A Model Specification Test

Knut Bartels *

August 3, 1998

Abstract

A general model specification test of a parametric model against a nonparametric or semiparametric alternative is studied. The test statistic employs a fixed kernel, not varying by a bandwidth. This test is proved to be consistent, the asymptotic distribution is derived and shown to be approximated by a bootstrap procedure.
Keywords: Model specification test, nonparametric or semiparametric alternative, nonlinear regression, bootstrap.

1 Introduction

Given is an i.i.d. random sample $Z_1 = (Y_1, X_1), \dots, Z_n = (Y_n, X_n)$ from a distribution D on $\mathbb{R} \times \mathbb{R}^d$. We want to test if the continuous regression function $\mu(x) := \mathsf{E}[Y|X = x]$ belongs to a parametric family of known real functions $f(x, \theta)$ on $\mathbb{R}^d \times \Theta$ where $\Theta \subseteq \mathbb{R}^p$ denotes the parameter space and $(Y, X) \sim \mathsf{D}$. Thus we consider the null hypothesis

$$\mathbf{H_0}: \quad \mathsf{P}\big[\mathsf{E}[Y|X] = f(X, \theta_0)\big] = 1 \text{ for some } \theta_0 \in \Theta \quad , \tag{1}$$

and the corresponding general alternative

$$\mathbf{H_1}: \quad \mathsf{P}\big[\mathsf{E}[Y|X] = f(X, \theta)\big] < 1 \text{ for all } \theta \in \Theta \quad , \tag{2}$$

where P is some probability measure dominating the marginal distribution D_X of X.

Several authors have contributed to this kind of problem already, some of them we will mention in the forthcoming. For further sources see the bibliography. A common method is to compare a parametric fit under $\mathbf{H_0}$ with a nonparametric estimate of μ (Eubank, Hart & LaRiccia, 1993). Here a

*The research for this paper was carried out within Sonderforschungsbereich 373 at the University of Potsdam and the Humboldt-University Berlin

bandwidth h is involved which is assumed to tend to zero at a certain rate. This assumption of a vanishing bandwidth is made in almost all of the literature concerning this test problem, see e.g. Fan & Li (1996), Härdle & Mammen (1993), Härdle & Horowitz (1994), Kozek (1991), Stute & Manteiga (1996) and Werwatz (1997). But for applications always a bandwidth has to be chosen and fixed for the further analysis. Also the convergence to the (normal) limiting distribution is known to be very slow and usually is bootstrapped using this chosen bandwidth (Härdle & Mammen, 1993). The aim of this paper is to study the asymptotic behavior of a large class of these tests with a nonvanishing *fixed* bandwidth. The results are similar to the test of Bierens (1990), which follows an idea different to the above. Bierens' test is included in the class of tests considered here by a special choice of a kernel function. The estimation of the critical values by bootstrap methods is easier to justify and less numerically intensive for the form the test statistic is presented here. Also the parametric estimator is not restricted to be derived by least squares here.
Principally we want to base our test on

$$T_n = T_n(\theta, Z_1, \dots, Z_n) := \frac{1}{n} \sum_{1 \le i < j \le n} U_i U_j K_{ij} \quad , \tag{3}$$

where

- $U_i := u(Y_i, X_i, \theta_0) = Y_i - f(X_i, \theta_0)$ are the true errors under $\mathbf{H_0}$,
- $K_{ij} := k(X_i, X_j, \theta_0)$ is a positive definite (see assumption **A3**) symmetric kernel, possibly depending on θ.

Since under $\mathbf{H_0}$ a true parameter θ_0, denoting a solution of (1), is not known, it has to be estimated from the data by $\hat{\theta}$. This leads to the actual test statistic

$$\hat{T}_n = T_n(\hat{\theta}, Z_1, \dots, Z_n) := \frac{1}{n} \sum_{1 \le i < j \le n} \hat{U}_i \hat{U}_j \hat{K}_{ij} \quad , \tag{4}$$

where $\hat{U}_i := u(Y_i, X_i, \hat{\theta}) = Y_i - f(X_i, \hat{\theta})$ are the parametrically estimated errors and $\hat{K}_{ij} := k(X_i, X_j, \hat{\theta})$.

2 Results

We introduce the following expressions:

$$v(x) := v(x;\theta_0) := \int_{\xi\in\mathbb{R}} k(x,\xi;\theta_0) f'(\xi;\theta_0)\,\mathrm{d}\mathsf{D}_X(\xi) \quad ,$$

such that for $i = 1,\ldots,n$ and $X \sim \mathsf{D}_X$ we have the $(1,p)$-vectors

$$v(X_i;\theta_0) \;= \mathsf{E}_X[k(X_i,X;\theta_0)f'(X;\theta_0)|X_i] \quad ;$$

m denotes the (p,p)-matrix

$$m := m(\theta_0) := \Bigl(\mathsf{E}\left[k(X_1,X_2;\theta_0)[f'(X_1;\theta_0)]_\alpha [f'(X_2;\theta_0)]_\beta\right]\Bigr)_{\alpha,\beta=1,\ldots,p} .$$

Now define the functional

$$\begin{aligned} q(z_1,z_2) &:= q(y_1,x_1,y_2,x_2;\theta_0) \\ &:= u(y_1,x_1;\theta_0)u(y_2,x_2;\theta_0)\cdot\bigl\{k(x_1,x_2;\theta_0) - v(x_1;\theta_0)w(x_2;\theta_0) \\ &\qquad - v(x_2;\theta_0)w(x_1;\theta_0) + w(x_1;\theta_0)^T m(\theta_0)\, w(x_2;\theta_0)\bigr\}. \end{aligned} \tag{5}$$

For sake of brevity, in the following we will drop indication of the dependency on θ_0 unless we want to stress it explicitly. For the assumptions made in the following theorem see section 3.

Theorem 2.1. *Under* $\mathbf{H_0}$ *and assumptions* **A0** *to* **A5**

$$\hat{T}_n = \tfrac{1}{n}\sum_{1\le i<j\le n} q(Z_i,Z_j) - \tfrac{1}{n}\sum_{i=1}^n u(Z_i)^2 v(X_i)w(X_i) + \tfrac{1}{2n}\sum_{i=1}^n u(Z_i)^2 w(X_i)^T m\, w(X_i) + o_p(1) \quad . \tag{6}$$

Hence $\hat{T}_n$ *is asymptotically distributed as*

$$c + \tfrac{1}{2}\sum_k \lambda_k(\chi_{1k}^2 - 1) \quad , \tag{7}$$

where

$$c := \tfrac{1}{2}\,\mathsf{E}\left[u(Z_1)^2 w(X_1)^T m\, w(X_1)\right] - \mathsf{E}\left[u(Z_1)^2 v(X_1)w(X_1)\right] \quad , \tag{8}$$

$\chi_{11}, \chi_{12}, \ldots$ *are independent standard normal random variables and* λ_k *are the eigenvalues of* $\mathcal{Q} : \phi(\cdot) \mapsto \int_{\mathbb{R}^{d+1}} q(\cdot,z)\phi(z)\,\mathrm{d}\mathsf{D}$, *which is a linear operator on* $L^2(\mathsf{D})$. *That is,* $\hat{T}_n - c$ *is asymptotically distributed as a weighted (in general infinite) sum of centered independent* χ_1^2*-statistics.*

Theorem 2.2. *Under* $\mathbf{H_1}$*, assumptions* **A0** *to* **A3** *and* **A6**

$$\hat{T}_n \underset{p}{\rightarrow} +\infty \quad . \tag{9}$$

Theorem 2.1 and 2.2 together imply the consistency of the test "reject $\mathbf{H_0}$ iff $\hat{T}_n > \tau_\alpha$" for any reasonably chosen $\tau_\alpha > 0$. Such a resonable choice would be an α-quantile of the limiting distribution of $\hat{T}_n$ under $\mathbf{H_0}$.
This distribution is case dependent but can be bootstrapped. For $i = 1, \ldots, n$ let $Z_i^* = (Y_i^*, X_i)$ with $Y_i^* = f(X_i, \hat{\theta}_n) + u_i^*$ be a bootstrap sample, where $u_i^* = \eta_i u(Y_i, X_i, \hat{\theta}_n)$ and the η_i's are distributed i.i.d. and independently of the Z_i's, with $\mathsf{E}[\eta_i] = 0$, $\mathsf{E}[\eta_i^2] = 1$ and $\mathsf{E}[\eta_i^4] < \infty$. Now $\hat{\theta}_n^*$ denotes the estimator based on the sample $Z_1^*, \ldots, Z_n^*$ and the bootstrapped test statistic is $\hat{T}_n^* := T_n(\hat{\theta}_n^*, Z_1^*, \ldots, Z_n^*)$.

Theorem 2.3. *Under* $\mathbf{H_0}$ *and assumptions* **A0** *to* **A5**

$$\sup_{x \in \mathbb{R}} \left| \mathsf{P}^*\{\hat{T}_n^* \leq x | Z_1, \ldots, Z_n\} - \mathsf{P}\{\hat{T}_n \leq x\} \right| \rightarrow 0 \quad (n \rightarrow \infty) \; . \tag{10}$$

We remark that very similar results hold for the test statistic including the diagonal

$$\hat{T}_n^{(v)} = T_n^{(v)}(\hat{\theta}, Z_1, \ldots, Z_n) := \frac{1}{n} \sum_{1 \leq i,j \leq n} \hat{U}_i \hat{U}_j \hat{K}_{ij} \quad . \tag{11}$$

The limiting distribution changes to the simpler form $\sum_k \lambda_k \chi_{1k}^2$, using the same notation. The explicit proofs for this version, modifications of the bootstrap procedure presented here, as well as simulation studies for the finite sample behavior will be presented in future works.

3 Assumptions

A0. $Z_1 = (Y_1, X_1), \ldots, Z_n = (Y_n, X_n)$ is an i.i.d. random sample from a distribution D on $\mathbb{R} \times \mathbb{R}^d$ with $\int_{\mathbb{R}^{d+1}} y^2 \, \mathsf{dD}(y,x) < \infty$. Let D_X denote the marginal distribution of X_1.

A1. The parameter space Θ is an open subset of $\mathbb{R}^p$.

A2. The model function $f : \mathbb{R}^d \times \Theta \to \mathbb{R}$ is measurable and twice continuously differentiable with respect to θ.
$\mathsf{E}[f(X,\theta)^2] < \infty$, $\mathsf{E}[[f'(X,\theta)]_{\alpha]}^2] < \infty$ and $\mathsf{E}[[f''(X,\theta)]_{\alpha\beta]}^2] < \infty$ for each fixed $\theta \in \Theta$, $\alpha, \beta = 1, \ldots, p$ and $X \sim \mathsf{D}_X$. (Shortly spoken, f and its partial derivatives up to second order are $L^2(D_X)$ for each fixed $\theta \in \Theta$.)
Further $f''_{\alpha\beta}$ is Lipschitz continuous for all $\alpha, \beta = 1, \ldots, p$.

A3. The kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ is twice continuously differentiable with respect to θ.
k is symmetric in the first two arguments and k, k'_α and $k''_{\alpha\beta}$ are bounded functions on $\mathbb{R}^d \times \mathbb{R}^d$ for each fixed $\theta \in \Theta$ and all $\alpha, \beta = 1, \ldots, p$.
Further $k''_{\alpha\beta}$ is Lipschitz continuous for all $\alpha, \beta = 1, \ldots, p$.
Define $\mathcal{K}_\theta : g(\cdot) \mapsto \int_{\mathbb{R}} k(\cdot, t, \theta) g(t) \, \mathsf{dD}_X$. For each fixed $\theta \in \Theta$ the Hilbert-Schmidt operator $\mathcal{K}_\theta$ on $L^2(D_X)$ is positive definite.

A4. Under $\mathbf{H_0}$

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} w(X_i, \theta_0) U_i + o_p(1) \quad ,$$

where $w(\cdot;\theta) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^p$ is a measurable function, continuous in θ, with $\mathsf{E}[[w(X_1;\theta)]_{\alpha]}^2] < \infty$ for each fixed $\theta \in \Theta$ and all $\alpha = 1, \ldots, p$, and $U_i = u(Y_i, X_i, \theta_0) = Y_i - f(X_i, \theta_0)$ are the true errors.

A5. $\mathsf{E}\big[u(Z_1,\theta_0)^2 [w(X_1;\theta_0)]_\alpha [w(X_1;\theta_0)]_\beta|\big] < \infty$ for all $\alpha, \beta = 1, \ldots, p$.

A6. Under $\mathbf{H_1}$ there exists a θ_1 in the interior of Θ such that
$n^{\frac{1}{2}}(\hat{\theta} - \theta_1) = O_p(1)$.

4 Proofs

4.1 Proof of theorem 2.1

We use the Taylor expansions

$$\begin{aligned}\hat{U}_i &:= Y_i - f(X_i,\hat{\theta}) = U_i - \Big(f(X_i,\hat{\theta}) - f(X_i,\theta_0)\Big) \\ &= U_i - f'(X_i,\theta_0)(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_0)^T f''(X_i,\theta_i^*)(\hat{\theta}-\theta_0) \\ &=: U_i - F_i'(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_0)^T F_i''(\theta_i^*)(\hat{\theta}-\theta_0) \qquad (12)\end{aligned}$$

and

$$\begin{aligned}\hat{K}_{ij} &:= k(X_i,X_j,\hat{\theta}) = k(X_i,X_j,\theta_0) \\ &\quad + k'(X_i,X_j,\theta_0)(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_0)^T k''(X_i,X_j,\theta_{ij}^*)(\hat{\theta}-\theta_0) \\ &=: K_{ij} + K_{ij}'(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_0)^T K_{ij}''(\theta_{ij}^*)(\hat{\theta}-\theta_0)\ . \qquad (13)\end{aligned}$$

Proposition 4.1. *Under* $\mathbf{H_0}$ *and assumptions* **A0** *to* **A4**

$$\begin{aligned}\hat{T}_n &= \tfrac{1}{n}\sum_{1\le i<j\le n} U_iU_jK_{ij} - \tfrac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}F_j'(\hat{\theta}-\theta_0) \qquad (14)\\ &\quad + \tfrac{1}{n}\sum_{1\le i<j\le n} K_{ij}F_i'(\hat{\theta}-\theta_0)F_j'(\hat{\theta}-\theta_0) + O_p(n^{-\frac{1}{2}})\ . \qquad (15)\end{aligned}$$

Proof. The expansions are now applied to $\hat{T}_n$, noting that **A4** forces each component of $(\hat{\theta}-\theta_0)$ to be $O_p(n^{-\frac{1}{2}})$:

$$\begin{aligned}\hat{T}_n &= \tfrac{1}{n}\sum_{1\le i<j\le n} \hat{U}_i\hat{U}_j\hat{K}_{ij} \\ &= \tfrac{1}{n}\sum_{1\le i<j\le n} U_iU_jK_{ij} \qquad (16)\\ &\quad + \tfrac{1}{n}\sum_{1\le i<j\le n} U_iU_j\big\{K_{ij}'(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_0)^T K_{ij}''(\theta_{ij}^*)(\hat{\theta}-\theta_0)\big\} \qquad (17)\\ &\quad - \tfrac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}F_j'(\hat{\theta}-\theta_0) \qquad (18)\\ &\quad + \tfrac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}(\hat{\theta}-\theta_0)^T F_j''(\theta_j^*)(\hat{\theta}-\theta_0) \qquad (19)\\ &\quad - \tfrac{1}{n}\sum_{1\le i\ne j\le n} U_i\big\{K_{ij}'(\hat{\theta}-\theta_0)F_j'(\hat{\theta}-\theta_0) + O_p(n^{-\frac{3}{2}})\big\} \qquad (20)\\ &\quad + \tfrac{1}{n}\sum_{1\le i<j\le n} K_{ij}\big\{F_i'(\hat{\theta}-\theta_0)F_j'(\hat{\theta}-\theta_0) + O_p(n^{-\frac{3}{2}})\big\} + O_p(n^{-\frac{1}{2}})\ . \qquad (21)\end{aligned}$$

Thus it remains to show that the expressions in (17), (19) and (20) are $O_p(n^{-\frac{1}{2}})$.

a) Analysis of (17):

$$\frac{1}{n} \sum_{1\le i<j\le n} U_i U_j K'_{ij}(\hat\theta - \theta_0) = \frac{1}{n} \sum_{\alpha=1}^{p} \sum_{1\le i<j\le n} U_i U_j [K'_{ij}]_\alpha [\hat\theta - \theta_0]_\alpha \quad . \qquad (22)$$

Since **A4** forces each component of $(\hat\theta - \theta_0)$ to be $O_p(n^{-\frac{1}{2}})$, it suffices to show $\frac{1}{n}\sum_{1\le i<j\le n} U_i U_j [K'_{ij}]_\alpha = O_p(1)$ for $\alpha = 1, \dots, p$. But this follows directly from lemma 5.1 (c) with $g = h = u$ and $b = [k']_\alpha$.

$$\frac{1}{n} \sum_{1\le i<j\le n} U_i U_j (\hat\theta - \theta_0)^T K''_{ij}(\theta^*_{ij})(\hat\theta - \theta_0) =$$

$$\sum_{\alpha,\beta=1}^{p} [\hat\theta - \theta_0]_\alpha [\hat\theta - \theta_0]_\beta \frac{1}{n} \sum_{1\le i<j\le n} U_i U_j [K''_{ij}(\theta^*_{ij})]_{\alpha,\beta} \quad . \qquad (23)$$

Again by **A4** it suffices to show $\frac{1}{n}\sum_{1\le i<j\le n} U_i U_j [K''_{ij}(\theta^*_{ij})]_{\alpha,\beta} = O_p(n^{\frac{1}{2}})$ for all $\alpha, \beta = 1 \dots, p$:

$$\left| \frac{1}{n} \sum_{1\le i<j\le n} U_i U_j [K''_{ij}(\theta^*_{ij})]_{\alpha,\beta} \right| \le \left| \frac{1}{n} \sum_{1\le i<j\le n} U_i U_j [K''_{ij}(\theta_0)]_{\alpha,\beta} \right| + \left(\frac{1}{n} \sum_{1\le i<j\le n} |U_i||U_j| \right) \lambda \|\hat\theta - \theta_0\| \, . \qquad (24)$$

Here we used that by **A3** there exists a Lipschitz constant λ such that

$$\|[K''_{ij}(\theta^*_{ij})]_{\alpha,\beta} - [K''_{ij}(\theta_0)]_{\alpha,\beta}\| \le \lambda \|\theta^*_{ij} - \theta_0\| \le \lambda \|\hat\theta - \theta_0\| \, . \qquad (25)$$

The first summand of the right hand side of (24) is $O_p(1)$ by lemma 5.1 (c) with $g = h = u$ and $b = [k'']_{\alpha,\beta}$. The second summand is $O_p(n^{\frac{1}{2}})$ because of **A4** and since $\frac{1}{n}\sum_{1\le i<j\le n} |U_i||U_j| = O_p(n)$ by 5.1 a) with $g = h = |u|$ and $b = 1$.

b) Analysis of (19):

$$\frac{1}{n} \sum_{1\le i\ne j\le n} U_i K_{ij} (\hat\theta - \theta_0)^T F''_j(\theta^*_i)(\hat\theta - \theta_0) =$$

$$\sum_{\alpha,\beta=1}^{p} [\hat\theta - \theta_0]_\alpha [\hat\theta - \theta_0]_\beta \frac{1}{n} \sum_{1\le i\ne j\le n} U_i K_{ij} [F''_j(\theta^*_i)]_{\alpha,\beta} \quad . \qquad (26)$$

Again by **A4** it suffices to show $\frac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}[F_j''(\theta_j^*)]_{\alpha,\beta} = O_p(n^{\frac{1}{2}})$ for all $\alpha, \beta = 1\ldots, p$:

$$\left|\frac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}[F''(\theta_j^*)]_{\alpha,\beta}\right| \le \left|\frac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}[F''(\theta_0)]_{\alpha,\beta}\right| + \left(\frac{1}{n}\sum_{1\le i\ne j\le n} |U_i||K_{ij}|\right)\lambda'\|\hat\theta-\theta_0\| \ . \tag{27}$$

Here we used that by **A2** there exists a Lipschitz constant λ' such that

$$\|[F_j''(\theta_j^*)]_{\alpha,\beta} - [F_j''(\theta_0)]_{\alpha,\beta}\| \le \lambda'\|\theta_j^*-\theta_0\| \le \lambda'\|\hat\theta-\theta_0\| \ . \tag{28}$$

The first summand of the right hand side of (27) is $O_p(n^{\frac{1}{2}})$ by lemma 5.1 (b) with $g = u$, $h = [f'']_{\alpha,\beta}$ and $b = k$. The second summand is $O_p(n^{\frac{1}{2}})$ because of **A4** and since $\frac{1}{n}\sum_{1\le i\ne j\le n}|U_i||K_{ij}| = O_p(n)$ by 5.1 a) with $g = |u|$, $h = 1$ and $b = |k|$.

c) Analysis of (20):

$$\frac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}'(\hat\theta-\theta_0)F_j'(\hat\theta-\theta_0) = \sum_{\alpha,\beta=1}^{p}[\hat\theta-\theta_0]_\alpha[\hat\theta-\theta_0]_\beta\frac{1}{n}\sum_{1\le i\ne j\le n} U_i[K_{ij}']_\alpha[F_j']_\beta \quad . \tag{29}$$

Again by **A4** it suffices to show $\frac{1}{n}\sum_{1\le i\ne j\le n} U_i[K_{ij}']_\alpha[F_j']_\beta = O_p(n^{\frac{1}{2}})$ for all $\alpha, \beta = 1\ldots, p$. But this is the case by lemma 5.1(b) with $g = u$, $h = [f']_\beta$ and $b = k'$.

□

In the following we use the shorthands $V_i := v(X_i, \theta_0)$ and $W_i := w(Z_i; \theta_0)$.

Proposition 4.2. *Under* $\mathbf{H_0}$ *and assumptions* **A0** *to* **A4**

$$\frac{1}{n}\sum_{1\le i\ne j\le n} U_iK_{ij}F_j'(\hat\theta-\theta_0) = \frac{1}{n}\sum_{1\le i<j\le n} U_iU_j\{V_iW_j + V_jW_i\} + \frac{1}{n}\sum_{i=1}^{n} U_i^2V_iW_i + o_p(1) \ . \tag{30}$$

Proof.

$$\frac{1}{n}\sum_{1\le i\ne j\le n} U_i K_{ij} F_j'(\hat\theta-\theta_0) = \sum_{\alpha=1}^{p}[\hat\theta-\theta_0]_\alpha \frac{1}{n}\sum_{i\ne j} U_i K_{ij}[F_j']_\alpha \ . \tag{31}$$

Consider

$$\begin{aligned}\frac{1}{n}\sum_{1\le i\ne j\le n} U_i K_{ij}[F_j']_\alpha &= \frac{1}{n}\sum_{i=1}^{n} U_i \sum_{j\ne i}(K_{ij}[F_j']_\alpha \\ &= \frac{1}{n}\sum_{i=1}^{n} U_i[V_i]_\alpha + \frac{1}{n}\sum_{i=1}^{n} U_i\left([V_i]_\alpha - \sum_{j\ne i} K_{ij}[F_j']_\alpha\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} U_i[V_i]_\alpha + O_p(1)\ ,\end{aligned} \tag{32}$$

since $\frac{1}{n}\sum_{j=1}^{n} K_{ij}[F_j']_\alpha \to [V_i]_\alpha$ and $\frac{1}{n}\sum_{i=1}^{n} U_i K_{ii}[F_i']_\alpha \to 0$ almost surely by the SLLN. Thus

$$\frac{1}{n}\sum_{1\le i\ne j\le n} U_i K_{ij} F_j'(\hat\theta-\theta_0) = \sum_{i=1}^{n} U_i V_i(\hat\theta-\theta_0) + O_p(n^{-\frac{1}{2}})\ . \tag{33}$$

Plugging the development of **A4** into (33) we obtain

$$\begin{aligned}\frac{1}{n}\sum_{1\le i\ne j\le n} U_i K_{ij} F_j'(\hat\theta-\theta_0) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} U_i V_i W_j U_j + o_p(1) \\ &= \frac{1}{n}\sum_{1\le i\ne j\le n} U_i V_i W_j U_j + \frac{1}{n}\sum_{i=1}^{n} U_i^2 V_i W_i + o_p(1) \\ &= \frac{1}{n}\sum_{1\le i<j\le n} U_i U_j\{V_i W_j + V_j W_i\} \qquad (34) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} U_i^2 V_i W_i + o_p(1)\ .\end{aligned}$$

□

Proposition 4.3. *Under* $\mathbf{H_0}$ *and assumptions* **A0** *to* **A4**

$$\frac{1}{n}\sum_{1\le i<j\le n} K_{ij} F_i'(\hat\theta-\theta_0) F_j'(\hat\theta-\theta_0) =$$

$$\frac{1}{n}\sum_{1\le i<j\le n} U_i U_j W_i^T m\, W_j + \frac{1}{2n}\sum_{i=1}^{n} U_i^2 W_i^T m\, W_i + o_p(1) \quad . \tag{35}$$

Proof. Since $F_i'(\hat\theta-\theta_0)\in I\!R$,

$$\begin{aligned}
&\sum_{1\le i<j\le n} K_{ij} F_i'(\hat\theta-\theta_0) F_j'(\hat\theta-\theta_0) \\
&= \sum_{1\le i<j\le n} K_{ij} (\hat\theta-\theta_0) F_i' F_j' (\hat\theta-\theta_0) \\
&= (\hat\theta-\theta_0)^T \underbrace{\left(\sum_{1\le i<j\le n} K_{ij} F_i' F_j'\right)}_{:=\tilde m_n} (\hat\theta-\theta_0) \ . \qquad (36)
\end{aligned}$$

Now consider $[\tilde m_n]_{\alpha,\beta} = \sum_{1\le i<j\le n} K_{ij}[F_i']_\alpha [F_j']_\beta$, the (α,β)-th component of the (p,p)-matrix $\tilde m_n$:

$$\begin{aligned}
[\tilde m_n]_{\alpha,\beta} &= \sum_{1\le i<j\le n} \left\{K_{ij}[F_i']_\alpha[F_j']_\beta - \mathsf{E}[K_{ij}[F_i']_\alpha[F_j']_\beta]\right\} + \sum_{1\le i<j\le n} \mathsf{E}[K_{ij}[F_i']_\alpha[F_j']_\beta] \qquad (37) \\
&= O_p(n^{\frac{3}{2}}) + \tfrac{n(n-1)}{2}\mathsf{E}[K_{12}[F_1']_\alpha[F_2']_\beta] \ ,
\end{aligned}$$

because the first sum of (37) is a centered nondegenerate U-statistic to which lemma 5.2 (a) applies. Applying this to each component of (36) and using **A4** we obtain

$$\tfrac{1}{n}\sum_{1\le i<j\le n} K_{ij} F_i'(\hat\theta-\theta_0) F_j'(\hat\theta-\theta_0) = \tfrac{1}{2} n(\hat\theta-\theta)^T m\,(\hat\theta-\theta) + O_p(n^{-\frac{1}{2}}) \ . \qquad (38)$$

Substituting the development of **A4** for $(\hat\theta-\theta)$ results in

$$\begin{aligned}
n(\hat\theta-\theta)^T m\,(\hat\theta-\theta) &= n^{-1}\left(\sum_{i=1}^n W_i U_i\right)^T m \left(\sum_{j=1}^n W_j U_j\right) + o_p(1) \\
&= n^{-1}\sum_{i,j=1}^n U_i U_j W_i^T m\, W_j + o_p(1) \\
&= 2n^{-1}\sum_{1\le i<j\le n} U_i U_j W_i^T m\, W_j \\
&\quad + n^{-1}\sum_{i=1}^n U_i^2 W_i^T m\, W_i + o_p(1) \ . \qquad (39)
\end{aligned}$$

Combining (38) and (39) now gives (35).

□

Now we can conclude with the proof of the theorem. Applying propositions 4.2 and 4.3 to the statement of proposition 4.1 we obtain

$$\begin{aligned}\hat{T}_n &= \tfrac{1}{n} \sum_{1\le i<j\le n} U_i U_j K_{ij} \\ &\quad - \tfrac{1}{n} \sum_{1\le i<j\le n} U_i U_j \{V_i W_j + V_j W_i\} - \tfrac{1}{n} \sum_{i=1}^{n} U_i^2 V_i W_i \\ &\quad + \tfrac{1}{n} \sum_{1\le i<j\le n} U_i U_j W_i^T m\, W_j + \tfrac{1}{2n} \sum_{i=1}^{n} U_I^2 W_i^T m\, W_i + o_p(1) \quad . \end{aligned} \tag{40}$$

Using **A2** and **A3** with $|k| < c$, by the Cauchy-Schwarz inequality we have $[V_1]_\alpha \le \big(c^2\, \mathsf{E}\big[[F_1']_\alpha^2]\big)^{\frac{1}{2}} =: c_\alpha < c_v < \infty$, say, and hence **A0** and **A5** give $\mathsf{E}[|U_1^2 [V_1]_\alpha [W_1]_\alpha|] \le c_v\, \mathsf{E}[U_1^2(1 + [W_1]_\alpha^2)] < \infty$ for $\alpha = 1, \dots, p$. Also $[m]_{\alpha,\beta} = \mathsf{E}\big[[F_1']_\alpha [V_2]_\beta\big] \le c_v\, \mathsf{E}\big[[F_1']_\alpha^2\big] < c_m$, say. By the law of large numbers

$$\tfrac{1}{n} \sum_{i=1}^{n} U_i^2 V_i W_i \xrightarrow{P} \mathsf{E}[U_1^2 V_1 W_1] \tag{41}$$

and also

$$\tfrac{1}{n} \sum_{i=1}^{n} U_i^2 W_i^T m\, W_i \xrightarrow{P} \mathsf{E}[U_1^2 W_1^T m\, W_1] \quad . \tag{42}$$

This proves (6). The remaining part of theorem 2.1 follows directly from lemma 5.2(b), if $\mathsf{E}[q(Z_1, Z_2)^2] < \infty$ and $\mathsf{E}[q(Z_1, Z_2)|Z_2] = 0$.

Proposition 4.4. *For $Z_1, Z_2 \sim \mathsf{D}$ the functional q as defined in* (5) *satisfies $\mathsf{E}[q(Z_1, Z_2)^2] < \infty$ and $\mathsf{E}[q(Z_1, Z_2)|Z_2] = 0$.*

Proof.

$$\begin{aligned}\mathsf{E}[q(Z_1, Z_2)|Z_2] &= \mathsf{E}\big[U_1 U_2 \{K_{12} - V_1 W_2 - V_2 W_1 + W_1^T m\, W_2\}\big|Z_2\big] \\ &= \mathsf{E}_{X_1}\big[\mathsf{E}[U_1|X_1] U_2 \{K_{12} - V_1 W_2 - V_2 W_1 + W_1^T m\, W_2\}\big] \\ &= 0 \quad , \end{aligned}$$

since $\mathsf{E}[U_1|X_1] = 0$.
Writing $\mathbf{1} := (1, \dots, 1)^T \in I\!R^p$ and using the upper bounds c_v and c_m defined above we have

$$
\begin{aligned}
&\mathsf{E}[q(Z_1,Z_2)^2] \\
&\quad = \mathsf{E}[U_1^2U_2^2\{K_{12} - V_1W_2 - V_2W_1 - W_1^T m\, W_2\}^2] \\
&\quad = \mathsf{E}[U_1^2U_2^2K_{12}^2] + \mathsf{E}[U_1^2U_2^2(V_1W_2)^2] + \mathsf{E}[U_1^2U_2^2(V_2W_1)^2] \\
&\qquad + \mathsf{E}[U_1^2U_2^2(W_1^T m\, W_2)^2] - 2\,\mathsf{E}[U_1^2U_2^2K_{12}V_1W_2] - 2\,\mathsf{E}[U_1^2U_2^2K_{12}V_2W_1] \\
&\qquad + 2\,\mathsf{E}[U_1^2U_2^2K_{12}W_1^T m\, W_2] + 2\,\mathsf{E}[U_1^2U_2^2V_1W_2V_2W_1] \\
&\qquad - 2\,\mathsf{E}[U_1^2U_2^2V_1W_2W_1^T m\, W_2] - 2\,\mathsf{E}[U_1^2U_2^2V_2W_1W_1^T m\, W_2] \\
&\quad < c^2\left(\mathsf{E}[U_1^2]\right)^2 + 2c_v^2\,\mathsf{E}[U_1^2]\,\mathsf{E}[U_2^2(1^TW_2)^2] \\
&\qquad + c_m^2\,\mathsf{E}[U_1^2U_2^2(W_1^TW_2)^2] + 4cc_v\,\mathsf{E}[U_1^2]\,\mathsf{E}[U_2^2|1^TW_2|] \\
&\qquad + 2cc_m\,\mathsf{E}[U_1^2U_2^2|W_1^TW_2|] + 2c_v^2\left(\mathsf{E}[U_1^2|1^TW_1|]\right)^2 \\
&\qquad + 4c_vc_m\,\mathsf{E}[U_1^2U_2^2|1^TW_1||W_1^TW_2|] \\
&\quad \le c^2\left(\mathsf{E}[U_1^2]\right)^2 + 2c_v^2\,\mathsf{E}[U_1^2]\sum_{\alpha,\beta=1}^{p}\mathsf{E}[U_2^2[W_2]_\alpha[W_2]_\beta] \\
&\qquad + c_m^2\sum_{\alpha,\beta=1}^{p}\left(\mathsf{E}[U_1^2[W_1]_\alpha[W_1]_\beta]\right)^2 + 4cc_v\,\mathsf{E}[U_1^2]\sum_{\alpha=1}^{p}\mathsf{E}[U_2^2(1+[W_2]_\alpha^2)] \\
&\qquad + 2cc_m\left(\sum_{\alpha=1}^{p}\mathsf{E}[U_1^2(1+[W_1]_\alpha^2)]\right)^2 + 2c_v^2\left(\sum_{\alpha=1}^{p}\mathsf{E}[U_1^2(1+[W_1]_\alpha^2)]\right)^2 \\
&\qquad + 4c_vc_m\sum_{\alpha,\beta=1}^{p}\mathsf{E}[U_1^2(1+[W_1]_\alpha^2)]\,\mathsf{E}[U_2^2[W_2]_\alpha[W_2]_\beta] \\
&\quad < \infty \quad , \qquad\qquad (43)
\end{aligned}
$$

since $\mathsf{E}[U_1^2(1+[W_1]_\alpha^2)] \le \mathsf{E}[U_1^2] + \mathsf{E}[U_1^2[W_1]_\alpha^2]$ and $\left(\mathsf{E}[U_1^2[W_1]_\alpha[W_1]_\beta]\right)^2 \le \mathsf{E}[U_1^2[W_1]_\alpha^2]\,\mathsf{E}[U_1^2[W_1]_\beta^2]$, which are bounded by assumptions **A5** and **A0**. □

4.2 Proof of theorem 2.2

If **A6** holds, we have $[\hat{\theta} - \theta_1]_\alpha = O_p(n^{-\frac{1}{2}})$ for each $\alpha = 1, \dots, p$. We denote $\Delta(x) := \mathsf{E}[Y_i|X_i = x] - f(x, \theta_1)$, $\Delta_i := \Delta(X_i)$ and $\varepsilon(x) := Y_i - \mathsf{E}[Y_i|X_i = x]$,

$\varepsilon_i := \varepsilon(X_i)$. Then we use the Taylor expansions

$$
\begin{aligned}
\hat{U}_i &:= Y_i - f(X_i, \hat{\theta}) && (44)\\
&= (Y_i - \mathsf{E}[Y_i|X_i]) + (\mathsf{E}[Y_i|X_i] - f(X_i,\theta_1)) - \Big(f(X_i,\hat{\theta}) - f(X_i,\theta_1)\Big)\\
&=: \varepsilon_i + \Delta_i - F_i'(\hat{\theta}-\theta_1) + (\hat{\theta}-\theta_1)^T F_i''(\theta_i^\dagger)(\hat{\theta}-\theta_1) && (45)
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{K}_{ij} &:= k(X_i, X_j, h, \hat{\theta}) = k(X_i, X_j; h, \theta_1)\\
&\quad + k'(X_i, X_j; h, \theta_1)(\hat{\theta}-\theta_0) + (\hat{\theta}-\theta_1)^T k''(X_i, X_j; h, \theta_{ij}^\dagger)(\hat{\theta}-\theta_1)\\
&=: K_{ij} + K_{ij}'(\hat{\theta}-\theta_1) + (\hat{\theta}-\theta_1)^T K_{ij}''(\theta_{ij}^\dagger)(\hat{\theta}-\theta_1)\ . && (46)
\end{aligned}
$$

Note that $\mathsf{E}[\Delta(X_1)^2] < \infty$ if **A0** and **A2** hold.

Proposition 4.5. *Under* $\mathbf{H_1}$, *assumptions* **A0** *to* **A3** *and* **A6**

$$\hat{T}_n = n^{-1} \sum_{1 \le i < j \le n} K_{ij}\Delta_i\Delta_j + O_p(n^{\frac{1}{2}})\ . \qquad (47)$$

Proof. Applying the expansions to $\hat{T}_n$ in the same fashion as in the proof of proposition 4.1 gives

$$
\begin{aligned}
\hat{T}_n = n^{-1} &\sum_{1 \le i < j \le n} K_{ij}\Delta_i\Delta_j && (48)\\
+ n^{-1} &\sum_{1 \le i \ne j \le n} \varepsilon_i K_{ij}\Delta_j && (49)\\
+ n^{-1} &\sum_{1 \le i \ne j \le n} K_{ij}F_i'(\hat{\theta}-\theta_1)\Delta_j && (50)\\
+ n^{-1} &\sum_{1 \le i < j \le n} K_{ij}'(\hat{\theta}-\theta_1)\Delta_i\Delta_j + O_p(1)\ . && (51)
\end{aligned}
$$

All terms not stated explicitly above are essentially the same as in (16) to (21) in the proof of proposition 4.1 with U_i and θ_0 replaced by ε_i and θ_1. These are there shown to be at least $O_p(1)$.
By lemma 5.1 b) with $g = \varepsilon$, $h = \Delta$ and $b = k$ (49) is $O_p(n^{\frac{1}{2}})$. Writing

$$n^{-1} \sum_{1 \le i \ne j \le n} K_{ij}F_i'(\hat{\theta}-\theta_1)\Delta_j = \sum_{\alpha=1}^{p} [\hat{\theta}-\theta_1]_\alpha n^{-1} \sum_{1 \le i \ne j \le n} K_{ij}[F_i']_\alpha \Delta_j \quad .$$

we see that (50) is $O_p(n^{\frac{1}{2}})$, since $\sum_{1\le i\ne j\le n} K_{ij}[F_i']_\alpha \Delta_j = O_p(n^2)$ by lemma 5.1 a) with $g = [f']_\alpha$, $h = \Delta$ and $b = k$. Similarily (51) is $O_p(n^{\frac{1}{2}})$, since $\sum_{1\le i<j\le n}[K_{ij}']_\alpha \Delta_i\Delta_j = O_p(n^2)$ by lemma 5.1 a) with $g = h = \Delta$ and $b = [k']_\alpha$. □

For proving theorem 2.2 it now suffices to show that $n^{-2}\sum_{1\le i<j\le n} K_{ij}\Delta_i\Delta_j$ converges in probability to a positive constant. By lemma 5.2 a) we have

$$\begin{aligned} n^{-2}\sum_{1\le i<j\le n} K_{ij}\Delta_i\Delta_j &= n^{-2}\sum_{1\le i<j\le n}\{K_{ij}\Delta_i\Delta_j - \mathsf{E}\Big[K_{ij}\Delta_i\Delta_j\Big]\} \\ &\quad + n^{-2}\sum_{1\le i<j\le n}\mathsf{E}\Big[K_{ij}\Delta_i\Delta_j\Big] \\ &= O_p(n^{-\frac{1}{2}}) + \mathsf{E}\Big[K_{12}\Delta_1\Delta_2\Big] \quad . \end{aligned}$$

Thus it remains to show $\mathsf{E}[K_{12}\Delta_1\Delta_2] > 0$. But this is a direct consequence of the positive definiteness of $\mathcal{K}_\theta$, demanded in **A3**.

4.3 Proof of theorem 2.3

By theorem 2.1 we have

$$\mathsf{P}\{\hat{T}_n \le x\} = \mathsf{P}\{Q_n + C_n \le x\}$$

with

$$Q_n = \tfrac{1}{n}\sum_{1\le i<j\le n} q(Z_i, Z_j; \theta_0)$$

and

$$\begin{aligned} C_n = \tfrac{1}{2n}\sum_{i=1}^{n} u(Z_i,\theta_0)^2 w(X_i;\theta_0)^T m(\theta_0)\, w(X_i;\theta_0) \\ - \tfrac{1}{n}\sum_{i=1}^{n} u(Z_i,\theta_0)^2 v(X_i;\theta_0) w(X_i;\theta_0) + o_p(1)\,, \end{aligned}$$

as well as

$$\mathsf{P}^*\{\hat{T}_n^* \le x | Z_1,\dots,Z_n\} = \mathsf{P}^*\{Q_n^* + C_n^* \le x | Z_1,\dots,Z_n\}$$

with

$$Q_n^* = \tfrac{1}{n}\sum_{1\le i<j\le n} q_n^*(Z_i^*, Z_j^*; \hat{\theta}_n)$$

and

$$
\begin{aligned}
C_n^* = \tfrac{1}{2n} \sum_{i=1}^{n} u(Z_i^*, \hat\theta_n)^2 w(X_i; \hat\theta_n)^T m_n^*(\hat\theta_n)\, w(X_i; \hat\theta_n) \\
- \tfrac{1}{n} \sum_{i=1}^{n} u(Z_i^*, \hat\theta_n)^2 v_n^*(X_i; \hat\theta_n) w(X_i; \hat\theta_n) + o_p(1) \,.
\end{aligned}
$$

Here

$$
\begin{aligned}
v_n^*(x;\theta) &:= \tfrac{1}{n} \sum_{j=1}^{n} k(x, X_j, \theta) f'(X_j, \theta) \,, \\
m_n^*(\theta) &:= \Big(\sum_{i,j=1}^{n} k(X_i, X_j; \theta) [f'(X_i;\theta)]_\alpha [f'(X_j;\theta)]_\beta \Big)_{\alpha,\beta=1,\dots,p} \,,
\end{aligned}
$$

and q_n^* is defined as q with v_n^* and m_n^* replacing v and m.
Now lemma 5.3 applies, and for each $\varepsilon > 0$

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big| \mathsf{P}^*\{Q_n^* + C_n^* \le x | Z_1, \dots, Z_n\} - \mathsf{P}\{Q_n + C_n \le x\} \Big| \\
&\quad \le \sup_{x\in\mathbb{R}} \Big| \mathsf{P}^*\{Q_n^* \le x | Z_1, \dots, Z_n\} - \mathsf{P}\{Q_n \le x\} \Big| && (52) \\
&\qquad + \sup_{x\in\mathbb{R}} \Big(\mathsf{P}\{Q_n \le x + \varepsilon\} - \mathsf{P}\{Q_n \le x - \varepsilon\} \Big) && (53) \\
&\qquad + \mathsf{P}\{|C_n - c| \ge \varepsilon\} && (54) \\
&\qquad + \mathsf{P}^*\{|C_n^* - c| \ge \varepsilon | Z_1, \dots, Z_n\} \,. && (55)
\end{aligned}
$$

Hence we have to show that the expressions (52) to (55) vanish or are at least of order $O(\varepsilon)$ for $n \to \infty$.

a) Analysis of (52):

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big| \mathsf{P}^*\{Q_n^* \le x | Z_1, \dots, Z_n\} - \mathsf{P}\{Q_n \le x\} \Big| \\
&\quad \le \sup_{x\in\mathbb{R}} \Big| \mathsf{P}^*\{\tfrac{1}{n} \sum_{1 \le i < j \le n} q_n^*(Z_i^*, Z_j^*; \hat\theta_n) \le x | Z_1, \dots, Z_n\} \\
&\qquad\qquad - \mathsf{P}^*\{\tfrac{1}{n} \sum_{1 \le i < j \le n} \eta_i \eta_j q(Z_i, Z_j; \theta_0) \le x | Z_1, \dots, Z_n\} \Big| && (56) \\
&\qquad + \sup_{x\in\mathbb{R}} \Big| \mathsf{P}^*\{\tfrac{1}{n} \sum_{1 \le i < j \le n} \eta_i \eta_j q(Z_i, Z_j; \theta_0) \le x | Z_1, \dots, Z_n\} \\
&\qquad\qquad - \mathsf{P}\{\tfrac{1}{n} \sum_{1 \le i < j \le n} q(Z_i, Z_j; \theta_0) \le x\} \Big| \,. && (57)
\end{aligned}
$$

The i.i.d. weighted bootstrap of quadratic forms as formulated by Dehling & Mikosch (1994, theorem 3.1) states exactly that (57) vanishes for $n \to \infty$.

By lemma 5.4 we obtain the almost sure pointwise convergences $v_n^*(x;\hat{\theta}_n) \to v(x;\theta_0)$ and $m_n^*(\hat{\theta}_n) \to m(\theta_0)$ as $n \to \infty$.
Also, by continuity, we have $u(z,\hat{\theta}_n) \to u(z,\theta_0)$ and $w(z,\hat{\theta}_n) \to w(z,\theta_0)$ almost surely. Thus we obtain for all $z_1, z_2 \in I\!R^{d+1}$

$$q_n^*(z_1,z_2;\hat{\theta}_n) \to q(z_1,z_2;\theta_0) \quad \text{almost surely} . \tag{58}$$

Note that by definition of Y_i^* and q_n^*

$$\frac{1}{n} \sum_{1 \le i<j \le n} q_n^*(Z_i^*, Z_j^*; \hat{\theta}_n) = \frac{1}{n} \sum_{1 \le i<j \le n} \eta_i \eta_j q_n^*(Z_i, Z_j; \hat{\theta}_n) \quad . \tag{59}$$

We now show the convergence in probability given $Z_1, \ldots, Z_n$:

$$R_n := \frac{1}{n} \sum_{1 \le i<j \le n} \eta_i \eta_j (q_n^*(Z_i, Z_j; \hat{\theta}_n) - q(Z_i, Z_j; \theta_0)) \xrightarrow{P^*} 0 \, . \tag{60}$$

Obviously $\mathsf{E}[R_n|Z_1, \ldots, Z_n] = 0$. By lemma 5.4 (a)

$$\mathsf{E}[R_n^2|Z_1, \ldots, Z_n] = \frac{1}{n^2} \sum_{1 \le i<j \le n} (q_n^*(Z_i, Z_j; \hat{\theta}_n) - q(Z_i, Z_j; \theta_0))^2 \to 0 \, , \tag{61}$$

if we show that $(q_n^*(z_1,z_2;\hat{\theta}_n) - q(z_1,z_2;\theta_0))^2 \le b(z_1,z_2)$ for some dominating function b with $\mathsf{E}[b(Z_1,Z_2)] < \infty$. Such a b exists by the triangle inequality, if there are dominating functions b_1 and b_2 for

$$(q_n^*(z_1,z_2;\hat{\theta}_n) - q(z_1,z_2;\hat{\theta}_n))^2 \tag{62}$$

and

$$(q(z_1,z_2;\hat{\theta}_n) - q(z_1,z_2;\theta_0))^2 \tag{63}$$

respectively. For (63) the existence of b_2 follows by the argument of lemma 5.4 b). So consider (62), which we can write as

$$\begin{aligned} u(z_1,\hat{\theta}_n)^2 u(z_2,\hat{\theta}_n)^2 \cdot \Big\{ & \big(v(x_1;\hat{\theta}_n) - v_n^*(x_1;\hat{\theta}_n)\big) w(x_2;\hat{\theta}_n) \\ & + \big(v(x_2;\hat{\theta}_n) - v_n^*(x_2;\hat{\theta}_n)\big) w(x_1;\hat{\theta}_n) \\ & - w(x_1;\hat{\theta}_n)^T \big(m(\hat{\theta}_n) - m_n^*(\hat{\theta}_n)\big) w(x_2;\hat{\theta}_n) \Big\}^2 \, . \end{aligned} \tag{64}$$

If $[v(x;\hat\theta_n)]_\alpha - [v_n^*(x;\hat\theta_n)]_\alpha$ and $[m(\hat\theta_n)]_{\alpha,\beta} - [m_n^*(\hat\theta_n)]_{\alpha,\beta}$ are uniformly bounded by constants c_v^* and c_m^*, we can proceed again as in the proof of lemma 5.4 b) to overcome the dependence on $\hat\theta_n$, and as in (43) to show the boundedness.
The existence of c_m^* follows directly, e.g. by lemma 5.4 b), for almost every sequence $\{Z_n\}$. For c_v^* we can argue almost surely as follows, assuming n sufficiently large:

$$
\begin{aligned}
&\big|[v(x;\hat\theta_n)]_\alpha - [v_n^*(x;\hat\theta_n)]_\alpha\big| \\
&\qquad \le \big|[v(x;\hat\theta_n)]_\alpha - [v(x;\theta_0)]_\alpha\big| + \big|[v_n^*(x;\theta_0)]_\alpha - [v_n^*(x;\hat\theta_n)]_\alpha\big| \\
&\qquad\qquad + \big|[v(x;\theta_0)]_\alpha - [v_n^*(x;\theta_0)]_\alpha\big| \\
&\qquad \le 2c\big|\mathsf{E}[[f''(X,\theta_0)]_\alpha] + \lambda^*\big|\delta_\alpha + 2c\Big|\tfrac{1}{n}\sum_{j=1}^n f''(X_j,\theta_j^*)(\hat\theta_n - \theta_0)\Big| \\
&\qquad\qquad + \tfrac{1}{\sqrt{2n\log\log n}}\Big|\Big(\sum_{j=1}^n k(x,X_j,\theta_0)[f'(X_j,\theta_0)]_\alpha - [v(x;\theta_0)]_\alpha\Big)\Big| \\
&\qquad \le 2c\big|\mathsf{E}[[f''(X,\theta_0)]_\alpha] + \lambda^*\big|\delta_\alpha + 2c\Big|\sum_\beta(\mathsf{E}[[f''(X,\theta_0)]_\beta] + \lambda^*)\Big|\delta_\alpha \\
&\qquad\qquad + \big(\mathsf{Var}[k(x,X,\theta_0)[f'(X,\theta_0)]_\alpha] + 1\big)^{\frac12} \\
&\qquad \le 2c\delta_\alpha(p+1)\big(\lambda^* + \max_\beta\big|\mathsf{E}[[f''(X,\theta_0)]_\beta]\big|\big) \\
&\qquad\qquad + \big(c^2\,\mathsf{E}[[f'(X,\theta_0)]_\alpha^2] - c_v^2 + 1\big)^{\frac12} \\
&\qquad =: c_v^* \quad .
\end{aligned}
$$

Here we have used the LIL, the previously introduced bound $|k| < c$, as well as $|[\hat\theta_n - \theta_0]_\alpha| < \delta_\alpha$ and a constant λ^*, greater than the Lipschitz bound for f''.
Thus we have proven (60), and now convergence in distribution, given the sequence $\{Z_n\}$, follows:

$$
\frac{1}{n}\sum_{1\le i<j\le n}\eta_i\eta_j q_n^*(Z_i,Z_j;\hat\theta_n) \rightsquigarrow \frac{1}{n}\sum_{1\le i<j\le n}\eta_i\eta_j q(Z_i,Z_j;\theta_0)) \ .
$$

Since the limiting distribution of this last U-statistic is continuous by theorem 2.1, the convergence of the according distribution functions is uniform, see e.g. Petrov (1995, page 15). Thus we have finally shown (56) to converge to zero.

b) Analysis of (53): By the continuity of the limiting distribution this term is of order $O(\varepsilon)$ for $n \to \infty$.

c) Analysis of (54): As in (41) and (42) $C_n \to c$ in probability, which was to show here.

d) Analysis of (55):

$$
\begin{aligned}
&\mathsf{P}^*\{|C_n^* - c| \geq \varepsilon | Z_1, \dots, Z_n\} \leq \\
&\quad \mathsf{P}^*\left\{ \left| \tfrac{1}{n} \sum_{i=1}^{n} \eta_i^2 \hat{U}_i^2 \hat{V}_i^* \hat{W}_i - \mathsf{E}[U_1^2 V_1 W_1] \right| \geq \tfrac{\varepsilon}{2} \,\middle|\, Z_1, \dots, Z_n \right\} \qquad (65) \\
&\quad + \mathsf{P}^*\left\{ \left| \sum_{i=1}^{n} \eta_i^2 \hat{U}_i^2 \hat{W}_i^T \hat{m}^* \hat{W}_i - \mathsf{E}[U_1^2 W_1^T m W_1] \right| \geq \tfrac{\varepsilon}{2} \,\middle|\, Z_1, \dots, Z_n \right\}, \qquad (66)
\end{aligned}
$$

using the shorthands $\hat{U}_i = u(Z_i, \hat{\theta}_n)$, $\hat{V}_i^* := v_n^*(X_i, \hat{\theta}_n)$, $\hat{W}_i = w(X_i, \hat{\theta}_n)$ and $\hat{m}^* := m_n^*(\hat{\theta}_n)$.
First consider (65). This is less or equal to

$$
\begin{aligned}
&\mathsf{P}^*\left\{ \left| \tfrac{1}{n} \sum_{i=1}^{n} (\eta_i^2 - 1) \hat{U}_i^2 \hat{V}_i^* \hat{W}_i \right| \geq \tfrac{\varepsilon}{4} \,\middle|\, Z_1, \dots, Z_n \right\} \\
&\qquad + \mathsf{P}^*\left\{ \left| \tfrac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2 \hat{V}_i^* \hat{W}_i - \mathsf{E}[U_1^2 V_1 W_1] \right| \geq \tfrac{\varepsilon}{4} \,\middle|\, Z_1, \dots, Z_n \right\}, \qquad (67)
\end{aligned}
$$

where the first summand vanishes by a version of the WLLN, see Petrov (1995), using **A5** and lemma 5.4, since $\mathsf{E}[\eta_1^2 - 1] = 0$. The second summand, which is nonrandom, eventually equals zero, since by lemma 5.4 $\frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2 \hat{V}_i^* \hat{W}_i \to \mathsf{E}[U_1^2 V_1 W_1]$ almost surely.
Finally (66) is handled completely analogously to (65).

5 Lemmas

Lemma 5.1. *Let* $Z_i = (Y_i, X_i)$, $i = 1, 2, \ldots,$ *be i.i.d. real random* $(1+d)$*-vectors and* $g, h : I\!R^{1+d} \to I\!R$ *two functions with* $\mathsf{E}[g(Z_1)^2] < \infty$, $\mathsf{E}[h(Z_1)^2] < \infty$. *Further assume that* $b : I\!R^d \times I\!R^d \to I\!R$ *is a bounded function with* $|b(\cdot,\cdot)| \le c$. *We abbreviate* $G_i = g(Z_i)$, $H_j = h(Z_j)$ *and* $B_{ij} := b(X_i, X_j)$.
a)

$$\sum_{1 \le i < j \le n} G_i B_{ij} H_j = O_p(n^2) \quad . \tag{68}$$

b) If $\mathsf{E}[G_1|X_1] = 0$ *then*

$$\sum_{1 \le i < j \le n} G_i B_{ij} H_j = O_p(n^{\frac{3}{2}}) \quad . \tag{69}$$

c) If $\mathsf{E}[G_1|X_1] = 0$ *and* $\mathsf{E}[H_1|X_1] = 0$ *then*

$$\sum_{1 \le i < j \le n} G_i B_{ij} H_j = O_p(n) \quad . \tag{70}$$

Proof. We show that in each of the three cases expectation and second moment of $\sum_{1 \le i < j \le n} G_i B_{ij} H_j$ divided by the claimed order are bounded; then the statements follow from Chebyshev's inequality.

With the Cauchy-Schwarz inequality and the boundedness of b we get

$$\begin{aligned}\left|\mathsf{E}\Big[\sum_{1 \le i < j \le n} G_i B_{ij} H_j\Big]\right| &= \left|\sum_{1 \le i < j \le n} \mathsf{E}_{X_i}\Big[\mathsf{E}[G_i|X_i]\,\mathsf{E}[B_{ij}H_j|X_i]\Big]\right| \\ &\le \binom{n}{2}\left|\mathsf{E}_{X_1}\Big[\mathsf{E}[G_1|X_1]\big(\mathsf{E}[B_{12}^2|X_1]\,\mathsf{E}[H_2^2|X_1]\big)^{-\frac{1}{2}}\Big]\right| \\ &\le c^2\binom{n}{2}\,\mathsf{E}_{X_1}\Big[|\,\mathsf{E}[G_1|X_1]|\big(\mathsf{E}[H_2^2|X_1]\big)^{-\frac{1}{2}}\Big] \, .\end{aligned} \tag{71}$$

Thus we have $\frac{1}{n^2}\,\mathsf{E}\left[\sum_{1 \le i < j \le n} G_i B_{ij} H_j\right] = O(1)$.
If $\mathsf{E}[G_1|X_1] = 0$ then (71) even gives $\mathsf{E}\left[\sum_{1 \le i < j \le n} G_i B_{ij} H_j\right] = 0$.

Now consider the second moments:

$$
\begin{aligned}
\mathsf{E}\Big[\Big(\sum_{1\le i<j\le n} G_i B_{ij} H_j\Big)^2\Big] &= \mathsf{E}\Big[\sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} G_i G_{i'} B_{ij} B_{i'j'} H_j H_{j'}\Big] \\
&= 4\sum_{1\le i<i'<j'<j\le n} \mathsf{E}\Big[G_i G_{i'} B_{ij} B_{i'j'} H_j H_{j'}\Big] \\
&\quad + 2\sum_{1\le i<j'<j\le n} \mathsf{E}\Big[G_i^2 B_{ij} B_{ij'} H_j H_{j'}\Big] \\
&\quad + 2\sum_{1\le i<i'<j\le n} \mathsf{E}\Big[G_i G_{i'} B_{ij} B_{ij'} H_j^2\Big] \\
&\quad + 2\sum_{1\le i<j<j'\le n} \mathsf{E}\Big[G_i G_j B_{ij} B_{jj'} H_j H_{j'}\Big] \\
&\quad + \sum_{1\le i<j\le n} \mathsf{E}\Big[G_i^2 B_{ij}^2 H_j^2\Big] \\
&= 4\sum_{1\le i<i'<j'<j\le n} \mathsf{E}_{X_i X_{i'}}\Big[\mathsf{E}[G_i|X_i]\,\mathsf{E}[G_{i'}|X_{i'}]\,\mathsf{E}[B_{ij}H_j|X_i]\,\mathsf{E}[B_{i'j'}H_{j'}|X_{i'}]\Big] \\
&\quad + 2\sum_{1\le i<j'<j\le n} \mathsf{E}_{X_i}\Big[\mathsf{E}[G_i^2|X_i]\,\mathsf{E}[B_{ij}H_j|X_i]\,\mathsf{E}[B_{ij'}H_{j'}|X_i]\Big] \\
&\quad + 2\sum_{1\le i<i'<j\le n} \mathsf{E}_{X_j}\Big[\mathsf{E}[G_i B_{ij}|X_j]\,\mathsf{E}[G_{i'}B_{i'j}|X_j]\,\mathsf{E}[H_j^2|X_j]\Big] \\
&\quad + 2\sum_{1\le i<j<j'\le n} \mathsf{E}_{X_j}\Big[\mathsf{E}[G_i B_{ij}|X_j]\,\mathsf{E}[B_{jj'}H_{j'}|X_j]\,\mathsf{E}[G_j H_j|X_j]\Big] \\
&\quad + \sum_{1\le i<j\le n} \mathsf{E}_{X_i X_j}\Big[\mathsf{E}[G_i^2|X_i] B_{ij}^2\,\mathsf{E}[H_j^2|X_j]\Big] \\
&= 4\tbinom{n}{4}\Big(\mathsf{E}_{X_1}\Big[\mathsf{E}[G_1|X_1]\,\mathsf{E}[B_{12}H_2|X_1]\Big]\Big)^2 \\
&\quad + 2\tbinom{n}{3}\,\mathsf{E}_{X_1}\Big[\mathsf{E}[G_1^2|X_1]\,(\mathsf{E}[B_{12}H_2|X_1])^2\Big] \\
&\quad + 2\tbinom{n}{3}\,\mathsf{E}_{X_2}\Big[(\mathsf{E}[G_1B_{12}|X_2])^2\,\mathsf{E}[H_2^2|X_2]\Big] \\
&\quad + 2\tbinom{n}{3}\,\mathsf{E}_{X_2}\Big[\mathsf{E}[G_1B_{12}|X_2]\,\mathsf{E}[B_{23}H_3|X_2]\,\mathsf{E}[G_2H_2|X_2]\Big] \\
&\quad + \tbinom{n}{2}\,\mathsf{E}_{X_1X_2}\Big[\mathsf{E}[G_1^2|X_1] B_{12}^2\,\mathsf{E}[H_2^2|X_2]\Big]
\end{aligned}
\tag{72}
$$

Applying the Cauchy-Schwarz inequality several times and using the boundedness of b we obtain from (72):

$$\begin{aligned}
\mathsf{E}\left[\left(n^{-2}\sum_{1\le i<j\le n} G_i B_{ij} H_j\right)^2\right] & \\
\le \tfrac{4}{n^4}\tbinom{n}{4}\,\mathsf{E}[G_1^2]c^2\,\mathsf{E}[H_2^2] & \\
+ \tfrac{2}{n^4}\tbinom{n}{3}\,\mathsf{E}[G_1^2]c^2\,\mathsf{E}[H_2^2] & \\
+ \tfrac{2}{n^4}\tbinom{n}{3}\,\mathsf{E}[G_1^2]c^2\,\mathsf{E}[H_2^2] & \\
+ \tfrac{2}{n^4}\tbinom{n}{3} c\left(\mathsf{E}[G_1^2]\right)^{\frac{1}{2}} c\left(\mathsf{E}[H_3^2]\right)^{\frac{1}{2}}\mathsf{E}[G_2H_2] & \\
+ \tfrac{1}{n^4}\tbinom{n}{2}\,\mathsf{E}[G_1^2]c^2\,\mathsf{E}[H_2^2] & \\
\le c^2\,\mathsf{E}[G_1^2]\,\mathsf{E}[H_2^2]\tfrac{n^3-4n+3}{6n^3} = O(1)\ . &
\end{aligned} \tag{73}$$

This proves part a).
If $\mathsf{E}[G_1|X_1]=0$ then also $\mathsf{E}[G_1B_{12}|X_2]=\mathsf{E}_{X_1}[\mathsf{E}[G_1|X_1]B_{12}|X_2]=0$. Hence from (72) we obtain as above

$$\begin{aligned}
\mathsf{E}\left[\left(n^{-\frac{3}{2}}\sum_{1\le i<j\le n} G_i B_{ij} H_j\right)^2\right] & \\
= \tfrac{2}{n^3}\tbinom{n}{3}\,\mathsf{E}_{X_1}\left[\mathsf{E}[G_1^2|X_1]\left(\mathsf{E}[B_{12}H_2|X_1]\right)^2\right] & \\
+ \tfrac{1}{n^3}\tbinom{n}{2}\,\mathsf{E}_{X_1X_2}\left[\mathsf{E}[G_1^2|X_1]B_{12}^2\,\mathsf{E}[H_2^2|X_2]\right] & \\
\le c^2\,\mathsf{E}[G_1^2]\,\mathsf{E}[H_2^2]\tfrac{2n^2-3n+1}{6n^2} = O(1)\ . &
\end{aligned} \tag{74}$$

This proves part b).
In part c) we have $\mathsf{E}[H_1|X_1]=0$, too. Then from (72) we conclude analogously

$$\begin{aligned}
\mathsf{E}\left[\left(n^{-1}\sum_{1\le i<j\le n} G_i B_{ij} H_j\right)^2\right] & \\
= \tfrac{1}{n^2}\tbinom{n}{2}\,\mathsf{E}_{X_1X_2}\left[\mathsf{E}[G_1^2|X_1]B_{12}^2\,\mathsf{E}[H_2^2|X_2]\right] & \\
\le c^2\,\mathsf{E}[G_1^2]\,\mathsf{E}[H_2^2]\tfrac{n-1}{2n} = O(1)\ . &
\end{aligned} \tag{75}$$

□

Lemma 5.2. *(Limit distributions of U-statistics)*
Let $Z_1,\dots,Z_n$ be independent real random l-vectors with common distribution D and let $s:\mathbb{R}^l\times\mathbb{R}^l\to\mathbb{R}$ be a symmetric function such that $\mathsf{E}[s(Z_1,Z_2)^2]<\infty$ and $\mathsf{E}[s(Z_1,Z_2)]=0$.

a) (nondegenerate case)
If $\mathsf{E}[s(Z_1, Z_2)|Z_2] \leq 0$*, then the U-statistic*

$$n^{-\frac{3}{2}} \sum_{1 \leq i < j \leq n} s(Z_i, Z_j) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

is asymptotically distributed normal with $\sigma^2 = \mathsf{Var}_Z \left[\mathsf{E}[s(Z, Z')|Z'] \right]$.
b) (degenerate case)
If $\mathsf{E}[s(Z_1, Z_2)|Z_2] = 0$*, then*

$$n^{-1} \sum_{1 \leq i < j \leq n} s(Z_i, Z_j) \rightsquigarrow \tfrac{1}{2} \sum_{k} \lambda_\nu (\chi^2_{1k} - 1)$$

where $\chi_{11}, \chi_{12}, \ldots$ *are independent standard normal random variables and* λ_k *are the eigenvalues of* $\mathcal{S} : \phi(\cdot) \mapsto \int_{\mathbb{R}^q} s(\cdot, z)\phi(z)\, \mathrm{d}\mathsf{D}$*, which is a linear operator on* $L^2(\mathsf{D})$.

Proof. For a proof of part a) see e.g. Hoeffding (1948) and for part b) e.g. Gregory (1977), or Serfling (1980, section 5.5) for both. □

Lemma 5.3. *Let* A*,* B *be two real random variables in a probability space with measure* P*, and let* A^**,* B^* *be two real random variables in a probability space with measure* P^**. Then for every* $\varepsilon > 0$ *and* $c \in \mathbb{R}$ *it holds*

$$\begin{aligned}
&\sup_{x \in \mathbb{R}} \left| \mathsf{P}[A + B \leq x] - \mathsf{P}^*[A^* + B^* \leq x] \right| \\
&\qquad \leq \sup_{x \in \mathbb{R}} \left| \mathsf{P}[A \leq x] - \mathsf{P}^*[A^* \leq x] \right| \\
&\qquad\quad + \sup_{x \in \mathbb{R}} \left(\mathsf{P}[A \leq x + \varepsilon] - \mathsf{P}^*[A^* \leq x - \varepsilon] \right) \\
&\qquad\quad + \mathsf{P}[|B - c| \geq \varepsilon] - \mathsf{P}^*[|B^* - c| \geq \varepsilon] \quad .
\end{aligned}$$

Proof. Fix $\varepsilon > 0$ and $c \in \mathbb{R}$ arbitrarily. From

$$\begin{aligned}
\mathsf{P}[A + c \leq x - \varepsilon] - \mathsf{P}[|B - c| \geq \varepsilon] &\leq \mathsf{P}[A + B \leq x] \\
&\leq \mathsf{P}[A + c \leq x + \varepsilon] + \mathsf{P}[|B - c| \geq \varepsilon]
\end{aligned}$$

and the analogous inequalities for P^* we deduce

$$\begin{aligned}
&\big|\mathsf{P}[A+B\le x]-\mathsf{P}^*[A^*+B^*\le x]\big|\\
&\qquad\le \max\big\{\mathsf{P}[A+c\le x+\varepsilon]-\mathsf{P}^*[A^*+c\le x-\varepsilon],\\
&\qquad\qquad \mathsf{P}^*[A^*+c\le x+\varepsilon]-\mathsf{P}[A+c\le x-\varepsilon]\big\}\\
&\qquad\quad +\mathsf{P}[|B-c|\ge\varepsilon]+\mathsf{P}^*[|B^*-c|\ge\varepsilon]\\
&\qquad= \max\big\{\mathsf{P}[A+c\le x-\varepsilon]-\mathsf{P}^*[A^*+c\le x-\varepsilon],\\
&\qquad\qquad \mathsf{P}^*[A^*+c\le x+\varepsilon]-\mathsf{P}[A+c\le x+\varepsilon]\big\}\\
&\qquad\quad +\mathsf{P}[A+c\le x+\varepsilon]-\mathsf{P}[A+c\le x-\varepsilon]\\
&\qquad\quad +\mathsf{P}[|B-c|\ge\varepsilon]+\mathsf{P}^*[|B^*-c|\ge\varepsilon]\\
&\qquad= \big|\mathsf{P}[A+c\le x-\varepsilon]-\mathsf{P}^*[A^*+c\le x-\varepsilon]\big|\\
&\qquad\quad +\mathsf{P}[A+c\le x+\varepsilon]-\mathsf{P}[A+c\le x-\varepsilon]\\
&\qquad\quad +\mathsf{P}[|B-c|\ge\varepsilon]+\mathsf{P}^*[|B^*-c|\ge\varepsilon]
\end{aligned}$$

Now taking the supremum over $x\in\mathbb{R}$ and cancelling superfluous constants, we obtain the statement of the lemma. □

Lemma 5.4. *Let $\{Z_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. real random d-vectors with common probability measure μ.*
a) Suppose $h_n:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ is a sequence of measurable functions that converge pointwise almost surely to h. Further there exists a dominating function b such that $|h_n|\le b$ and $\mathsf{E}_{Z_1Z_2}[|b(Z_1,Z_2)|]<\infty$. Then

$$\tfrac{1}{n^2}\sum_{i,j=1}^{n}h_n(Z_i,Z_j)\to\mathsf{E}_{Z_1Z_2}[h(Z_1,Z_2)]\quad\textit{almost surely}. \tag{76}$$

b) Let $\{\theta_n(Z_1,\dots,Z_n)\}_{n=1}^{\infty}$ be a random sequence in $\Theta\subseteq\mathbb{R}^p$, with $\theta_n\to\theta_0$ almost surely, θ_0 in the interior of Θ. Suppose $h:\mathbb{R}^d\times\mathbb{R}^d\times\Theta\to\mathbb{R}$ is continuously differentiable with respect to θ and $\mathsf{E}_{Z_1Z_2}[|h(Z_1,Z_2,\theta_0)|]<\infty$. Then $h_n=h(\cdot,\cdot,\theta_n)$ satisfies the conditions of part a) and we have

$$\tfrac{1}{n^2}\sum_{i,j=1}^{n}h(Z_i,Z_j,\theta_n)\to\mathsf{E}_{Z_1Z_2}[h(Z_1,Z_2,\theta_0)]\quad\textit{almost surely}. \tag{77}$$

Proof. These are variation and special case of a generalized bounded convergence theorem, see e.g. Royden (1988, page 270). Instead of describing the adaptions we give a complete proof here.

For almost every sequence $\{Z_i\}_{i=1}^{\infty}$ we have $\theta_n \to \theta_0$ and

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \phi(Z_i, Z_j) \to \mathsf{E}_{Z_1 Z_2}[\phi(Z_1, Z_2)] \,, \tag{78}$$

if ϕ is any bounded μ-measurable function. This SLLN for V-statistics can be found e.g. in Serfling (1980, chapter5); note that ϕ need not be symmetric. In the following we consider such a sequence $\{Z_i\}_{i=1}^{\infty}$.

a) Our first goal is to show

$$\mathsf{E}_{Z_1 Z_2}[g(Z_1, Z_2)] \le \liminf_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} g_n(Z_i, Z_j) \tag{79}$$

for any sequence of nonnegative measurable functions $\{g_n\}$ that converge pointwise to g. It suffices to prove

$$\mathsf{E}_{Z_1 Z_2}[\phi(Z_1, Z_2)] \le \liminf_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} g_n(Z_i, Z_j) \tag{80}$$

for every bounded measurable function $\phi \le g$.
We choose one such ϕ and $\varepsilon > 0$ and keep them fixed in the following. Define $\Phi_n := \{(z_1, z_2) : g_n(z_1, z_2) \ge \phi(z_1, z_2)\}$. Φ_n is an increasing sequence of sets whose union contains $\Phi := I\!R^d \times I\!R^d$. Hence $\Phi\neg\Phi_n$ is a decreasing sequence of measurable sets with empty intersection and there exists a m such that $\mu(\Phi\neg\Phi_m) < \varepsilon$. Since $\mu(\Phi\neg\Phi_m) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} I_{\Phi\neg\Phi_m}(Z_i, Z_j)$ - apply (78) to the indicator function $I_{\Phi\neg\Phi_m}$ - we may choose $n_0 \ge m$ so that $\frac{1}{n^2} \sum_{i,j=1}^{n} I_{\Phi\neg\Phi_m}(Z_i, Z_j) < \varepsilon$ for all $n \ge n_0$. Since $\Phi\neg\Phi_n \subseteq \Phi\neg\Phi_m$ for all $n \ge m$ we have $\frac{1}{n^2} \sum_{i,j=1}^{n} I_{\Phi\neg\Phi_n}(Z_i, Z_j) < \varepsilon$ for all $n \ge n_0$. Thus

$$\begin{aligned}
\frac{1}{n^2} \sum_{i,j=1}^{n} g_n(Z_i, Z_j) &\ge \frac{1}{n^2} \sum_{i,j=1}^{n} g_n(Z_i, Z_j) I_{\Phi_n}(Z_i, Z_j) \\
&\ge \frac{1}{n^2} \sum_{i,j=1}^{n} \phi(Z_i, Z_j) I_{\Phi_n}(Z_i, Z_j) \\
&= \frac{1}{n^2} \sum_{i,j=1}^{n} \phi(Z_i, Z_j) - \frac{1}{n^2} \sum_{i,j=1}^{n} \phi(Z_i, Z_j) I_{\Phi\neg\Phi_n}(Z_i, Z_j) \\
&\ge \frac{1}{n^2} \sum_{i,j=1}^{n} \phi(Z_i, Z_j) - \varepsilon \sup_{\Phi}(\phi) \,.
\end{aligned}$$

Since ε was arbitrary, we obtain (80) by taking $\liminf_{n\to\infty}$ on both sides and using (78). Finally we apply (79) to $b(Z_1, Z_2) + h_n(Z_1, Z_2)$ and $b(Z_1, Z_2) - h_n(Z_1, Z_2)$ to end the proof of part a).

b) Since $\theta_n \to \theta_0$ the continuity of h gives $h(z_1, z_2, \theta_n) \to h(z_1, z_2, \theta_0)$ for all $z_1, z_2 \in I\!R^{d+1}$. We now merely have to show that there exists a dominating function b for almost all n.
We choose $\delta > 0$ such that $\Psi := \{(z_1, z_2) : |\theta_n - \theta_0| < \delta\} \subset \Theta$ and find a n_1 such that $\theta_n \in \Psi$ for all $n \geq n_1$. Now

$$\begin{aligned}|h(z_1, z_2, \theta_n)| &= |h(z_1, z_2, \theta_0) + h'(z_1, z_2, \theta')(\theta_n - \theta_0)| \\ &\leq |h(z_1, z_2, \theta_0)| + \sup_\Psi |h'(z_1, z_2, \theta)|\delta \\ &=: b(z_1, z_2) \quad .\end{aligned}$$

Since h' is continuous with respect to θ, $\sup_\Psi |h'(z_1, z_2, \theta)|$ is bounded, and thus $\mathsf{E}_{Z_1 Z_2}[b(Z_1, Z_2)] < \infty$. □

6 Notation

The following general principles apply to the notations in this paper:

- Expectations are taken with respect to the indexed random variables. If no index is given, they are taken with respect to all free random variables in the argument.
- Capital roman letters always denote random variables.
- Deterministic functions and constants are denoted by lower case roman letters. If these functions occur with random variables $Z_1, Z_2, \ldots$ as arguments - and thus become random variables themselves - they may be abbreviated by the respective capital letters with indices according to the original random variables involved.
- Lower case greek letters may denote all kinds of objects. Roman letters stand for the same or same kind of object throughout the paper while greek letters may change their meaning.
- Throughout the paper $Z_i = (Y_i, X_i), Z = (Y, X) \sim \mathsf{D}$ and $z = (y, x)$ are assumed without further mentioning.

- For a vector $v = (v_1, \dots, v_p)^T$ or a matrix $m = (m_{\alpha\beta})$ indexed brackets refer to the indicated entry: $[v]_\alpha = [v^T]_\alpha = v_\alpha$, $[m]_{\alpha\beta} = m_{\alpha\beta}$.
- For a function $f(\theta)$ we denote the derivatives as follows: $f' := (f'_1, \dots, f'_p)$ and $f'' := (f''_{\alpha\beta})_{\alpha,\beta=1,\dots,p}$ with partial derivatives $f'_\alpha := [f'(x,\theta)]_\alpha = \frac{\partial}{\partial\theta_\alpha} f(x,\theta)$ and $f''_{\alpha\beta} := [f''(x,\theta)]_{\alpha\beta} = \frac{\partial^2}{\partial\theta_\alpha\partial\theta_\beta} f(x,\theta)$ where $\theta = (\theta_1, \dots, \theta_p)^T$, $\alpha, \beta = 1, \dots, p$.

References

Arcones, M. A. and Giné, E. (1992). On the bootstrap of U and V statistics. *The Annals of Statistics 20*(2), 655–674.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory of the bootstrap. *The Annals of Statistics 9*(6), 1196–1217.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica 58*(6), 1443–1458.

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica 65*(5), 1129–1152.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for u-statistics. *Journal of Multivariate Analysis 51*, 392–413.

Eubank, R. L., Hart, J. D., and LaRiccia, V. N. (1993). Testing goodnessof fit via nonparametric function estimation techniques. *Commun. Statst. Theory Meth. 22*, 3327–3354.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: nonparametric versus Bierens' test. Working paper, Department of Economics, University of Windsor.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *The Annals of Statistics 5*(1), 110–123.

Härdle, W. and Horowitz, J. L. (1994). Testing a parametric model against a semiparametric alternative. *Econometric theory 10*, 821–848.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics 21*(4), 1926–1947.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics 19*, 293–325.

Kozek, A. S. (1991). A nonparametric test of fit of a parametric model. *Journal of Multivariate Analysis 37*, 66–75.

Li, Q. (1995). A consistent test for linearity in partially linear regression models. Manuscript, Department of Economics, University of Guelph.

Mikosch, T. (1994). Almost sure convergence of bootstrapped means and u-statistics. *Journal of Statistical Planning and Inference 41*, 1–19.

Petrov, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford Studies in Probability 4. Oxford: Clarendon Press.

Rodrigues-Campos, M. C., Manteiga, W. G., and Cao, R. (1998). Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. *Journal of Statistical Planning and Inference 67*, 99–122.

Rotar, V. I. (1979). Limit theorems for polylinear forms. *Journal of Multivariate Analysis 9*, 511–530.

Royden, H. L. (1988). *Real Analysis* (Third Edition ed.). New York: Macmillan.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons.

Stute, W. and Manteiga, W. G. (1996). Nn goodness-of-fit tests for linear models. *Journal of Statistical Planning and Inference 53*, 75–92.

Werwatz, A. (1997). A consistent test for misspecification in polychotomous response models. Discussion paper, SFB 373, Humboldt Universit;at zu Berlin.

Wu, C.-F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics 14*, 1261–1350.