\tau_j = j^{-2}$, $j \geqslant 1$, is used. In particular, S_n^{0np} is much more powerful than the test of Horowitz [2012].

2.7 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 τ, 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.

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 $\sum_{i'<i} = \sum_{i=1}^n \sum_{i'=1}^{i-1}$. Further, we denote $\|F_{m_n} T\phi\|_\tau^2 = \sum_{j=1}^{m_n} \tau_j \, \mathbb{E}[\phi(Z) f_j(W)]^2$ and $\|T\phi\|_\tau^2 = \sum_{j=1}^{\infty} \tau_j \, \mathbb{E}[\phi(Z) f_j(W)]^2$ for any $\phi \in L_Z^2$.

Proofs of Section 2.2

Proof of Theorem 2.2.1. Under H_0 we have $(Y_i - \varphi_0(Z_i)) f_{\underline{m}}^\tau(W_i) = U_i f_{\underline{m}}^\tau(W_i)$ for all $m \geqslant 1$ and consequently we observe

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

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

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

By using Assumptions 2.1 and 2.2, i.e., $\sup_{j\in\mathbb{N}}\mathbb{E}\,|f_j(W)|^4 \leqslant \eta_f\,\eta_p$ and $\mathbb{E}[U^4|W] \leqslant \sigma^4$, we conclude

$$\mathbb{E}\,\|U f^{\tau}_{\underline{m}}(W)\|^4 \leqslant \max_{1\leqslant j\leqslant m} \mathbb{E}\,|U f_j(W)|^4 \Big(\sum_{j=1}^{m}\tau_j\Big)^2 \leqslant \eta_f\,\eta_p\,\sigma^4 \Big(\sum_{j=1}^{m}\tau_j\Big)^2. \tag{2.24}$$

Let $m = m_n$ satisfy condition (2.5) then $\mathbb{E}\,\|U f^{\tau}_{\underline{m_n}}(W)\|^4 = o(n\varsigma^2_{m_n})$. Therefore, it is sufficient to prove

$$\sqrt{2}(\varsigma_{m_n} n)^{-1}\sum_{i\neq i'}\sum_{j=1}^{m_n} U_i U_{i'} f^{\tau}_j(W_i) f^{\tau}_j(W_{i'}) \xrightarrow{d} \mathcal{N}(0,1). \tag{2.25}$$

Since $\varsigma_{m_n} = o(1)$ this follows from Lemma 2.7.2 and thus, completes the proof. □

Proof of Theorem 2.2.2. Similarly to the proof of Theorem 2.2.1 it is sufficient to study the asympotic behavior of $n^{-1}\sum_{j=1}^{m_n}\sum_{i\neq i'} U_i U_{i'} f^{\tau}_j(W_i) f^{\tau}_j(W_{i'})$. For any finite $m \geqslant 1$ we obtain

$$\begin{aligned}
&\mathbb{E}\Big|\frac{1}{n}\sum_{j=1}^{m}\sum_{i\neq i'} U_i U_{i'} f^{\tau}_j(W_i) f^{\tau}_j(W_{i'}) - \frac{1}{n}\sum_{j=1}^{\infty}\sum_{i\neq i'} U_i U_{i'} f^{\tau}_j(W_i) f^{\tau}_j(W_{i'})\Big|^2 \\
&\leqslant \mathbb{E}\Big[\,\mathbb{E}[U_1^2 U_2^2|W_1,W_2]\Big(\sum_{j>m} f^{\tau}_j(W_1) f^{\tau}_j(W_2)\Big)^2\Big] \leqslant \sigma^4\eta_p \sum_{j,l>m}\Big(\int_{\mathcal{W}} f^{\tau}_j(s) f^{\tau}_l(s)\nu(ds)\Big)^2 \\
&\qquad\qquad \leqslant \sigma^4\eta_p\sum_{j\geqslant m}\tau_j^2
\end{aligned}$$

which, since $\sum_{j\geqslant 1}\tau_j^2 = O(1)$, becomes sufficiently small (depending on m). Note that $(\frac{1}{\sqrt{n}}\sum_i U_i f^{\tau}_1(W_i),\dots,\frac{1}{\sqrt{n}}\sum_i U_i f^{\tau}_m(W_i)) \xrightarrow{d} \mathcal{N}(0,\Sigma_m)$. Hence, for any finite $m \geqslant 1$ we have

$$\sum_{j=1}^{m}\Big|\frac{1}{\sqrt{n}}\sum_i U_i f^{\tau}_j(W_i)\Big|^2 \xrightarrow{d} \sum_{j=1}^{m}\lambda_j\chi^2_{1j}$$

with λ_j, $1\leqslant j\leqslant m$, being eigenvalues of Σ_m. Moreover, we conclude for $m \geqslant 1$

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

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

Proof of Proposition 2.2.3. For ease of notation let $\delta_n(\cdot):=\varsigma_{m_n}^{1/2}n^{-1/2}\delta(\cdot)$. Under the sequence of alternatives (2.7) the following decomposition holds true

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

Due to Theorem 2.2.1 we have $(\sqrt{2}\varsigma_{m_n})^{-1}(n\,I_n-\mu_{m_n})\xrightarrow{d}\mathcal{N}(0,1)$. Consider II_n. We observe

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

From the definition of δ_n and condition (2.5) we infer that $n\,\mathbb{E}\,|II_n|=o(\varsigma_{m_n})$. Consider III_n. Employing again the definition of δ_n it is easily seen that $n\varsigma_{m_n}^{-1}III_n=\sum_{j=1}^{m_n}\tau_j[T\delta]_j^2+o_p(1)$. We conclude $(\sqrt{2}\varsigma_{m_n})^{-1}nIII_n=2^{-1/2}\sum_{j\geqslant1}\delta_j^2+o_p(1)$, which completes the proof. $\square$

Proof of Proposition 2.2.4. Let $\delta_n(\cdot):=n^{-1/2}\delta(\cdot)$. Similarly to the proof of Theorem 2.2.2 it is straightforward to see that under the sequence of alternatives (2.8) it holds

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

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

Proof of Proposition 2.2.5. If H_0 fails we observe that $\|T(\varphi-\varphi_0)\|_\tau^2=\int_{\mathcal{W}}|LT(\varphi-\varphi_0)(w)p_W(w)/\nu(w)|^2\nu(dw)\geqslant C\|LT(\varphi-\varphi_0)\|_W^2>0$ since p_W/ν 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 $S_n = \|T(\varphi - \varphi_0)\|_\tau^2 + o_p(1)$. We make use of the decomposition

$$S_n = \sum_{j=1}^{m_n} \tau_j \big|n^{-1} \sum_i (Y_i - \varphi_0(Z_i)) f_j(W_i) - [T(\varphi - \varphi_0)]_j\big|^2$$
$$+2\sum_{j=1}^{m_n} \tau_j \big(n^{-1} \sum_i (Y_i - \varphi_0(Z_i)) f_j(W_i) - [T(\varphi-\varphi_0)]_j\big)[T(\varphi-\varphi_0)]_j + \|F_{m_n} T(\varphi-\varphi_0)\|_\tau^2$$
$$= I_n + II_n + III_n.$$

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

Proof of Proposition 2.2.6. We make use of the decomposition

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}(n\, S_n - \mu_{m_n}) > q_{1-\alpha}\Big)$$
$$\geqslant \mathbb{P}\Big(\|n^{-1/2} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f^\tau_{\underline{m_n}}(W_i)\|^2 + \|n^{-1/2} \sum_i U_i f^\tau_{\underline{m_n}}(W_i)\|^2 - \mu_{m_n}$$
$$> \sqrt{2}\,\varsigma_{m_n} q_{1-\alpha} + 2|\langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f^\tau_{\underline{m_n}}(W_i), \sum_i U_i f^\tau_{\underline{m_n}}(W_i)\rangle|\Big).$$

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

$$\langle n^{-1} \sum_i (\varphi(Z_i) - \varphi_0(Z_i)) f^\tau_{\underline{m_n}}(W_i), \sum_i U_i f^\tau_{\underline{m_n}}(W_i)\rangle = O_p\big(\max(\sqrt{n}\|LT(\varphi - \varphi_0)\|_W, \varsigma_{m_n})\big).$$

Indeed, this is easily seen from

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

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

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

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

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}(n\,S_n - \mu_{m_n}) > q_{1-\alpha}\Big)$$
$$\geqslant \mathbb{P}\Big(\|n^{-1/2}\sum_i(\varphi(Z_i) - \varphi_0(Z_i))f^{\tau}_{\underline{m_n}}(W_i)\|^2 + \|n^{-1/2}\sum_i U_i f^{\tau}_{\underline{m_n}}(W_i)\|^2 - \mu_{m_n}$$
$$> \sqrt{2}\,\varsigma_{m_n} q_{1-\alpha} + C\max(\sqrt{n}\|LT(\varphi - \varphi_0)\|_W, \varsigma_{m_n})\Big) - \varepsilon'.$$

Note that $\|n^{-1/2}\sum_i U_i f^{\tau}_{\underline{m_n}}(W_i)\|^2 = \mu_{m_n} + O_p(\varsigma_{m_n})$ due to Theorem 2.2.1. Moreover,

$$\|n^{-1/2}\sum_i(\varphi(Z_i) - \varphi_0(Z_i))f^{\tau}_{\underline{m_n}}(W_i)\|^2 \geqslant n\|F_{m_n}T(\varphi - \varphi_0)\|^2_\tau$$
$$-2|\langle\sum_i(\varphi(Z_i)-\varphi_0(Z_i))f^{\tau}_{\underline{m_n}}(W_i) - n[LT(\varphi-\varphi_0)]_{\underline{m_n}}, [LT(\varphi-\varphi_0)]_{\underline{m_n}}\rangle = I_n + II_n.$$

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

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

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

Proofs of Section 2.3

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 2.3.1. The proof is based on the decomposition under H_p

$$S^\mathrm{p}_n = \|\frac{1}{n}\sum_i U_i f^{\tau}_{\underline{m_n}}(W_i)\|^2 + 2\langle\frac{1}{n}\sum_i U_i f^{\tau}_{\underline{m_n}}(W_i), \frac{1}{n}\sum_i(\phi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n))f^{\tau}_{\underline{m_n}}(W_i)\rangle$$
$$+ \|\frac{1}{n}\sum_i(\phi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n))f^{\tau}_{\underline{m_n}}(W_i)\|^2 = I_n + 2II_n + III_n. \quad (2.26)$$

Due to Theorem 2.2.1 it holds $(\sqrt{2}\varsigma_{m_n})^{-1}(nI_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0,1)$. Consider III_n. Since $\phi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n) = \phi_\vartheta(Z_i)^t(\vartheta_0 - \widehat{\vartheta}_n) + o_p(n^{-1})$ observe

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

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

$$\begin{aligned}\sum_{j=1}^{m_n}[T\phi_{\vartheta_l}]_j^2 = \sum_{j=1}^{m_n}\Big(\int_{\mathcal{W}}(T\phi_{\vartheta_l})(w) f_j(w) p_w(w) dw\Big)^2 \\ \leqslant \int_{\mathcal{W}} |(T\phi_{\vartheta_l})(w) p_w(w)/\nu(w)|^2 \nu(dw) \leqslant \eta_p \|T\phi_{\vartheta_l}\|_W^2 \\ \leqslant \eta_p \,\mathbb{E}\,|\phi_{\vartheta_l}(Z,\vartheta_0)|^2 \leqslant \eta_p \eta_\phi \quad (2.27)\end{aligned}$$

by applying Jensen's inequality. Moreover, we calculate

$$\sum_{l=1}^{k}\sum_{j=1}^{m_n} \mathbb{E}\,\Big|\frac{1}{n}\sum_i \phi_{\vartheta_l}(Z_i) f_j(W_i) - [T\phi_{\vartheta_l}]_j\Big|^2 \leqslant \frac{km_n}{n} \sup_{j,l\geqslant 1} \mathbb{E}\,|\phi_{\vartheta_l}(Z) f_j(W)|^2 \leqslant \eta^4 \frac{km_n}{n}. \quad (2.28)$$

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 $1 \leqslant l \leqslant k$

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

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

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

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

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

which implies $nII_n = o_p(\varsigma_{m_n})$ and thus, in light of decomposition (3.23), completes the proof. □

PROOF OF THEOREM 2.3.2. For $1 \leqslant j \leqslant m_n$ we make use of the following decomposition

$$
\begin{aligned}
n^{-1/2}\sum_i f_j(W_i)\Big(U_i+\phi(Z_i)-\phi(Z_i,\widehat{\vartheta}_n)\Big) &= n^{-1/2}\sum_i \Big(f_j(W_i)U_i-\sum_{l=1}^k [T\phi_{\vartheta_l}]_j h_l(V_i)\Big)\\
&+\sum_{l=1}^k \Big(n^{-1}\sum_i f_j(W_i)\phi_{\vartheta_l}(Z_i)-[T\phi_{\vartheta_l}]_j\Big)\Big(n^{-1/2}\sum_i h_l(V_i)\Big)\\
+\sum_{l=1}^k n^{-1}\sum_i f_j(W_i)\phi_{\vartheta_l}(Z_i)r_l+o_p(1) &= A_{nj}+B_{nj}+C_{nj}+o_p(1) \qquad (2.29)
\end{aligned}
$$

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

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

Clearly, for all $1\leqslant i\leqslant n$ the random variables $U_i f_j^\tau(W_i)+\mathbb{E}\,[f_j^\tau(W)\phi_\vartheta(Z)^t]h_{\underline{k}}(V_i)$, $1\leqslant j\leqslant m_n$, are centered with bounded second moment. Due to the proof of Theorem 2.2.2 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 (2.28) yields $\sum_{j=1}^{m_n}B_{nj}^2=o_p(1)$. Since $\sum_{j=1}^{m_n}[T\phi_\vartheta]_j^2\leqslant\eta_p\eta_\phi$ we have $\|\,\mathbb{E}[f_{\underline{m_n}}(W)\phi_\vartheta(Z)^t]r_{\underline{k}}\|^2\leqslant k\,\eta_p\eta_\phi\|r_{\underline{k}}\|^2=o_p(1)$ and hence $\sum_{j=1}^{m_n}C_{nj}^2=o_p(1)$. Finally, Cauchy Schwarz's inequality implies that $\sum_{j=1}^{m_n}\tau_j A_{nj}(B_{nj}+C_{nj})=o_p(1)$, which completes the proof. □

PROOF OF PROPOSITION 2.3.3. Consider the case $\varsigma_{m_n}^{-1}=o(1)$. Under the sequence of alternatives (2.7) the following decomposition holds true

$$
\begin{aligned}
S_n^{\mathrm{p}}=S_n+2\langle n^{-1}\sum_i(U_i+\varsigma_{m_n}^{1/2}n^{-1/2}\delta(Z_i))f_{\underline{m_n}}^\tau(W_i),n^{-1}\sum_i(\phi(Z_i)-\phi(Z_i,\widehat{\vartheta}_n))f_{\underline{m_n}}^\tau(W_i)\rangle\\
+\|n^{-1}\sum_i(\phi(Z_i)-\phi(Z_i,\widehat{\vartheta}_n))f_{\underline{m_n}}^\tau(W_i)\|^2.
\end{aligned}
$$

Due to Proposition 2.2.3 and the proof of Theorem 2.3.1 it is sufficient to show

$$
\langle n^{-1}\sum_i\delta(Z_i)f_{\underline{m_n}}^\tau(W_i),n^{-1/2}\sum_i(\phi(Z_i)-\phi(Z_i,\widehat{\vartheta}_n))f_{\underline{m_n}}^\tau(W_i)\rangle=o_p(\sqrt{\varsigma_{m_n}}). \qquad (2.30)
$$

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

$$\sum_{j=1}^{m_n} \delta_j n^{-1/2} \sum_i (\phi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n)) f_j(W_i) = \sqrt{n}(\vartheta_0 - \widehat{\vartheta}_n)^t \sum_{j=1}^{m_n} \delta_j\, \mathbb{E}[\phi_\vartheta(Z) f_j(W)] + o_p(1)$$
$$\leqslant \eta_p\, \eta_\phi\, \sqrt{n} \|\vartheta_0 - \widehat{\vartheta}_n\| \sum_{j=1}^{\infty} \delta_j^2 + o_p(1) = O_p(1)$$

and hence (2.30) holds true.

Consider the case $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$. We make use of decomposition (2.29) where U_i is replaced by $U_i + n^{-1/2}\delta(Z_i)$. Similarly to the proof of Proposition 2.2.4 it is easily seen that $\sum_{j=1}^{m_n} \tau_j A_{nj}^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^{p}\, \chi_{1j}^2(\delta_j/\lambda_j^{\mathrm{p}})$. Thereby, due to the proof of Theorem 2.3.2, the assertion follows. □

PROOF OF PROPOSITION 2.3.4. It is sufficient to prove $S_n^{\mathrm{p}} = \|T(\varphi - \phi(\cdot, \vartheta_0))\|_\tau^2 + o_p(1)$. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Since $\|n^{-1} \sum_i (\phi(Z_i, \vartheta_0) - \phi(Z_i, \widehat{\vartheta}_n)) f_{\underline{m_n}}^\tau(W_i)\|^2 = o_p(1)$ (cf. proof of Theorem 2.3.1) and $\|n^{-1} \sum_i (Y_i - \phi(Z_i, \vartheta_0)) f_{\underline{m_n}}^\tau(W_i)\|^2 = \|T(\varphi - \phi(\cdot, \vartheta_0))\|_\tau^2 + o_p(1)$ (cf. proof of Proposition 2.2.5) the result follows. In case of $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$, we infer similarly that $S_n^{\mathrm{p}} = \sum_{j=1}^{m_n} \tau_j |n^{-1} \sum_i ((Y_i - \phi(Z_i, \vartheta_0)) f_j(W_i)|^2 + o_p(1)$ and hence, $S_n^{\mathrm{p}} = \|T(\varphi - \varphi_0)\|_\tau^2 + o_p(1)$. □

PROOF OF PROPOSITION 2.3.5. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. The basic inequality $(a-b)^2 \geqslant a^2/2 - b^2$, $a, b \in \mathbb{R}$, yields

$$\begin{aligned}
&\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}(n\, S_n^{\mathrm{p}} - \mu_{m_n}) > q_{1-\alpha}\Big) \\
&\geqslant \mathbb{P}\Big(1/2 \|n^{-1/2} \sum_i (\varphi(Z_i) - \phi(Z_i, \vartheta_0)) f_{\underline{m_n}}^\tau(W_i)\|^2 + \|n^{-1/2} \sum_i U_i f_{\underline{m_n}}^\tau(W_i)\|^2 - \mu_{m_n} \\
&\qquad > \sqrt{2}\,\varsigma_{m_n} q_{1-\alpha} + 2|\langle n^{-1} \sum_i (\varphi(Z_i) - \phi(Z_i, \widehat{\vartheta}_n)) f_{\underline{m_n}}^\tau(W_i), \sum_i U_i f_{\underline{m_n}}^\tau(W_i)\rangle| \\
&\qquad + \|n^{-1/2} \sum_i (\phi(Z_i, \vartheta_0) - \phi(Z_i, \widehat{\vartheta}_n)) f_{\underline{m_n}}^\tau(W_i)\|^2\Big). \qquad (2.31)
\end{aligned}$$

From the proof of Theorem 2.3.1 we infer $\|n^{-1/2} \sum_i (\phi(Z_i, \widehat{\vartheta}_n) - \phi(Z_i, \vartheta_0)) f_{\underline{m_n}}^\tau(W_i)\|^2 = o_p(\varsigma_{m_n})$ and

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

Thus, following line by line the proof of Proposition 2.2.6, the assertion follows. In case of $\sum_{j=1}^{m_n}\tau_j^2 = O(1)$ the assertion follows similarly. $\square$

Proofs of Section 2.4.

In the following, we denote $Q = \mathbb{E}[e_{\underline{k_n}}(Z)e_{\underline{k_n}}(Z)^t]$ and $\widehat{Q} = n^{-1}\sum_{i=1}^n e_{\underline{k_n}}(Z_i)e_{\underline{k_n}}(Z_i)^t$. By Assumption 2.7, the eigenvalues of $Q = \mathbb{E}[e_{\underline{k_n}}(Z)e_{\underline{k_n}}(Z)^t]$ are bounded away from zero and hence, it may be assumed that $Q = I_{k_n}$ (cf. Newey [1997], p. 161).

Proof of Theorem 2.4.1. The proof is based on the decomposition (3.23) where the estimator $\phi(\cdot,\widehat{\vartheta}_n)$ is replaced by $\overline{\varphi}_{k_n}(\cdot)$ given in (2.10). It holds $nIII_n = o_p(\varsigma_{m_n})$, which can be seen as follows. We make use of

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

Consider A_{n1}. We observe

$$A_{n1} \leqslant 2\|T(E_{k_n}\varphi_0 - \overline{\varphi}_{k_n})\|_W^2 + 2\|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2 \sum_{j=1}^{m_n}\tau_j \sum_{l=1}^{k_n}\Big|n^{-1}\sum_i e_l(Z_i)f_j(W_i) - [T]_{jl}\Big|^2$$
$$=: 2B_{n1} + 2B_{n2}. \quad (2.32)$$

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

$$B_{n1} \leqslant \|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2 \leqslant 2\Big\|[\widehat{Q}]_{k_n}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_i Y_i e_{\underline{k_n}}(Z_i)\Big\|^2$$
$$+ 2\big\|[\widehat{Q}]_{\underline{k_n}} - I_{k_n}\big\|^2 \,\big\|[\widehat{Q}]_{\underline{k_n}}^{-1}\big\|^2 \,\Big\|[\widehat{Q}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_i Y_i e_{\underline{k_n}}(Z_i)\Big\|^2.$$

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

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

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

$$\begin{aligned}\mathbb{E}\,\big\|[\widehat{Q}]_{\underline{k_n}}[\varphi_0]_{\underline{k_n}} - n^{-1}\sum_i Y_i e_{\underline{k_n}}(Z_i)\big\|^2 \leqslant n^{-1}\sum_{j=1}^{k_n}\mathbb{E}\,|e_j(Z)E_{k_n}\varphi_0(Z) - Ye_j(Z)|^2\\ \leqslant n^{-1}\sup_{z\in\mathcal{Z}}\|e_{k_n}(z)\|^2\,\mathbb{E}\,|E_{k_n}\varphi_0(Z)-Y|^2 \leqslant 4\eta_e n^{-1}k_n\big(\sup_{z\in\mathcal{Z}}|E_{k_n}^\perp\varphi_0(z)|+\|\varphi_0\|_Z^2+\mathbb{E}\,Y^2\big)\\ = O(n^{-1}k_n)\end{aligned}$$

Moreover, since the difference of eigenvalues of $[\widehat{Q}]_{\underline{k_n}}$ and I_{k_n} is bounded by $\|[\widehat{Q}]_{\underline{k_n}} - I_{k_n}\|$, the smallest eigenvalue of $[\widehat{Q}]_{\underline{k_n}}$ converges in probability to one and thereby, $\|[\widehat{Q}]_{\underline{k_n}}^{-1}\|^2 = 1 + o_p(1)$. Consequently,

$$n\|E_{k_n}\varphi_0 - \overline{\varphi}_{k_n}\|_Z^2 = O_p(k_n) \tag{2.33}$$

and since $k_n = o(\varsigma_{m_n})$ we proved $nB_{n1} = o_p(\varsigma_{m_n})$. In addition, applying inequality (2.28) together with equation (2.33) yields $nB_{n2} = o_p(\varsigma_{m_n})$. Consequently, $nA_{n1} = o(\varsigma_{m_n})$. Consider A_{n2}. Similar to the derivation of (2.27) we obtain

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

We have

$$\sum_{j=1}^{m_n}\tau_j\,\mathbb{E}\,|(E_{k_n}^\perp\varphi_0)(Z)f_j(W)|^2 = O\Big(\gamma_{k_n}^{-1}\sum_{j=1}^{m_n}\tau_j\Big) = o(\varsigma_{m_n}) \tag{2.34}$$

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

$$\begin{aligned}nII_n \leqslant \Big|\sum_{j=1}^{m_n}\tau_j\sum_i U_if_j(W_i)([\varphi_0]_{\underline{k_n}}-[\overline{\varphi}]_{\underline{k_n}})^t\Big(n^{-1}\sum_i e_{\underline{k_n}}(Z_i)f_j(W_i)-\mathbb{E}\,[e_{\underline{k_n}}(Z)f_j(W)]\Big)\Big|\\ +\Big|\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}([\varphi_0]_l - [\overline{\varphi}]_l)\Big(\sum_i U_if_j(W_i)[T]_{jl}\Big)\Big|\\ +\Big|\sum_{j=1}^{m_n}\tau_j\Big(\sum_i U_if_j(W_i)\Big)\Big(n^{-1}\sum_i E_{k_n}^\perp\varphi_0(Z_i)f_j(W_i) - \mathbb{E}[E_{k_n}^\perp\varphi_0(Z)f_j(W)]\Big)\Big|\\ +\Big|\sum_{j=1}^{m_n}\tau_j\Big(\sum_i U_if_j(W_i)\Big)\,\mathbb{E}[E_{k_n}^\perp\varphi_0(Z)f_j(W)]\Big| = C_{n1}+C_{n2}+C_{n3}+C_{n4}.\end{aligned} \tag{2.35}$$

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

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

From $\mathbb{E}\,|\sum_i U_i f_j(W_i)|^2 \leqslant n\,\eta_f \sigma^2$, relation (2.33), and inequality (2.28) we infer $C_{n1} = o_p(\varsigma_{m_n})$ due to condition (2.11). For C_{n2} we evaluate

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

Now $\sum_{j=1}^{m_n}\sum_{l=1}^{k_n}[T]_{jl}^2 = O(k_n)$ together with (2.33) yields $C_{n2} = o_p(1)$. Consider C_{n3}. Since $\mathbb{E}[U^2|W] \leqslant \sigma^2$ we conclude similarly as in inequality (2.34) that

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

Consider C_{n4}. We calculate

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

Consequently, in light of decomposition (2.35) we obtain $nII_n = o(\varsigma_{m_n})$, which completes the proof. □

Proof of Theorem 2.4.2. Employing the equality $[\widehat{Q}]_{\underline{k_n}}^{-1} = I_{k_n} - [\widehat{Q}]_{\underline{k_n}}^{-1}([\widehat{Q}]_{\underline{k_n}} - I_{k_n})$ we obtain for all $1 \leqslant j \leqslant m_n$

$$\begin{aligned}
&n^{-1/2}\sum_i f_j(W_i)\Big(U_i + \varphi_0(Z_i) - \overline{\varphi}_{k_n}(Z_i)\Big)\\
&= n^{-1/2}\sum_i \Big(f_j(W_i)U_i + \mathbb{E}[f_j(W)e_{\underline{k_n}}(Z)^t]e_{\underline{k_n}}(Z_i)(\varphi_0(Z_i) - Y_i)\Big)\\
&\qquad + n^{-1/2}\sum_i \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][\widehat{Q}]_{\underline{k_n}}^{-1}([\widehat{Q}]_{\underline{k_n}} - I_{k_n})(E_{k_n}\varphi_0(Z_i) - Y_i)\\
&\qquad + \Big(n^{-1}\sum_i f_j(W_i)e_l(Z_i) - \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right]\Big)\sqrt{n}([\varphi_0]_{\underline{k_n}} - [\overline{\varphi}]_{\underline{k_n}})\\
&+ n^{-1/2}\sum_i \mathbb{E}[f_j(W)e_{\underline{k_n}}(Z)^t]e_{\underline{k_n}}(Z_i)(f_j(W_i) - (E_{k_n}T^*f_j)(Z_i)) = A_{nj} + B_{nj} + C_{nj} + D_{nj}.
\end{aligned} \tag{2.36}$$

Due to Assumption 2.7 (ii) we may assume that $\{e_1, \ldots, e_k\}$ 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, $\sum_{l\geqslant 1}\mathbb{E}[f_j(W)e_l(Z)]e_l(\cdot)$ belongs to $\mathcal{L}_Z^2$ for all $j \geqslant 1$. Now following line by line the proof of Theorem 2.2.2 we deduce

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

Moreover, we see similarly to the proof of Theorem 2.4.1 that $\sum_{j=1}^{m_n}\tau_j(B_{nj}^2+C_{nj}^2+D_{nj}^2)=o_p(1)$, which completes the proof. □

Proof of Lemma 2.4.3. Note that the squared Frobenius norm of $\widehat{\Sigma}_{m_n}-\Sigma_{m_n}$ is given by

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

by using relation (2.33). 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. □

Proof of Proposition 2.4.4. Consider the case $\varsigma_{m_n}^{-1}=o(1)$. Due to condition (2.14) and similarly to the proof of Theorem 2.4.1 we observe

$$\begin{aligned}
&\sum_{j=1}^{m_n}\Big|n^{-1/2}\sum_i(\varphi(Z_i)-\overline{\varphi}_{k_n}(Z_i))f_j^\tau(W_i)\Big|^2\\
&=\sum_{j=1}^{m_n}\frac{1}{n}\Big|\sum_i\Big(\varphi(Z_i)-\varphi^{[n]}(Z_i)-\frac{\sqrt{n}}{\varsigma_{m_n}}\delta(Z_i)\Big)f_j^\tau(W_i)\Big|^2+\sum_{j=1}^{m_n}\Big|\frac{1}{n}\sum_i\delta(Z_i)f_j^\tau(W_i)\Big|^2+o_p(\varsigma_{m_n})\\
&\qquad=\sum_{j=1}^{m_n}\delta_j^2/\varsigma_{m_n}+o_p(\varsigma_{m_n}).
\end{aligned}$$

Consequently, the result follows as in the proof of Theorem 2.4.1. For $\sum_{j=1}^{m_n}\tau_j^2=O(1)$ we conclude similarly. □

PROOF OF PROPOSITION 2.4.5. Similar to the proof of Proposition 2.3.4. □

PROOF OF PROPOSITION 2.4.6. We make use of inequality (2.31) where $\phi(\cdot, \widehat{\vartheta}_n)$ is replaced by $\overline{\varphi}_{k_n}$. From the proof of Theorem 2.4.1 we conclude that $\|n^{-1/2}\sum_i(\overline{\varphi}_{k_n}(Z_i) - \varphi_0(Z_i))f^\tau_{\underline{m_n}}(W_i)\|^2 = o_p(\varsigma_{m_n})$ and

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

uniformly over all $\varphi \in \mathcal{I}^\rho_n$. Thus, following line by line the proof of Proposition 2.2.6, the assertion follows. □

Proofs of Section 2.5.

Recall that $[T]_{\underline{k}} = \mathbb{E}[e_{\underline{k}}(W)e_{\underline{k}}(Z)^t]$. In the following, we introduce the function $\varphi_{k_n}(\cdot) := e_{\underline{k_n}}(\cdot)^t[T]^{-1}_{\underline{k_n}}[g]_{\underline{k_n}}$ which belongs to $\mathcal{L}^2_Z$. For all $k \geqslant 1$ let us denote $\Omega_k := \{\|[\widehat{T}]^{-1}_{\underline{k}}\| \leqslant \sqrt{n}\}$ and $\mho_k := \{\|Q_k\|\|[T]^{-1}_{\underline{k}}\| \leqslant 1/2\}$ where $Q_k = [\widehat{T}]_{\underline{k}} - [T]_{\underline{k}}$. Note that $\mathbb{E}\,\mathbb{1}_{\Omega^c_{k_n}} = \mathbb{P}(\Omega^c_{k_n}) = o(1)$ (cf. proof of Proposition 1.3.1) and, hence $\mathbb{1}_{\Omega_{k_n}} = 1 + o_p(1)$.

PROOF OF THEOREM 2.5.1. For the proof we make use of decomposition (3.23) where the estimator $\phi(\cdot, \widehat{\vartheta}_n)$ is replaced by $\widehat{\varphi}_{k_n}$ given in (2.16). Consider III_n. Observe

$$III_n \leqslant 2\|n^{-1}\sum_i(\varphi_{k_n}(Z_i) - \widehat{\varphi}_{k_n}(Z_i))f^\tau_{\underline{m_n}}(W_i)\|^2 \\ + 2\|n^{-1}\sum_i \left(\varphi_{k_n}(Z_i) - \varphi_0(Z_i)\right)f^\tau_{\underline{m_n}}(W_i)\|^2 = 2A_{n1} + 2A_{n2}. \quad (2.37)$$

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

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

The link condition $T \in \mathcal{T}^\upsilon_{d,D}$ yields $\|T(\widehat{\varphi}_{k_n} - \varphi_{k_n})\|^2_W \leqslant d\|\widehat{\varphi}_{k_n} - \varphi_{k_n}\|^2_\upsilon$. From Theorem 2.6 of Johannes and Schwarz [2010] and condition (2.17) we infer $n\|\widehat{\varphi}_{k_n} - \varphi_{k_n}\|^2_\upsilon = O_p\big(\max(n\upsilon_{k_n}\gamma^{-1}_{k_n}, k_n)\big) = o_p(\varsigma_{m_n})$. This together with estimate (2.28) implies $nA_{n1} =$

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

$$\begin{aligned}\mathbb{E}\, A_{n2} &\leqslant 2\|T(\varphi_{k_n}-\varphi_0)\|_W^2 + 2n^{-1}\,\mathbb{E}\,\|(\varphi_{k_n}(Z)-\varphi_0(Z))f^{\tau}_{\underline{m_n}}(W)\|^2\\ &\leqslant 2d\|\varphi_{k_n}-\varphi_0\|_\upsilon^2 + 2n^{-1}\sum_{l\geqslant 1} l^2([\varphi_{k_n}]_l-[\varphi_0]_l)^2\sum_{j=1}^{m_n}\tau_j\sum_{l\geqslant 1} l^{-2}\,\mathbb{E}\,|e_l(Z)f_j(W)|^2\\ &\leqslant 8Dd^2\rho\Big(\frac{\upsilon_{k_n}}{\gamma_{k_n}}\|\varphi_{k_n}-\varphi_0\|_\gamma^2 + \frac{\pi^2}{6}\eta^4\|\varphi_{k_n}-\varphi_0\|_\gamma^2\frac{k_n^2}{n\gamma_{k_n}}\sum_{j=1}^{m_n}\tau_j\Big). \qquad (2.38)\end{aligned}$$

where we used Lemma A.2 of Johannes and Schwarz [2010], i.e., $\|\varphi_{k_n}-\varphi_0\|_w^2 \leqslant 4Dd\rho w_{k_n}\gamma_{k_n}^{-1}$ for a nondecreasing sequence w. Condition (2.17) together with the estimate $k_n^2 \leqslant \sigma^4\sum_{j=1}^{m_n}\tau_j$ for n sufficiently large implies $nA_{n2} = o_p(\varsigma_{m_n})$. Consequently, due to (2.37) we have shown $nIII_n = o_p(\varsigma_{m_n})$. The proof of $nII_n = o_p(\varsigma_{m_n})$ is based on decomposition (2.35) where $\overline{\varphi}_{k_n}$ and $E_{k_n}^{\perp}\varphi_0$ are replaced by $\widehat{\varphi}_{k_n}$ and $\varphi_{k_n}-\varphi_0$, respectively. Consider C_{n1}. We calculate

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

Since $\sqrt{n}\|\widehat{\varphi}_{k_n}-\varphi_{k_n}\|_\upsilon = o_p(\varsigma_{m_n}^{1/2})$ we obtain, similarly as in the proof of Theorem 2.4.1, $C_{n1} = o_p(\varsigma_{m_n})$. Consider C_{n2}. Again similarly to the proof of Theorem 2.4.1 we observe

$$\begin{aligned}C_{n2} &= \Big|\sum_{j=1}^{m_n}\tau_j\sum_{l=1}^{k_n}[T]_{jl}([\widehat{\varphi}_{k_n}]_l-[\varphi_{k_n}]_l)\Big(\sum_i U_i f_j(W_i)\Big)\Big|\\ &\leqslant \big(n\|\widehat{\varphi}_{k_n}-\varphi_{k_n}\|_\upsilon^2\big)^{1/2}\Big(\sigma^2\sum_{l=1}^{k_n}\upsilon_l^{-1}\sum_{j=1}^{m_n}[T]_{jl}^2\Big)^{1/2} + o_p(1) = o(\varsigma_{m_n})\end{aligned}$$

by exploiting $\sum_{j=1}^{m_n}[T]_{jl}^2 = \|Te_l\|_W^2 \leqslant d\upsilon_l$. Consider C_{n3}. Since $\mathbb{E}[U^2|W] \leqslant \sigma^2$ we conclude similarly as in inequality (2.34) using Lemma A.2 of Johannes and Schwarz [2010]

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

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

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

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

PROOF OF THEOREM 2.5.2. Observe $[\widehat{T}]_{\underline{k_n}}[\varphi_{k_n}]_{\underline{k_n}} - [\widehat{g}]_{\underline{k_n}} = n^{-1}\sum_i f_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)$ and hence, for all $1 \leqslant j \leqslant m_n$

$$\begin{aligned}
n^{-1/2}\sum_i f_j(W_i)(U_i + \varphi_0(Z_i) - \widehat{\varphi}_{k_n}(Z_i))& \\
= n^{-1/2}\sum_i \Big(f_j(W_i)U_i + \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} e_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\Big)& \\
- n^{-1/2}\sum_i \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} Q_{k_n}[\widehat{T}]_{\underline{k_n}}^{-1} e_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)& \\
+\Big(n^{-1}\sum_i f_j(W_i)e_{\underline{k_n}}(Z_i)^t - \mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right]\Big)[\widehat{T}]_{\underline{k_n}}^{-1}\Big(n^{-1/2}\sum_i e_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\Big)& \\
+ n^{-1/2}\sum_i (\varphi_0(Z_i) - \varphi_{k_n}(Z_i)) f_j(W_i) = A_{nj} + B_{nj} + C_{nj} + D_{nj}.& \qquad (2.39)
\end{aligned}$$

Consider A_{nj}. The entries of the inverse of the matrix $(\mathbb{E}[f_j(W)e_l(Z)])_{j,l\geqslant 1}$, which is equivalent to $(\mathbb{E}[e_j(W)e_l(Z)])_{j,l\geqslant 1}$, are denoted by t_{jl} for all $j, l \geqslant 1$. For each $1 \leqslant j \leqslant m_n$, note that $\mathbb{E}\left[f_j(W)e_{\underline{k_n}}(Z)^t\right][T]_{\underline{k_n}}^{-1} e_{\underline{k_n}}(\cdot)$ converges in probability to $\sum_{l,k\geqslant 1}[T]_{jl}t_{lk}e_k(\cdot) = e_j(\cdot)$. This proves

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

Consider B_{nj}. By employing $\|[\widehat{T}]_{\underline{k}}^{-1}\|\,\mathbb{1}_{\mho_k} \leqslant 2\|[T]_{\underline{k}}^{-1}\|$ and $\|[\widehat{T}]_{\underline{k}}^{-1}\|^2\,\mathbb{1}_{\Omega_k} \leqslant n$ for all

$k \geqslant 1$ it follows

$$\sum_{j=1}^{m_n} \tau_j B_{nj}^2 \mathbb{1}_{\Omega_{k_n}} = \sum_{j=1}^{m_n} \tau_j B_{nj}^2 \mathbb{1}_{\Omega_{k_n}} (\mathbb{1}_{\mho_{k_n}} + \mathbb{1}_{\mho_{k_n}^c})$$
$$\leqslant \| \mathbb{E}[f_{\underline{m_n}}^\tau(W) e_{\underline{k_n}}(Z)^t][T]_{\underline{k_n}}^{-1} \|^2 \Big(4 \|[T]_{\underline{k_n}}^{-1}\|^2 \|Q_{k_n}\|^2 \|n^{-1/2} \sum_i f_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\|^2$$
$$+ n\|Q_{k_n}\|^2 \|n^{-1/2} \sum_i f_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\|^2 \mathbb{1}_{\mho_{k_n}^c} \Big).$$

As above we have $\| \mathbb{E} f_{\underline{m_n}}^\tau(W) e_{\underline{k_n}}(Z)^t [T]_{\underline{k_n}}^{-1} \|^2 = O(1)$. Moreover, $n\|Q_{k_n}\|^2 = O_p(k_n^2)$ and $\|n^{-1/2} \sum_i e_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\|^2 = O_p(k_n)$ due to Lemma 1.5.1. In addition, similarly to the proof of Proposition 1.3.1, it can be seen that $n\|Q_{k_n}\|^2 \|n^{-1/2} \sum_i e_{\underline{k_n}}(W_i)(\varphi_{k_n}(Z_i) - Y_i)\|^2 \mathbb{1}_{\mho_{k_n}^c} = o_p(1)$. Consequently, $\sum_{j=1}^{m_n} \tau_j B_{nj}^2 = o_p(1)$ using $\mathbb{1}_{\Omega_{k_n}} = 1 + o_p(1)$. Similarly, it is easily seen that $\sum_{j=1}^{m_n} \tau_j C_{nj}^2 = o_p(1)$ and $\sum_{j=1}^{m_n} \tau_j D_{nj}^2 = o_p(1)$, which proves the result. □

PROOF OF PROPOSITION 2.5.3. Consider the case $\varsigma_{m_n}^{-1} = o(1)$. Similar to the proof of Proposition 2.3.3 it is sufficient to show

$$\langle n^{-1} \sum_i \delta(Z_i) f_{\underline{m_n}}^\tau(W_i), n^{-1/2} \sum_i (\varphi_0(Z_i) - \widehat{\varphi}_{k_n}(Z_i)) f_{\underline{m_n}}^\tau(W_i) \rangle = o_p(\sqrt{\varsigma_{m_n}}). \quad (2.40)$$

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

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

As in the proof of Theorem 2.5.1 it can be seen $\sum_{j=1}^{m_n} \tau_j [T\delta]_j \sum_i (\varphi_0(Z_i) - \varphi_{k_n}(Z_i)) f_j(W_i) = o_p(\sqrt{n}\varsigma_{m_n})$ and, hence equation (2.40) holds true. Consider the case $\sum_{j=1}^{m_n} \tau_j^2 = O(1)$. We make use of decomposition (2.39) where U_i is replaced by $U_i + n^{-1/2}\delta(Z_i)$. Similarly to the proof of Proposition 2.2.4 it is seen that $\sum_{j=1}^{m_n} \tau_j A_{nj}^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j^{np} \chi_{1j}^2(\delta_j/\lambda_j^{np})$. Thereby, due to the proof of Theorem 2.3.2, the assertion follows. □

PROOF OF PROPOSITION 2.5.4. Similar to the proof of Proposition 2.3.4. □

PROOF OF PROPOSITION 2.5.5. We make use of inequality (2.31) where $\phi(\cdot, \widehat{\vartheta}_n)$ is replaced by $\widehat{\varphi}_{k_n}$. From the proof of Proposition 2.5.1 we infer $\|n^{-1/2} \sum_i (\widehat{\varphi}_{k_n}(Z_i) -$

$\varphi_0(Z_i))f^\tau_{\underline{m_n}}(W_i)\|^2 = o_p(\varsigma_{m_n})$ and

$$\langle n^{-1}\sum_i(\varphi(Z_i) - \widehat{\varphi}_{k_n}(Z_i))f^\tau_{\underline{m_n}}(W_i), \sum_i U_i f^\tau_{\underline{m_n}}(W_i)\rangle$$
$$= \langle n^{-1}\sum_i(\varphi(Z_i) - \varphi_0(Z_i))f^\tau_{\underline{m_n}}(W_i), \sum_i U_i f^\tau_{\underline{m_n}}(W_i)\rangle + o_p(\varsigma_{m_n})$$

uniformly over all $\varphi \in \mathcal{J}^\rho_n$. Consequently, following line by line the proof of Proposition 2.2.6, the assertion follows. □

Technical assertions.

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

$$Q_{ni} := \begin{cases} \sum_{l=1}^{i-1} X_{li}, & \text{for } i = 2,\dots,n,\\ 0, & \text{for } i = 1 \text{ and } i > n.\end{cases} \tag{2.41}$$

Then clearly

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

Let $\mathcal{B}_{ni} := \mathcal{B}((Z_1, Y_1, W_1), \dots, (Z_i, Y_i, W_i))$, $1 \leqslant i \leqslant n$, $n \geqslant 1$, be the σ-algebra generated by $(Z_1, Y_1, W_1), \dots, (Z_i, Y_i, W_i)$. Since $U_i f^\tau_j(W_i)$, $1 \leqslant i \leqslant n$, are centered random variables it follows that $\{(\sum_{i'=1}^i Q_{ni'}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale for each $n \geqslant 1$ and hence $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale difference array for each $n \geqslant 1$. Moreover, it satisfies the conditions of Proposition 2.7.1 as shown in the following technical result.

Proposition 2.7.1. *If $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale difference array for each $n \geqslant 1$ satisfying conditions*

$$\sum_{i=1}^\infty \mathbb{E}\,|Q_{ni}|^2 \leqslant 1 \quad \textit{for all } n \geqslant 1, \tag{2.42}$$

$$\sum_{i=1}^\infty Q_{ni}^2 = 1 + o_p(1), \tag{2.43}$$

$$\sup_{i\geqslant 1} |Q_{ni}| = o_p(1) \tag{2.44}$$

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

Proof. See Awad [1981]. □

Note that this result has been also applied by Ghorai [1980] to establish asymptotic normality of an orthogonal series type density estimator. Indeed, the following proof is similar to the proof of Lemma 2 of Ghorai [1980].

Lemma 2.7.2. *Let* Q_{ni} *be defined as in* (2.41). *Let Assumptions 2.1–2.4 be satisfied and assume* $\big(\sum_{j=1}^{m_n} \tau_j\big)^3 = o(n)$. *Then conditions* (2.42)–(2.44) *hold true.*

Proof. Proof of (2.42). Observe that $\mathbb{E}[X_{1i}X_{1i'}] = 0$ for $i \neq i'$ and thus, for $i = 2, \dots, n$ we have

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

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

$$\sum_{i=1}^{n}\mathbb{E}\,|Q_{ni}|^2 = \frac{2}{n^2}\sum_{i=1}^{n-1} i = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n} \tag{2.45}$$

which proves (2.42).

Proof of (2.43). Using relation (2.45) we observe

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

Consider I_n. Observe that

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

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

$$I_n \leqslant \frac{4}{n^4 \varsigma_{m_n}^4}\Big(\sum_{j=1}^{m_n}\tau_j\Big)^3\Big(\frac{n(n-1)}{2}\sum_{j=1}^{m_n}\tau_j(\mathbb{E}\,|Uf_j(W)|^4)^2 + n(n-1)(n-2)\sum_{j=1}^{m_n} s_{jj}^2\,\mathbb{E}\,|Uf_j(W)|^4\Big)$$

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

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

Consider $A_{ii'}$. Exploiting relation (2.45) 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

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

Moreover, applying Cauchy Schwarz's inequality twice gives

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

Thereby, it holds $2\sum_{i<i'}\mathbb{E}\,A_{ii'} = 1 + o(1)$. Now consider $B_{ii'}$. Since $\{f_l\}_{l\geqslant 1}$ forms an

orthonormal basis on the support of W we obtain

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

This, together with relation (2.45), 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

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

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

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

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

Proof of (2.44). Note that $\mathbb{P}\big(\sup_{i\geqslant 1}|Q_{ni}| > \varepsilon\big) \leqslant \sum_{i=1}^n \mathbb{P}(Q_{ni}^2 > \varepsilon^2)$ and, hence the assertion follows from the Markov inequality. $\square$

3 Specification Testing in Nonparametric Instrumental Quantile Regression

3.1 Introduction

Regression models that involve instrumental variables are widely used in economics to overcome endogeneity problems. In these models, assuming additive separable structural disturbances can often not be justified by the data. This is why their nonseparable extension has been studied extensively in the recent past. Under certain key conditions the nonseparable model implies a conditional quantile restriction. This quantile representation is used in the literature to obtain identification and estimation results. If one of the key conditions is violated, however, the quantile regression representation is potentially misspecified as we illustrate below. This paper contributes to the literature a specification test in nonparametric instrumental quantile regression. In addition, we develop several tests to justify model simplification.

We consider the nonseparable model

$$Y = h(Z, V) \tag{3.1}$$

where Y is a scalar dependent variable, Z a vector of regressors, and V an unobserved scalar variable. Here Z contains potentially endogenous entries in the sense that Z and V may not be independently distributed. Typically in the literature, one finds the following key conditions: an instrumental variable W is available being independent of V, the function h is strictly monotonic in its second argument (wlog strictly monotonically increasing), and V is continuously distributed (wlog $V \sim \mathcal{U}[0,1]$). Given these conditions, for any $0 < q < 1$ the quantile structural effect $\varphi(\cdot) := h(\cdot, q)$ satisfies

$$\mathbb{P}(Y \leqslant \varphi(Z)|W) = \mathbb{P}(h(Z,V) \leqslant h(Z,q)|W) = \mathbb{P}(V \leqslant q|W) = q,$$

and hence we obtain the following quantile regression representation (Horowitz and

Lee [2007])

$$Y = \varphi(Z) + U \quad \text{with} \quad \mathbb{P}(U \leqslant 0|W) = q. \tag{3.2}$$

Research on identification and estimation in nonparametric instrumental quantile regression has been active in the last decade. Chesher [2003] investigated nonparametric identification of derivatives of the unknown functions in a triangular array structure. Chernozhukov and Hansen [2005] and Chernozhukov et al. [2007] give identification conditions and develop a nonparametric minimum distance estimator. In that model, Horowitz and Lee [2007] propose an estimator based on Tikhonov regularization. Chen and Pouzo [2012] study penalized sieve minimum distance estimator. Dunker et al. [2011] consider an iteratively regularized Gauß-Newton method to solve nonlinear inverse problems involving instrumental variables. Gagliardini and Scaillet [2012b] study the asymptotic distribution of a Tikhonov regularized estimator. All these papers assume that a solution to the model equation exists.

Specification tests in instrumental variable models is a subject of considerable literature (the earliest work goes back to Sargan [1958] which together with Hansen [1982] are known as the test of overidentifying restrictions). In the context of nonparametric instrumental mean regression, tests for correct specification have been proposed by Gagliardini and Scaillet [2007], Horowitz [2012], and see also Chapter 2. On the other hand, Horowitz and Lee [2009] established a test of parametric specification of φ in model (3.2). But, as we far as we know, its nonparametric extension, that is whether a function φ solving (3.2) exists, has not yet been addressed in the literature.

Indeed, there exist several sources for misspecification in model (3.2). Model (3.1) need not to imply the quantile regression model (3.2) if the instrument W is not valid (that is W is not independent of V), h is not strictly monotonic in its second argument or V is not continuously distributed. Under violation of one of the key conditions the correct quantile structural representation is given by

$$Y = \varphi(Z) + U \quad \text{with} \quad \mathbb{P}(U \leqslant 0|W) = q + \xi(W) \tag{3.3}$$

for some function ξ. If h is not strictly monotonic in its second argument we have $\xi(W) = \mathbb{P}(h(Z,V) \leqslant h(Z,q)|W) - q$. If h satisfies the monotonicity condition but the instrument W is not valid then $\xi(W) = \mathbb{P}(V \leqslant q|W) - q$. Moreover, if h is monotonic and W is a valid instrument but V is not continuously distributed then $\xi(W) \equiv \mathbb{P}(V \leqslant q) - q$. Consequently, any estimator of φ in the misspecified model (3.2) does not account for the additive term $\xi(W)$ and hence might not converge to the true structural quantile effect. The aim of this paper is to test whether model (3.2) is correctly specified, that is $\xi \equiv 0$.

We also provide extensions of our results concerning model simplification. We establish a test of exogeneity of the regressors Z in the quantile regression model (3.2), that is whether $\mathbb{P}(U \leqslant 0|Z) = q$. Falsely assuming exogeneity of the regressors leads to inconsistent estimators whereas treating exogenous regressors as if they were endogenous can lower the accuracy of estimation dramatically. Moreover, we propose a test of dimension reduction, that is whether certain regressors can be omitted from the quantile structural effect φ. A test of dimension reduction may reduce the complexity of the model and help to increase the accuracy of estimators of the structural effect φ. In particular, by justifying a withdrawal of regressors that are only weakly correlated to the instrument one might obtain identification in the restricted specification while φ is not identified in the original model. Further, a test of additivity of the structural function is established.

Our test statistic is based on the $\mathcal{L}^2$ norm of the empirical conditional quantile restriction. We establish the asymptotic distribution of our test statistic and its consistency against fixed alternatives. Also uniform consistency over certain classes of functions can be obtained. By Monte Carlo simulations we demonstrate the power properties of our test in finite samples. As an empirical illustration, we study a nonparametric median regression model of the effects of class size on test scores of 4th grade students in Israel. We reject the hypothesis of exogeneity of class size at the 0.5–quantile but fail to reject the hypothesis of correct specification of an instrumental median regression model.

3.2 The test statistic and its asymptotic properties

This section begins with the definition of the test statistic and states assumptions required to obtain its asymptotic distribution under the null hypothesis. Further, we show that the penalized sieve estimator of Chen and Pouzo [2012] can be used to estimate the structural effect φ. Moreover, we study power and consistency properties of our test.

3.2.1 Definition of the test statistic

The quantile regression model (1.1a) leads to a nonlinear operator equation, as we see in the following. Let Φ be a Banach space and let us introduce the Hilbert space $\mathcal{L}^2_W := \{\psi : \|\psi\|^2_W := \mathbb{E}\,|\psi(W)|^2 < \infty\}$. Then we define the nonlinear operator $\mathcal{T} : \Phi \to \mathcal{L}^2_W$ with

$$\mathcal{T}\phi = \mathbb{E}[\mathbb{1}\,\{Y \leqslant \phi(Z)\} - q|W] \tag{3.4}$$

for any $\phi \in \Phi$ where $\mathbb{1}$ denotes the indicator function. Thereby, model (1.1a) can be rewritten as the operator equation $\mathcal{T}\varphi = 0$. Throughout the paper, we assume that the function φ defined by the operator equation is identified.

In many economic applications, for instance when estimating a demand function or Engel curves, the structural function of interest may be assumed to be smooth. This *a priori* knowledge is captured by a set $\mathcal{B}$ which we introduce below. Thereby, we consider the null hypothesis

$$H_0: \text{ there exists a function } \varphi \in \mathcal{B} \text{ such that } \mathcal{T}\varphi = 0. \tag{3.5}$$

The alternative is that for any $\varphi \in \mathcal{B}$ it holds $\mathcal{T}\varphi = \xi$ for some nonzero, bounded function ξ.

To motivate the test statistic assume that the support of W is contained in $[0,1]$. Let p_W denote the marginal density of W. By carrying out an additional smoothing step one might obtain better power properties as we discuss below. Let L be a nonsingular smoothing operator L acting on $\mathcal{L}^2_W$. Then $(\mathcal{T}\varphi)(w) = 0$ if and only if $(L\mathcal{T}\varphi)(w)p_W(w) = 0$ for all w in the support of W. Our test statistic is based on a sample analog of $\int_0^1 |(L\mathcal{T}\varphi)(w)p_W(w)|^2 dw$. Further, let us introduce approximating functions $\{f_j\}_{j\geqslant 1}$ which are assumed to form an orthonormal basis $\mathcal{L}^2_{[0,1]}$. In what follows, let $\{\tau_j, f_j\}_{j\geqslant 1}$ be the eigenvalue decomposition of the smoothing operator L with $\tau := (\tau_j)_{j\geqslant 1}$ being a positive nonincreasing sequence. Further, due to Parseval's identity the following representation holds true

$$\int_0^1 |(L\mathcal{T}\varphi)(w)p_W(w)|^2 dw = \sum_{j=1}^{\infty} \tau_j \, \mathbb{E}\left[\left(\mathbb{1}\{Y \leqslant \varphi(Z)\} - q\right) f_j(W)\right]^2. \tag{3.6}$$

Throughout the paper, we assume that an independent and identically distributed n-sample of (Y, Z, W) is available. Further, we truncate the infinite sum on the right hand side of (3.6) and replace the expectation by sample mean. Our test statistic is then given by

$$S_n := \sum_{j=1}^{m_n} \tau_j \Big| n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n(Z_i)\} - q\right) f_j(W_i)\Big|^2 \tag{3.7}$$

where m_n tends to infinity as $n \to \infty$ and $\widehat{\varphi}_n$ is an estimator of φ. Here, $\widehat{\varphi}_n$ can be any consistent estimator of φ satisfying a certain convergence condition which we specify below. We will see that this condition is satisfied by the estimator of Chen and Pouzo [2012]. We reject the hypothesis H_0 if nS_n becomes too large. Further note that if no additional smoothing is carried out, that is L is the identity, then $\tau_j = 1$ for all $j \geqslant 1$.

Remark 3.2.1 (Parametric Specification). *Given some finite integer k assume that functions $e_1, \dots, e_k$ are available such that $\varphi(\cdot) = \sum_{j=1}^{k} \vartheta_j e_j(\cdot)$ for some unknown coefficients $\vartheta_1, \dots, \vartheta_k$. To test the parametric specification $\mathcal{T}\varphi = 0$ one may proceed as in Horowitz and Lee [2009] where the vector of coefficients is estimated via the instrumental quantile regression estimator of Chernozhukov and Hansen [2006]. In the general case, however, a finite dimensional approximation might be not accurate for consistently estimating the structural function φ. Therefore, we require the dimension parameter k to tend to infinity as the sample size increases.* $\square$

3.2.2 Assumptions and notation.

To obtain asymptotic normality as $m_n \to \infty$ we need to standardize our test statistic S_n by appropriate mean and variance, which we introduce in the following definition.

Definition 3.2.1. *For all $m \geqslant 1$ let Σ_m denote a $m \times m$ matrix with the entries $s_{jl} = \sqrt{\tau_j \tau_l}\, \mathbb{E}\left[f_j(W) f_l(W)\right]$, $1 \leqslant j, l \leqslant m$. Then the trace and the Frobenius norm of Σ_m are respectively denoted by*

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

In order to obtain our asymptotic result we state the following assumptions. Our first assumption gathers conditions which we require for the basis $\{f_j\}_{j \geqslant 1}$ in $\mathcal{L}^2_W$. In the following, let $\{f_j\}_{j \geqslant 1}$ form an orthonormal basis on the support $\mathcal{W}$ of W with respect to the Lebesgue measure ν (that is, $\int_{\mathcal{W}} f_j(w) f_l(w) \nu(dw) = 1$ if $j = l$ and zero otherwise).

Assumption 3.1. *There exists some constants $\eta_f, \eta_p \geqslant 1$ such that*

$$\sup_{l \geqslant 1} \Big(\int_{\mathcal{W}} |f_l(s)|^4 \nu(ds) \Big) \leqslant \eta_f \quad \textit{and} \quad \sup_{w \in \mathcal{W}} p_W(w) \leqslant \eta_p.$$

Assumption 3.1 holds for sufficiently large η_f if the basis $\{f_l\}_{l \geqslant 1}$ is uniformly bounded, such as trigonometric bases or B-splines that have been orthogonalized. Moreover, Assumption 3.1 holds in case of Hermite polynomials.

Our methodology requires conditions specifying the relation of the nonlinear operator $\mathcal{T}$ to its linearization. In the following, we denote the Fréchet derivative of $\mathcal{T}$ at φ by

$$T\phi := \int_{\mathcal{Z}} p_{Y,Z|W}(\varphi(z), z, W) \phi(z) dz$$

where $p_{Y,Z|W}$ denotes the density of (Y, Z) given W. Let $D_{\mathcal{B}}$ denote a distance on $\mathcal{B}$. Throughout the paper, we assume that $D_{\mathcal{B}}$ is bounded by the supremum norm. Let us specify an upper bound on the Taylor remainder of $\mathcal{T}$.

Assumption 3.2. *There exists some constant $\eta > 0$ such that for all functions $\phi \in \mathcal{U}_\varepsilon(\varphi) = \{\phi \in \mathcal{B} : \, D_{\mathcal{B}}(\phi, \varphi) \leqslant \varepsilon\}$ for some $\varepsilon > 0$ it holds*

$$\|\mathcal{T}\phi - \mathcal{T}\varphi - T(\phi - \varphi)\|_W^2 \leqslant \eta \|\mathcal{T}\phi - \mathcal{T}\varphi\|_W^2.$$

Assumption 3.2 is also known as the tangential cone condition and frequently used in the analysis of nonlinear operator equations (for a further discussion and examples we refer to Hanke et al. [1995]).

Assumption 3.3. *Let $\mathcal{B}_n := \{\phi \in \mathcal{B} : \, D_{\mathcal{B}}(\phi, \varphi) \leqslant \mathcal{R}_n\}$ then we assume that there exist constants $C > 0$ and $\kappa \in (0, 1]$ such that*

$$\max_{1 \leqslant j \leqslant m_n} \mathbb{E}\Big[\sup_{\phi \in \mathcal{B}_n} \big| \mathbb{1}\{Y \leqslant \varphi(Z)\} - \mathbb{1}\{Y \leqslant \phi(Z)\}\big|^2 f_j^2(W)\Big] \leqslant C\mathcal{R}_n^{2\kappa}. \tag{3.8}$$

Assumption 3.3 states that the function $\varphi \mapsto (\mathbb{1}\{Y \leqslant \varphi(Z)\} - q) f_j(W)$, $1 \leqslant j \leqslant m_n$, is locally uniformly $\mathcal{L}_W^2$ continuous. This condition has also been exploited by Chen et al. [2003] (Theorem 3), Chen [2007] (Lemma 4.2 (i)) or Chen and Pouzo [2012] (Remark c.1).

Assumption 3.4. *There exists an estimator $\widehat{\varphi}_n \in \mathcal{B}$ of φ based on an iid. n-sample of (Y, Z, W) from the model* (1.1a) *such that (i) $D_{\mathcal{B}}(\widehat{\varphi}_n, \varphi) = o_p(\mathcal{R}_n)$ with $\mu_{m_n} \mathcal{R}_n^{\kappa} = o(\varsigma_{m_n})$ where κ is as in* (3.8) *and (ii) $n\|T(\widehat{\varphi}_n - \varphi)\|_W^2 = o_p(\varsigma_{m_n})$.*

In Assumption 3.4, condition (i) requires to choose m_n such that the $D_{\mathcal{B}}$–rate of convergence of $\widehat{\varphi}_n$ has a stronger decay than $\varsigma_{m_n} \mu_{m_n}^{-1}$. Assumption 3.4 (ii) ensures that the difference of $\widehat{\varphi}_n - \varphi$ in our test statistic is asymptotically negligible. As we see in subsection 3.2.4, Assumption 3.4 (ii) is satisfied by the estimator of Chen and Pouzo [2012].

In the following, we describe smoothness conditions imposed on the structural function φ which is captured by the set $\mathcal{B}$. Let Z have support $\mathcal{Z} \subset \mathbb{R}^d$ and for any vector of nonnegative integers $k = (k_1, \dots, k_d)$ define $|k| = \sum_{j=1}^d k_j$ and $D^k = \delta^{|k|}/(\delta z_1^{k_1} \dots \delta z_d^{k_d})$. For some integer $p > 0$ we define the norms

$$\|\phi\|_{\alpha,p} = \Big(\sum_{|k| \leqslant \alpha + \alpha_0} \int_{\mathcal{Z}} |D^k \phi(z)|^p dz\Big)^{1/p}, \quad \|\phi\|_{\alpha,\infty} = \max_{|k| \leqslant \alpha} \sup_{z \in \mathcal{Z}} |D^k \phi(z)|$$

where α and α_0 are positive integers. We denote the Sobolev spaces associated with the norm $\|\cdot\|_{\alpha,p}$ by

$$W^{\alpha,p} := \{\phi : \mathcal{Z} \to \mathbb{R} : \|\phi\|_{\alpha,p} < \infty\}. \tag{3.9}$$

For some constant $\rho > 0$, define $\mathcal{B}$ as the Sobolev ellipsoid of radius ρ given by

$$\mathcal{B} := \{\phi \in W^{\alpha,p} : \|\phi\|_{\alpha,p} \leqslant \rho\}. \tag{3.10}$$

The following assumption gathers regularity conditions imposed on the structural functions φ and the supports $\mathcal{Z}$ of Z and $\mathcal{W}$ of W.

Assumption 3.5. *(i) Let $\alpha_0 > d/p$ and $\alpha > \min(d/\kappa, d/2)$ where the constant $\kappa > 0$ satisfies equation* (3.8). *(ii) $\mathcal{Z}$ is bounded, convex and satisfies a uniform cone property. (iii) $\mathcal{W}$ is bounded and the marginal density p_W is uniformly bounded away from zero on $\mathcal{W}$.*

Assumption 3.5 (i) requires α to be large if (3.8) holds only for small $\kappa > 0$ or the dimension d is large. Assumption 3.5 (ii) imposes a weak regularity condition on the shape of $\mathcal{Z}$. For the uniform cone property see, for instance, Paragraph 4.4 in Adams and Fournier [2003]. This property was also used by Santos [2012].

Example 3.2.1. *Let $F_{Y|ZW}$ denote the cumulative distribution function of Y given (Z, W) and assume that it is Lipschitz continuous with constant $C_L > 0$, i.e., $|F_{Y|ZW}(y) - F_{Y|ZW}(y')| \leqslant C_L|y - y'|$ for all (y, y'). Due Assumption 3.5 the Sobolev space $W^{\alpha,p}$ can be embedded in $W^{\alpha,\infty}$ (cf. Adams and Fournier [2003]). In particular, $D_{\mathcal{B}}(\phi, \varphi) = \|\phi - \varphi\|_\infty = \sup_{z\in\mathcal{Z}} |\phi(z) - \varphi(z)|$ is bounded on $\mathcal{B}$ and moreover, Assumption 3.3 holds true. Indeed, by following Chen et al. [2003] (page 1599 – 1600) we observe for each $1 \leqslant j \leqslant m_n$*

$$\begin{aligned}
&\mathbb{E}\,[\sup_{\phi\in\mathcal{B}_n} |\,\mathbb{1}\{Y \leqslant \varphi(Z)\} - \mathbb{1}\{Y \leqslant \phi(Z)\}|^2 f_j^2(W)] \\
&\qquad\qquad \leqslant \mathbb{E}\left[\left(\,\mathbb{1}\{Y \leqslant \varphi(Z) + \mathcal{R}_n\} - \mathbb{1}\{Y \leqslant \varphi(Z) - \mathcal{R}_n\}\right) f_j^2(W)\right] \\
&\qquad\qquad = \mathbb{E}\left[\left(F_{Y|ZW}(\varphi(Z) + \mathcal{R}_n) - F_{Y|ZW}(\varphi(Z) - \mathcal{R}_n)\right) f_j^2(W)\right] \leqslant C_L\mathcal{R}_n
\end{aligned}$$

which implies Assumption 3.3 with $\kappa = 1/2$. In addition, if L is the identity operator then condition 3.4 (i) is equivalent to $\mathcal{R}_n = o(m_n^{-1})$. □

3.2.3 Asymptotic distribution under the null hypothesis

In this section, we establish asymptotic normality of our test statistic. The following theorem establishes asymptotic normality of S_n after standardization by μ_{m_n} and ς_{m_n}.

Theorem 3.2.1. *Let Assumptions 3.1–3.5 be satisfied. Further, if*

$$\varsigma_{m_n}^{-1} = o(1) \quad \textit{and} \quad \Big(\sum_{j=1}^{m_n} \tau_j\Big)^3 = o(n) \tag{3.11}$$

then we have under H_0

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

3.2.4 Estimating procedure

In the previous section, we derived the asymptotic distribution of our test statistic under the condition $n\|T(\widehat{\varphi}_n - \varphi)\|_W^2 = o_p(\varsigma_{m_n})$. In this section, we present an estimation method for the structural function φ that satisfies this condition. We discuss a penalized sieve minimum distance estimator suggested by Chen and Pouzo [2012].

Chen and Pouzo [2012] propose a class of estimators, which are minimizers of a penalized minimum distance criterion over a collection of sieve spaces. Given a sequence of known basis functions $\{e_j\}_{j\geqslant 1}$ the series least square estimator of $\mathcal{T}\phi$ is given by

$$(\widehat{\mathcal{T}}\phi)(\cdot) = e_{\underline{l_n}}(\cdot)'\Big(\sum_{i=1}^n e_{\underline{l_n}}(W_i) e_{\underline{l_n}}(W_i)^t\Big)^{-1} \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q) e_{\underline{l_n}}(W_i) \tag{3.12}$$

where the integer l_n grows slowly with the sample size n. Chen and Pouzo [2012] consider the following penalized sieve minimum distance estimator, $\widehat{\varphi}_n$, defined as

$$\widehat{Q}_n(\widehat{\varphi}_n) \leqslant \inf_{\phi\in\mathcal{B}_{k_n}} \widehat{Q}_n(\phi) \tag{3.13}$$

where

$$\widehat{Q}_n(\phi) = n^{-1}\sum_{i=1}^n (\widehat{\mathcal{T}}\phi)(W_i)'\,(\widehat{\mathcal{T}}\phi)(W_i) + \lambda_n \mathcal{P}(\phi). \tag{3.14}$$

Here $\mathcal{B}_{k_n}$ is a closed sieve parameter space whose complexity (denoted as $k_n := \dim(\mathcal{B}_{k_n})$) grows with sample size n and becomes dense in the original function space $\mathcal{B}$ ($\mathcal{B}_k \subseteq \mathcal{B}_{k+1} \subseteq \mathcal{B}$), $\lambda_n \geqslant 0$ is a penalization parameter such that $\lambda_n \to 0$ as $n \to \infty$, and penalty function $\mathcal{P} : \mathcal{B} \to [0, \infty)$. One could also generalize the results to an empirical analog of the penalty. Let $D_{\mathcal{B}}$ be a norm on $\mathcal{B}$ and hence, we write $D_{\mathcal{B}}(\phi, \varphi) = \|\phi - \varphi\|_{\mathcal{B}}$ for any $\phi \in \Phi$. In the following, let $E_{k_n}\varphi \in \mathcal{B}_{k_n}$ such that $\|E_{k_n}\varphi - \varphi\|_{\mathcal{B}} = o(1)$ and assume $\|\mathcal{T} E_{k_n}\varphi\|_W = o(1)$.

Chen and Pouzo [2012] establish consistency of the estimator $\widehat{\varphi}_n$ given in (3.13) at a certain convergence rate. In the following, we assume that $\|\widehat{\varphi}_n - \varphi\|_{\mathcal{B}} = o_p(\mathcal{R}_n)$ where $\mu_{m_n}\mathcal{R}_n = o(\varsigma_{m_n})$ which can be ensured by a moderate choice of m_n. Further, we restrict the class of functions to the shrinking sets

$$\mathcal{B}_o = \{\phi \in \mathcal{B} : \|\phi - \varphi\|_{\mathcal{B}} \leqslant \mathcal{R}_n, \|\phi\|_{\mathcal{B}} \leqslant C_1, \lambda_n\mathcal{P}(\phi) \leqslant \lambda_n C_2\} \quad \text{and} \quad \mathcal{B}_{ok_n} = \mathcal{B}_o \cap \mathcal{B}_{k_n}$$

for some constants $C_1, C_2 > 0$. In the following, $\|\cdot\|$ denotes the euclidean norm.

Assumption 3.6. *(i) $\mathcal{B}_o$ and $\mathcal{B}_{ok_n}$ are convex; (ii) there is a constant $C > 0$ such that $\sup_y \sup_{(z,w)\in\mathcal{Z}\times\mathcal{W}} p_{Y|Z,W}(y, z, w) \leqslant C$; (iii) $\mathcal{Z}$ is compact with Lipschitz continuous boundary, and the marginal density of Z is bounded and bounded away from zero over $\mathcal{Z}$. (iv) $\sup_{j\geqslant 1} E|e_j(Z)|^2 \leqslant \eta$ for some constant $\eta > 0$, the smallest eigenvalue of $E[e_{\underline{k}}(Z)e_{\underline{k}}(Z)^t]$ is bounded away from zero for all $k \geqslant 1$, and $\sup_{z\in\mathcal{Z}} \|e_{\underline{k_n}}(z)\|^2 = o(n/k_n)$.*

Let us introduce a sequence $(\delta_n)_{n\geqslant 1}$ satisfying $\delta_n^2 = \max\left(n^{-1}l_n, b_{l_n}\right)$ where b_{l_n} is the order of the bias of the least square estimator $\widehat{\mathcal{T}}$. In the following result we show that the estimator $\widehat{\varphi}_n$ given in (3.13) converges sufficiently fast to φ in the pseudo norm induced by T. Note that sufficient conditions for consistency of $\widehat{\varphi}_n$ are given in Section 3 of Chen and Pouzo [2012].

Proposition 3.2.2. *Consider the estimator $\widehat{\varphi}_n$ given in* (3.13) *with $\|\widehat{\varphi}_n - \varphi\|_{\mathcal{B}} = o_p(\mathcal{R}_n)$. Let $\varphi \in \mathcal{B}_o$ and $E_{k_n}\varphi \in \mathcal{B}_{ok_n}$, and let Assumptions 3.2 with $0 < \eta < 1$ and 3.6 hold. If*

$$n \max\left(\delta_n^2, \lambda_n, \|T(E_{k_n}\varphi - \varphi)\|_W^2\right) = o(\varsigma_{m_n}) \tag{3.15}$$

then we have

$$n\|T(\widehat{\varphi}_n - \varphi)\|_W^2 = o_p(\varsigma_{m_n}).$$

Remark 3.2.2. *Chen and Pouzo [2012] prove under mild regularity assumptions on the joint joint distribution of (Y, Z, W) in Lemma C.2 that $\delta_n^2 = \max\left(n^{-1}l_n, l_n^{-2\beta/\dim(W)}\right)$ as long as $\mathcal{T}\phi$ belongs to a Hölder space with Hölder parameter β. In this case, condition $n\delta_n^2 = o(\varsigma_{m_n})$ is satisfied if $l_n = o(\varsigma_{m_n})$ and $n = o(\varsigma_{m_n} l_n^{2\beta/\dim(W)})$. On the other hand, Chen and Pouzo [2012] assume in the Hilbert space case the link condition $\|T(E_{k_n}\varphi -$*

$\varphi)\|_W^2 \leqslant \upsilon_{k_n}\|E_{k_n}\varphi - \varphi\|_{\mathcal{B}}^2$ for some positive nonincreasing sequence $(\upsilon_j)_{j\geqslant 1}$. Moreover, with $\varphi \in W^{\alpha,2}$ and the norm $\|\phi\|_{\mathcal{B}} = \big(\int_0^1 \phi^2(z)dz\big)^{1/2}$ for all $\phi \in \mathcal{B}$ it is well known that $\|E_{k_n}\varphi - \varphi\|_{\mathcal{B}} = O(k_n^{-\alpha/d})$ for splines, wavelets, power series, and Fourier series bases. Hence, $n\|T(E_{k_n}\varphi - \varphi)\|_W^2 = o(\varsigma_{m_n})$ holds true if $n\upsilon_{k_n} = o(\varsigma_{m_n} k_n^{2\alpha/d})$. $\square$

Below, we write $a_n \sim b_n$ when there exist constants $c, c' > 0$ such that $cb_n \leqslant a_n \leqslant c'b_n$ for sufficiently large n.

Example 3.2.2. *Consider the Hilbert space setting of Remark 3.2.2 with no additional smoothing, that is $L = Id$. Further, we have $\sqrt{m_n} = O(\varsigma_{m_n})$ and let $\max\{\delta_n^2, \lambda_n\} = \delta_n^2 = k_n/n$ within a constant. Further, let $k_n \sim n^{\chi}$ where $\chi > 0$ is specified in the following two cases.*

(i) Mildly ill-posed case: If $\upsilon_{k_n} \sim k_n^{-2\zeta/d}$ for some $\zeta \geqslant 0$ then in order for (3.15) *to hold we require $m_n \sim n^{\iota}$ with $0 < \iota < 1/3$ and*

$$d(1 - \iota/2)/(2\alpha + 2\zeta) < \chi < \iota/2.$$

(ii) Severely ill-posed case: If $\upsilon_{k_n} \sim \exp\big(-k_n^{2\zeta/d}\big)$ for some $\zeta > 0$ then condition (3.15) *is satisfied if, for example, m_n satisfies $m_n = o(k_n^{\alpha})$ and $k_n^2 = o(m_n)$ where $k_n \sim \big(\log n - \log(m_n^{3/2})\big)^{d/(2\zeta)}$.* $\square$

3.2.5 Limiting behavior under local alternatives

As illustrated in the introduction, a violation of our *key conditions* leads to an additional additive term in the conditional quantile restriction. In the following, we study the power of the test, i.e., the probability to reject a false hypothesis, against a sequence of linear local alternatives that tends to zero as $n \to \infty$. It is shown that the power of our tests essentially relies on the choice of the weighting sequence τ.

We consider the following sequence of linear local alternatives

$$Y = \varphi(Z) + U \quad \text{where} \quad \mathbb{P}(U \leqslant 0|W) = q + \varsigma_{m_n}^{1/2} n^{-1/2}\xi(W) \tag{3.16}$$

for some bounded function ξ which is strictly positive on $\mathcal{W}$. The null hypothesis H_0 occurs if the function ξ vanishes. The next result establishes asymptotic normality for the standardized test statistic S_n. Let us denote $\xi_j := \sqrt{\tau_j}\,\mathbb{E}[\xi(W)f_j(W)]$.

Proposition 3.2.3. *Given the conditions of Theorem 3.2.1 it holds under* (3.16)

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

From Proposition 3.2.3 we see that our test can detect local linear alternatives at a rate which becomes arbitrarily close to $n^{-1/2}$ as the degree of smoothing is increased.

3.2.6 Consistency against a fixed alternative

Let us first establish consistency when H_0 does not hold, that is, there exists no solution to (3.5) belonging to $\mathcal{B}$. 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. Note that since p_W bounded away from zero over $\mathcal{W}$ we have $\|L\mathcal{T}\phi\|_W > 0$ for any alternative function ϕ.

Proposition 3.2.4. *Assume that H_0 does not hold. Consider a sequence $(\gamma_n)_{n\geqslant 1}$ satisfying $\gamma_n = o(n\varsigma_{m_n}^{-1})$. Under the conditions of Theorem 3.2.1 we have*

$$\mathbb{P}\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}\Big(\frac{n}{q(1-q)}\,S_n - \mu_{m_n}\Big) > \gamma_n\Big) = 1 + o(1).$$

3.2.7 Uniform consistency

In the following, we show that our test is consistent uniformly over some appropriate class of functions. This implies that there are no alternative functions in this class over which our test has low power. Let us introduce the class of functions

$$\mathcal{G}_n^\rho = \Big\{\phi \in \varPhi : \|L\mathcal{T}\phi\|_W^2 \geqslant \rho\,\varsigma_{m_n} n^{-1} \text{ and } \phi \text{ is bounded on } \mathcal{Z}\Big\}.$$

Assume that the tangential cone condition (cf. Assumption 3.2 (ii)) holds and consider the Hilbert space case, that is $\varPhi = L_Z^2$, then $\|\mathcal{T}\phi\|_W^2 \leqslant (1+\eta)\|T(\phi-\varphi)\|_W^2 \leqslant (1+\eta)\|\phi-\varphi\|_Z^2$ within a constant. We conclude that $\mathcal{G}_n^\rho$ contains all functions whose L_Z^2-distance to the structural function φ is at least $n^{-1}\varsigma_{m_n}$ within a constant. For the next result let q_α denote the $1-\alpha$ quantile of $\mathcal{N}(0,1)$.

Proposition 3.2.5. *Under the conditions of Theorem 3.2.1 we have for any $\varepsilon > 0$, any $0 < \alpha < 1$, and any sufficiently large constant $\rho > 0$ that*

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

3.3 Extensions

As we see in this section, our testing procedure can potentially be applied to a much wider range of situations. We now discuss several corollaries that generalize the previous results in various ways.

3.3.1 Testing exogeneity

In this subsection, we propose a test whether the vector of regressors Z is exogenous at a quantile $0 < q < 1$, that is $H_0^e : \mathbb{P}(U \leqslant 0|Z) = q$. The null hypothesis H_0^e holds true if and only if there exists a function φ^e such that $\mathbb{P}(Y \leqslant \varphi^e(Z)|Z) = q$ or, equivalently, $\varphi = \varphi^e$ where φ satisfies $\mathcal{T}\varphi = 0$. Further, due to nonsingularity of the operator $\mathcal{T}$ hypothesis H_0^e is equivalent to

$$\mathcal{T}\varphi^e = 0. \tag{3.17}$$

Let us now propose an estimator for the conditional quantile function φ^e. For each $k \geqslant 1$ let $\pi(t)$ be a k–dimensional vector with entries $e_j(t)$ for $1 \leqslant j \leqslant k$. As basis $\{e_j\}_{j\geqslant 1}$ we use B-splines. Then our estimator of φ is given for all $k \geqslant 1$ by

$$\widehat{\varphi}_k^{\mathrm{e}}(\cdot) := e_{k_n}(\cdot)^t \widehat{\beta}_k \quad \text{where} \quad \widehat{\beta}_k = \arg\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n \varrho_q(Y_i - \pi(Z_i)^t \beta) \tag{3.18}$$

where $\varrho_q(u) = |u| - (2q-1)u$ is the check function. This estimator was studied by He and Shi [1994].

Assumption 3.7. *(i) Assume that Z is scalar and continuously distributed with $\mathcal{Z} \subset [0,1]$. (ii) There exist constants C, $c > 0$ such that $c \leqslant p_Z(z) \leqslant C$ for all $z \in [0,1]$. (iii) Under H_0^{e} the random variable U has a density function which is strictly positive at zero. (iv) For some constant $C > 0$ it holds $\sup_y \sup_{(z,w)\in\mathcal{Z}\times\mathcal{W}} p_{Y|Z,W}(y,z,w) \leqslant C$.*

Our test statistic to validate the null hypothesis H_0^e is given by S_n but where the estimator $\widehat{\varphi}_n$ is replaced by the estimator $\widehat{\varphi}_k^{\mathrm{e}}$ of the conditional quantile function. That is

$$S_n^{\mathrm{e}} := \sum_{j=1}^{m_n} \tau_j \Big|n^{-1} \sum_{i=1}^n \big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n^e(Z_i')\} - q\big) f_j(W_i)\Big|^2.$$

We reject the hypothesis H_0^e if nS_n^{e} becomes too large. The next result establishes asymptotic normality of our test statistic S_n^{e} under the null hypothesis.

Corollary 3.3.1. *Let Assumptions 3.1–3.3, 3.5, and 3.7 hold true. Let m_n satisfy condition* (3.11) *of Theorem 3.2.1. Consider the estimator $\widehat{\varphi}_{k_n}^e$ given in* (3.18) *where k_n satisfies*

$$k_n = o(\varsigma_{m_n}), \quad \mu_{m_n} = o_p(k_n^{r\kappa}\varsigma_{m_n}) \quad \textit{and} \quad n = o_p(k_n^{2r}\varsigma_{m_n}) \tag{3.19}$$

where $r = \alpha - 1/p$. Then we have under H_0^e

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

Example 3.3.1. *Let us illustrate when condition* (3.19) *holds true. We consider the case when no additional smoothing is applied, that is L is the identity. Hence, we have $\sqrt{m_n} = O(\varsigma_{m_n})$. Further, let $m_n \sim n^\iota$ with $0 < \iota < 1/3$. Then for* (3.19) *to hold let $k_n \sim n^\chi$ where $\chi > 0$ satisfies*

$$\max\Big(\frac{1-\iota/2}{2r}, \frac{\iota}{2r\kappa}\Big) < \chi < \iota/2.$$

Hence we require $r > 1/\kappa$ which a slightly stronger restriction than Assumption 3.5 (i). In addition $r > 1/\iota - 1/2$ is required. □

3.3.2 Testing dimension reduction

In applications one might want to reduce the complexity of the model by omitting certain regressors from the structural function φ. In this sense, we propose a test of dimension reduction: Let Z' be a vector containing only entries of Z with $\dim(Z') < \dim(Z)$ then the hypothesis under consideration is given by

$$H_0' : \text{there exists a function } \varphi \in \mathcal{B} \text{ only depending on } Z' \text{ such that } \mathcal{T}\varphi = 0. \quad (3.20)$$

The alternative is that there exists no function in $\mathcal{B}$ depending on Z' and being root of $\mathcal{T}\varphi = 0$. In order to validate the null hypothesis H_0' we consider the test statistic

$$S_n' := \sum_{j=1}^{m_n} \big|n^{-1}\sum_{i=1}^{n}\big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n(Z_i')\} - q\big) f_j^{\tau}(W_i)\big|^2$$

where $\widehat{\varphi}_n$ is an estimator of φ based on an iid. sample $(Y_1, Z_1', W_1), \ldots, (Y_n, Z_n', W_n)$ of (Y, Z', W). It is clear that by reducing the dimension of the regressor Z we can weaken conditions on the instruments in order to obtain identification (cf. Example 2.5.3 in case of mean regression). The next asymptotic normality result is a direct consequence of Theorem 3.2.1 and hence its proof is omitted.

Corollary 3.3.2. *Given the condition of Theorem 3.2.1 we have under H_0'*

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

Let us now provide two examples which demonstrate the usefulness of our dimension reduction test.

Example 3.3.2 (Test of exogeneity). *Let* $Z = (Z', Z'')$ *where* Z' *is exogenous, that is* $\mathbb{P}(U \leqslant 0|Z') = q$. *Via a test of hypothesis* H_0' *one may justify whether the potentially endogenous part of* Z *can be omitted from the structural function* φ. *To be more precise, the testing procedure is based on the test statistic* S_n' *where* $\widehat{\varphi}$ *could be the b-spline estimator* (3.18). □

Example 3.3.3 (Identification). *Assume there exists some* $\lambda \in (0,1)$ *such that*

$$\mathcal{T}\phi - \mathcal{T}\psi = \mathbb{E}\left[h(Z,W)(\phi - \psi)(Z)\big|W\right]$$

where $h(Z,W) = p_{Y|ZW}\big((1-\lambda)\psi(Z) + \lambda\phi(Z), Z, W\big)$ *and* $p_{Y|ZW}$ *denotes the density of* Y *conditional on* Z *and* W. *Let* $Z = (Z', Z'')$ *where* Z' *is independent of* Z'' *and* Z'' *is independent of* W. *Let* ϕ *and* ψ *only depend on* Z''. *Assume that there exist non-negative functions* ν_1 *and* ν_2 *such that* $h(Z,W) = \nu_1(Z',W)\nu_2(Z'')$ *and* $\nu_2(Z'') = -\nu_2(-Z'')$ *then*

$$\begin{aligned}\mathcal{T}\phi - \mathcal{T}\psi &= \mathbb{E}\left[\nu_1(Z',W)\,\mathbb{E}[\nu_2(Z'')(\phi-\psi)(Z'')|W,Z']|W\right]\\ &= \mathbb{E}\left[\nu_1(Z',W)\,\mathbb{E}[\nu_2(Z'')(\phi-\psi)(Z'')]\big|W\right] = 0\end{aligned}$$

for any distribution of Z'' *being symmetric to zero and any symmetric function* $\phi - \psi$. *We conclude that* φ *is not identified in model* (1.1a). *Applying the test of dimension reduction may justify a withdrawal of the regressor* Z'' *and might lead to identification.* □

3.3.3 Testing additivity

By assuming an additive structure of φ one might reduce the effect of dimensionality of the regressors on the convergence rate of an estimator (cf. Chen and Pouzo [2012] in case of instrumental quantile regression). Applying this structure leads, however, to inconsistent estimators in general if the function φ does not obey an additive form. Our aim in the following is to test whether

$$H_0^{add}: \text{there exist functions } \varphi_1, \varphi_2 \in \mathcal{B} \text{ such that } \mathbb{P}(Y \leqslant \varphi_1(Z') + \varphi_2(Z'')|W) = q. \tag{3.21}$$

Similarly as above we obtain the test statistic

$$S_n^{add} := \sum_{j=1}^{m_n} \left| n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}\left\{ Y_i \leqslant \widehat{\varphi}_{1n}(Z_i') + \widehat{\varphi}_{2n}(Z_i') \right\} - q \right) f_j^{\tau}(W_i) \right|^2$$

where the estimator $(\widehat{\varphi}_{1n}, \widehat{\varphi}_{2n})$ of $\varphi = (\varphi_1, \varphi_2)$ is given by 3.13. The next asymptotic normality result is a direct consequence of Theorem 3.2.1 and hence its proof is omitted.

Corollary 3.3.3. *Given the conditions of Theorem 3.2.1 we have under H_0^{add}*

$$\left(\sqrt{2}\varsigma_{m_n}\right)^{-1}\left(\frac{n}{q(1-q)}\, S_n^{add} - \mu_{m_n}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

3.4 Monte Carlo simulation

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo simulation. The sample size is 1000 and there are 1000 Monte Carlo replications in each experiment. Results are presented for the nominal levels 0.05. Let Φ denote the cumulative standard normal. Throughout this simulation study, realizations (Z, W) were generated by $Z = \Phi(\zeta\omega + (1 - \zeta^2)\varepsilon)$ and $W = \Phi(\omega)$ where ω, $\varepsilon \sim N(0, 1)$, Here, the constant $\zeta > 0$ measures the correlation of Z to W and is varied in the experiments. Realizations of Y were generated from

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

where $U = \vartheta\,\varepsilon + \sqrt{1-\vartheta^2}\,\varepsilon$ with $\varepsilon \sim N(0,1)$ and where the constants $c_U > 0$, $\vartheta > 0$ are varied in the experiments. As basis $\{f_j\}_{j\geqslant 1}$ we choose cosine basis functions given by

$$f_j(t) = \sqrt{2}\cos(\pi j t) \quad \text{for } j = 1, 2, \dots$$

Testing Exogeneity The realizations (Y, Z, W) are generated as described above with $c_U = 0.5$ and structural effect $\varphi_1(z) = \sum_{j=1}^{\infty}(-1)^{j+1} j^{-2}\sin(j\pi z)$. For computational reasons we truncate the infinite sum at $K = 100$. The resulting function is displayed in Figure 1. We estimate the structural relationship using Lagrange polynomials. Note that ϑ measures the degree of endogeneity of Z and is varied among the experiments. The null hypothesis H_0 holds true if $\vartheta = 0$ and is false otherwise.

In Table 2 we depict the empirical rejection probabilities when using either no smoothing or additional smoothing with $\tau_j = j^{-.25}$ or $\tau_j = j^{-.5}$, $j \geqslant 1$, which we denote by S_n^{0e}, $S_n^{0.25e}$ or $S_n^{0.5e}$, respectively. In the simulation study we choose m_n such that

ζ	ϑ	*Empirical Rejection probability using* S_n^{0e}	$S_n^{0.25e}$	$S_n^{0.5e}$
0.4	0.0	0.048	0.045	0.037
	0.2	0.137	0.175	0.208
	0.25	0.254	0.287	0.346
	0.3	0.387	0.446	0.508
	0.35	0.565	0.627	0.690
0.7	0.0	0.034	0.031	0.032
	0.2	0.248	0.298	0.376
	0.25	0.492	0.548	0.652
	0.3	0.764	0.818	0.876
	0.35	0.941	0.956	0.984

Table 3.1: Empirical Rejection probabilities for testing exogeneity

$\sum_{j=1}^{m_n} \tau_j \approx n^{1/3}$. Hence, without additional smoothing the number of basis functions used is $m_n = 10$ $(= n^{1/3})$. On the other hand, with additional smoothing in case, of $\tau_j = j^{-.25}$ we have $m_n = 20$ whereas if $\tau_j = j^{-.5}$ we let $m_n = 100$. We like to emphasize that, especially in the case of additional smoothing, the results of our test statistic are not sensitive to the choice of the number of basis functions. As we see from Table 2, our test becomes slightly with additional smoothing.

Testing a Nonparametric Specification In case of nonparametric specification, we consider the structural function $\varphi_2(z) = \sum_{j=1}^{\infty} j^{-4}\cos(j\pi z)$. Again, for computational reasons we truncate the infinite sum at $K = 100$. The resulting functions are displayed in Figure 1. To estimate the structural function we apply the procedure of Chen and Pouzo [2012] given in (3.12) with b-splines as approximation basis functions. That is, for the sieve space $\mathcal{B}_{k_n}$ we use b-splines of order 2 with 5 knots (hence $k_n = 5$) and for $\widehat{\mathcal{T}}$ we use b-splines of order 6 with 11 knots (hence $l_n = 15$).

If H_0 is false, then $\mathbb{P}(Y \leqslant \varphi(Z)|W) = q + \xi(W)$ for some function ξ. In our experiments, we consider $\xi(W) = -\mathbb{P}(\varphi(Z) < Y \leqslant \varphi(Z) + \rho(Z)|W)$ for some function ρ which we specify below. The definition of ξ implies $\mathbb{P}(Y \leqslant \varphi(Z) + \rho(Z)|W) = q + \xi(W)$. Consequently, when H_0 is false we generate realizations of Y from

$$Y = \varphi(Z) + c_j\,\rho_j(Z) + U$$

for $j = 1, 2, 3, 4$, where

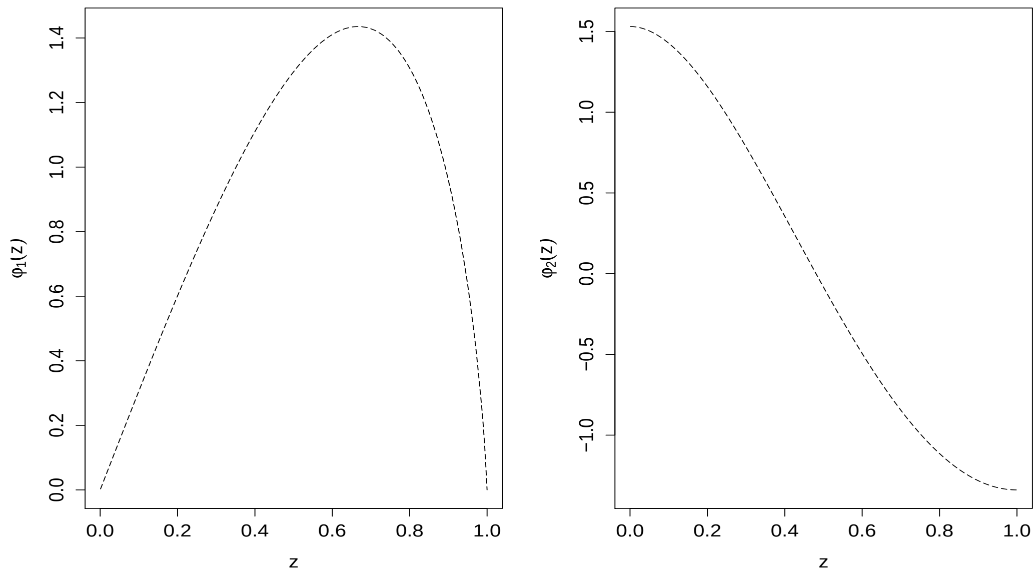

Figure 3.1: Graph of φ_1 and φ_2

$$\rho_1(z) = 1 - (2z - 1)^2,$$

$$\rho_2(z) = z\,\mathbb{1}\{z \leqslant 1/2\} + (1 - z)\,\mathbb{1}\{z > 1/2\},$$

$$\rho_3(z) = \exp(2z)\,\mathbb{1}\{z \leqslant 1/2\} + \exp(2(1 - z))\,\mathbb{1}\{z > 1/2\} - 1,$$

$$\rho_4(z) = \exp(4z)\,\mathbb{1}\{z \leqslant 1/2\} + \exp(4(1 - z))\,\mathbb{1}\{z > 1/2\} - 1,$$

and $c_j > 0$ is a normalizing constant such that $\int_0^1 \rho_j(z)dz = 0.5$ for $j = 1, 2, 3, 4$.

In Table 2, we depict the empirical rejection probabilities when using S_n^{np} with either no smoothing or additional smoothing $\tau_j = j^{-0.25}$, $j \geqslant 1$, or $\tau_j = j^{-5}$, $j \geqslant 1$, which we denote by S_n^{0np}, $S_n^{0.25np}$, or $S_n^{0.5np}$, respectively. The number of cosine basis functions f_j to construct our test statistic is exactly the same as in the setting of the test of exogeneity as described above.

The results of the experiments are shown in Table 2. We see that the empirical rejection probability increases as the function ρ becomes more and more irregular. Interestingly, although ρ_1 is a smooth function we reject in this case the null hypothesis more often than in every second Monte Carlo iteration. The reason is that adding ρ_1 on the structural function φ_2 increases the nonlinearity and hence the number of spline basis function is not accurate anymore.

Model	*Empirical Rejection probability*		
	S_n^0	$S_n^{0.25}$	$S_n^{0.5}$
H_0 true	0.041	0.052	0.056
ρ_1	0.664	0.657	0.699
ρ_2	0.789	0.805	0.847
ρ_3	0.829	0.839	0.879
ρ_4	0.871	0.885	0.909

Table 3.2: Empirical Rejection probabilities

3.5 An empirical application

To illustrate our testing procedure, we present an empirical application concerning estimation of the effects of class size on students' performances on standardized tests. Angrist and Lavy [1999] studied the effects of class size on test scores of 4th and 5th grade students in Israel. In this empirical illustration, we focus on 4th grade reading comprehension which was also considered by Horowitz [2011b].

In this empirical example we study the model

$$Y_{sc} = \varphi(Z_{sc}) + V_s + U_{sc} \quad \text{with} \quad \mathbb{P}(V_s + U_{sc} \leqslant 0 | W_{sc}) = q \tag{3.22}$$

where Y_{sc} is the average reading comprehension test score of 4th grade students in class c of school s, Z_{sc} is the number of students in class c of school s, V_s is an unobserved school-specific random effect, and U_{sc} is an unobserved, independently over classes and schools distributed random variable. We introduce the instrument W_{sc} below. In contrast to an additively separable model, the quantile regression model permits for unobserved heterogeneity among different classes. In this empirical illustration, we focus on the median structural effect of class size on test scores and hence, let $q = 0.5$.

The class size Z_{sc} may be endogenous, for instance, due to the socioeconomic background of the students. To identify the causal effect of class size on scholar achievement Angrist and Lavy [1999] use *Maimonides' rule* as instruments. According to this administrative rule, maximum class size is given by 40 pupils and will be split if the number of enrolled students exceeds this number. More precisely, assuming that cohorts are divided into classes of equal size, Maimonides' rule is described by

$$W_{sc} = E_s / \lceil 1 + (E_s - 1)/40 \rceil$$

where E_s denotes enrollment in school s and $\lceil x \rceil$ denotes the largest integer less or equal to x. In addition, information on the fraction of disadvantaged students in class c of school s is available. In this empirical application, we restrict our sample to classes where the fraction of disadvantaged students does not exceed 7%. In doing so, we have a sample of $n = 861$ observations which is sufficient to provide significant nonparametric results, as we show below. Note that Horowitz [2011b] could show that a linear relation between class size and scholar achievement as used by Angrist and Lavy [1999] is misspecified.

(order, knots)	(2,6)	(2,7)	(2,8)	(3,4)	(3,5)	(3,6)	(4,3)	(4,4)	(4,5)
Value of S_n^e	3.278	2.872	3.368	3.316	2.300	2.513	3.397	2.437	3.090

Table 3.3: Values of the statistic S_n^e using weights $\tau_j = j^{-0.25}$

In the following we want to test nonparametrically whether class size is endogenous at the 0.5–quantile. The null hypothesis is that $\mathbb{P}(V_s + U_{sc} \leqslant 0 | Z_{sc}) = 0.5$. We construct our test statistic with additional smoothing as for the Monte Carlo experiments in the previous section with $\tau_j = j^{-0.25}$, $j \geqslant 1$. That is, again the number of cosine basis functions is round about $n^{2/3} \approx 90$. Table 3 depicts the results of our test statistic S_n^e with the b-spline estimator given in (3.18) for different choices of orders and knots. As we see from Table 3, our test statistic exceeds for each choice of b-spline basis functions the 0.05-level critical value 1.960 of the standard normal distribution. Consequently, we may conclude that class size does not satisfy the conditional quantile restriction $\mathbb{P}(V_s + U_{sc} \leqslant 0 | Z_{sc}) = 0.5$ and hence, the hypothesis of exogeneity fails.

We now perform a nonparametric specification test of model (3.22). To estimate the structural effect in model (3.22) we proceed as in the previous section. That is we use Chen and Pouzo [2012]'s method and the b-spline basis as approximating functions. For the sieve space $\mathcal{B}_{k_n}$ we use b-splines basis functions of order 3 with 5 knots (hence $k_n = 6$) and for the estimator $\widehat{T}$ we use b-splines basis functions of order 2 with 17 knots (hence $l_n = 17$).

degree of smoothing	no smoothing	$\tau_j = -0.25$	$\tau_j = -0.5$
Value of S_n	-0.511	0.825	1.600

Table 3.4: Values of the statistic S_n and critical values

The result of our test statistic for different degrees of weighting are depicted in Table

4. We construct our test statistics as in the Monte Carlo investigation of the previous section. As we see from Table 4, for each degree of weighting the null hypothesis of correct specification of the quantile regression model (3.22) cannot be rejected. This is not the case if k_n is chosen too small or too large and hence the specification test could also be used to exclude inappropriate choices of the basis functions.

3.6 Conclusion

In this paper, we developed a nonparametric specification test for the quantile regression model (1.1a). We established the asymptotic distribution of our test under the null hypothesis. Our test is consistent against a fixed alternative and we study its power properties by considering a sequence of local alternatives. We also illustrated several extensions of our test theory. Thereby, we developed a test of exogeneity, a test of dimension reduction, and finally a test for additivity. We demonstrated via a Monte Carlo simulation study that our testing procedure performs well in finite samples. The usefulness of our testing procedure is illustrated by an empirical example. Our testing methodology fails to reject a median regression model of Angrist and Lavy [1999]'s data for reading comprehension but reject exogeneity at the 0.5–quantile of the class size.

Appendix

Proofs of Section 3.2

In the appendix, $\underline{f_{m_n}^\tau}$ denotes the m_n dimensional vector with entries $\sqrt{\tau_j} f_j$ for $1 \leqslant j \leqslant M_n$. Moreover, $\|\cdot\|$ is the usual Euclidean norm. We further denote $\mathcal{T}_{m_n} := F_{m_n}\mathcal{T}$ and $T_{m_n} := F_{m_n}T$. For ease of notation, let $\mathbf{X}_i = (Y_i, Z_i, W_i)$ for $1 \leqslant i \leqslant n$ with realizations $\mathbf{x} = (y, z, w) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{W}$. Let $\mathcal{H}$ be a class of measurable functions with a measurable envelope function H. Then $N(\varepsilon, \mathcal{H}, L_X^2)$ and $N_{[]}(\varepsilon, \mathcal{H}, L_X^2)$, respectively, denote the covering and bracketing numbers for the set $\mathcal{H}$. In addition, let $J_{[]}(1, \mathcal{H}, L_X^2)$ denote a bracketing integral of $\mathcal{H}$, that is,

$$J_{[]}(1, \mathcal{H}, L_X^2) = \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon \|H\|_X, \mathcal{H}, L_X^2)} d\varepsilon.$$

Throughout the proofs, we will use $C > 0$ to denote a generic finite constant that may be different in different uses. Further, for ease of notation we write $\sum_i$ for $\sum_{i=1}^n$ and $\sum_{i'<i}$ for $\sum_{i=1}^n \sum_{i'=1}^{i-1}$.

PROOF OF THEOREM 3.2.1. The proof is based on the decomposition

$$S_n = \|n^{-1}\sum_i (\mathbb{1}\{U_i \leqslant 0\} - q) f^\tau_{\underline{m_n}}(W_i)\|^2$$
$$-2\langle n^{-1}\sum_i (\mathbb{1}\{U_i \leqslant 0\} - q) f^\tau_{\underline{m_n}}(W_i), n^{-1}\sum_i \big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n(Z_i)\} - \mathbb{1}\{U_i \leqslant 0\}\big) f^\tau_{\underline{m_n}}(W_i)\rangle$$
$$+\|n^{-1}\sum_i \big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n(Z_i)\} - \mathbb{1}\{U_i \leqslant 0\}\big) f^\tau_{\underline{m_n}}(W_i)\|^2 = I_n + 2II_n + III_n. \tag{3.23}$$

Consider I_n. We calculate further

$$\varsigma_{m_n}^{-1}(nI_n - \mu_{m_n}) = \frac{1}{q(1-q)\varsigma_{m_n} n}\sum_i \sum_{j=1}^{m_n} \big(|(\mathbb{1}\{U_i \leqslant 0\} - q) f_j^\tau(W_i)|^2 - q(1-q)s_{jj}\big)$$
$$+ \frac{1}{q(1-q)\varsigma_{m_n} n}\sum_{i\neq i'}\sum_{j=1}^{m_n}(\mathbb{1}\{U_i \leqslant 0\} - q)(\mathbb{1}\{U_{i'} \leqslant 0\} - q) f_j^\tau(W_i) f_j^\tau(W_{i'})$$

where the first summand tends in probability to zero as $n \to \infty$. Indeed, condition $\mathbb{P}(U \leqslant 0|W) = q$ yields $\mathbb{E}\,|(\mathbb{1}\{U \leqslant 0\} - q) f_j^\tau(W)|^2 = q(1-q)s_{jj}$, $j \geqslant 1$, and hence

$$\mathbb{E}\,\Big|\frac{1}{\varsigma_{m_n} n}\sum_i\sum_{j=1}^{m_n} |(\mathbb{1}\{U_i \leqslant 0\} - q) f_j^\tau(W_i)|^2 - q(1-q)s_{jj}\Big|^2$$
$$= \frac{1}{\varsigma_{m_n}^2 n}\mathbb{E}\,\Big|\sum_{j=1}^{m_n} |(\mathbb{1}\{U \leqslant 0\} - q) f_j^\tau(W)|^2 - q(1-q)s_{jj}\Big|^2 \leqslant \frac{(1+q)^4\eta_f\eta_p}{\varsigma_{m_n}^2 n}\Big(\sum_{j=1}^{m_n}\tau_j\Big)^2 = o(1)$$

by using $\sup_{j\in\mathbb{N}}\mathbb{E}\,|f_j(W)|^4 \leqslant \eta_f\eta_p$. Therefore, to establish $(\sqrt{2}\varsigma_{m_n})^{-1}(nI_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0,1)$ it is sufficient to show

$$\frac{\sqrt{2}}{q(1-q)\varsigma_{m_n} n}\sum_{i\neq i'}\sum_{j=1}^{m_n}(\mathbb{1}\{U_i \leqslant 0\} - q)(\mathbb{1}\{U_{i'} \leqslant 0\} - q) f_j^\tau(W_i) f_j^\tau(W_{i'}) \xrightarrow{d} \mathcal{N}(0,1). \tag{3.24}$$

Since $\varsigma_{m_n} = o(1)$ this follows from Lemma 2.7.2. Consider III_n. Let us define for $1 \leqslant j \leqslant m_n$ and $1 \leqslant i \leqslant n$

$$h_j(\mathbf{X}_i, \phi) = \big(\mathbb{1}\{Y_i \leqslant \varphi(Z_i)\} - \mathbb{1}\{Y_i \leqslant \phi(Z_i)\}\big) f_j(W_i)$$

and the classes $\mathcal{H}_{jn} = \{h_j(\cdot, \phi) : \phi \in \mathcal{B}_n\}$ and $\mathcal{H}_j = \{h_j(\cdot, \phi) : \phi \in \mathcal{B}\}$. We observe

$$III_n = \sum_{j=1}^{m_n} \tau_j \big| n^{-1} \sum_i h_j(\mathbf{X}_i, \widehat{\varphi}_n)\big|^2$$
$$\leqslant 2\eta_p \|\mathcal{T}\widehat{\varphi}_n - \mathcal{T}\varphi\|_W^2 + 2\sum_{j=1}^{m_n} \tau_j \big| n^{-1} \sum_i h_j(\mathbf{X}_i, \widehat{\varphi}_n) - \mathbb{E}\, h_j(\mathbf{X}, \widehat{\varphi}_n)\big|^2$$

From the tangential cone condition together with Assumption 3.4 we infer $n\|\mathcal{T}\widehat{\varphi}_n - \mathcal{T}\varphi\|_W^2 \leqslant n(1+\eta)\|T(\widehat{\varphi}_n - \varphi)\|_W^2 = o_p(\varsigma_{m_n})$. Recall the definition $\mathcal{B}_n := \{\phi \in \mathcal{B} : D_{\mathcal{B}}(\phi, \varphi) \leqslant \mathcal{R}_n\}$. We observe for every $\phi \in \mathcal{B}_n$ that

$$\big|h_j(\mathbf{X}_i, \phi)\big|^2 \leqslant \sup_{\phi \in \mathcal{B}_n} \big|\big(\mathbb{1}\{Y_i \leqslant \varphi(Z_i)\} - \mathbb{1}\{Y_i \leqslant \phi(Z_i)\}\big) f_j(W_i)\big|^2 =: H_j^2(\mathbf{X}_i)$$

and hence, H_j is an envelope function of the class $\mathcal{H}_{jn}$ and due to Assumption 3.3 we have $\mathbb{E}[H_j^2(\mathbf{X})] \leqslant C\mathcal{R}_n^\kappa$ for n sufficiently large. Moreover, condition $D_{\mathcal{B}}(\widehat{\varphi}_n, \varphi) = o_p(\mathcal{R}_n)$ implies

$$\mathbb{P}\Big(\sum_{j=1}^{m_n} \tau_j \big|n^{-1/2} \sum_i h_j(\mathbf{X}_i, \widehat{\varphi}_n) - \mathbb{E}\, h_j(\mathbf{X}, \widehat{\varphi}_n)\big|^2 > \varepsilon\Big)$$
$$\leqslant \sum_{j=1}^{m_n} \tau_j \varepsilon^{-1}\, \mathbb{E} \sup_{\phi \in \mathcal{B}_n} \Big|n^{-1/2} \sum_i h_j(\mathbf{X}_i, \phi) - \mathbb{E}\, h_j(\mathbf{X}, \phi)\Big|^2 + o(1)$$
$$\leqslant \sum_{j=1}^{m_n} \tau_j \varepsilon^{-1} \Big(\mathbb{E} \sup_{\phi \in \mathcal{B}_n} \Big|n^{-1/2} \sum_i h_j(\mathbf{X}_i, \phi) - \mathbb{E}\, h_j(\mathbf{X}, \phi)\Big| + \big(\mathbb{E}\,|H_j(\mathbf{X})|^2\big)^{1/2}\Big)^2 + o(1)$$

where the last inequality is due to Theorem 2.14.5 of van der Vaart and Wellner [2000]. We further conclude by applying the last display of Theorem 2.14.2 of van der Vaart and Wellner [2000]

$$\mathbb{E} \sup_{\phi \in \mathcal{B}_n} \Big|n^{-1/2} \sum_i h_j(\mathbf{X}_i, \phi) - \mathbb{E}\, h_j(\mathbf{X}, \phi)\Big| \leqslant C J_{[]}(1, \mathcal{H}_{jn}, L^2_{\mathbf{X}}) \big(\mathbb{E}\,|H_j(\mathbf{X})|^2\big)^{1/2}.$$

Now since $\max_{1 \leqslant j \leqslant m_n} \mathbb{E}\,|H_j(\mathbf{X})|^2 \leqslant C\mathcal{R}_n^\kappa$ for n sufficiently large and $\mu_{m_n}\mathcal{R}_n^\kappa = o(\varsigma_{m_n})$ it is sufficient to show that $\max_{1 \leqslant j \leqslant m_n} J_{[]}(1, \mathcal{H}_{jn}, L^2_{\mathbf{X}}) < \infty$. Further, Lemma 4.2 (i) of

Chen [2007] yields

$$\begin{aligned} N_{[]}(\varepsilon(\mathbb{E}\,|H_j(\mathbf{X})|^2)^{1/2}, \mathcal{H}_{jn}, L^2_{\mathbf{X}}) &\leqslant N_{[]}\Big(\varepsilon, (\mathbb{E}\,|H_j(\mathbf{X})|^2)^{-1/2}\,\mathcal{H}_{jn}, L^2_{\mathbf{X}}\Big)\\ &\leqslant N_{[]}\Big(\varepsilon, \mathcal{H}_j, L^2_{\mathbf{X}}\Big)\\ &\leqslant N\Big(\Big(\frac{\varepsilon}{2C}\Big)^{2/\kappa}, \mathcal{B}, D_{\mathcal{B}}\Big)\\ &\leqslant N\Big(\Big(\frac{\varepsilon}{2C}\Big)^{2/\kappa}, \mathcal{B}, \|\cdot\|_\infty\Big) \end{aligned}$$

where we used $D_{\mathcal{B}}(\phi,\varphi) \leqslant \|\phi-\varphi\|_\infty$ for any $\phi \in \mathcal{B}$. Employing condition $\alpha_0 > d/p$ and Theorem 6.2 Part II of Adams and Fournier [2003] yields that $W^{\alpha+\alpha_0,p}$ is compactly embedded in $W^{\alpha,\infty}$. Thereby, $\mathcal{B} \subset W^{\alpha,p}$ is totally bounded in $W^{\alpha,\infty}$ which implies $\|\phi\|_{\alpha,\infty} \leqslant C$ for all $\phi \in \mathcal{B}$. Let $W^{\alpha,\infty}_C := \{W^{\alpha,\infty} : \|\phi\|_{\alpha,\infty} \leqslant C\}$. Now Theorem 2.7.1 of van der Vaart and Wellner [2000] gives

$$\log N\big(\varepsilon^{2/\kappa}, \mathcal{B}, \|\cdot\|_\infty\big) \leqslant \log N\big(\varepsilon^{2/\kappa}, W^{\alpha,\infty}_C, \|\cdot\|_\infty\big) \leqslant C\varepsilon^{-2d/(\alpha\kappa)}$$

where C depends on the diameter of $\mathcal{Z}$. Now due to Assumption 3.5 (i) it is straightforward to see that $\max_{1\leqslant j\leqslant m_n} J_{[]}(1, \mathcal{H}_{jn}, L^2_{\mathbf{X}}) < \infty$ and hence, $nIII_n = o_p(\varsigma_{m_n})$.

Consider II_n. We observe

$$\begin{aligned} nII_n &= \sum_{j=1}^{m_n}\tau_j\big(\sum_i(\mathbb{1}\{U_i\leqslant 0\}-q)f_j(W_i)\big)\Big(n^{-1}\sum_i h_j(\mathbf{X}_i,\widehat{\varphi}_n)\Big) + o_p(\varsigma_{m_n})\\ &= \sum_{j=1}^{m_n}\tau_j\big(\sum_i(\mathbb{1}\{U_i\leqslant 0\}-q)f_j(W_i)\big)\Big(n^{-1}\sum_i h_j(\mathbf{X}_i,\widehat{\varphi}_n) - \mathbb{E}\,h_j(\mathbf{X},\widehat{\varphi}_n)\Big)\\ &\quad + \sum_{j=1}^{m_n}\tau_j\big(\sum_i(\mathbb{1}\{U_i\leqslant 0\}-q)f_j(W_i)\big)\,\mathbb{E}\,h_j(\mathbf{X},\widehat{\varphi}_n) + o_p(\varsigma_{m_n})\\ &= C_{n1} + C_{n2} + o_p(\varsigma_{m_n}). \end{aligned}$$

Condition $\mu_{m_n}\mathcal{R}^\kappa_n = o(\varsigma_{m_n})$ yields $nC_{n1} = o_p(\varsigma_{m_n})$ by applying the Cauchy Schwarz inequality. Consider C_{n2}. Define $t_j = \mathbb{E}\,h_j(\mathbf{X},\widehat{\varphi}_n)\big(\sum_{j=1}^{m_n}(\mathbb{E}\,h_j(\mathbf{X},\widehat{\varphi}_n))^2\big)^{-1/2}$ for $1 \leqslant j \leqslant m_n$ then $\sum_{j=1}^{m_n} t_j^2 = 1$ and hence

$$\begin{aligned} \mathbb{E}\,|C_{n2}|^2 &\leqslant \eta_p n\|\mathcal{T}\widehat{\varphi}_n - \mathcal{T}\varphi\|^2_W\,\mathbb{E}\Big|\sum_{j=1}^{m_n}\tau_j t_j(\mathbb{1}\{U\leqslant 0\}-q)f_j(W)\Big|^2\\ &\leqslant \eta_p^2 n\|\mathcal{T}\widehat{\varphi}_n - \mathcal{T}\varphi\|^2_W\int_{\mathcal{W}}\Big|\sum_{j=1}^{m_n}\tau_j t_j f_j(w)\Big|^2\nu(dw) \leqslant \eta_p^2 n\|\mathcal{T}\widehat{\varphi}_n - \mathcal{T}\varphi\|^2_W = o(\varsigma_{m_n}) \end{aligned}$$

which completes the proof. $\square$

PROOF OF PROPOSITION 3.2.2. We first check whether the conditions of Lemma B.1 of Chen and Pouzo [2012] are satisfied. The tangential cone condition (Assumption 3.2 (ii)) with $0 < \eta < 1$ yields $(1-\eta)\|T(\phi-\varphi)\|_W \leqslant \|\mathcal{T}\phi - \mathcal{T}\varphi\|_W \leqslant \|\mathcal{T}\phi\|_W$ for all $\phi \in \mathcal{B}$ and moreover, $\|\mathcal{T}(E_{k_n}\varphi)\|_W \leqslant \|\mathcal{T}(E_{k_n}\varphi) - \mathcal{T}\varphi\|_W \leqslant (1+\eta)\|T(E_{k_n}\varphi - \varphi)\|_W$. Consequently, Assumption 4.1 (ii) of Chen and Pouzo [2012] follows.

Due to Assumption 3.6 and since $\{f_j\}_{j\geqslant 1}$ forms an orthonormal basis in L^2_W we may apply Lemma C.2 of Chen and Pouzo [2012] which implies that the least squares estimator $\widehat{\mathcal{T}}$ of $\mathcal{T}$ given in (3.12) satisfies $n^{-1}\sum_i |(\widehat{\mathcal{T}}E_{k_n}\varphi)(W_i)|^2 \leqslant C\|\mathcal{T}E_{k_n}\varphi\|^2_W + O_p(\delta_n^2)$ and moreover, $n^{-1}\sum_i |(\widehat{\mathcal{T}}\phi)(W_i)|^2 \geqslant C\|\mathcal{T}\phi\|^2_W - O_p(\delta_n^2)$ uniformly over $\mathcal{B}^{\rho}_{k_n}$ for a finite constant $C > 0$. Thereby, Assumption 3.3 of Chen and Pouzo [2012] holds true. In addition, since $\mathcal{H}_{ojn} := \{h_j(\cdot,\phi) : \phi \in \mathcal{B}_{ok_n}\} \subset \mathcal{H}_{jn}$ (defined in the proof of Theorem 3.2.1) we infer from the proof of Theorem 3.2.1 that $\max_{1\leqslant j\leqslant m_n} J_{[]}(1, \mathcal{H}_{ojn}, L^2_{\mathbf{X}}) \leqslant C < \infty$ and hence, Assumption C.2 of Chen and Pouzo [2012] holds true.

Consequently, we may apply Lemma B.1 (ii) of Chen and Pouzo [2012] which yields $\|T(\widehat{\varphi}_n - \varphi)\|_W = O_p(r_n)$ where $r_n = \max\big(\delta_n, o(\sqrt{\lambda_n}), \|T(E_{k_n}\varphi - \varphi)\|_W\big)$. Now condition (3.15) yields $n\|T(\widehat{\varphi}_n - \varphi)\|^2_W = o_p(\varsigma_{m_n})$, which completes the proof. $\square$

PROOF OF PROPOSITION 3.2.3. Since $\|n^{-1}\sum_i \xi(W_i) f^{\tau}_{\underline{m_n}}(W_i)\|^2 = O_p\big(\sum_{j=1}^{\infty}\xi_j^2\big)$ we see that

$$
\begin{aligned}
&(\sqrt{2}\varsigma_{m_n})^{-1}\Big\|n^{-1/2}\sum_i(\mathbb{1}\{U_i \leqslant 0\} - q - \varsigma_{m_n}^{1/2} n^{-1/2}\xi(W_i)) f^{\tau}_{\underline{m_n}}(W_i)\Big\|^2\\
&= (\sqrt{2}\varsigma_{m_n})^{-1}\Big\|n^{-1/2}\sum_i(\mathbb{1}\{U_i \leqslant 0\} - q) f^{\tau}_{\underline{m_n}}(W_i)\Big\|^2\\
&+ (\sqrt{2\varsigma_{m_n}})^{-1}\langle n^{-1}\sum_i(\mathbb{1}\{U_i \leqslant 0\} - q) f^{\tau}_{\underline{m_n}}(W_i), \mathbb{E}[\xi(W) f^{\tau}_{\underline{m}}(W)]\rangle + O_p\Big(\sum_{j=1}^{\infty}\xi_j^2\Big)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = I_n + II_n + O_p\Big(\sum_{j=1}^{\infty}\xi_j^2\Big).
\end{aligned}
$$

Further, since $\mathbb{E}\,|II_n|^2 \leqslant \eta^2 n^{-1}\sum_{j=1}^{\infty}\xi_j^2 = o(1)$ the assertion follows similarly to the proof of Theorem 3.2.1. $\square$

PROOF OF PROPOSITION 3.2.4. The null hypothesis fails if $\mathcal{T}\varphi = \xi$ for some non zero function $\xi \in \mathcal{L}^2_W$. For the proof it is sufficient to show $S_n = \|L\xi\|^2_W + o_p(1)$. Since $\|n^{-1}\sum_i(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_n(Z_i)\} - \mathbb{1}\{Y_i \leqslant \varphi(Z_i)\}) f^{\tau}_{\underline{m_n}}(W_i)\|^2 = o_p(1)$ (cf. proof of Theorem

3.2.1) and

$$\|n^{-1}\sum_i(\mathbb{1}\{Y_i \leqslant \varphi(Z_i)\}-q)f^{\tau}_{\underline{m_n}}(W_i)\|^2 \geqslant \int_{\mathcal{W}} |(L\mathcal{T}\varphi)(w)p_W(w)/\nu(w)|^2\nu(dw)+o_p(1)$$
$$\geqslant C\|L\xi\|^2_W + o_p(1)$$

which proves the result. □

Proof of Proposition 3.2.5. Let us denote $\overline{\mathcal{T}}_n\phi := n^{-1}\sum_i(\mathbb{1}\{Y_i \leqslant \phi(Z_i)\}-q)f^{\tau}_{\underline{m_n}}(W_i)$ and $\overline{\mathcal{T}}\phi := \mathbb{E}(\mathbb{1}\{Y \leqslant \phi(Z)\} - q)f^{\tau}_{\underline{m_n}}(W)$ for all $\phi \in \Phi$. The basic inequality $(a-b)^2 \geqslant a^2/2 - b^2$, $a, b \in \mathbb{R}$, yields for all $\phi \in \mathcal{G}^{\rho}_n$ and φ with $\mathcal{T}\varphi = 0$ that

$$P\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}(n\,S_n - \mu_{m_n}) > q_\alpha\Big)$$
$$\geqslant P\Big(\frac{n}{2}\|\overline{\mathcal{T}}_n\phi - \overline{\mathcal{T}}_n\varphi\|^2 + n\|\overline{\mathcal{T}}_n\varphi\|^2 - \mu_{m_n} > \sqrt{2}\,\varsigma_{m_n}q_\alpha + 2n|\langle\overline{\mathcal{T}}_n\widehat{\varphi}_n - \overline{\mathcal{T}}_n\varphi, \overline{\mathcal{T}}_n\varphi\rangle|$$
$$+ n\|\overline{\mathcal{T}}_n\widehat{\varphi}_n - \overline{\mathcal{T}}_n\varphi\|^2\Big).$$

From the proof of Theorem 3.2.1 we see $n\langle\overline{\mathcal{T}}_n\widehat{\varphi}_n - \overline{\mathcal{T}}_n\varphi, \overline{\mathcal{T}}_n\varphi\rangle = o_p(\varsigma_{m_n})$, $n\|\overline{\mathcal{T}}_n\widehat{\varphi}_n - \overline{\mathcal{T}}_n\varphi\|^2 = o(\varsigma_{m_n})$ and $n\|\overline{\mathcal{T}}_n\varphi\|^2_W - \mu_{m_n} = O_p(\varsigma_{m_n})$. Moreover, we observe

$$n\|\overline{\mathcal{T}}_n\phi - \overline{\mathcal{T}}_n\varphi\|^2_W \geqslant n\|\overline{\mathcal{T}}\phi\|^2 - 2n|\langle\overline{\mathcal{T}}_n\phi - \overline{\mathcal{T}}_n\varphi - \overline{\mathcal{T}}\phi, \overline{\mathcal{T}}\phi\rangle| = I_n - 2II_n.$$

Note that $\|L\mathcal{T}\phi\|^2_W \leqslant C$ for all $\phi \in \mathcal{G}^{\rho}_n$ we have $I_n \geqslant Cn\|L\mathcal{T}\phi\|^2_W$ for n sufficiently large. Consider II_n. For $1 \leqslant j \leqslant m_n$ let $s_j = \mathbb{E}[(\mathcal{T}\phi)(W)f^{\tau}_j(W)]/\|\overline{\mathcal{T}}\phi\|$ then clearly $\sum_{j=1}^{m_n} s_j^2 = 1$ and thus $\mathbb{E}\,|\sum_{j=1}^{m_n} s_j f_j(W)|^2 \leqslant \eta_f\eta_p$. We evaluate

$$\mathbb{E}\,|II_n|^2 \leqslant n\|\overline{\mathcal{T}}\phi\|^2\,\mathbb{E}\Big|\sum_{j=1}^{m_n} s_j(\mathbb{1}\{Y \leqslant \phi(Z)\}-\mathbb{1}\{Y \leqslant \varphi(Z)\})f^{\tau}_j(W)\Big|^2 \leqslant n\,\eta_f\eta_p\|L\mathcal{T}\phi\|^2_W.$$

and hence $II_n = O_p(\sqrt{n}\|L\mathcal{T}\phi\|_W)$. Consequently, for all $0 < \varepsilon < 1$ and n sufficiently large we have

$$P\Big((\sqrt{2}\,\varsigma_{m_n})^{-1}(n\,S_n - \mu_{m_n}) > q_\alpha\Big) \geqslant P\Big(\frac{n}{2}\|L\mathcal{T}\phi\|^2_W > \sqrt{2}\,\varsigma_{m_n}q_\alpha\Big) - \varepsilon,$$

which proves the assertion. □

Proof of Corollary 3.3.1. It is sufficient to show that Assumption 3.4 is satisfied.

Since $\sup_y \sup_{(z,w)\in\mathcal{Z}\times\mathcal{W}} p_{Y|Z,W}(y,z,w) \leqslant C$ we have

$$\|T(\widehat{\varphi}^{\mathrm{e}}_{k_n} - \varphi_0)\|^2_W = \mathbb{E}\,|\,\mathbb{E}[p_{Y|Z,W}(\varphi_0(Z),Z,W)(\widehat{\varphi}^{\mathrm{e}}_{k_n} - \varphi_0)(Z)|W]|^2 \leqslant C\|\widehat{\varphi}^{\mathrm{e}}_{k_n} - \varphi_0\|^2_Z.$$

Under the conditions of Assumption 3.7, He and Shi [1994] (proof of Theorem 2.1 equation (3.11) and (3.12)) establish that $\|\widehat{\varphi}^{\mathrm{e}}_{k_n} - \varphi_0\|^2_Z = O_p(n^{-1}k_n + k_n^{-2r})$. Consequently, $n\|T(\widehat{\varphi}^{\mathrm{e}}_{k_n} - \varphi_0)\|^2_W = O_p(k_n + nk_n^{-2r}) = o_p(\varsigma_{m_n})$ and hence Assumption 3.4 holds true. $\square$

Bibliography

R. A. Adams and J. J. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier/Academic Press, Amsterdam, 2003.

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

C. Ai and X. Chen. Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141:5–43, 2007.

Y. Ait-Sahalia, P. J. Bickel, and T. M. Stoker. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105(2): 363–412, 2001.

J. D. Angrist and A. B. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

J. D. Angrist and V. C. Lavy. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575, 1999.

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

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113(3):301–413, 1999.

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.

D. Bosq. *Nonparametric statistics for stochastic processes*. Springer, New York, 1998.

J. Bound, D. A. Jaeger, and R. M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450, 1995.

C. Breunig and J. Johannes. On rate optimal local estimation in nonparametric instrumental regression. *arXiv preprint arXiv:0902.2103*, 2009.

L. D. Brown and M. G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *Annals of Statistics*, 24(6):2524–2535, 1996.

H. Cardot and J. Johannes. Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, (101):395–408, 2010.

M. Carrasco, J.-P. Florens, and E. Renault. Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 6. North Holland, 2006.

X. Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 2007.

X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 01 2012.

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

X. Chen, O. Linton, and I. Van Keilegom. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71:1591–1608, 2003.

V. Chernozhukov and C. Hansen. An iv model of quantile treatment effects. *Econometrica*, 73:245–261, 2005.

V. Chernozhukov and C. Hansen. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525, June 2006.

V. Chernozhukov, G. Imbens, and W. K. Newey. Instrumental variable estimation of nonseparable models. *Journal of Econometrics*, 139(1):4–14, 2007.

A. Chesher. Identification in nonseparable models. *Econometrica*, 71(5):1405–1441, 2003.

S. Darolles, J. P. Florens, and E. Renault. Nonparametric instrumental regression. Working paper 05-2002, CRDE, 2002.

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.

D. Donoho. Statistical estimation and optimal recovery. *Annals of Statistics*, 22:238–270, 1994.

D. Donoho and M. Low. Renormalization exponents and optimal pointwise rates of convergence. *Annals of Statistics*, 20:944–970, 1992.

F. Dunker, J.-P. Florens, T. Hohage, J. Johannes, and E. Mammen. Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. Technical report, University of Göttingen, 2011.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic, Dordrecht, 2000.

J. P. Florens. Inverse problems and structural econometrics: The example of instrumental variables. Discussion Paper # 0537, Institut de statistique, Université catholique de Louvain, 2002.

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(1):61 – 75, 2012a.

P. Gagliardini and O. Scaillet. Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica*, 80(4):1533–1562, 2012b.

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.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Annals of Statistics*, 39(3):1608–1632, 2011.

A. Goldenshluger and S. V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118:169–186, 2000.

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

M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72(1):21–37, Nov. 1995. doi: 10.1007/s002110050158.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, 1982.

W. Härdle and E. Mammen. Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21(4):1926–1947, 1993.

X. He and P. Shi. Convergence rate of b-spline estimators of nonparametric conditional quantile functionsâĹŮ. *Journaltitle of Nonparametric Statistics*, 3(3-4):299–308, 1994.

N. Hearst, T. Newman, et al. A randomized natural experiment. *New England Journal of Medicine*, 314:620–624, 1986.

E. Heinz. Beiträge zur störungstheorie der spektralzerlegung. *Math. Anal.*, 123:415–438, 1951.

M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. Discussion Paper # 990, WIAS, Berlin, 2004.

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. Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. 2011a.

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

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

J. L. Horowitz and S. Lee. Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75:1191–1208, 2007.

J. L. Horowitz and S. Lee. Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152(2):141–152, 2009.

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.

I. Ibragimov and R. Has'minskii. On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory of Probability and its Applications*, 29: 18–32, 1984.

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

J. Johannes, S. Van Bellegem, and A. Vanhems. Convergence rates for ill-posed inverse problems with an unknown operator. *Econometric Theory*, 27(03):522–545, 2011.

T. Kawata. *Fourier analysis in probability theory.* Academic Press, New York, 1972.

S. Krein and Y. I. Petunin. Scales of banach spaces. In *Russian Math. Surveys*, volume 21, pages 85–169, 1966.

O. Lepskij. On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.*, 35(3):454–466, 1990.

K. Li. Minimaxity of the method of regularization of stochastic processes. *Annals of Statistics*, 10:937–942, 1982.

J.-M. Loubes and C. Marteau. Oracle inequalities for instrumental variable regression. Technical report, Toulouse, 2009.

B. A. Mair. Tikhonov regularization for finitely and infinitely smoothing operators. *SIAM J. Appl. Anal.*, 25:135–147, 1994.

P. Massart. *Concentration inequalities and model selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003.* Lecture Notes in Mathematics 1896. Berlin: Springer. xiv, 337 p. , 2007.

E. Miguel, S. Satyanath, and E. Sergenti. Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, 112(4):725–753, 2004.

F. Natterer. Error bounds for tikhonov regularization in hilbert scales. *Appl. Anal.*, 18: 29–37, 1984.

A. Neubauer. An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM Journal of Numerical Analysis*, 25 (6):1313–1326, 1988.

W. K. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837, 1990.

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.

V. V. Petrov. *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press., Oxford, 4. edition, 1995.

A. Santos. Instrumental variable methods for recovering continuous linear functionals. *Journal of Econometrics*, 2011.

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

J. D. Sargan. The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415, 1958.

R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 1981.

T. A. Severini and G. Tripathi. Efficiency bounds for estimating linear functionals of non-parametric regression models with endogenous regressors. *Journal of Econometrics*, 2010.

P. Speckman. Minimax estimation of linear functionals in a Hilbert space. Unpublished manuscript., 1979.

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

A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer, corrected edition, Nov. 2000. ISBN 0387946403.

Lebenslauf

seit 10.2009	Promotion in Volkswirtschaftslehre an der Universität Mannheim, Mitglied der Graduiertenschule CDSE
04.2009	Diplom in Mathematik mit Nebenfach Wirtschaftswissenschaften an der Ruprecht–Karls–Universität Heidelberg
10.2005 – 03.2006	Auslandsstudium an der University of York, England
07.2005	Vordiplom in Mathematik mit Nebenfach Wirtschaftswissenschaften an der Julius–Maximilians–Universität Würzburg
10.2002 – 07.2003	Zivildienst am Zentrum für Körperbehinderte in Würzburg
05.2002	Abitur am Matthias–Grünewald–Gymnasium Würzburg