

A discusión

THE TWO-SAMPLE PROBLEM WITH REGRESSION ERRORS: AN EMPIRICAL PROCESS APPROACH*

Juan Mora and Natalie Neumeyer**

WP-AD 2005-18

Corresponding author: Juan Mora, Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Apartado de Correos 99, 03080 Alicante, Spain. E-mail: juan@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Mayo 2005
Depósito Legal: V-2407-2005

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

* We are grateful to M.A. Delgado and H. Dette for helpful comments. Financial support from Spanish DGI and Deutsche Forschungsgemeinschaft (SFB 475) is gratefully acknowledged.

** J. Mora: Universidad de Alicante; N. Neumeyer: Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.

THE TWO-SAMPLE PROBLEM WITH REGRESSION ERRORS: AN EMPIRICAL PROCESS APPROACH

Juan Mora and Natalie Neumeyer**

ABSTRACT

We describe how to test the null hypothesis that errors from two parametrically specified regression models have the same distribution versus a general alternative. First we obtain the asymptotic properties of test-statistics derived from the difference between the two residual-based empirical distribution functions. Under the null distribution they are not asymptotically distribution free and, hence, a consistent bootstrap procedure is proposed to compute critical values. As an alternative, we describe how to perform the test with statistics based on martingale-transformed empirical processes, which are asymptotically distribution free. Some Monte Carlo experiments are performed to compare the behaviour of all statistics with moderate sample sizes.

KEYWORDS: Two-Sample Problem; Residual-Based Empirical Process; Smooth Bootstrap; Martingale Transform.

1 Introduction

Consider two regression models, specified as

$$Y_{ji} = \mu_j(X_{ji}, \theta_j) + \sigma_j(X_{ji}, \theta_j)\varepsilon_{ji}, \quad j = 1, 2, \quad i = 1, \dots, n_j, \quad (1)$$

where $\{(Y_{ji}, X'_{ji})'\}_{i=1}^{n_j}$ are independent and identically distributed (i.i.d.) observations, $\mu_j : \mathbb{R}^{p_j} \times \mathbb{R}^{k_j} \rightarrow \mathbb{R}$, $\sigma_j : \mathbb{R}^{p_j} \times \mathbb{R}^{k_j} \rightarrow \mathbb{R}$ are known functions, $\theta_j \in \Theta_j \subset \mathbb{R}^{k_j}$ are unknown parameter vectors, and the errors $\{\varepsilon_{ji}\}_{i=1}^{n_j}$ are such that $E(\varepsilon_{ji} | X_{ji}) = E(\varepsilon_{ji}) = 0$, $E(\varepsilon_{ji}^2 | X_{ji}) = E(\varepsilon_{ji}^2) = 1$. Assuming independence between the two samples, the objective of this paper is to propose statistics to test

$$H_0 : F_1(\cdot) = F_2(\cdot) \quad \text{versus} \quad H_1 : F_1(\cdot) \neq F_2(\cdot),$$

where $F_j(\cdot)$ is the distribution function of ε_{ji} , which is assumed to be continuous, but unspecified.

If regression errors were observable, the problem that we consider here would be the classical two-sample problem. In fact, our test can be thought of as an extension of the two-sample problem. Suppose that the distribution functions of two observable variables Y_{1i} and Y_{2i} are compared using a classical nonparametric test, such as the Kolmogorov-Smirnov test. One of the drawbacks of nonparametric tests in this context is that when the null hypothesis is rejected the statistic gives no intuition about the cause of the rejection. To explore why the null has been rejected, it would be of interest to test whether the distribution functions of Y_{1i} and Y_{2i} differ only by a shift in location, with or without regressors; this test is not a specific case of our problem, but it can be treated in an entirely similar way with obvious changes. If the null hypothesis were rejected again, one might be interested in going one step further and testing whether the distribution functions of Y_{1i} and Y_{2i} are the same except for differences in mean and variance, which might depend on regressors -and this is precisely the problem that we consider. Thus, the testing procedures that we describe here can be used as a tool to explore

whether the reason why the null hypothesis is rejected in a two-sample problem is the presence of significant differences in the first or second order moments.

The testing problem that we study in this paper also arises directly in many contexts in applied work. In Economics, for example, the productivity of a firm is defined as the error from a regression model, and the researcher is often interested in comparing the distribution functions of productivity of firms from two different groups. In applied medical studies, the researcher is sometimes interested in comparing the distribution functions of certain standardized variables with data from healthy and unhealthy individuals. In many other areas it is often of interest to test whether two observable variables belong to the same location-scale family, which is also a specific case of the test that we study. In all these situations, the usual approach to test for the equality of the distribution functions is to test for the equality of just some moments (third, fourth and so on) or, with a parametric approach, to propose parametric models for the errors and then test whether the parameters estimated are equal. Instead, we propose to compare the entire distribution functions without assuming any parametric form for them.

The test statistics that we consider here are based on the comparison between estimates of the distribution functions $F_j(\cdot)$. If errors were observable, we could use the well-known Kolmogorov-Smirnov statistic $K_{n_1, n_2} := [(n_1 n_2 / (n_1 + n_2))]^{1/2} \sup_{x \in \mathbb{R}} |F_{1n_1}(x) - F_{2n_2}(x)|$, where $F_{jn_j}(\cdot)$ denotes the empirical distribution function based on $\{\varepsilon_{ji}\}_{i=1}^{n_j}$. Another popular alternative would be the Cramér-von Mises statistic $C_{n_1, n_2} := [(n_1 n_2 / (n_1 + n_2))^2] [\sum_{j=1}^2 \sum_{i=1}^{n_j} \{F_{1n_1}(\varepsilon_{ji}) - F_{2n_2}(\varepsilon_{ji})\}^2]$. If H_0 is true and the distribution function $F_j(\cdot)$ is continuous, these statistics are distribution-free and their asymptotic behavior is known; hence any of them could be used to perform a consistent test (for details, see e.g. Shorack and Wellner 1986, Section 9.9). In our context, we do not observe ε_{1i} and ε_{2i} , but we assume that

well-behaved estimates $\widehat{\theta}_1, \widehat{\theta}_2$ are available, and hence we can construct residuals

$$\widehat{\varepsilon}_{ji} = \{Y_{ji} - \mu_j(X_{ji}, \widehat{\theta}_j)\} / \sigma_j(X_{ji}, \widehat{\theta}_j), \quad j = 1, 2, \quad i = 1, \dots, n_j,$$

and the residual-based test statistics

$$\begin{aligned} \widehat{K}_{n_1, n_2} &:= \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{1/2} \sup_{x \in \mathbb{R}} \left| \widehat{F}_{1n_1}(x) - \widehat{F}_{2n_2}(x) \right|, \\ \widehat{C}_{n_1, n_2} &:= \frac{n_1 n_2}{(n_1 + n_2)^2} \sum_{j=1}^2 \sum_{i=1}^{n_j} \{ \widehat{F}_{1n_1}(\widehat{\varepsilon}_{ji}) - \widehat{F}_{2n_2}(\widehat{\varepsilon}_{ji}) \}^2, \end{aligned}$$

where $\widehat{F}_{jn_j}(\cdot)$ denotes the empirical distribution function based on $\{\widehat{\varepsilon}_{ji}\}_{i=1}^{n_j}$. These are the test statistics that we first study in this paper.

Many papers have studied the consequences of replacing errors by residuals in test statistics based on empirical distribution functions. In a one-sample context, Pierce and Kopecky (1979), Loynes (1980), Bai (1994) and Koul (1996), among others, have derived the asymptotic distribution of residual-based goodness-of-fit statistics. In a two-sample context, Koul and Sen (1985) consider a problem similar to ours, but they assume linearity for $\mu_j(\cdot, \cdot)$ and no scale estimation. In their framework, they prove that if the mean of the regressors is zero then the Kolmogorov-Smirnov statistic is asymptotically distribution-free; but this property does not hold if the mean of the regressors is not zero, or if a scale function is estimated. Koul (1996) considers the problem of testing whether the distribution functions of errors before and after a known change point are the same, and proved that the Kolmogorov-Smirnov statistic is then asymptotically distribution-free. But there is a crucial difference between our problem and the change point problem: in our context, it is natural to assume that the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are independent, whereas in the change point problem all $n_1 + n_2$ residuals are constructed with the same estimator. Bai and Ng (2001) propose a statistic to test for symmetry of regression errors which compares the empirical distribution functions of positive and negative residuals in the same fashion as we do. In their setup,

the two samples are not independent, which again makes their problem different from ours. In the related problem of testing for independence between errors from two independent regression models, Delgado and Mora (2000) prove that residual-based statistics have the same asymptotic distribution as those based on errors. However, this property does not hold here. As we prove below, in the two-sample problem that we consider the residual-based statistics \widehat{K}_{n_1, n_2} and \widehat{C}_{n_1, n_2} do not have the same asymptotic behavior as the ones based on errors K_{n_1, n_2} and C_{n_1, n_2} . Moreover, \widehat{K}_{n_1, n_2} and \widehat{C}_{n_1, n_2} are not distribution-free, even asymptotically; hence, it is not possible to derive asymptotic critical values valid for any situation. To overcome this problem, two different approaches can be followed: approximating critical values by bootstrap methods, or using statistics based on transformed empirical processes.

The usefulness of bootstrap methods in nonparametric distance tests was first highlighted by Romano (1988). Applications of bootstrap methods similar to the one we consider here have been proposed by Stute, González-Manteiga and Presedo-Quindimil (1998) and González-Manteiga and Delgado (2001), among many others. Recently, Neumeyer, Nagel and Dette (2005a, 2005b) have considered a symmetric wild bootstrap and a parametric bootstrap procedure in the context of goodness-of-fit tests for error distribution in linear models. In this paper we follow the same approach as in Koul and Lahiri (1994), and propose a bootstrap resampling scheme based on a nonparametric kernel estimate of the error distribution function. However, in contrast to Koul and Lahiri (1994), our regression models are not assumed to be linear, which gives rise to higher technical complexity in the proof of consistency.

Khmaladze (1981, 1993) proposed an alternative way to circumvent the problems that replacing errors by residuals causes when using statistics based on empirical processes. Under weak assumptions, he proved that certain martingale transforms of residual empirical processes converge weakly to Brownian motions

and, hence, test statistics based on them are asymptotically distribution-free, and approximate critical values can be obtained without bootstrap or simulation methods. These results have been fruitfully exploited in nonparametric specification tests, see e.g. Koul and Stute (1999), Bai and Ng (2001), Stute and Zhu (2002), Bai (2003) and Khmaladze and Koul (2004). In this paper we discuss how martingale-transformed processes can be used in our context to derive asymptotically distribution-free test statistics.

The rest of this paper is organized as follows. In Section 2 we derive the asymptotic properties of \widehat{K}_{n_1, n_2} or \widehat{C}_{n_1, n_2} , propose a bootstrap procedure to approximate their distribution and prove that this bootstrap procedure is consistent. In Section 3 we consider statistics based on martingale transforms of the residual-based empirical processes and derive their asymptotic properties. In Section 4 we report the results of two Monte Carlo experiments that illustrate the performance of the statistics with moderate sample sizes. Some concluding remarks are provided in Section 5. All proofs are relegated to Section 6.

2 Statistics based on residual empirical processes

The asymptotic behavior of \widehat{K}_{n_1, n_2} and \widehat{C}_{n_1, n_2} can be derived by studying the weak convergence of the residual empirical process on which they are based. To analyze this process, the following assumptions are required.

Assumption 1: Both distribution functions $F_j(\cdot)$ have density functions $f_j(\cdot)$

which are continuously differentiable and strictly positive. Additionally,

$\sup_{x \in \mathbb{R}} |x f_j(x)| < \infty$, $\sup_{x \in \mathbb{R}} |x^2 \dot{f}_j(x)| < \infty$, where $\dot{f}_j(\cdot)$ denotes the derivative of $f_j(\cdot)$, and $E\{(1 + \varepsilon_{ji}^2) \dot{f}_j(\varepsilon_{ji})^2 / f_j(\varepsilon_{ji})^2\} < \infty$.

Assumption 2: Both $\mu_j(\cdot, \cdot)$ and $\sigma_j(\cdot, \cdot)$ have continuous derivatives with respect to the second argument $\dot{\mu}_j(\cdot, \cdot)$ and $\dot{\sigma}_j(\cdot, \cdot)$, and all these functions are Lip-

schitz continuous with respect to the second argument, i.e., there exist a function $M_{1j}(\cdot)$ and a positive value α_{1j} such that $|\mu_j(x, u) - \mu_j(x, v)| \leq M_{1j}(x) \|u - v\|^{\alpha_{1j}}$, and $\sigma_j(\cdot, \cdot)$, $\dot{\mu}_j(\cdot, \cdot)$, $\dot{\sigma}_j(\cdot, \cdot)$ satisfy similar inequalities for certain functions $M_{2j}(\cdot)$, $M_{3j}(\cdot)$, $M_{4j}(\cdot)$ and positive values α_{2j} , α_{3j} , α_{4j} , respectively. Additionally, $\sigma_j(\cdot, u) > S_j > 0$ for all u in a neighborhood of θ_j , $E\{\|\dot{\mu}_j(X_{ji}, \theta_j)\|^2\} < \infty$, $E\{\|\dot{\sigma}_j(X_{ji}, \theta_j)\|^2\} < \infty$ and $E\{M_{lj}(X_{ji})^2\} < \infty$, for $l = 1, \dots, 4$.

Assumption 3: There exist functions $\psi_j(\cdot, \cdot, \cdot)$ such that the estimators $\widehat{\theta}_j$ satisfy that $n_j^{1/2}(\widehat{\theta}_j - \theta_j) = n_j^{-1/2} \sum_{i=1}^{n_j} \psi_j(X_{ji}, \varepsilon_{ji}, \theta_j) + o_p(1)$. Additionally, $E\{\psi_j(X_{ji}, \varepsilon_{ji}, \theta_j) \mid X_{ji}\} = 0$ and $E\{\|\psi_j(X_{ji}, \varepsilon_{ji}, \theta_j)\|^2\} < \infty$.

Assumption 1 is a technical condition for studying residual empirical processes using mean-value arguments. Observe that this assumption implies that both $f_j(\cdot)$ and $\dot{f}_j(\cdot)$ are bounded. The differentiability condition required in assumption 2 is relatively standard in nonlinear estimations, whereas the Lipschitz-continuity condition allows us to handle the supremum conditions which typically appear in the literature on residual empirical processes. Assumption 3 ensures that the estimators are root- n -consistent and allows us to derive the covariances of the limit process. The moment conditions introduced in assumptions 2 and 3 ensure that the expectations which appear below are finite.

To derive the asymptotic behavior of our test statistics, first we present a proposition that establishes an “oscillation-like” result between error-based empirical processes and residual-based ones in our context. For $t \in [0, 1]$, define

$$V_{jn_j}(t) := n_j^{-1/2} \sum_{i=1}^{n_j} [I\{\varepsilon_{ji} \leq F_j^{-1}(t)\} - t],$$

where $I\{\cdot\}$ is the indicator function, and define $\widehat{V}_{jn_j}(t)$ in the same way as $V_{jn_j}(t)$ but replacing ε_{ji} by $\widehat{\varepsilon}_{ji}$.

Proposition 1: *If (1) and assumptions 1-3 hold, then*

$$\sup_{t \in [0,1]} \left| \widehat{V}_{jn_j}(t) - \{V_{jn_j}(t) + g_j(t)' \xi_{jn_j}\} \right| = o_p(1), \quad (2)$$

where $g_j(t) := f_j\{F_j^{-1}(t)\}(1, F_j^{-1}(t))'$, $\xi_{jn_j} := n_j^{-1/2} \sum_{l=1}^{n_j} \omega'_j \psi_j(X_{jl}, \varepsilon_{jl}, \theta_j)$
and $\omega_j := (E\{\dot{\mu}_j(X_{ji}, \theta_j)/\sigma_j(X_{ji}, \theta_j)\}, E\{\dot{\sigma}_j(X_{ji}, \theta_j)/\sigma_j(X_{ji}, \theta_j)\})$.

Theorem 1: *If (1) and assumptions 1-3 hold, and $n_2 = \lambda n_1$ for a fixed λ , then:*

a) *Under H_0 ,*

$$\widehat{K}_{n_1, n_2} \xrightarrow{d} \sup_{t \in [0,1]} |D^{(1)}(t)| \quad \text{and} \quad \widehat{C}_{n_1, n_2} \xrightarrow{d} \int_0^1 \{D^{(1)}(t)\}^2 dt,$$

where $D^{(1)}(\cdot)$ is a Gaussian process on $[0, 1]$ with $E\{D^{(1)}(t)\} = 0$,

$$\begin{aligned} \text{Cov}\{D^{(1)}(s), D^{(1)}(t)\} &= \min(s, t) - st + \\ &\quad \{\lambda/(\lambda + 1)\} \Lambda_1(s, t, \theta_1) + \{1/(\lambda + 1)\} \Lambda_2(s, t, \theta_2), \end{aligned}$$

and $\Lambda_j(s, t, \theta_j) := g_j(s)' \omega'_j E[I\{\varepsilon_{ji} \leq F_j^{-1}(t)\} \psi_j(X_{ji}, \varepsilon_{ji}, \theta_j)] + g_j(t)' \omega'_j E[I\{\varepsilon_{ji} \leq F_j^{-1}(s)\} \psi_j(X_{ji}, \varepsilon_{ji}, \theta_j)] + g_j(s)' \omega'_j E\{\psi_j(X_{ji}, \varepsilon_{ji}, \theta_j) \psi_j(X_{ji}, \varepsilon_{ji}, \theta_j)'\} \omega_j g_j(t)$.

b) *Under H_1 , for all $x \in \mathbb{R}$,*

$$P(\widehat{K}_{n_1, n_2} > x) \rightarrow 1 \quad \text{and} \quad P(\widehat{C}_{n_1, n_2} > x) \rightarrow 1.$$

If the distributions of $\sup_{t \in [0,1]} |D^{(1)}(t)|$ (or $\int_0^1 \{D^{(1)}(t)\}^2 dt$) were known, according to this theorem \widehat{K}_{n_1, n_2} (or \widehat{C}_{n_1, n_2}) could be used as a statistic to perform a consistent test. Unfortunately, the covariance structure of $D^{(1)}(\cdot)$ depends, in general, on the unspecified distribution function $F_j(\cdot)$, the unknown parameters θ_j and other expectations. Hence, it is not possible to obtain asymptotic critical values valid for any situation. For this reason, we propose to approximate critical values with a resampling procedure.

In our context, at first sight one might think that a correct bootstrap p-value could be obtained by simply drawing i.i.d. bootstrap samples from the empirical

distribution function based on the centered residuals. However, this is not the case because, as Koul and Lahiri (1994) point out, the asymptotic distribution of the statistics depends crucially on the assumption of continuity, and hence the bootstrap samples must be generated from a continuous distribution. Accordingly, based on the sample $\mathcal{Y}_{n_1, n_2} = \{(Y_{ji}, X'_{ji}) \mid i = 1, \dots, n_j, j = 1, 2\}$, let $\tilde{f}_{n_1, n_2}(\cdot)$ be a kernel density estimate computed with the centered residuals, i.e.

$$\tilde{f}_{n_1, n_2}(x) = \frac{1}{(n_1 + n_2)h_{n_1, n_2}} \sum_{j=1}^2 \sum_{i=1}^{n_j} \varphi\left\{\frac{x - (\hat{\varepsilon}_{ji} - \bar{\varepsilon})}{h_{n_1, n_2}}\right\},$$

where $\varphi(\cdot)$ is a kernel density function, h_{n_1, n_2} is a smoothing value and $\bar{\varepsilon} := (\sum_{j=1}^2 \sum_{i=1}^{n_j} \hat{\varepsilon}_{ji}) / (n_1 + n_2)$. Let $\tilde{F}_{n_1, n_2}(\cdot)$ denote the corresponding distribution function. It is then possible to generate i.i.d. random variables $\{U_{ji}^*\}_{i=1}^{n_j}$ with uniform distribution on $[0, 1]$ and define bootstrap errors $\varepsilon_{ji}^* := \tilde{F}_{n_1, n_2}^{-1}(U_{ji}^*)$ for $i = 1, \dots, n_j$. Conditionally on the sample \mathcal{Y}_{n_1, n_2} the random variables $\{\varepsilon_{ji}^*\}_{i=1}^{n_j}$ are i.i.d. from a distribution with density $\tilde{f}_{n_1, n_2}(\cdot)$. Now define the bootstrap observations

$$Y_{ji}^* := \mu_j(X_{ji}, \hat{\theta}_j) + \sigma_j(X_{ji}, \hat{\theta}_j)\varepsilon_{ji}^*, \quad i = 1, \dots, n_j,$$

let $\hat{\theta}_j^*$ be the bootstrap estimate of θ_j computed with $\{(Y_{ji}^*, X'_{ji})'\}_{i=1}^{n_j}$ and consider

$$\hat{\varepsilon}_{ji}^* := \{Y_{ji}^* - \mu_j(X_{ji}, \hat{\theta}_j^*)\} / \sigma_j(X_{ji}, \hat{\theta}_j^*), \quad i = 1, \dots, n_j.$$

Finally, we can define the bootstrap test statistics

$$\begin{aligned} \hat{K}_{n_1, n_2}^* &:= \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} \sup_{x \in \mathbb{R}} \left| \hat{F}_{1n_1}^*(x) - \hat{F}_{2n_2}^*(x) \right|, \\ \hat{C}_{n_1, n_2}^* &:= \frac{n_1 n_2}{(n_1 + n_2)^2} \sum_{j=1}^2 \sum_{i=1}^{n_j} \{ \hat{F}_{1n_1}^*(\hat{\varepsilon}_{ji}^*) - \hat{F}_{2n_2}^*(\hat{\varepsilon}_{ji}^*) \}^2, \end{aligned}$$

where $\hat{F}_{jn_j}^*(\cdot)$ denotes the empirical distribution function based on $\{\hat{\varepsilon}_{ji}^*\}_{i=1}^{n_j}$. With these statistics, the bootstrap procedure that we propose to use to perform the test works as follows: based on the sample \mathcal{Y}_{n_1, n_2} , generate bootstrap data and compute the bootstrap statistic \hat{T}_{n_1, n_2}^* , (where $T = K$ or C); repeat this process

B times and then reject H_0 with significance level α if $\widehat{T}_{n_1, n_2} > T_\alpha^*$, where T_α^* is the $1 - \alpha$ sample quantile from $\{\widehat{T}_{n_1, n_2, b}^*\}_{b=1}^B$. To prove the consistency of this bootstrap procedure, the following additional assumptions are required.

Assumption 4: The kernel function $\varphi(\cdot)$ is a positive, symmetric and twice continuously differentiable probability density, such that $\int x\varphi(x)dx = 0$, $\int x^4\varphi(x)dx < \infty$ and $\sup_{x \in \mathbb{R}} \dot{\varphi}(x)^2/\varphi(x) < \infty$.

Assumption 5: The smoothing value is such that $h_{n_1, n_2} + (n_1 + n_2)^{-a}h_{n_1, n_2}^{-1} = o(1)$, for some $a \in (0, 1/4)$.

Assumption 6: The bootstrap estimators $\widehat{\theta}_j^*$ are such that, for every $\epsilon > 0$, $P\{\|n_j^{1/2}(\widehat{\theta}_j^* - \widehat{\theta}_j) - n_j^{-1/2} \sum_{i=1}^{n_j} \psi_j(X_{ji}, \varepsilon_{ji}^*, \widehat{\theta}_j)\| > \epsilon \mid \mathcal{Y}_{n_1, n_2}\} = o_p(1)$. Additionally, $E[\psi_j(X_{ji}, \varepsilon_{ji}^*, \widehat{\theta}_j) \mid \mathcal{Y}_{n_1, n_2}] = 0$, $\psi_j(\cdot, \cdot, \cdot)$ is continuous with respect to the third argument and $E\{\|\psi_j(X_{ji}, \varepsilon_{ji}, u)\|^2\}$ is uniformly bounded for u in a neighborhood of θ_j .

Assumptions 4 and 5 ensure uniform convergence of $\widetilde{f}_{n_1, n_2}(\cdot)$ in probability to the mixture density $f_M(\cdot) := \{\lambda/(\lambda + 1)\}f_1(\cdot) + \{1/(\lambda + 1)\}f_2(\cdot)$, whereas Assumption 6 guarantees that the bootstrap estimator behaves properly (see Koul and Lahiri, 1994). In analogy to the original processes, define the bootstrap processes on $[0, 1]$

$$V_{jn_j}^*(t) := n_j^{-1/2} \sum_{i=1}^{n_j} [I\{\varepsilon_{ji}^* \leq \widetilde{F}_{n_1, n_2}^{-1}(t)\} - t]$$

and $\widehat{V}_{jn_j}^*(t)$ in the same way as $V_{jn_j}^*(t)$ but replacing ε_{ji}^* by $\widehat{\varepsilon}_{ji}^*$. Before proving the consistency of the bootstrap procedure in our problem, we prove two properties about the relationship between bootstrap empirical processes and residual-based ones. Hereafter, $F_{X_j}(\cdot)$ denotes the distribution function of X_{ji} .

Proposition 2: *If (1) and assumptions 1-6 are satisfied then, for all $\epsilon > 0$,*

$$P\left(\sup_{t \in [0, 1]} \left| \widehat{V}_{jn_j}^*(t) - \{V_{jn_j}^*(t) + \widetilde{V}_{jn_j}^*(t)\} \right| > \epsilon \mid \mathcal{Y}_{n_1, n_2}\right) = o_p(1),$$

where $\tilde{V}_{jn_j}^*(t) := n_j^{-1/2} \sum_{i=1}^{n_j} (\tilde{F}_{n_1, n_2}^{-1} \{ [\tilde{F}_{n_1, n_2}^{-1}(t) \sigma_j(X_{ji}, \hat{\theta}_j^*) + \mu_j(X_{ji}, \hat{\theta}_j^*) - \mu_j(X_{ji}, \hat{\theta}_j)] / \sigma_j(X_{ji}, \hat{\theta}_j) \} - t)$.

Proposition 3: *If (1) and assumptions 1-6 are satisfied then, for all $\epsilon > 0$,*

$$P(\sup_{t \in [0,1]} |\hat{V}_{jn_j}^*(t) - \{V_{jn_j}^*(t) + \tilde{g}_j(t) \hat{\xi}_{jn_j}^*\}| > \epsilon \mid \mathcal{Y}_{n_1, n_2}) = o_p(1),$$

where $\tilde{g}_j(t) := \tilde{f}_{n_1, n_2} \{ \tilde{F}_{n_1, n_2}^{-1}(t) \} (1, \tilde{F}_{n_1, n_2}^{-1}(t))'$, $\hat{\xi}_{jn_j}^* := n_j^{-1/2} \sum_{l=1}^{n_j} \omega_j^* \psi_j(X_{jl}, \varepsilon_{jl}^*, \hat{\theta}_j)$ and $\omega_j^* := (\int \dot{\mu}_j(x, \hat{\theta}_j) / \sigma_j(x, \hat{\theta}_j) dF_{X_j}(x), \int \dot{\sigma}_j(x, \hat{\theta}_j) / \sigma_j(x, \hat{\theta}_j) dF_{X_j}(x))$.

Theorem 2: *If (1) and assumptions 1-6 hold, and $n_2 = \lambda n_1$ for a fixed λ then, for all $x \in \mathbb{R}$,*

$$\begin{aligned} P(\hat{K}_{n_1 n_2}^* \leq x \mid \mathcal{Y}_{n_1, n_2}) &\longrightarrow P(\sup_{t \in [0,1]} |D^{(2)}(t)| \leq x) \text{ and} \\ P(\hat{C}_{n_1, n_2}^* \leq x \mid \mathcal{Y}_{n_1, n_2}) &\longrightarrow P(\int_0^1 \{D^{(2)}(t)\}^2 dt \leq x), \end{aligned}$$

in probability, where $D^{(2)}(\cdot)$ is a Gaussian process on $[0, 1]$ with $E\{D^{(2)}(t)\} = 0$ and with the same covariances as $D^{(1)}(\cdot)$, but replacing $F_j(\cdot)$ by the mixture distribution function $F_M(\cdot) := \{1/(1+\lambda)\}F_1(\cdot) + \{\lambda/(1+\lambda)\}F_2(\cdot)$, and ε_{ji} by a random variable ε_i with distribution function $F_M(\cdot)$.

Observe that, under H_0 , $D^{(2)}(\cdot)$ has the same distribution as $D^{(1)}(\cdot)$; thus, the bootstrap critical values correctly approximate the asymptotic ones. Under H_1 , the bootstrap critical values converge to a fixed value; hence, the test performed with the bootstrap critical values is consistent.

3 Statistics based on martingale transforms of residual empirical processes

As Khmaladze (1981) points out in a seminal paper, the theoretical problems which stem from the replacement of errors by residuals in goodness-of-fit tests can

be circumvented using martingale transformation methods. Specifically, Khmaladze (1981) considers the problem of testing the null hypothesis “the distribution function of the error terms $\{\varepsilon_{ji}\}_{i=1}^{n_j}$ is $F_0(\cdot)$ ”, where $F_0(\cdot)$ is a known distribution function. He proves that if the standard residual-based empirical process has an asymptotic representation such as (2), then it is possible, by means of a martingale transformation, to derive a residual-based process that converges weakly to a standard Brownian motion on $[0, 1]$. Hence, goodness-of-fit statistics based on the martingale-transformed process prove to be asymptotically distribution-free. Therefore, they are a very appealing alternative to test statistics based on the standard residual-based empirical processes.

Let us see how these results apply in our context. Observe that our null hypothesis “ $H_0 : F_1(\cdot) = F_2(\cdot)$ ” is true if and only if the hypothesis “ $H_0^* : \text{the distribution function of the error terms } \{\varepsilon_{1i}\}_{i=1}^{n_1} \text{ is } F_2(\cdot)$ ” is true, and this property also holds if the role of the samples is interchanged in H_0^* . Thus, our test is equivalent to either of these two goodness-of-fit tests. If the distribution functions $F_j(\cdot)$ were known, we could then derive the martingale-transformed processes for these goodness-of-fit tests and then test our null hypothesis with any of them; but in our context $F_1(\cdot)$ and $F_2(\cdot)$ are not known. However, as Bai (2003) points out, under very mild conditions the replacement of unknown quantities by suitable estimators in martingale-transformed processes does not affect the limit distributions. This is the approach that we follow here.

As before, we assume that we can obtain well-behaved residuals $\widehat{\varepsilon}_{ji}$ and the residual-based empirical distribution functions $\widehat{F}_{jn_j}(\cdot)$. The martingale-transformed process that should be used to test whether the distribution function of the error terms $\{\varepsilon_{ji}\}_{i=1}^{n_1}$ is $F_{3-j}(\cdot)$, if $F_{3-j}(\cdot)$ is known, is defined for $x \in \mathbb{R}$ as

$$\widehat{W}_{jn_j}(x) := n_j^{1/2} [\widehat{F}_{jn_j}(x) - \int_{-\infty}^x q_{3-j}(y)' C_{3-j}(y)^{-1} \{ \int_y^{\infty} q_{3-j}(z) d\widehat{F}_{jn_j}(z) \} f_{3-j}(y) dy],$$

where $q_j(y) := (1, \dot{f}_j(y)/f_j(y), 1+y\dot{f}_j(y)/f_j(y))'$ and $C_j(y) := \int_y^\infty q_j(w)q_j(w)'f_j(w)dw$. Therefore, to derive a feasible martingale-transformed process we require estimates of $f_j(y)$ and $\dot{f}_j(y)/f_j(y)$. We propose to use kernel estimators for $f_j(y)$ and $\dot{f}_j(y)$ but, for technical reasons, trimmed kernel estimators for $\dot{f}_j(y)/f_j(y)$. Thus, we define $\widehat{f}_j(y) := \sum_{i=1}^{n_j} \varphi\{(y - \widehat{\varepsilon}_{ji})/h_{jn_j}\}/(n_j h_{jn_j})$, $\widehat{\dot{f}}_j(y) := \sum_{i=1}^{n_j} \dot{\varphi}\{(y - \widehat{\varepsilon}_{ji})/h_{jn_j}\}/(n_j h_{jn_j}^2)$, where, as before, $\varphi(\cdot)$ is a kernel density function and h_{jn_j} are smoothing values,

$$\widehat{G}_j(y) := \begin{cases} \widehat{\dot{f}}_j(y)/\widehat{f}_j(y) & \text{if } |y| \leq a_{jn_j}, \widehat{f}_j(y) \geq b_{jn_j}, \left| \widehat{\dot{f}}_j(y) \right| \leq c_{jn_j} \widehat{f}_j(y) \\ 0 & \text{otherwise,} \end{cases}$$

where a_{jn_j} , b_{jn_j} and c_{jn_j} are trimming values, $\widehat{q}_j(y) := (1, \widehat{G}_j(y), 1 + y\widehat{G}_j(y))'$ and $\widehat{C}_j(y) := \int_y^\infty \widehat{q}_j(w)\widehat{q}_j(w)'\widehat{f}_j(w)dw$. With these estimates, we can construct the estimated martingale-transformed process $\widetilde{\widehat{W}}_{jn_j}(\cdot)$, which is defined in the same way as $\widehat{\widehat{W}}_{jn_j}(\cdot)$, but replacing $q_{3-j}(\cdot)$, $C_{3-j}(\cdot)$ and $f_{3-j}(\cdot)$ by $\widehat{q}_{3-j}(\cdot)$, $\widehat{C}_{3-j}(\cdot)$ and $\widehat{f}_{3-j}(\cdot)$. Using these processes we can obtain two Kolmogorov-Smirnov statistics and two Cramér-von Mises ones. To define these statistics the supremum (in the Kolmogorov-Smirnov case) and the integral (in the Cramér-von Mises case) are not taken with respect to \mathbb{R} , because the asymptotic equivalence between the original martingale-transformed process $\widehat{\widehat{W}}_{jn_j}(\cdot)$ and the estimated martingale-transformed process $\widetilde{\widehat{W}}_{jn_j}(\cdot)$ is only proved at intervals $(-\infty, x_0]$, with $x_0 \in \mathbb{R}$ (see Theorem 4 in Bai, 2003). Therefore, we consider the Kolmogorov-Smirnov martingale-transformed statistics

$$\overline{K}_{n_1, n_2, x_0}^{(j)} := \widehat{F}_{jn_j}(x_0)^{-1/2} \sup_{x \in (-\infty, x_0]} \left| \widetilde{\widehat{W}}_{jn_j}(x) \right|,$$

and the Cramér-von Mises ones

$$\overline{C}_{n_1, n_2, x_0}^{(j)} := \widehat{F}_{jn_j}(x_0)^{-2} n_j^{-1} \sum_{i=1}^{n_j} I(\widehat{\varepsilon}_{ji} \leq x_0) \widetilde{\widehat{W}}_{jn_j}(\widehat{\varepsilon}_{ji})^2,$$

where x_0 is any fixed real number. The factor $\widehat{F}_{jn_j}(x_0)$ is introduced in these statistics in order to obtain an asymptotic distribution which does not depend on

x_0 . To derive the asymptotic properties of these statistics, the following additional assumptions are required.

Assumption 7: The derivatives of the density functions $\dot{f}_j(\cdot)$ are Lipschitz continuous of order $d_j > 0$, and $C_j(y)$ are non-singular matrices for every $y \in [-\infty, +\infty)$.

Assumption 8: The kernel function $\varphi(\cdot)$ and its derivative $\dot{\varphi}(\cdot)$ have bounded total variation.

Assumption 9: The smoothing and trimming values satisfy that $h_{jn_j}^2 = o(b_{jn_j})$, $a_{jn_j}^{-1} = o(1)$, $b_{jn_j} = o(1)$, $c_{jn_j}^{-1} = o(1)$, $h_{jn_j}c_{jn_j} = o(1)$, $n_j^{-1}h_{jn_j}^{-3}a_{jn_j}^3 = o(1)$, $\log(h_{jn_j}^{-1})/(n_jh_{jn_j}) = o(b_{jn_j}^2)$, $a_{jn_j} = o(n_j^{1/2}h_{jn_j}^2b_{jn_j})$, $h_{jn_j} \log n_j = o(1)$ and $a_{jn_j}h_{jn_j}^{2d_j} \log n_j = o(b_{jn_j})$.

The assumption about matrices $C_j(y)$ ensures that the martingale transformation can be performed. Assumptions 8 and 9 ensure that the replacement of the density functions and their derivatives by nonparametric estimates does not affect the limit distributions. Note that assumption 9 allows us to choose the optimal smoothing value $h_{jn_j} = M_j n_j^{-1/5}$ for a fixed M_j , whereas there is plenty of freedom for choosing the rates of convergence of the trimming values. Before deriving the asymptotic properties of the statistics, we derive two properties of the nonparametric estimators that are required later.

Proposition 4: *If (1) and assumptions 1-4, 7-9 hold, $\alpha_{1j} \geq 1$ and $\alpha_{2j} \geq 1$, then*

$$\int_{-\infty}^{\infty} \|\hat{q}_j(y) - q_j(y)\|^2 f_j(y) dy = o_p(1). \quad (3)$$

Proposition 5: *If (1) and assumptions 1-4, 7-9 hold, $\alpha_{1j} \geq 1$ and $\alpha_{2j} \geq 1$, then*

$$\sup_{x \in \mathbb{R}} \left\| n_j^{-1/2} \sum_{i=1}^{n_j} [I(\varepsilon_{ji} \geq x) \{\hat{q}_j(\varepsilon_{ji}) - q_j(\varepsilon_{ji})\} - \int_x^{\infty} \{\hat{q}_j(y) - q_j(y)\} f_j(y) dy] \right\| = o_p(1) \quad (4)$$

Theorem 3: *If (1) and assumptions 1-4, 7-9 hold, $\alpha_{1j} \geq 1$, $\alpha_{2j} \geq 1$ and $n_2 = \lambda n_1$ for a fixed λ , then:*

a) Under H_0 , if $F_j(x_0)$ is in $(0, 1)$,

$$\overline{K}_{n_1, n_2, x_0}^{(j)} \xrightarrow{d} \sup_{t \in [0, 1]} |W(t)| \quad \text{and} \quad \overline{C}_{n_1, n_2, x_0}^{(j)} \xrightarrow{d} \int_0^1 \{W(t)\}^2 dt,$$

where $W(\cdot)$ is a standard Brownian motion on $[0, 1]$.

b) Under H_1 , if $E(\varepsilon_{ji}^3) < \infty$, there exists $x_ \in \mathbb{R}$ such that if $x_0 \geq x_*$ then for all $x \in \mathbb{R}$,*

$$P\left(\overline{K}_{n_1, n_2, x_0}^{(j)} > x\right) \rightarrow 1 \quad \text{and} \quad P\left(\overline{C}_{n_1, n_2, x_0}^{(j)} > x\right) \rightarrow 1.$$

Theorem 3 suggests that one can use either process $\widetilde{W}_{1n_1}(\cdot)$ or $\widetilde{W}_{2n_2}(\cdot)$ to obtain asymptotically distribution-free statistics which are consistent against any alternative, as long as a large enough x_0 is selected. However, the behavior of test statistics based on $\widetilde{W}_{1n_1}(\cdot)$ and $\widetilde{W}_{2n_2}(\cdot)$ may not be similar because their power functions may be very different, as is shown by the simulation results that we report below. For this reason, the test should be performed combining statistics based on both processes; in this case, the following corollary applies.

Corollary: *Let $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a continuous function. If the assumptions in Theorem 3 hold, then:*

a) Under H_0 , if $F_j(x_0)$ is in $(0, 1)$,

$$G(\overline{K}_{n_1, n_2, x_0}^{(1)}, \overline{K}_{n_1, n_2, x_0}^{(2)}) \xrightarrow{d} G(\varsigma_1, \varsigma_2) \quad \text{and} \\ G(\overline{C}_{n_1, n_2, x_0}^{(1)}, \overline{C}_{n_1, n_2, x_0}^{(2)}) \xrightarrow{d} G(\zeta_1, \zeta_2),$$

where ς_1, ς_2 are independent random variables both with the same distribution as $\sup_{t \in [0, 1]} |W(t)|$, ζ_1, ζ_2 are independent random variables both with the same distribution as $\int_0^1 \{W(t)\}^2 dt$, and $W(\cdot)$ is a standard Brownian motion on $[0, 1]$.

b) Under H_1 , if $E(\varepsilon_{ji}^3) < \infty$ and $\lim_{\min(y,z) \rightarrow \infty} G(y, z) = \infty$, then there exists $x_* \in \mathbb{R}$ such that, if $x_0 \geq x_*$, for all $x \in \mathbb{R}$

$$P\{G(\overline{K}_{n_1, n_2, x_0}^{(1)}, \overline{K}_{n_1, n_2, x_0}^{(2)}) > x\} \rightarrow 1 \text{ and } P\{G(\overline{C}_{n_1, n_2, x_0}^{(1)}, \overline{C}_{n_1, n_2, x_0}^{(2)}) > x\} \rightarrow 1.$$

In the simulations that we report below, we choose the maximum as function $G(\cdot, \cdot)$, i.e. we consider

$$\begin{aligned} \overline{K}_{n_1, n_2, x_0} & : = \max\{\overline{K}_{n_1, n_2, x_0}^{(1)}, \overline{K}_{n_1, n_2, x_0}^{(2)}\}, \\ \overline{C}_{n_1, n_2, x_0} & : = \max\{\overline{C}_{n_1, n_2, x_0}^{(1)}, \overline{C}_{n_1, n_2, x_0}^{(2)}\}. \end{aligned}$$

Other sensible choices would be $G(x_1, x_2) = x_1^2 + x_2^2$ or $G(x_1, x_2) = |x_1| + |x_2|$. The crucial point is that the asymptotic null distributions do not depend on any characteristic of the data, as long as the assumptions are met.

Asymptotic critical values for the Kolmogorov-Smirnov statistics can be derived taking into account that the distribution function of $\sup_{t \in [0,1]} |W(t)|$ is $(4/\pi) \sum_{j=0}^{\infty} (-1)^j \exp\{-(2j+1)^2 \pi^2 / (8x^2)\} / (2j+1)$, see e.g. Shorack and Wellner (1986, p. 34). From here it follows that the asymptotic critical values for $\overline{K}_{n_1, n_2, x_0}^{(1)}$ or $\overline{K}_{n_1, n_2, x_0}^{(2)}$ at the 10%, 5% and 1% significance levels are 1.960, 2.24 and 2.807; and the asymptotic critical values for $\overline{K}_{n_1, n_2, x_0}$ at those levels are 2.231, 2.493 and 3.023. Asymptotic critical values for the Cramér-von Mises statistics can be derived taking into account that the distribution function of $\int_0^1 \{W(t)\}^2 dt$ is $2^{3/2} \sum_{j=0}^{\infty} (-1)^j [1 - \Phi\{(4j+1)/(2x^{1/2})\}] \exp[\ln\{(2j)!\} - 2 \ln(j!)] / 4^j$, see e.g. Rothman and Woodroffe (1972). From here it follows that the asymptotic critical values for $\overline{C}_{n_1, n_2, x_0}^{(1)}$ or $\overline{C}_{n_1, n_2, x_0}^{(2)}$ at the 10%, 5% and 1% significance levels are 1.196, 1.656 and 2.787; and the asymptotic critical values for $\overline{C}_{n_1, n_2, x_0}$ at those levels are 1.638, 2.126 and 3.290.

4 Simulations

In order to check the behavior of the statistics, we perform two Monte Carlo experiments. In both cases we test the null hypothesis that the zero-mean unit-variance errors ε_{1i} and ε_{2i} have the same distribution function at the 5% significance level.

In Experiment 1 we test whether two samples come from the same location-scale model. Specifically, we generate i.i.d. observations $\{Y_{1i}\}_{i=1}^{n_1}$, each of them defined as $Y_{1i} = \mu_1 + \sigma_1\varepsilon_{1i}$, where $\mu_1 = 1$, $\sigma_1 = 1$, $\varepsilon_{1i} = \{V_{1i} - E(V_{1i})\}/\text{Var}(V_{1i})^{1/2}$, and V_{1i} is generated from an extreme-value distribution with density function $f_{V_1}(x) = \exp\{x - \exp(x)\}$; and we generate i.i.d. observations $\{Y_{2i}\}_{i=1}^{n_2}$, each of them defined as $Y_{2i} = \mu_2 + \sigma_2\varepsilon_{2i}$, where $\mu_2 = 2$, $\sigma_2 = 2$, $\varepsilon_{2i} = \{V_{2i} - E(V_{2i})\}/\text{Var}(V_{2i})^{1/2}$, and V_{2i} is generated from a log-gamma distribution with density function $f_{V_2}(x) = \exp\{(1 + \delta)x - \exp(x)\}/\Gamma(1 + \delta)$. The value of δ varies from one simulation to another; we consider $\delta = 0, 1, 2, 3, 4, 5$. Observe that H_0 holds if and only if $\delta = 0$; as δ grows, the distribution of ε_{2i} becomes closer to the standard normal. The null hypothesis here amounts to saying that the distribution functions of Y_{1i} and Y_{2i} are the same except for changes in location and scale. To compute the statistics, μ_j and σ_j are estimated by the sample mean and variance of $\{Y_{ji}\}_{i=1}^{n_j}$.

In Experiment 2 we compare a normal distribution and a Student's t distribution in a multiple regression with homoskedastic errors. Specifically, we generate i.i.d. observations $\{(Y_{1i}, X'_{1i})'\}_{i=1}^{n_1}$, where $Y_{1i} = \beta_{11} + \beta_{12}X_{11i} + \beta_{13}X_{12i} + \beta_{14}(X_{11i}^2 - 1)(X_{12i}^2 - 1) + \sigma_1\varepsilon_{1i}$, X_{1i} , X_{2i} and ε_{1i} are all independent with standard normal distribution, $\beta_{11} = 0$, $\beta_{12} = \beta_{13} = 1$, $\beta_{14} = 0.2$, and $\sigma_1 = 1$; and we generate i.i.d. $\{(Y_{2i}, X'_{2i})'\}_{i=1}^{n_2}$ with the same characteristics as the first sample, except that $\varepsilon_{2i} = V_{2i}/\text{Var}(V_{2i})^{1/2}$, and the distribution of V_{2i} is Student's t with δ^{-1} degrees of freedom. The values of δ that we consider are $\delta = 0, 1/9, 1/7, 1/5, 1/4, 1/3$ (if $\delta = 0$, ε_{2i} is generated from a standard normal distribution). Again, H_0 is true

if and only if $\delta = 0$. To compute the statistics, residuals are based on the least squares estimates obtained within each sample.

In all nonparametric estimations, the standard normal density is used as the kernel function. In order to examine the effect of undersmoothing or oversmoothing, we use three different smoothing values. To compute \widehat{K}_{n_1, n_2} and \widehat{C}_{n_1, n_2} , we consider $h_{n_1, n_2}^{(l)} = M_l(n_1 + n_2)^{-1/5}$, for $l = 1, 2, 3$ and, for each sample size and experiment, M_2 is chosen after graphical inspection of some preliminary estimates, $M_1 = 2M_2/3$ and $M_3 = 4M_2/3$. To compute $\overline{K}_{n_1, n_2, x_0}^{(j)}$ and $\overline{C}_{n_1, n_2, x_0}^{(j)}$, we consider $h_{jn_j}^{(l)} = M_l n_j^{-1/5}$, for $l = 1, 2, 3$ and M_1, M_2 and M_3 are selected as above. When using \widehat{K}_{n_1, n_2} and \widehat{C}_{n_1, n_2} , the critical values are computed with $B = 500$ bootstrap replications. When using $\overline{K}_{n_1, n_2, x_0}^{(j)}$ and $\overline{C}_{n_1, n_2, x_0}^{(j)}$, x_0 is always chosen as the 95% quantile of the residuals from the j -th sample. All the integrals that have to be computed to obtain martingale-based statistics are approximated as follows: for a given function $H : \mathbb{R} \rightarrow \mathbb{R}$,

$$\int_A H(x) dx \approx \sum_{l=1}^m (y_l - y_{l-1}) H\left(\frac{y_l + y_{l-1}}{2}\right) I\left(\frac{y_l + y_{l-1}}{2} \in A\right),$$

where $y_l := -8 + \Delta l$, $\Delta := 0.0025$ and $m = 6400$; we have checked that this approximation yields very accurate results in all cases. For the sake of simplicity, no trimming values are used when computing $\widehat{G}_j(\cdot)$.

In Tables 1 and 2 we report the proportion of rejections of H_0 in 1000 simulation runs using the Cramér-von Mises statistics. For the sake of brevity, we do not include the results for Kolmogorov-Smirnov statistics, which are quite similar. However, we do include the results for the infeasible Cramér-von Mises statistic $\overline{C}_{n_1, n_2, x_0}^{(IN)}$ that is obtained when the martingale transformation is performed with the true density functions of errors, i.e., $\overline{C}_{n_1, n_2, x_0}^{(IN)} := \max\{\overline{C}_{n_1, n_2, x_0}^{(IN, 1)}, \overline{C}_{n_1, n_2, x_0}^{(IN, 2)}\}$, where $\overline{C}_{n_1, n_2, x_0}^{(IN, j)} := \widehat{F}_{jn_j}(x_0)^{-2} n_j^{-1} \sum_{i=1}^{n_j} I(\widehat{\varepsilon}_{ji} \leq x_0) \widehat{W}_{jn_j}(\widehat{\varepsilon}_{ji})^2$.

TABLE 1

Proportion of Rejections of H_0 in Experiment 1, $\alpha = 0.05$

δ	\widehat{C}_{n_1, n_2}			$\overline{C}_{n_1, n_2, x_0}$			$\overline{C}_{n_1, n_2, x_0}^{(IN)}$
	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	
$n_1 = n_2 = 50$							
0	0.048	0.057	0.059	0.012	0.132	0.476	0.128
1	0.073	0.080	0.087	0.011	0.138	0.538	0.217
2	0.077	0.088	0.093	0.013	0.152	0.684	0.295
3	0.082	0.088	0.094	0.017	0.184	0.792	0.365
4	0.085	0.097	0.103	0.038	0.281	0.992	0.411
5	0.107	0.115	0.118	0.047	0.356	0.998	0.430
$n_1 = n_2 = 150$							
0	0.047	0.054	0.059	0.021	0.103	0.385	0.101
1	0.117	0.121	0.127	0.024	0.106	0.437	0.353
2	0.183	0.185	0.194	0.023	0.136	0.823	0.566
3	0.219	0.226	0.233	0.029	0.148	0.938	0.663
4	0.255	0.260	0.268	0.074	0.268	0.991	0.724
5	0.323	0.331	0.341	0.121	0.523	1.000	0.755
$n_1 = n_2 = 250$							
0	0.052	0.053	0.056	0.027	0.086	0.243	0.089
1	0.136	0.138	0.138	0.032	0.361	0.644	0.473
2	0.259	0.262	0.268	0.038	0.568	0.975	0.706
3	0.336	0.341	0.346	0.082	0.753	0.994	0.834
4	0.423	0.429	0.434	0.117	0.847	1.000	0.895
5	0.463	0.466	0.470	0.179	0.885	1.000	0.929

TABLE 2

Proportion of Rejections of H_0 in Experiment 2, $\alpha = 0.05$

δ	\hat{C}_{n_1, n_2}			\bar{C}_{n_1, n_2, x_0}			$\bar{C}_{n_1, n_2, x_0}^{(IN)}$
	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	
$n_1 = n_2 = 50$							
0	0.051	0.049	0.056	0.105	0.035	0.016	0.034
1/9	0.058	0.061	0.059	0.168	0.081	0.069	0.099
1/7	0.081	0.095	0.090	0.163	0.098	0.074	0.112
1/5	0.092	0.115	0.122	0.221	0.152	0.130	0.176
1/4	0.160	0.189	0.201	0.287	0.230	0.219	0.250
1/3	0.253	0.291	0.308	0.380	0.323	0.317	0.343
$n_1 = n_2 = 150$							
0	0.045	0.047	0.047	0.103	0.039	0.035	0.036
1/9	0.121	0.123	0.120	0.275	0.219	0.245	0.222
1/7	0.152	0.161	0.162	0.331	0.292	0.334	0.303
1/5	0.296	0.306	0.305	0.447	0.429	0.502	0.470
1/4	0.485	0.505	0.505	0.577	0.589	0.683	0.614
1/3	0.724	0.735	0.737	0.695	0.745	0.824	0.744
$n_1 = n_2 = 250$							
0	0.054	0.051	0.053	0.106	0.052	0.085	0.048
1/9	0.153	0.169	0.173	0.360	0.330	0.460	0.312
1/7	0.270	0.278	0.273	0.411	0.420	0.537	0.440
1/5	0.472	0.490	0.490	0.586	0.610	0.785	0.625
1/4	0.690	0.702	0.702	0.747	0.800	0.896	0.788
1/3	0.913	0.920	0.922	0.874	0.920	0.978	0.917

The results in these tables do not allow us to give a clear-cut answer to the question of what test statistic should be preferred. Bootstrap-based statistics perform reasonably well in terms of size and power in both experiments. Moreover, bandwidth selection does not play a crucial role in their behavior. On the other hand, the infeasible statistic based on martingale processes behaves extremely well in terms of power, but the asymptotic critical value is not very accurate, which results in a certain discrepancy between empirical size and significance level if the sample sizes are not large enough. The feasible statistic based on martingale processes continues to be very powerful, but now bandwidth selection does play a crucial role, to the extent that slight deviations from the correct value may lead to wrong decisions.

Another relevant conclusion that follows from our experiments is that there may be big differences between $\overline{C}_{n_1, n_2, x_0}^{(1)}$ and $\overline{C}_{n_1, n_2, x_0}^{(2)}$, even if infeasible statistics are used. For example, in Experiment 2 with $n_1 = n_2 = 150$, if we generate ε_1 from a standard normal distribution and test the null hypothesis “ H_0^* : the distribution of ε_1 is a standardized Student’s t distribution with δ^{-1} degrees of freedom” with significance level 0.05, then the proportion of rejections is always below 0.10; but if we generate ε_2 from a standardized Student’s t distribution with δ^{-1} degrees of freedom and test the null hypothesis “ H_0^{**} : the distribution of ε_2 is a standard normal” with significance level 0.05, then the proportion of rejection ranges from 0.25 (when $\delta^{-1} = 9$) to 0.75 (when $\delta^{-1} = 3$). So it is important to consider a test statistic that combines $\overline{C}_{n_1, n_2, x_0}^{(1)}$ and $\overline{C}_{n_1, n_2, x_0}^{(2)}$, since in practice we cannot know in advance which of them would lead to a more powerful test.

5 Concluding Remarks

In this paper we suggest two alternative procedures for comparing the distribution functions of errors from two regression models that specify parametrically

the conditional mean and variance. Firstly, we propose using statistics based on residual empirical processes and approximating critical values with a smooth bootstrap method. We derive conditions under which this bootstrap method is consistent. Secondly, we propose using statistics based on martingale transforms of the residual empirical processes, replacing the unspecified functions by non-parametric estimates. We give conditions under which this replacement has no effect on the asymptotic null distribution of the statistics. We compare the performance of the two alternative procedures with two Monte Carlo experiments. The results of these experiments suggest that the statistics based on martingale transforms behave much better in terms of power, but they are too sensitive to bandwidth selection and use asymptotic critical values that are not very accurate with small sample sizes.

Two natural extensions stem from our work. The methodology that we describe here can also be used when the distribution functions of k regression errors are to be compared, with $k > 2$, using $k - 1$ pairwise comparisons as suggested in Neumeyer and Dette (2003, Remark 2.5); our results may well continue to hold in this context, at the expense of some more complex notation. On the other hand, we could also consider a purely nonparametric framework, i.e., comparing the distribution functions of errors from two nonparametrically specified regression models. However, the extension of our results to this context is far from trivial. It might be possible to use the results derived in Akritas and Van Keilegom (2001) to derive the asymptotic properties of statistics based on empirical processes constructed with nonparametric residuals, but their results are valid only for models with a single explanatory variable. Additionally, in nonparametric regression, the oscillation-like result which relates error-based and residual-based empirical processes does not guarantee that a suitable martingale transform exists.

6 Proofs

The following lemma is required in the proof of Proposition 1.

Lemma 1: *Let $f(\cdot)$ be a continuous real function such that $\sup_{x \in \mathbb{R}} |xf(x)| < \infty$. If the sequence of real functions $\{\xi_n(\cdot)\}_{n \in \mathbb{N}}$ satisfies that $|\xi_n(x)| \leq 1$ and $\{a_n\}_{n \in \mathbb{N}}$ is a sequence such that $a_n = o(1)$, then $\sup_{x \in \mathbb{R}} |xf\{x[1 + \xi_n(x)a_n]\}| = O(1)$.*

Proof: As $h(x_1, x_2) = |x_1 f(x_2)|$ is a continuous function, it is bounded in a neighborhood of $(0, 0)'$. If x and $x_n := x[1 + \xi_n(x)a_n]$ are not in that neighborhood then $|xf(x_n)| = |x/x_n| |x_n f(x_n)| \leq [1 + \xi_n(x)a_n]^{-1} \sup_{x \in \mathbb{R}} |xf(x)| = O(1)$. ■

Proof of Proposition 1: We apply Theorem 1 in Bai (2003) to our j -th regression model. The relationship between the notation in Bai (2003) and our notation is as follows: $F_i(x \mid \Omega_i, \theta) \equiv F_j[\{x - \mu_j(X_{ji}, \theta_j)\}/\sigma_j(X_{ji}, \theta_j)]$, $U_i \equiv F_j(\varepsilon_{ji})$, $\hat{U}_i \equiv F_j(\hat{\varepsilon}_{ji})$, $\hat{V}_n(r) \equiv \hat{V}_{jn_j}(r)$, $V_n(r) \equiv V_{jn_j}(r)$ and $\bar{g}(r) \equiv -\omega_j g_j(r)$. To check that our assumptions imply that assumption A1 in Bai (2003) holds, note that for fixed $M > 0$, if \sup_u denotes the supremum for $\|u - \theta_j\| \leq Mn_j^{-1/2}$, then

$$\begin{aligned} \sup_{x \in \mathbb{R}} \sup_u \left\| \frac{\partial F_i}{\partial \theta}(x \mid \Omega_i, u) \right\|^2 &\leq 2 \sup_{x \in \mathbb{R}} \left| f_j \left\{ \frac{x - \mu_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right|^2 \sup_u \left\| \frac{\dot{\mu}_j(X_{ji}, u)}{\sigma_j(X_{ji}, u)} \right\|^2 \\ &+ 2 \sup_{x \in \mathbb{R}} \left| \frac{x - \mu_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} f_j \left\{ \frac{x - \mu_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right|^2 \sup_u \left\| \frac{\dot{\sigma}_j(X_{ji}, u)}{\sigma_j(X_{ji}, u)} \right\|. \end{aligned}$$

Now, from our assumption 2,

$$E \left\{ \sup_u \left\| \frac{\dot{\mu}_j(X_{ji}, u)}{\sigma_j(X_{ji}, u)} \right\|^2 \right\} \leq \frac{2E\{M_3(X_{ji})^2\}(Mn_j^{-1/2})^{2\alpha_{3j}} + 2E\{\|\dot{\mu}_j(X_{ji}, \theta_j)\|^2\}}{S_j^2},$$

and a similar inequality holds for $E\{\sup_u \|\dot{\sigma}_j(X_{ji}, u)/\sigma_j(X_{ji}, u)\|^2\}$. From here it follows that $E\{\sup_{x \in \mathbb{R}} \sup_u \|\frac{\partial F_i}{\partial \theta}(x \mid \Omega_i, u)\|^2\}$ is bounded, and all other conditions of assumption A1 in Bai (2003) readily follow from our assumptions 1 and 2.

To check that our assumptions imply that assumption A2 in Bai (2003) holds, note that if we define $\eta_{ji}(t, u, v) := \{F_j^{-1}(t)\sigma_j(X_{ji}, u) + \mu_j(X_{ji}, u) - \mu_j(X_{ji}, v)\} / \sigma_j(X_{ji}, v)$ and $h_j(x) := xf_j(x)$, then

$$\begin{aligned} & \left\| n^{-1} \sum_{i=1}^n \frac{\partial F_i}{\partial \theta} \{F_i^{-1}(t \mid \Omega_i, u) \mid \Omega_i, v\} - \bar{g}(r) \right\| = \\ & \left\| n_j^{-1} \sum_{i=1}^{n_j} f_j \{ \eta_{ji}(t, u, v) \} \left\{ \frac{\dot{\mu}_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} + \eta_{ji}(t, u, v) \frac{\dot{\sigma}_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} \right\} + \bar{g}(r) \right\| \leq \\ & \left\| n_j^{-1} \sum_{i=1}^{n_j} f_j \{ \eta_{ji}(t, u, v) \} \left[\frac{\dot{\mu}_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} - E \left\{ \frac{\dot{\mu}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right] \right\| + \\ & \left\| n_j^{-1} \sum_{i=1}^{n_j} [f_j \{ \eta_{ji}(t, u, v) \} - f_j \{ F_j^{-1}(t) \}] \right\| \left\| E \left\{ \frac{\dot{\mu}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right\| + \\ & \left\| n_j^{-1} \sum_{i=1}^{n_j} h_j \{ \eta_{ji}(t, u, v) \} \left[\frac{\dot{\sigma}_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} - E \left\{ \frac{\dot{\sigma}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right] \right\| + \\ & \left\| n_j^{-1} \sum_{i=1}^{n_j} [h_j \{ \eta_{ji}(t, u, v) \} - h_j \{ F_j^{-1}(t) \}] \right\| \left\| E \left\{ \frac{\dot{\sigma}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right\| = \end{aligned}$$

(I) + (II) + (III) + (IV), say.

Now observe that if C_{1j} is a bound for $f_j(\cdot)$ and, for fixed M , $\sup_{u,v}$ denotes the supremum for $\|u - \theta_j\| \leq Mn_j^{-1/2}$ and $\|v - \theta_j\| \leq Mn_j^{-1/2}$, then

$$\begin{aligned} \sup_{t \in [0,1]} \sup_{u,v} (I) & \leq C_{1j} \left\{ \sup_{u,v} \left\| n_j^{-1} \sum_{i=1}^{n_j} \frac{\dot{\mu}_j(X_{ji}, v) - \dot{\mu}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, v)} \right\| \right\} + \\ & \sup_{u,v} \left\| n_j^{-1} \sum_{i=1}^{n_j} \dot{\mu}_j(X_{ji}, \theta_j) \frac{\sigma_j(X_{ji}, \theta_j) - \sigma_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)\sigma_j(X_{ji}, \theta_j)} \right\| + \end{aligned}$$

$$\left\| n_j^{-1} \sum_{i=1}^{n_j} \left[\frac{\dot{\mu}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} - E \left\{ \frac{\dot{\mu}_j(X_{ji}, \theta_j)}{\sigma_j(X_{ji}, \theta_j)} \right\} \right] \right\|.$$

Now using that $\sigma_j(\cdot, v)$ is bounded away from zero in a neighborhood of θ_j and the Lipschitz-continuity of $\dot{\mu}_j(\cdot, \cdot)$ and $\sigma_j(\cdot, \cdot)$, it follows that the first two terms on the right-hand side of the previous inequality are $O_p(1)o(1)$, whereas the third term is $o_p(1)$ by the weak law of large numbers; hence $\sup_{t \in [0,1]} \sup_{u,v} (I) = o_p(1)$. As $h_j(\cdot)$ is also bounded and $\dot{\sigma}_j(\cdot, \cdot)$ is also Lipschitz-continuous, with the same reasoning it follows that $\sup_{t \in [0,1]} \sup_{u,v} (III) = o_p(1)$. To analyze (II), note that if we add and subtract $f_j\{F_j^{-1}(t)\sigma_j(X_{ji}, u)/\sigma_j(X_{ji}, v)\}$ and apply the mean-value theorem twice then

$$\begin{aligned} & f_j\{\eta_{ji}(t, u, v)\} - f_j\{F_j^{-1}(t)\} \\ = & \dot{f}_j \left\{ \frac{F_j^{-1}(t)\sigma_j(X_{ji}, u)}{\sigma_j(X_{ji}, v)} + \xi_1 \frac{\mu_j(X_{ji}, u) - \mu_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} \right\} \frac{\mu_j(X_{ji}, u) - \mu_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} \\ & + \dot{f}_j \left\{ F_j^{-1}(t) \left[1 + \xi_2 \frac{\sigma_j(X_{ji}, u) - \sigma_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)} \right] \right\} F_j^{-1}(t) \frac{\sigma_j(X_{ji}, u) - \sigma_j(X_{ji}, v)}{\sigma_j(X_{ji}, v)}, \end{aligned}$$

where ξ_1, ξ_2 are in $[0, 1]$. Again, using Lipschitz-continuity and the lower bound for $\sigma_j(\cdot, v)$ it follows that $\sup_{u,v} n_j^{-1} \sum_{i=1}^{n_j} |\mu_j(X_{ji}, u) - \mu_j(X_{ji}, v)| / \sigma_j(X_{ji}, v) = O_p(1)o(1)$, and $\sup_{u,v} n_j^{-1} \sum_{i=1}^{n_j} |\sigma_j(X_{ji}, u) - \sigma_j(X_{ji}, v)| / \sigma_j(X_{ji}, v) = O_p(1)o(1)$; hence, if C_{2j} is a bound for $\dot{f}_j(\cdot)$ and C_{3j} is the bound for $x\dot{f}_j(x[1 + \xi_2\{\sigma_j(X_{ji}, u) - \sigma_j(X_{ji}, v)\}/\sigma_j(X_{ji}, v)])$ which is obtained by applying Lemma 1 above, it follows that $\sup_{t \in [0,1]} \sup_{u,v} (II) \leq \{C_{2j}O_p(1)o(1) + C_{3j}O_p(1)o(1)\} E\{\|\dot{\mu}_j(X_{ji}, \theta_j)\|\} S_j^{-1} = o_p(1)$. And since $\dot{h}_j(\cdot)$ is also bounded and satisfies the conditions required in Lemma 1, with the same reasoning it also follows that $\sup_{t \in [0,1]} \sup_{u,v} (IV) = o_p(1)$. On the other hand, $\int_0^1 \|(1, \dot{g}(r)')\|^2 dr \leq 1 + \|E\{\dot{\mu}_j(X_{ji}, \theta_j)/\sigma_j(X_{ji}, \theta_j)\}\|^2 E\{\dot{f}_j(\varepsilon_{ji})^2/f_j(\varepsilon_{ji})^2\} + \|E\{\dot{\sigma}_j(X_{ji}, \theta_j)/\sigma_j(X_{ji}, \theta_j)\}\|^2 [2 + 2E\{\varepsilon_{ji}^2 \dot{f}_j(\varepsilon_{ji})^2/f_j(\varepsilon_{ji})^2\}]$, which is finite by our assumptions 1 and 2. This completes the proof that all assertions of the assumption A2 in Bai (2003) hold, except (possibly) for the condition on $C(s)$, which in fact is not required for his Theorem 1 to hold.

Finally, note that our assumption 3 readily implies that assumption A3 in Bai (2003) holds, whereas his assumption A4 is not required in our context because there is no information truncation. Thus we can apply Theorem 1 in Bai (2003), and then (2) follows from our assumption 3 and equation (2) in Bai (2003). ■

Proof of Theorem 1: First we prove the theorem for \widehat{K}_{n_1, n_2} . Note that, under H_0 , $\widehat{K}_{n_1, n_2} = \sup_{t \in [0, 1]} \left| \widehat{D}_{n_1, n_2}(t) \right|$, where for $t \in [0, 1]$ we define

$$\widehat{D}_{n_1, n_2}(t) := \{\lambda / (\lambda + 1)\}^{1/2} \widehat{V}_{1n_1}(t) - \{1 / (\lambda + 1)\}^{1/2} \widehat{V}_{2n_2}(t), \quad (5)$$

Thus, it suffices to prove that $\widehat{D}_{n_1, n_2}(\cdot)$ converges weakly to $D^{(1)}(\cdot)$. From (2) and (5), it follows that $\widehat{D}_{n_1, n_2}(\cdot)$ has the same asymptotic behavior as $\widetilde{D}_{n_1, n_2}(\cdot) := \{\lambda / (\lambda + 1)\}^{1/2} \{V_{1n_1}(\cdot) + g_1(\cdot)' \xi_{1n_1}\} - \{1 / (\lambda + 1)\}^{1/2} \{V_{2n_2}(\cdot) + g_2(\cdot)' \xi_{2n_2}\}$. Now observe that $E\{\widetilde{D}_{n_1, n_2}(t)\} = 0$, and routine calculations yield that $\lim_{n_j \rightarrow \infty} \text{Cov}\{V_{jn_j}(s) + g_j(s)' \xi_{jn_j}, V_{jn_j}(t) + g_j(t)' \xi_{jn_j}\} = \min(s, t) - st + \Lambda_j(s, t, \theta_j)$. Hence, the covariance function of $\widetilde{D}_{n_1, n_2}(\cdot)$ converges to that of $D^{(1)}(\cdot)$. Using a standard multivariate central limit theorem it follows then that the finite dimensional distributions of $\widetilde{D}_{n_1, n_2}(\cdot)$ converge to those of $D^{(1)}(\cdot)$. Additionally, as $g_j(t)' \xi_{jn_j}$ only depends on t through $f_j\{F_j^{-1}(t)\}$ and $F_j^{-1}(t)$, from assumption 1 it readily follows that $\widetilde{D}_{n_1, n_2}(\cdot)$ is tight, which completes the proof of part **a**. On the other hand, under our assumptions $\sup_{x \in \mathbb{R}} \left| \widehat{F}_{jn_j}(x) - F_j(x) \right| = o_p(1)$, and hence $\sup_{x \in \mathbb{R}} \left| \widehat{F}_{1n_1}(x) - \widehat{F}_{2n_2}(x) \right|$ converges in probability to $\sup_{z \in \mathbb{R}} |F_1(x) - F_2(x)|$. Under H_1 , $\sup_{z \in \mathbb{R}} |F_1(x) - F_2(x)| > 0$, and part **b** follows from there.

As regards \widehat{C}_{n_1, n_2} , first note that if $\widehat{F}_{n_1, n_2}(\cdot)$ denotes the empirical distribution function based on the $n_1 + n_2$ residuals, then $\widehat{C}_{n_1, n_2} = \{n_1 n_2 / (n_1 + n_2)\} \int \{\widehat{F}_{1n_1}(x) - \widehat{F}_{2n_2}(x)\}^2 d\widehat{F}_{n_1, n_2}(x)$. Using similar arguments to those in part c of Proposition A1 in Delgado and Mora (2000), it follows that $\widehat{C}_{n_1, n_2} = \widetilde{C}_{n_1, n_2} + o_p(1)$, where $\widetilde{C}_{n_1, n_2} := \{n_1 n_2 / (n_1 + n_2)\} \int \{\widehat{F}_{1n_1}(x) - \widehat{F}_{2n_2}(x)\}^2 dF_M(x)$, and $F_M(\cdot) := \{1 / (1 + \lambda)\} F_1(\cdot) + \{\lambda / (1 + \lambda)\} F_2(\cdot)$. Now, under H_0 , $\widetilde{C}_{n_1, n_2} = \int_0^1 \{\widehat{D}_{n_1, n_2}(t)\}^2 dt$, and part **a** follows from there as before. On the other hand, $\int \{\widehat{F}_{1n_1}(x) - \widehat{F}_{2n_2}(x)\}^2 dF_M(x)$

converges in probability to $\int \{F_1(x) - F_2(x)\}^2 dF_M(x)$; under H_1 this integral is positive, which completes the proof of part **b**. ■

Proof of Proposition 2: The proof is similar to the proof of Theorem 2.6 in Rao and Sethuraman (1975) and the proof of Lemma 1 in Loynes (1980). However, it is more complicated due to the fact that we consider a random (conditional) probability measure $P_{n_1, n_2}^*(\cdot) = P(\cdot \mid \mathcal{Y}_{n_1, n_2})$ and the random variables ε_{ji}^* are not i.i.d. but have (conditional) distribution function \tilde{F}_{n_1, n_2} . Hence we present the proof in detail. For ease of notation we write $s_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) := \sigma_j(X_{ji}, \hat{\theta}_j^*)/\sigma_j(X_{ji}, \hat{\theta}_j)$, $m_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) := \{\mu_j(X_{ji}, \hat{\theta}_j^*) - \mu_j(X_{ji}, \hat{\theta}_j)\}/\sigma_j(X_{ji}, \hat{\theta}_j)$,

$$R_{n_1, n_2}(y, \hat{\theta}_j^*) := n_j^{-1/2} \sum_{i=1}^{n_j} \left[I \left\{ \varepsilon_{ji}^* \leq y s_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) + m_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) \right\} - \tilde{F}_{n_1, n_2} \left(y s_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) + m_j(X_{ji}, \hat{\theta}_j, \hat{\theta}_j^*) \right) - I \{ \varepsilon_{ji}^* \leq y \} + \tilde{F}_{n_1, n_2}(y) \right].$$

To prove the proposition we show that for every fixed $M > 0$ and for all $\alpha > 0$,

$$P \left(\sup_{y \in \mathbb{R}} \sup_{\eta: \|\eta - \hat{\theta}_j\| \leq \frac{M}{\sqrt{n_j}}} |R_{n_1, n_2}(y, \eta)| > \alpha \mid \mathcal{Y}_{n_1, n_2} \right) = o_p(1). \quad (6)$$

Note that from assumption 6 it follows that there exists $M > 0$, such that

$$P \left(\|\hat{\theta}_j^* - \hat{\theta}_j\| \leq \frac{M}{\sqrt{n_j}} \mid \mathcal{Y}_{n_1, n_2} \right) \longrightarrow 1 \quad \text{in probability.}$$

To derive (6), first of all cover $\{\eta \in \Theta_j \mid \|\eta - \hat{\theta}_j\| \leq M/\sqrt{n_j}\}$ using $K = O(\epsilon^{-k_j})$ balls B_1, \dots, B_K with centers η_1, \dots, η_K and radius $\epsilon/\sqrt{n_j}$, where the constant ϵ will be specified later. Applying assumption 2 we obtain for all $\eta \in B_k$, $|s_j(x, \hat{\theta}_j, \eta) - s_j(x, \hat{\theta}_j, \eta_k)| \leq M_{2j}(x)\epsilon/(S_j\sqrt{n_j})$ and $|m_j(x, \hat{\theta}_j, \eta) - m_j(x, \hat{\theta}_j, \eta_k)| \leq M_{1j}(x)\epsilon/(S_j\sqrt{n_j})$. With the definitions

$$\begin{aligned} x_{i, n_j, k}^L(y) &= y s_j(X_{ji}, \hat{\theta}_j, \eta_k) + m_j(X_{ji}, \hat{\theta}_j, \eta_k) - \frac{\epsilon}{\sqrt{n_j}} (y M_{2j}(X_{ij}) + M_{1j}(X_{ji}))/S_j \\ x_{i, n_j, k}^U(y) &= y s_j(X_{ji}, \hat{\theta}_j, \eta_k) + m_j(X_{ji}, \hat{\theta}_j, \eta_k) + \frac{\epsilon}{\sqrt{n_j}} (y M_{2j}(X_{ij}) + M_{1j}(X_{ji}))/S_j \end{aligned}$$

we have the bracketing

$$x_{i,n_j,k}^L(y) \leq y s_j(X_{ji}, \hat{\theta}_j, \eta) + m_j(X_{ji}, \hat{\theta}_j, \eta) \leq x_{i,n_j,k}^U(y)$$

and therefore

$$I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^L(y)\} \leq I\{\varepsilon_{ji}^* \leq y s_j(X_{ji}, \hat{\theta}_j, \eta) + m_j(X_{ji}, \hat{\theta}_j, \eta)\} \leq I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(y)\}$$

for all $\eta \in B_k$. In the following we concentrate only on the upper bound. The lower bound is treated exactly in the same way and we then use the argumentation that from a bracketing $a_t \leq x_t \leq b_t$ follows $\sup_t |x_t| \leq \sup_t |a_t| + \sup_t |b_t|$. To estimate $\sup_{y \in \mathbb{R}} \sup_{\|\eta - \hat{\theta}_j\| \leq M/\sqrt{n_j}} |R_{n_1, n_2}(y, \eta)|$ in (6) we now only have to consider

$$\begin{aligned} \max_{k=1, \dots, K} \sup_{y \in \mathbb{R}} \left| n_j^{-1/2} \sum_{i=1}^{n_j} [I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(y)\} - I\{\varepsilon_{ji}^* \leq y\}] \right. \\ \left. - \tilde{F}_{n_1, n_2}(x_{i,n_j,k}^U(y)) + \tilde{F}_{n_1, n_2}(y) \right|. \end{aligned} \quad (7)$$

The remainder is estimated as follows using a Taylor expansion (where $\eta \in B_k$),

$$\begin{aligned} n_j^{-1/2} \sum_{i=1}^{n_j} \left| \tilde{F}_{n_1, n_2}(y s_j(X_{ji}, \hat{\theta}_j, \eta) + m_j(X_{ji}, \hat{\theta}_j, \eta)) - \tilde{F}_{n_1, n_2}(x_{i,n_j,k}^U(y)) \right| \\ \leq 2n_j^{-1/2} \sum_{i=1}^{n_j} \left[\sup_{y \in \mathbb{R}} |\tilde{f}_{n_1, n_2}(y)| \frac{M_{1j}(X_{ji})\epsilon}{S_j \sqrt{n_j}} + \sup_{y \in \mathbb{R}} |y \tilde{f}_{n_1, n_2}(y^*)| \frac{M_{2j}(X_{ji})\epsilon}{S_j \sqrt{n_j}} \right], \end{aligned}$$

where y^* converges in probability to y as $n_j \rightarrow \infty$. From the uniform convergence of \tilde{f}_{n_1, n_2} to f_M in probability, Lemma 1, the law of large numbers and Assumption 2, we obtain the bound $\epsilon O_p(1)$, which can be made arbitrarily small for a proper choice of ϵ (in probability).

Next we split the interval $[0, 1]$ into $L = \sqrt{n_j}/\lambda$ intervals of length $\lambda/\sqrt{n_j}$ using points $0 = t_0 < t_1 < \dots < t_L = 1$. For $\ell = 0, \dots, L$ we define y_ℓ and \tilde{y}_ℓ via $F_M(y_\ell) = t_\ell$ and $F_M(x_{i,n_j,k}^U(\tilde{y}_\ell)) = t_\ell$ (we suppress the dependence from i, n_j, k for ease of notation). For every $y \in \mathbb{R}$ there are ℓ and $\tilde{\ell}$, such that $y_{\ell-1} \leq y \leq y_\ell$ and $\tilde{y}_{\tilde{\ell}-1} \leq y \leq \tilde{y}_{\tilde{\ell}}$. These indices ℓ and $\tilde{\ell}$ are linked because $F_M(y_\ell) = t_\ell =$

$F_M(x_{i,n_j,k}^U(\tilde{y}_\ell)) = F_M(\tilde{y}_\ell) + o_p(1)$. Therefore in the following we assume $\ell = \tilde{\ell}$ to simplify the demonstration of the proof. From the monotonicity of the functions $y \mapsto \tilde{F}_{n_1,n_2}(y)$ and $y \mapsto \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(y))$ we obtain the bracketing

$$\begin{aligned} & I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(\tilde{y}_{\ell-1})\} - I\{\varepsilon_{ji}^* \leq y_\ell\} - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_\ell)) + \tilde{F}_{n_1,n_2}(y_{\ell-1}) \\ & \leq I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(y)\} - I\{\varepsilon_{ji}^* \leq y\} - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(y)) + \tilde{F}_{n_1,n_2}(y) \\ & \leq I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(\tilde{y}_\ell)\} - I\{\varepsilon_{ji}^* \leq y_{\ell-1}\} - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_{\ell-1})) + \tilde{F}_{n_1,n_2}(y_\ell). \end{aligned}$$

As before, we restrict our considerations to the upper bound. Instead of (7) it suffices to estimate

$$\begin{aligned} \max_{\substack{k=1,\dots,K \\ \ell=1,\dots,L}} \left| n_j^{-1/2} \sum_{i=1}^{n_j} [I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(\tilde{y}_\ell)\} - I\{\varepsilon_{ji}^* \leq y_{\ell-1}\}] \right. \\ \left. - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_\ell)) + \tilde{F}_{n_1,n_2}(y_{\ell-1}) \right|. \end{aligned} \quad (8)$$

The remainder is treated as follows,

$$\begin{aligned} & n_j^{1/2} |\tilde{F}_{n_1,n_2}(y_\ell) - \tilde{F}_{n_1,n_2}(y_{\ell-1})| \\ & + n_j^{-1/2} \sum_{i=1}^{n_j} |\tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_\ell)) - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_{\ell-1}))| \\ & \leq n_j^{1/2} \sup_{t \in [0,1]} \left| \frac{\partial}{\partial t} (\tilde{F}_{n_1,n_2} \circ F_M^{-1})(t) \right| \left\{ |F_M(y_\ell) - F_M(y_{\ell-1})| \right. \\ & \quad \left. + n_j^{-1} \sum_{i=1}^{n_j} |F_M(x_{i,n_j,k}^U(\tilde{y}_\ell)) - F_M(x_{i,n_j,k}^U(\tilde{y}_{\ell-1}))| \right\} \\ & = 2\lambda \sup_{y \in \mathbb{R}} \left| \frac{\tilde{f}_{n_1,n_2}(y)}{f_M(y)} \right| = \lambda O_p(1) \end{aligned} \quad (9)$$

and can be made arbitrarily small by proper choice of λ , in probability.

From now on we can follow the proof of Rao and Sethuraman (1975, p. 307) or Loynes (1980, p. 293) by writing (8) as

$$\max_{\substack{k=1,\dots,K \\ \ell=1,\dots,L}} \left| n_j^{-1/2} \sum_{i=1}^{n_j} W_{i,n_j,k,\ell} \right|,$$

where $W_{i,n_j,k,\ell} = (B_{i,n_j,k,\ell} - |p_{i,n_j,k,\ell}|) \text{sign}(p_{i,n_j,k,\ell})$, $p_{i,n_j,k,\ell} = \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_\ell)) - \tilde{F}_{n_1,n_2}(y_{\ell-1})$, and under the conditional probability measure P_{n_1,n_2}^* the random

variables $B_{i,n_j,k,\ell} = |I\{\varepsilon_{ji}^* \leq x_{i,n_j,k}^U(\tilde{y}_\ell)\} - I\{\varepsilon_{ji}^* \leq y_{\ell-1}\}|$ are independent and Bernoulli distributed with probability $|p_{i,n_j,k,\ell}|$. From the proofs cited we see that we can derive our assertion (6) by proving $n_j^{-1/2} \sum_{i=1}^{n_j} |p_{i,n_j,k,\ell}| = O_p(1)$ uniformly in k, ℓ . We have with $y_{\ell-1} \leq y \leq y_\ell$ and $\tilde{y}_{\ell-1} \leq y \leq \tilde{y}_\ell$,

$$\begin{aligned} n_j^{-1/2} \sum_{i=1}^{n_j} |p_{i,n_j,k,\ell}| &\leq n_j^{-1/2} \sum_{i=1}^{n_j} \left\{ |\tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(\tilde{y}_\ell)) - \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(y))| \right. \\ &\quad \left. + |\tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(y)) - \tilde{F}_{n_1,n_2}(y)| + |\tilde{F}_{n_1,n_2}(y) - \tilde{F}_{n_1,n_2}(y_{\ell-1})| \right\}. \end{aligned}$$

The first and last terms can be treated in exactly the same way as (9) due to the monotonicity of $y \mapsto \tilde{F}_{n_1,n_2}(y)$ and $y \mapsto \tilde{F}_{n_1,n_2}(x_{i,n_j,k}^U(y))$. The asymptotic order is $\lambda O_p(1)$. By the definition of $x_{i,n_j,k}^U(y)$ and a Taylor expansion the middle term is equal to

$$\begin{aligned} &\frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} \left| \tilde{F}_{n_1,n_2} \left(y s_j(X_{ji}, \hat{\theta}_j, \eta_k) + m_j(X_{ji}, \hat{\theta}_j, \eta_k) + \frac{(y M_{2j}(X_{ij}) + M_{1j}(X_{ji}))\epsilon}{S_j \sqrt{n_j}} \right) \right. \\ &\quad \left. - \tilde{F}_{n_1,n_2}(y) \right| \\ &\leq \sup_{y \in \mathbb{R}} |\tilde{f}_{n_1,n_2}(y)| \left[\frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} |m_j(X_{ji}, \hat{\theta}_j, \eta_k)| + \frac{\epsilon}{n_j S_j} \sum_{i=1}^{n_j} M_{1j}(X_{ji}) \right] \\ &\quad + \frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} \sup_{y \in \mathbb{R}} |\tilde{f}_{n_1,n_2}(y^*) y| \left\{ |s_j(X_{ji}, \hat{\theta}_j, \eta_k) - 1| + \frac{\epsilon}{n_j S_j} \sum_{i=1}^{n_j} M_{2j}(X_{ji}) \right\} \\ &\leq O_p(1) \left(\frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} (M_{1j}(X_{ji}) + M_{2j}(X_{ji})) \|\hat{\theta}_j - \eta_k\| + \epsilon O_p(1) \right) \\ &= O_p(1)(M O_p(1) + \epsilon O_p(1)) = O_p(1) \end{aligned}$$

and this concludes the proof. ■

Proof of Proposition 3: From Proposition 2 it follows that $\widehat{V}_{j n_j}^*$ and $V_{j n_j} + \tilde{V}_{j n_j}^*$, where $\tilde{V}_{j n_j}^*$ is as defined in Proposition 2, are asymptotically equivalent (in terms of conditional weak convergence in probability). We further define $\check{V}_{j n_j}^*(t) := n_j^{1/2} \int G_{n_1,n_2}(x, t) dF_{X_j}(x)$, where

$$G_{n_1,n_2}(x, t) = \tilde{F}_{n_1,n_2}^{-1} \left(\tilde{F}_{n_1,n_2}^{-1}(t) \frac{\sigma_j(x, \hat{\theta}_j^*)}{\sigma_j(x, \hat{\theta}_j)} + \frac{\mu_j(x, \hat{\theta}_j^*) - \mu_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} \right) - t = o_p(1)$$

for all x , uniformly in t (this follows applying a Taylor expansion; compare the proof below). Denote the empirical distribution function of X_{j1}, \dots, X_{jn_j} by F_{X_j, n_j} ; then $n_j^{1/2}(F_{X_j, n_j} - F_{X_j})$ converges to a Gaussian process and hence

$$\begin{aligned}\tilde{V}_{jn_j}^*(t) - \check{V}_{jn_j}^*(t) &= \int G_{n_1, n_2}(x, t) d(\sqrt{n_j}(F_{X_j, n_j} - F_{X_j}))(x) \\ &= o_p(1) = o_{P_{n_1, n_2}^*}(1) \text{ in probability,}\end{aligned}\tag{10}$$

where $P_{n_1, n_2}^*(\cdot)$ denotes the conditional probability measure $P(\cdot \mid \mathcal{Y}_{n_1, n_2})$ and the notation “ $Z_{n_1, n_2} = o_{P_{n_1, n_2}^*}(1)$ in probability” means $P(Z_{n_1, n_2} > \epsilon \mid \mathcal{Y}_{n_1, n_2}) = o_p(1)$ for all $\epsilon > 0$. The last equality in (10) follows from Markov’s inequality. By a Taylor expansion of $\tilde{F}_{n_1, n_2}^{-1}$ we further obtain the asymptotic equivalence of $\check{V}_{jn_j}^*(t)$ and the process

$$\sqrt{n_j} \int \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t^*)) \left(\tilde{F}_{n_1, n_2}^{-1}(t) \frac{\sigma_j(x, \hat{\theta}_j^*) - \sigma_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} + \frac{\mu_j(x, \hat{\theta}_j^*) - \mu_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} \right) dF_{X_j}(x),$$

where $\tilde{F}_{n_1, n_2}^{-1}(t^*)$ lies between $\tilde{F}_{n_1, n_2}^{-1}(t)$ and

$$\begin{aligned}&\tilde{F}_{n_1, n_2}^{-1}(t) \frac{\sigma_j(x, \hat{\theta}_j^*) - \sigma_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} + \frac{\mu_j(x, \hat{\theta}_j^*) - \mu_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} \\ &= \left(\tilde{F}_{n_1, n_2}^{-1}(t) \frac{\dot{\sigma}_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} + \frac{\dot{\mu}_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} \right) (\hat{\theta}_j^* - \hat{\theta}_j) + o_{P_{n_1, n_2}^*} \left(\frac{1}{\sqrt{n_j}} \right) = O_{P_{n_1, n_2}^*} \left(\frac{1}{\sqrt{n_j}} \right)\end{aligned}$$

from assumptions 2 and 6. Therefore $\tilde{F}_{n_1, n_2}^{-1}(t^*) - \tilde{F}_{n_1, n_2}^{-1}(t)$ converges to zero (in probability) and we obtain

$$\begin{aligned}&|\tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t^*)) - \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t))| \\ &\leq |f_M(\tilde{F}_{n_1, n_2}^{-1}(t^*)) - f_M(\tilde{F}_{n_1, n_2}^{-1}(t))| + 2 \sup_y |\tilde{f}_{n_1, n_2}(y) - f_M(y)|.\end{aligned}$$

These terms converge to zero uniformly in t (in probability) by the uniform continuity of f_M (assumption 1) and by the uniform convergence of \tilde{f}_{n_1, n_2} (compare our assumptions 4 and 5 and Koul and Lahiri, 1994). We obtain the asymptotic

equivalence of $\check{V}_{jn_j}^*(t)$ and

$$\int \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t)) \left(\tilde{F}_{n_1, n_2}^{-1}(t) \frac{\dot{\sigma}_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} (\hat{\theta}_j^* - \hat{\theta}_j) + \frac{\dot{\mu}_j(x, \hat{\theta}_j)}{\sigma_j(x, \hat{\theta}_j)} (\hat{\theta}_j^* - \hat{\theta}_j) \right) dF_{X_j}(x)$$

uniformly in $t \in [0, 1]$ in probability. The assertion of the Lemma now follows from assumption 6. ■

Proof of Theorem 2: As convergence in probability is equivalent to the existence of almost surely convergent subsequences in every subsequence, we restrict our proof to the case of almost sure convergence of $\hat{\theta}_j$ to θ_j and show almost sure convergence of $P(\hat{K}_{n_1 n_2}^* \leq x \mid \mathcal{Y}_{n_1, n_2})$. From Proposition 3 we have asymptotic equivalence in terms of weak convergence, conditionally on the sample \mathcal{Y}_{n_1, n_2} , of the test statistic $\hat{K}_{n_1 n_2}^*$ and $\sup_{t \in [0, 1]} |\tilde{D}_{n_1 n_2}^*(t)|$, where

$$\begin{aligned} \tilde{D}_{n_1 n_2}^*(t) &= \sqrt{\frac{\lambda}{1+\lambda}} (V_{1n_1}^*(t) + \tilde{g}_1(t) \hat{\xi}_{1n_1}^*) - \sqrt{\frac{1}{1+\lambda}} (V_{2n_2}^*(t) + \tilde{g}_2(t) \hat{\xi}_{2n_2}^*) \\ &= \sum_{j=1}^2 \frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} \alpha_j \Delta_j^*(X_{ji}, \varepsilon_{ji}^*, t) \end{aligned}$$

and $\Delta_j^*(z, u, t) = I\{u \leq \tilde{F}_{n_1, n_2}^{-1}(t)\} - t + \tilde{g}_j(t) \omega_j^{*'} \psi_j(z, u, \hat{\theta}_j)$, $\alpha_1 = \sqrt{\frac{\lambda}{1+\lambda}}$, $\alpha_2 = -\sqrt{\frac{1}{1+\lambda}}$. Note that this process is by assumption centered with respect to the conditional expectation. In the following we show weak convergence (conditionally on \mathcal{Y}_{n_1, n_2}) of $\tilde{D}_{n_1 n_2}^*$ to the Gaussian process $D^{(1)}$ for almost all samples \mathcal{Y}_{n_1, n_2} (compare the proof of Theorem 1, Stute et al., 1998). First we show conditional convergence of the finite dimensional distributions. To keep the proof more readable we demonstrate the validity of Lindeberg's condition only for the statistic

$$\begin{aligned} Z_{n_j, 1}^*(t) &= n_j^{-1/2} \sum_{i=1}^{n_j} \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t)) \int \frac{\dot{\mu}_j(x, \hat{\theta}_j)'}{\sigma_j(x, \hat{\theta}_j)} dF_{X_j}(x) \psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j) \\ &= n_j^{-1/2} \sum_{i=1}^{n_j} \Delta_{j, 1}^*(X_{ji}, \varepsilon_{ji}^*, t), \end{aligned}$$

where $t \in [0, 1]$ is fixed. The sequence $\tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t))$ converges almost surely to $f_M(F_M^{-1}(t))$ (this follows from Koul and Lahiri, 1994, our assumptions 4 and 5 and the additional assumption made in this section that $\hat{\theta}_j$ converges a.s.) and is therefore bounded; $\int \dot{\mu}_j(x, \hat{\theta}_j)/\sigma_j(x, \hat{\theta}_j) dF_{X_j}(x)$ is also bounded because of the almost sure convergence of $\hat{\theta}_j$ to θ_j and assumption 2. We then have for a constant C that

$$\begin{aligned} L_{n_j}(\delta) &= \frac{1}{n_j} \sum_{i=1}^{n_j} E \left[\left(\Delta_{j,1}^*(X_{ji}, \varepsilon_{ji}^*, t) \right)^2 I \{ |\Delta_{j,1}^*(X_{ji}, \varepsilon_{ji}^*, t)| > n_j^{-1/2} \delta \} \middle| \mathcal{Y}_{n_1, n_2} \right] \\ &\leq C^2 E \left[\|\psi_j(X_{j1}, \varepsilon_{j1}^*, \hat{\theta}_j)\|^2 I \{ \|\psi_j(X_{j1}, \varepsilon_{j1}^*, \hat{\theta}_j)\| > n_j^{1/2} \delta / C \} \middle| \mathcal{Y}_{n_1, n_2} \right] \quad \text{a.s.} \\ &= C^2 \int \int \|\psi_j(x, u, \hat{\theta}_j)\|^2 I \{ \|\psi_j(x, u, \hat{\theta}_j)\| > n_j^{1/2} \delta / C \} \tilde{f}_{n_1, n_2}(u) du dF_{X_j}(x) \end{aligned}$$

converges to zero almost surely for $n_j \rightarrow \infty$ (compare Lemma 3.2 and the proof of Theorem 2.1 in Koul and Lahiri, 1994).

To show conditional tightness almost surely, we also confine ourselves to the process $Z_{n_j, 1}^*$ defined above. Tightness of

$$Z_{n_j, 2}^*(t) = n_j^{-1/2} \sum_{i=1}^{n_j} \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t)) \tilde{F}_{n_1, n_2}^{-1}(t) \int \frac{\dot{\sigma}_j(x, \hat{\theta}_j)'}{\sigma_j(x, \hat{\theta}_j)} dF_{X_j}(x) \psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j)$$

follows in a similar manner and tightness of the uniform process $n_j^{-1/2} \sum_{i=1}^{n_j} [I\{\varepsilon_{ji}^* \leq \tilde{F}_{n_1, n_2}^{-1}(t)\} - t] = n_j^{-1/2} \sum_{i=1}^{n_j} [I\{U_{ji}^* \leq t\} - t]$, where the U_{ji}^* are i.i.d. uniform on $[0, 1]$, is evident. Considering the stochastic equicontinuity condition for $Z_{n_j, 1}^*$ we obtain the following for $\epsilon > 0$, $\delta > 0$, applying Chebyshev's inequality:

$$\begin{aligned} &P \left(\sup_{\substack{s, t \in [0, 1] \\ |s-t| < \delta}} |Z_{n_j, 1}^*(s) - Z_{n_j, 1}^*(t)| > \epsilon \middle| \mathcal{Y}_{n_1, n_2} \right) \\ &\leq \frac{C^2}{\epsilon^2} \int \int \|\psi_j(x, u, \hat{\theta}_j)\|^2 \tilde{f}_{n_1, n_2}(u) du dF_{X_j}(x) \\ &\quad \times \left(\sup_{\substack{s, t \in [0, 1] \\ |s-t| < \delta}} \left| \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t)) - \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(s)) \right| \right)^2 \quad \text{a.s.} \end{aligned}$$

The integral is almost surely bounded. We further estimate

$$\begin{aligned} & \sup_{\substack{s, t \in [0, 1] \\ |s-t| < \delta}} \left| \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(t)) - \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(s)) \right| \\ & \leq \sup_{\substack{s, t \in [0, 1] \\ |s-t| < \delta}} |f_M(F_M^{-1}(t)) - f_M(F_M^{-1}(s))| + 2 \sup_{s \in [0, 1]} \left| \tilde{f}_{n_1, n_2}(\tilde{F}_{n_1, n_2}^{-1}(s)) - f_M(F_M^{-1}(s)) \right|. \end{aligned}$$

The first term converges to zero as $\delta \searrow 0$ because $f_M\{F_M^{-1}(\cdot)\}$ is uniformly continuous by assumption 1. The second term converges to zero almost surely as $n_j \rightarrow \infty$ analogously to Lemma 3.3 in Koul and Lahiri (1994). This proves conditional tightness of $Z_{n_j, 1}^*$ in the sense that for almost all sequences \mathcal{Y}_{n_1, n_2} and for all $\epsilon > 0$,

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P\left(\sup_{\substack{s, t \in [0, 1] \\ |s-t| < \delta}} |Z_{n_j, 1}^*(s) - Z_{n_j, 1}^*(t)| > \epsilon \mid \mathcal{Y}_{n_1, n_2} \right) = 0.$$

Finally, the conditional covariances of the process $\tilde{D}_{n_1 n_2}^*$ are

$$\begin{aligned} & \text{Cov}(\tilde{D}_{n_1 n_2}^*(s), \tilde{D}_{n_1 n_2}^*(t) \mid \mathcal{Y}_{n_1, n_2}) \\ & = \sum_{j=1}^2 \frac{1}{n_j} \sum_{i=1}^{n_j} \alpha_j^2 E[\Delta_j^*(X_{ji}, \varepsilon_{ji}^*, s) \Delta_j^*(X_{ji}, \varepsilon_{ji}^*, t) \mid \mathcal{Y}_{n_1, n_2}] \\ & = \min(s, t) - st + \{\lambda/(\lambda + 1)\} \Lambda_1^*(s, t) + \{1/(\lambda + 1)\} \Lambda_2^*(s, t), \end{aligned}$$

where

$$\begin{aligned} \Lambda_j^*(s, t) & = \tilde{g}_j(s)' \omega_j^{*'} E[\psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j) I\{\varepsilon_{ji}^* \leq \tilde{F}_{n_1, n_2}^{-1}(t)\} \mid \mathcal{Y}_{n_1, n_2}] \\ & \quad + \tilde{g}_j(t)' \omega_j^{*'} E[\psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j) I\{\varepsilon_{ji}^* \leq \tilde{F}_{n_1, n_2}^{-1}(s)\} \mid \mathcal{Y}_{n_1, n_2}] \\ & \quad + \tilde{g}_j(s)' \omega_j^{*'} E[\psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j) \psi_j(X_{ji}, \varepsilon_{ji}^*, \hat{\theta}_j)' \mid \mathcal{Y}_{n_1, n_2}] \omega_j^* \tilde{g}_j(t). \end{aligned}$$

Taking into account the almost sure convergence of $\tilde{f}_{n_1, n_2}(\cdot)$ to $f_M(\cdot)$, $\hat{\theta}_j$ to θ_j and applying the dominated convergence theorem to the integrals, it follows that this conditional covariance converges almost surely to the covariance $\text{Cov}(D^{(2)}(s), D^{(2)}(t))$ defined in Theorem 2. Thus, the assertion of Theorem 2 for the test statistic $\hat{K}_{n_1 n_2}^*$ follows from the continuous mapping theorem.

When $\widehat{F}_{n_1, n_2}(\cdot)$ denotes the empirical distribution function of all $n_1 + n_2$ residuals $\widehat{\varepsilon}_{ji}$, for the second test statistic we have

$$\begin{aligned}\widehat{C}_{n_1 n_2}^* &= \frac{n_1 n_2}{n_1 + n_2} \int (\widehat{F}_{1n_1}^*(y) - \widehat{F}_{2n_2}^*(y))^2 d\widehat{F}_{n_1, n_2}(y) \\ &= \int_0^1 (\widehat{D}_{n_1, n_2}^*(t))^2 dt + \int_0^1 (\widetilde{D}_{n_1, n_2}^*(t))^2 d((\widehat{F}_{n_1, n_2} - \widetilde{F}_{n_1, n_2}) \circ \widetilde{F}_{n_1, n_2}^{-1}(t)),\end{aligned}$$

where $\widehat{D}_{n_1, n_2}^* = \sqrt{\lambda/(1+\lambda)}\widehat{V}_{1n_1}^* - 1/\sqrt{1+\lambda}\widehat{V}_{2n_2}^*$. From the already shown asymptotic equivalence of \widehat{D}_{n_1, n_2}^* and $\widetilde{D}_{n_1, n_2}^*$, the weak convergence of $\widetilde{D}_{n_1, n_2}^*$ and the almost sure convergence of $\sup_{y \in \mathbb{R}} |\widehat{F}_{n_1, n_2}(y) - \widetilde{F}_{n_1, n_2}(y)|$ to zero we can derive that $\widehat{C}_{n_1 n_2}^*$ is asymptotically equivalent to $\int_0^1 (\widetilde{D}_{n_1, n_2}^*(t))^2 dt$. ■

The following three lemmas are required in the proof of Proposition 4. Lemma 2 extends a well-known inequality in nonparametric estimation (see e.g. Hájek and Sidák 1967, p. 17). Lemma 3 extends the results of contiguous measures in location and scale derived in Hájek and Sidák (1967), Sections VI.1 and VI.2.

Lemma 2: *If assumptions 1, 4 hold and $h_{jn_j} = o(1)$, then*

$$\int_{-\infty}^{\infty} (1+y^2)\overline{f}_j(y)^2/\overline{f}_j(y)dy = O(1), \quad (11)$$

where $\overline{f}_j(y) := h_{jn_j}^{-1} \int_{-\infty}^{\infty} f_j(w)\varphi\{(y-w)/h_{jn_j}\}dw$ and $\dot{\overline{f}}_j(y) := h_{jn_j}^{-2} \int_{-\infty}^{\infty} f_j(w)\dot{\varphi}\{(y-w)/h_{jn_j}\}dw$.

Proof: First observe that $\overline{f}_j(y)$ and $\dot{\overline{f}}_j(y)$ can also be expressed as $\int_{-\infty}^{\infty} f_j(y - h_{jn_j}z)\varphi(z)dz$ and $\int_{-\infty}^{\infty} \dot{f}_j(y - h_{jn_j}z)\varphi(z)dz$, respectively. Thus,

$$\overline{f}_j(y)^2 \leq \left\{ \int_{-\infty}^{\infty} \left| \dot{f}_j(y - h_{jn_j}z) \right| \varphi(z) dz \right\}^2 \leq$$

$$\overline{f}_j(y) \int_{-\infty}^{\infty} \frac{\dot{f}_j(y - h_{jn_j}z)^2}{f_j(y - h_{jn_j}z)} \varphi(z) dz = \overline{f}_j(y) \int_{-\infty}^{\infty} \frac{\dot{f}_j(w)^2}{f_j(w)} \varphi\left(\frac{y-w}{h_{jn_j}}\right) \frac{1}{h_{jn_j}} dw,$$

where the second inequality follows by applying Cauchy-Schwarz inequality with $X = f_j(y - h_{jn_j}Z)^{1/2}$ and $Y = \dot{f}_j(y - h_{jn_j}Z)/f_j(y - h_{jn_j}Z)^{1/2}$, for a random variable

Z with density function $\varphi(\cdot)$. Hence, the left-hand side of (11) is bounded by

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1+y^2) \frac{\dot{f}_j(w)^2}{f_j(w)} \varphi\left(\frac{y-w}{h_{jn_j}}\right) \frac{1}{h_{jn_j}} dw dy \leq \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1+2w^2) \frac{\dot{f}_j(w)^2}{f_j(w)} \varphi\left(\frac{y-w}{h_{jn_j}}\right) \frac{1}{h_{jn_j}} dw dy + \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2(y-w)^2 \frac{\dot{f}_j(w)^2}{f_j(w)} \varphi\left(\frac{y-w}{h_{jn_j}}\right) \frac{1}{h_{jn_j}} dw dy, \end{aligned}$$

where the inequality follows because $y^2 \leq 2(y-w)^2 + 2w^2$. The first term in the latter expression equals $E\{(1+2\varepsilon_{ji}^2)\dot{f}_j(\varepsilon_{ji})^2/f_j(\varepsilon_{ji})^2\} = O(1)$, and the second term equals $2h_{jn_j}^2 E\{\dot{f}_j(\varepsilon_{ji})^2/f_j(\varepsilon_{ji})^2\} \int_{-\infty}^{\infty} v^2 \varphi(v) dv = o(1)$. ■

Lemma 3: Let $\{Z_i\}_{i=1}^n$ be i.i.d. real random variables with a distribution function $F(\cdot)$ that admits a density function $f(\cdot)$ such that $E\{(1+Z_i^2)\dot{f}(Z_i)^2/f(Z_i)^2\} < \infty$. If $\{d_{ni}\}_{i=1}^n, \{e_{ni}\}_{i=1}^n$ are constants such that $\max_{1 \leq i \leq n} d_{ni}^2 = o(1)$, $\sum_{i=1}^n d_{ni}^2 = O(1)$, $\max_{1 \leq i \leq n} (e_{ni}-1)^2 = o(1)$ and $\sum_{i=1}^n (e_{ni}-1)^2 = O(1)$, then the n -dimensional product measure induced by $\{d_{n1} + e_{n1}Z_1, \dots, d_{nn} + e_{nn}Z_n\}$ is contiguous to the n -dimensional product measure induced by $\{Z_1, \dots, Z_n\}$.

Proof: First assume that $E\{(1+Z_i^2)\dot{f}(Z_i)^2/f(Z_i)^2\} > 0$ and $\sum_{i=1}^n (d_{ni}, e_{ni} - 1)'(d_{ni}, e_{ni} - 1) \rightarrow B$, where B is a positive definite (p.d.) matrix. Define $W_{ni} := f\{(Z_i - d_{ni})/e_{ni}\}/\{e_{ni}f(Z_i)\}$, $W_\nu := 2 \sum_{i=1}^n (W_{ni}^{1/2} - 1)$ and $s(x) := f(x)^{1/2}$; then, reasoning as in Lemma VI.2.1.a of Hájek and Sidák (1967), and using a first-order Taylor expansion of $r(h_1, h_2) := s\{(x - h_1)/h_2\}/h_2^{1/2}$ at the point $(0, 1)'$, for a fixed x , it follows that $E(W_{ni}^{1/2} - 1) = -\frac{1}{2} \int [s\{(x - d_{ni})/e_{ni}\}/e_{ni}^{1/2} - s(x)]^2 dx = -\frac{1}{2} [d_{ni}^2 \int \dot{s}(x)^2 dx + (e_{ni} - 1)^2 \int \{s(x)/2 + x\dot{s}(x)\}^2 dx + 2d_{ni}(e_{ni} - 1) \int \dot{s}(x)\{s(x)/2 + x\dot{s}(x)\} dx] + o(1)$; hence,

$$E(W_\nu) \longrightarrow -\text{Tr}(I_{01}B)/4, \quad (12)$$

where $I_{01} := E\{(\dot{f}(Z_i)/f(Z_i), 1+Z_i\dot{f}(Z_i)/f(Z_i))'(\dot{f}(Z_i)/f(Z_i), 1+Z_i\dot{f}(Z_i)/f(Z_i))'\}$ and $\text{Tr}(\cdot)$ stands for the trace operator. On the other hand, if we define $T_{ni} :=$

$d_{ni}\dot{s}(Z_i)/s(Z_i) + (e_{ni} - 1)\{1/2 + Z_i\dot{s}(Z_i)/s(Z_i)\}$ and $T_\nu := -2\sum_{i=1}^n T_{ni}$, using the central limit theorem in Hájek and Sidák (1967, p. 153), it follows that

$$T_\nu \xrightarrow{d} N(0, \text{Tr}(I_{01}B)); \quad (13)$$

and with a similar reasoning to that used in Lemma VI.2.1.b of Hájek and Sidák (1967) it follows that $\text{Var}(W_\nu - T_\nu) \leq 4\sum_{i=1}^n \int [s\{(x - d_{ni})/e_{ni}\}/e_{ni}^{1/2} - s(x) + d_{ni}\dot{s}(x) + (e_{ni} - 1)\{s(x)/2 + x\dot{s}(x)\}]^2 dx$; hence, using a first-order Taylor expansion as before, we derive that

$$\text{Var}(W_\nu - T_\nu) = o(1). \quad (14)$$

From (12), (13) and (14), it follows that $W_\nu \xrightarrow{d} N(-\text{Tr}(I_{01}B)/4, \text{Tr}(I_{01}B))$; this implies, by LeCam's second lemma, that $\{Z_1, \dots, Z_n\}$ and $\{d_{n1} + e_{n1}Z_1, \dots, d_{nn} + e_{nn}Z_n\}$ are contiguous. Finally, using the same argument as in Hájek and Sidák (1967, p. 219), it readily follows that the assumptions $E\{(1 + Z_i^2)\dot{f}(Z_i)^2/f(Z_i)^2\} > 0$ and $\sum_{i=1}^n (d_{ni}, e_{ni} - 1)'(d_{ni}, e_{ni} - 1) \rightarrow B$ p.d., can be replaced by $E\{(1 + Z_i^2)\dot{f}(Z_i)^2/f(Z_i)^2\} < \infty$ and $\sum_{i=1}^n d_{ni}^2 = O(1)$, $\sum_{i=1}^n (e_{ni} - 1)^2 = O(1)$. ■

Lemma 4: If assumption 2 holds, then $\max_{1 \leq i \leq n_j} |M_{lj}(X_{ji})| = o_p(n_j^{1/2})$.

Proof: Let $\eta > 0$ be arbitrary, then

$$\begin{aligned} P(\max_{1 \leq i \leq n_j} |M_{lj}(X_{ji})| > \eta n_j^{1/2}) &\leq \sum_{i=1}^{n_j} P(|M_{lj}(X_{ji})| > \eta n_j^{1/2}) \\ &\leq n_j E \left[\left(\frac{|M_{lj}(X_{ji})|}{\eta n_j^{1/2}} \right)^2 I\{|M_{lj}(X_{ji})| > \eta n_j^{1/2}\} \right] \\ &= \frac{1}{\eta^2} E [M_{lj}(X_{ji})^2 I\{M_{lj}(X_{ji})^2 > \eta^2 n_j\}]. \end{aligned}$$

The last term is $o(1)$ because $E[M_{lj}(X_{ji})^2] < \infty$ by assumption 2. ■

Proof of Proposition 4: Denote $G_j(y) := \dot{f}_j(y)/f_j(y)$ and define $\tilde{q}_j(y)$, $\tilde{f}_j(y)$, $\tilde{\hat{f}}_j(y)$ and $\tilde{G}_j(y)$ in the same way as $\hat{q}_j(y)$, $\hat{f}_j(y)$, $\hat{\hat{f}}_j(y)$ and $\hat{G}_j(y)$, but replacing residuals $\hat{\varepsilon}_{ji}$ by errors ε_{ji} . We first prove that

$$\int_{-\infty}^{\infty} \|\tilde{q}_j(y) - q_j(y)\|^2 f_j(y) dy = o_p(1). \quad (15)$$

Observe that $\|\tilde{q}_j(y) - q_j(y)\|^2 = (1 + y^2)\{\tilde{G}_j(y) - G_j(y)\}^2$. Let $\bar{f}_j(y)$, $\tilde{f}_j(y)$ be as in Lemma 2; then $\{\tilde{G}_j(y) - G_j(y)\}^2 \leq 3\{\tilde{G}_j(y) - \tilde{G}_j(y)\bar{f}_j(y)^{1/2}/f_j(y)^{1/2}\}^2 + 3\{\tilde{G}_j(y)\bar{f}_j(y)^{1/2}/f_j(y)^{1/2} - \bar{f}_j(y)/[f_j(y)\bar{f}_j(y)]^{1/2}\}^2 + 3\{\bar{f}_j(y)/[f_j(y)\bar{f}_j(y)]^{1/2} - G_j(y)\}^2$. Therefore, if \mathcal{A}_{jn_j} denotes the event $\{y \in \mathbb{R} : |y| \leq a_{jn_j}, \tilde{f}_j(y) \geq b_{jn_j}, |\tilde{f}_j(y)| \leq c_{jn_j}\tilde{f}_j(y)\}$,

$$\begin{aligned} & \int_{-\infty}^{\infty} \|\tilde{q}_j(y) - q_j(y)\|^2 f_j(y) dy \leq \\ & 3 \int_{-\infty}^{\infty} (1 + y^2) \tilde{G}_j(y)^2 \{f_j(y)^{1/2} - \bar{f}_j(y)^{1/2}\}^2 dy + \\ & 3 \int_{\mathcal{A}_{jn_j}} (1 + y^2) \{\tilde{G}_j(y) - \bar{f}_j(y)/\bar{f}_j(y)\}^2 \bar{f}_j(y) dy + \\ & 3 \int_{\mathcal{A}_{jn_j}^c} (1 + y^2) \{\tilde{G}_j(y) - \bar{f}_j(y)/\bar{f}_j(y)\}^2 \bar{f}_j(y) dy + \\ & 3 \int_{-\infty}^{\infty} (1 + y^2) \{\bar{f}_j(y)/\bar{f}_j(y)^{1/2} - \dot{f}_j(y)/f_j(y)^{1/2}\}^2 dy \\ & \equiv 3\{(I) + (II) + (III) + (IV)\}, \text{ say.} \end{aligned}$$

To prove that (I) is $o_p(1)$ note that, with our notation, we can rewrite inequality (6.22) in Bickel (1982) as follows:

$$\{f_j(y)^{1/2} - \bar{f}_j(y)^{1/2}\}^2 \leq \frac{h_{jn_j}^2}{4} \int_0^1 \int_{-\infty}^{\infty} \frac{z^2 \dot{f}_j(y - \lambda h_{jn_j} z)^2}{f_j(y - \lambda h_{jn_j} z)} \varphi(z) dz d\lambda.$$

Hence, as $\tilde{G}_j(y)^2 \leq c_{jn_j}^2$,

$$\begin{aligned} (I) & \leq \frac{(c_{jn_j} h_{jn_j})^2}{4} \int_{-\infty}^{\infty} (1 + y^2) \int_0^1 \int_{-\infty}^{\infty} \frac{z^2 \dot{f}_j(y - \lambda h_{jn_j} z)^2}{f_j(y - \lambda h_{jn_j} z)} \varphi(z) dz d\lambda dy = \\ & \frac{(c_{jn_j} h_{jn_j})^2}{4} \int_{-\infty}^{\infty} \frac{\dot{f}_j(w)^2}{f_j(w)} \int_0^1 \int_{-\infty}^{\infty} (1 + y^2) \left(\frac{y - w}{\lambda h_{jn_j}}\right)^2 \varphi\left(\frac{y - w}{\lambda h_{jn_j}}\right) \frac{1}{\lambda h_{jn_j}} dy d\lambda dw. \end{aligned}$$

Now, with the change of variable $y = w + v\lambda h_{jn_j}$ in the integral with respect to y , and since $1 + (w + v\lambda h_{jn_j})^2 \leq 1 + 2w^2 + 2v^2\lambda^2 h_{jn_j}^2$, it also follows that

$$\begin{aligned} (I) & \leq \frac{(c_{jn_j} h_{jn_j})^2}{4} \left\{ \int_{-\infty}^{\infty} \frac{\dot{f}_j(w)^2}{f_j(w)} (1 + 2w^2) dw \int_0^1 d\lambda \int_{-\infty}^{\infty} v^2 \varphi(v) dv + \right. \\ & \left. 2h_{jn_j}^2 \int_{-\infty}^{\infty} \frac{\dot{f}_j(w)^2}{f_j(w)} dw \int_0^1 \lambda^2 d\lambda \int_{-\infty}^{\infty} v^4 \varphi(v) dv \right\} = o(1)\{O(1) + o(1)O(1)\} = o(1). \end{aligned}$$

As (II) is non-negative, to prove that (II) is $o_p(1)$ it suffices to prove that $E(II) = o(1)$. Note that if $y \in \mathcal{A}_{jn_j}$ then $\tilde{G}_j(y) - \bar{f}_j(y)/\bar{f}_j(y) = \tilde{G}_j(y)\{1 - \tilde{f}_j(y)/\bar{f}_j(y)\} + \{\tilde{f}_j(y) - \bar{f}_j(y)\}/\bar{f}_j(y)$; hence, as $\bar{f}_j(y)$ is non-random,

$$\begin{aligned} E(II) &\leq 2c_{jn_j}^2 \int_{\mathcal{A}_{jn_j}} (1+y^2) E[\{\bar{f}_j(y) - \tilde{f}_j(y)\}^2] / \bar{f}_j(y) dy + \\ &2 \int_{\mathcal{A}_{jn_j}} (1+y^2) E[\{\tilde{f}_j(y) - \bar{f}_j(y)\}^2] / \bar{f}_j(y) dy \leq \\ &2 \frac{1}{n_j h_{jn_j}} (c_{jn_j}^2 \kappa_0 + \frac{\kappa_1}{h_{jn_j}^2}) \int_{-\infty}^{\infty} (1+y^2) I(y \in \mathcal{A}_{jn_j}) dy, \end{aligned}$$

where the last inequality follows because $E[\{\bar{f}_j(y) - \tilde{f}_j(y)\}^2] = \text{Var}\{\tilde{f}_j(y)\} \leq E[\varphi\{(y - \varepsilon_{ij})/h_{jn_j}\}^2] / (n_j h_{jn_j}^2) \leq \kappa_0 \bar{f}_j(y) / (n_j h_{jn_j})$, for $\kappa_0 := \sup_{x \in \mathbb{R}} \varphi(x)$, and similarly $E[\{\tilde{f}_j(y) - \bar{f}_j(y)\}^2] \leq \kappa_1 \bar{f}_j(y) / (n_j h_{jn_j}^3)$, for $\kappa_1 := \sup_{x \in \mathbb{R}} \dot{\varphi}(x)^2 / \varphi(x)$. If $y \in \mathcal{A}_{jn_j}$, then $|y| \leq a_{jn_j}$; thus, the integral in the latter expression is bounded by $\int_{-a_{jn_j}}^{a_{jn_j}} (1+y^2) dy$, and hence

$$E(II) \leq 4 \frac{1}{n_j h_{jn_j}} (c_{jn_j}^2 \kappa_0 + \frac{\kappa_1}{h_{jn_j}^2}) (a_{jn_j} + \frac{a_{jn_j}^3}{3}) = o(1).$$

To prove that (III) is $o_p(1)$ it suffices to prove that $E(III) = o(1)$. Note that

$$E(III) = \int_{-\infty}^{\infty} (1+y^2) \bar{f}_j(y)^2 / \bar{f}_j(y) P(y \in \mathcal{A}_{jn_j}^C) dy.$$

Under our assumptions, for fixed y it follows readily that $(1+y^2) \bar{f}_j(y)^2 / \bar{f}_j(y) \rightarrow (1+y^2) \dot{f}_j(y)^2 / f_j(y)$. Moreover, $\tilde{f}_j(y) \xrightarrow{p} f_j(y)$, $\dot{\tilde{f}}_j(y) \xrightarrow{p} \dot{f}_j(y)$; hence, $P(y \in \mathcal{A}_{jn_j}^C) \leq P\{\tilde{f}_j(y) < b_{jn_j}\} + I(|y| > a_{jn_j}) + P\{|\dot{\tilde{f}}_j(y)| > c_{jn_j} \tilde{f}_j(y)\} = o(1)$. Additionally, (11) ensures that the uniform integrability results apply, and hence $E(III) = o(1)$. Finally note that, for fixed y , $(1+y^2) \{\bar{f}_j(y)/\bar{f}_j(y)^{1/2} - \dot{f}_j(y)/f_j(y)^{1/2}\}^2 \rightarrow 0$; now (11) and assumption 1 ensure that the uniform integrability results apply and hence (IV) = $o(1)$. Thus, the proof of (15) is now complete.

To prove that (3) follows from (15), we use the contiguity result derived in Lemma 3. Define now $d_{jn_j i}(u) := \{\mu_j(X_{ji}, \theta_j) - \mu_j(X_{ji}, u)\} / \sigma_j(X_{ji}, u)$ and $e_{jn_j i}(u) := \sigma_j(X_{ji}, \theta_j) / \sigma_j(X_{ji}, u)$. From assumption 2 and Lemma 4 we obtain

$$\sup_u \max_{1 \leq i \leq n_j} d_{jn_j i}^2(u) \leq (M^{2\alpha_{1j}} / S_j^2) \max_{1 \leq i \leq n_j} M_{1j}(X_{ji})^2 / n_j^{\alpha_{1j}} = O(1) o_p(n_j^{1-\alpha_{1j}}) = o_p(1),$$

where, as above, \sup_u denotes the supremum for $\|u - \theta_j\| \leq Mn_j^{-1/2}$, for a fixed M . Similarly,

$$\sup_u \max_{1 \leq i \leq n_j} \{e_{jn_j i}(u) - 1\}^2 \leq (M^{2\alpha_{2j}} / S_j^2) \max_{1 \leq i \leq n_j} M_{2j}(X_{ji})^2 / n_j^{\alpha_{2j}} = o_p(1).$$

On the other hand,

$$\sup_u \sum_{i=1}^{n_j} d_{jn_j i}^2(u) \leq \{n_j^{-1} \sum_{i=1}^{n_j} M_{1j}(X_{ji})^2\} M^{2\alpha_{1j}} n_j^{1-\alpha_{1j}} / S_j^2 = O_p(1) O(1) = O_p(1),$$

and similarly $\sup_u \sum_{i=1}^{n_j} \{e_{jn_j i}(u) - 1\}^2 = O_p(1)$. Then, taking into account the root- n_j -consistency of $\widehat{\theta}_j$, the result derived in Lemma 3 and the relationship between ε_{jn_j} and $\widehat{\varepsilon}_{jn_j}$, using a similar argument to that in Bickel (1982, p. 657), it follows that the measure induced by $\{\widehat{\varepsilon}_{j1}, \dots, \widehat{\varepsilon}_{jn_j}\}$ is contiguous to the measure induced by $\{\varepsilon_{j1}, \dots, \varepsilon_{jn_j}\}$; therefore, (3) follows from (15). ■

Proof of Proposition 5: Let $\widetilde{q}_j(y)$, $\widetilde{f}_j(y)$, $\widetilde{\dot{f}}_j(y)$ and $\widetilde{G}_j(y)$ be as defined in Proposition 4. Reasoning as in the proof of Proposition 4, we only have to show:

$$\sup_{x \in \mathbb{R}} \left\| n_j^{-1/2} \sum_{i=1}^{n_j} [I(\varepsilon_{ji} \geq x) \{\widetilde{q}_j(\varepsilon_{ji}) - q_j(\varepsilon_{ji})\}] - \int_x^\infty \{\widetilde{q}_j(y) - q_j(y)\} f_j(y) dy \right\| = o_p(1).$$

To keep the proof readable we only present the line of argument for the second component of $\widetilde{q}_j(y) - q_j(y)$, that is

$$\frac{\widetilde{\dot{f}}_j(y)}{\widetilde{f}_j(y)} - \frac{\dot{f}_j(y)}{f_j(y)} = \frac{\widetilde{\dot{f}}_j(y) f_j(y) - \dot{f}_j(y) \widetilde{f}_j(y)}{f_j(y)^2} (1 + r_{jn_j}(y)),$$

where $r_{jn_j}(y) = (f_j(y) - \tilde{f}_j(y))/\tilde{f}_j(y)$ and we assume in the following $|y| \leq a_{jn_j}$, $\tilde{f}_j(y) \geq b_{jn_j}$, $|\tilde{f}_j(y)| \leq c_{jn_j}\tilde{f}_j(y)$ (i.e. $y \in \mathcal{A}_{jn_j}$ with the definition from the proof of Proposition 4, where $P(y \in \mathcal{A}_{jn_j}^C) = o(1)$ was shown). For the remainder we have uniformly in $y \in \mathbb{R}$,

$$|r_{jn_j}(y)| = O\left(\frac{h_{jn_j}^2}{b_{jn_j}}\right) + O\left(\frac{(\log h_{jn_j}^{-1})^{1/2}}{(n_j h_{jn_j})^{1/2} b_{jn_j}}\right) = o(1) \quad (16)$$

(see e.g. Silverman, 1978, for uniform rates of kernel density estimators). Thus, in the following we concentrate on proving $\sup_{x \in \mathbb{R}} |V_{jn_j}(x)| = o_p(1)$, where

$$\begin{aligned} V_{jn_j}(x) &= n_j^{-1/2} \sum_{i=1}^{n_j} \left[I(\varepsilon_{ji} \geq x) \frac{\dot{f}_j(\varepsilon_{ji}) f_j(\varepsilon_{ji}) - \dot{\tilde{f}}_j(\varepsilon_{ji}) \tilde{f}_j(\varepsilon_{ji})}{f_j^2(\varepsilon_{ji})} I(\tilde{f}_j(\varepsilon_{ji}) \geq b_{jn_j}) \right. \\ &\quad \left. - \int_x^\infty \frac{\tilde{f}_j(y) f_j(y) - \dot{\tilde{f}}_j(y) \tilde{f}_j(y)}{f_j^2(y)} I(\tilde{f}_j(y) \geq b_{jn_j}) f_j(y) dy \right] \\ &= \frac{1}{n_j \sqrt{n_j}} \sum_{i=1}^{n_j} \sum_{k=1}^{n_j} \left[I(\varepsilon_{ji} \geq x) \left(\frac{1}{h_{jn_j}^2} \dot{\varphi}\left(\frac{\varepsilon_{ji} - \varepsilon_{jk}}{h_{jn_j}}\right) \right. \right. \\ &\quad \left. \left. - \frac{1}{h_{jn_j}} \varphi\left(\frac{\varepsilon_{ji} - \varepsilon_{jk}}{h_{jn_j}}\right) \frac{\dot{f}_j(\varepsilon_{ji})}{f_j(\varepsilon_{ji})} \right) \frac{1}{f_j(\varepsilon_{ji})} I(\tilde{f}_j(\varepsilon_{ji}) \geq b_{jn_j}) \right. \\ &\quad \left. - \int I(y \geq x) \left(\frac{1}{h_{jn_j}^2} \dot{\varphi}\left(\frac{y - \varepsilon_{jk}}{h_{jn_j}}\right) - \frac{1}{h_{jn_j}} \varphi\left(\frac{y - \varepsilon_{jk}}{h_{jn_j}}\right) \frac{\dot{f}_j(y)}{f_j(y)} \right) \right. \\ &\quad \left. \times I(\tilde{f}_j(y) \geq b_{jn_j}) dy \right]. \end{aligned}$$

With some straightforward calculations one can show $E\{V_{jn_j}(x)\} = o(1)$ and $\text{Var}(V_{jn_j}(x)) = o(1)$ for each fixed x . However, to derive the result uniformly in $x \in \mathbb{R}$ we will use empirical process and U-process theory.

The indicators $I(\tilde{f}_j \geq b_{jn_j})$ assure the existence of expectations. Note that $f_j \geq \tilde{f}_j - |\tilde{f}_j - f_j|$. The term in absolute value is uniformly almost surely of order $o(b_{jn_j})$ (compare (16)). Therefore on the event $\{\tilde{f}_j \geq b_{jn_j}\}$ we can assume $f_j \geq b_{jn_j}/2$ and obtain, e. g., for the expectation of the term appearing in the above double sum for the special case $i = k$,

$$\frac{1}{\sqrt{n_j}} E \left[\left| \left(\frac{\dot{\varphi}(0)}{h_{jn_j}^2} - \frac{\varphi(0) \dot{f}_j(\varepsilon_{ji})}{h_{jn_j} f_j(\varepsilon_{ji})} \right) \frac{1}{f_j(\varepsilon_{ji})} I(\tilde{f}_j(\varepsilon_{ji}) \geq b_{jn_j}) \right| \right] \leq O\left(\frac{1}{n_j^{1/2} h_{jn_j}^2 b_{jn_j}}\right) = o(1).$$

For ease of notation in the following we omit the indicators and implicitly assume that we only consider sets where $\tilde{f}_j \geq b_{jn_j}$ and $f_j \geq b_{jn_j}/2$ is valid. We concentrate our considerations on the following (symmetrized) U-process (compare the definition of $V_{jn_j}(x)$),

$$\frac{1}{2h_{jn_j}^2} U_{jn_j}(g_{x,h_{jn_j}}) = \frac{1}{2h_{jn_j}^2} \frac{1}{n_j \sqrt{n_j}} \sum_{i=1}^{n_j} \sum_{\substack{k=1 \\ k \neq i}}^{n_j} g_{x,h_{jn_j}}(\varepsilon_{ji}, \varepsilon_{jk}),$$

where

$$\begin{aligned} g_{x,h}(u, v) &= I(u \geq x) \left(\dot{\varphi}\left(\frac{u-v}{h}\right) - h\varphi\left(\frac{u-v}{h}\right) \frac{\dot{f}_j(u)}{f_j(u)} \right) \frac{1}{f_j(u)} \\ &\quad - \int I(y \geq x) \left(\dot{\varphi}\left(\frac{y-v}{h}\right) - h\varphi\left(\frac{y-v}{h}\right) \frac{\dot{f}_j(y)}{f_j(y)} \right) dy \\ &\quad + I(v \geq x) \left(\dot{\varphi}\left(\frac{v-u}{h}\right) - h\varphi\left(\frac{v-u}{h}\right) \frac{\dot{f}_j(v)}{f_j(v)} \right) \frac{1}{f_j(v)} \\ &\quad - \int I(y \geq x) \left(\dot{\varphi}\left(\frac{y-u}{h}\right) - h\varphi\left(\frac{y-u}{h}\right) \frac{\dot{f}_j(y)}{f_j(y)} \right) dy. \end{aligned}$$

We denote by \mathcal{G} the function class $\{g_{x,h} | x \in \mathbb{R}, h \in (0, 1)\}$ and by $\tilde{\mathcal{G}}$ the class of conditional expectations $\{\tilde{g}_{x,h} | x \in \mathbb{R}, h \in (0, 1)\}$, where

$$\begin{aligned} \tilde{g}_{x,h}(v) &= E[g_{x,h}(\varepsilon_{ji}, v)] \\ &= I(v \geq x) \int \left(\dot{\varphi}\left(\frac{v-u}{h}\right) - h\varphi\left(\frac{v-u}{h}\right) \frac{\dot{f}_j(v)}{f_j(v)} \right) \frac{1}{f_j(v)} f_j(u) du \\ &\quad - \int \int I(y \geq x) \left(\dot{\varphi}\left(\frac{y-u}{h}\right) - h\varphi\left(\frac{y-u}{h}\right) \frac{\dot{f}_j(y)}{f_j(y)} \right) dy f_j(u) du. \end{aligned}$$

Hoeffding's decomposition gives $U_{jn_j}(g_{x,h}) = U_{jn_j}(\bar{g}_{x,h}) + 2n_j^{-1/2} \sum_{i=1}^{n_j} \tilde{g}_{x,h}(\varepsilon_{ji})$, where $\bar{g}_{x,h}(u, v) = g_{x,h}(u, v) - \tilde{g}_{x,h}(u) - \tilde{g}_{x,h}(v)$ and $U_{jn_j}(\bar{g}_{x,h})$ is a degenerate U-process. With some tedious calculations using Lemma 22(ii), Corollary 17 about sums of Euclidean classes, a similar result on products of bounded Euclidean classes, and Corollary 21 of Nolan and Pollard (1987) it can be shown that the function classes \mathcal{G} and $\tilde{\mathcal{G}}$ are Euclidean (see Definition 8 in Nolan and Pollard, 1987, p. 789). Then it is easy to see that the assumptions of Theorem 5 in Nolan and Pollard (1988, p. 1294) are satisfied and from the proof of this Theorem it

can be concluded that $\sup_{x \in \mathbb{R}, h \in (0,1)} |U_{jn_j}(\bar{g}_{x,h})| = O_p(n_j^{-1/2})$ (compare the proof of Lemma 5.3.a, Neumeyer and Dette (2003) for a similar argument). Therefore we obtain

$$\frac{1}{2h_{jn_j}^2} \sup_{x \in \mathbb{R}} |U_{jn_j}(\bar{g}_{x,h_{jn_j}})| = O_p(n_j^{-1/2} h_{jn_j}^{-2}) = o_p(1)$$

and in the following we only have to consider the empirical process part of Hoeffding's decomposition, that is

$$n_j^{-1/2} h_{jn_j}^{-2} \sum_{i=1}^{n_j} \tilde{g}_{x,h_{jn_j}}(\varepsilon_{ji}) = h_{jn_j}^{-2} \sqrt{n_j} \left(P_{jn_j}(\tilde{g}_{x,h_{jn_j}}) - P(\tilde{g}_{x,h_{jn_j}}) \right),$$

where P and P_{jn_j} denote the measure and empirical measure of ε_{ji} , respectively. The rest of our proof uses ideas from Theorem 37 in Pollard (1984, p. 34). We cannot apply this theorem directly because our function class $\tilde{\mathcal{G}}$ has no constant envelope. For the complete line of arguments we refer to the proof of Lemma 5.3.a in Neumeyer and Dette (2003). Note that the function class $\tilde{\mathcal{G}}$ is Euclidean. We define the following sequences of real numbers, $\alpha_{jn_j} = n_j^{-1/2} h_{jn_j}^{-2-2d_j} b_{jn_j}$ and $\delta_{jn_j}^2 = h_{jn_j}^{4+2d_j} b_{jn_j}^{-1}$ (using d_j from assumption 7) such that

$$\begin{aligned} & P(\tilde{g}_{x,h_{jn_j}}^2) \\ & \leq E \left[\left(I(\varepsilon_{ji} \geq x) \int \left(\dot{\varphi}\left(\frac{\varepsilon_{ji} - u}{h_{jn_j}}\right) - h_{jn_j} \varphi\left(\frac{\varepsilon_{ji} - u}{h_{jn_j}}\right) \frac{\dot{f}_j(\varepsilon_{ji})}{f_j(\varepsilon_{ji})} \right) \frac{1}{f_j(\varepsilon_{ji})} f_j(u) du \right)^2 \right] \\ & = h_{jn_j}^4 \int_x^\infty \left(\int \frac{1}{h_{jn_j}^2} \dot{\varphi}\left(\frac{y - u}{h_{jn_j}}\right) f_j(u) du - \int \frac{1}{h_{jn_j}} \varphi\left(\frac{y - u}{h_{jn_j}}\right) f_j(u) du \frac{\dot{f}_j(y)}{f_j(y)} \right)^2 \frac{dy}{f_j(y)} \\ & \leq h_{jn_j}^4 \int_{-\infty}^\infty \left(\int \dot{f}_j(y - h_{jn_j} z) \varphi(z) dz - \int f_j(y - h_{jn_j} z) \varphi(z) dz \frac{\dot{f}_j(y)}{f_j(y)} \right)^2 \frac{dy}{f_j(y)} \\ & = o(\delta_{jn_j}^2) \end{aligned}$$

uniformly in $x \in \mathbb{R}$ (using our implicit assumption $f_j \geq b_{jn_j}/2$). Additionally, we have $\log n_j = o(n_j \delta_{jn_j}^2 \alpha_{jn_j}^2)$ and $n_j \delta_{jn_j}^2 \alpha_{jn_j} \rightarrow \infty$. Using the argument in Lemma 5.3.a, Neumeyer and Dette (2003), we obtain

$$\sup_{x \in \mathbb{R}} |P_{jn_j}(\tilde{g}_{x,h_{jn_j}}) - P(\tilde{g}_{x,h_{jn_j}})| = o_p(\delta_{jn_j}^2 \alpha_{jn_j})$$

and conclude the assertion from $h_{jn_j}^{-2} \sqrt{n_j} o(\delta_{jn_j}^2 \alpha_{jn_j}) = o(1)$. ■

Proof of Theorem 3: First we prove that under H_0 $\widehat{W}_{jn_j}(\cdot) := \widehat{W}_{jn_j}\{F_j^{-1}(\cdot)\}$ converges weakly to a standard Brownian motion. Let $D[0, b]$ ($b > 0$) denote the space of cadlag functions endowed with the Skorohod metric, and define the mapping $\Gamma_j(\cdot)$ from $D[0, 1]$ to $D[0, 1]$ as follows:

$$\Gamma_j\{\gamma(\cdot)\}(t) := \int_0^t q_j\{F_j^{-1}(s)\}' C_j\{F_j^{-1}(s)\}^{-1} \left[\int_s^1 q_j\{F_j^{-1}(r)\} d\gamma(r) \right] ds.$$

It is easy to check that $\Gamma_j(\cdot)$ is a linear mapping. Note that $q_j\{F_j^{-1}(\cdot)\}$ is the derivative of $Q_j(\cdot) := (Q_{1j}(\cdot), Q_{2j}(\cdot), Q_{3j}(\cdot))'$, where $Q_{1j}(t) := t$, $Q_{2j}(t) := f_j\{F_j^{-1}(t)\}$, $Q_{3j}(t) := F_j^{-1}(t) f_j\{F_j^{-1}(t)\}$. With this notation, observe that

$$\Gamma_j\{Q_{lj}(\cdot)\} = Q_{lj}(\cdot), \quad (17)$$

for $l = 1, 2, 3$; this follows because $C_j\{F_j^{-1}(s)\}^{-1} C_j\{F_j^{-1}(s)\} = \mathbf{I}_3$ implies that $C_j\{F_j^{-1}(s)\}^{-1} \int_s^1 \dot{Q}_j(r) dQ_{1j}(r) = (1, 0, 0)'$; hence $\Gamma_j\{Q_{1j}(\cdot)\}(t) = \int_0^t \dot{Q}_j(s) (1, 0, 0)' ds = Q_{1j}(t)$, and a similar reasoning applies to $Q_{2j}(\cdot)$, $Q_{3j}(\cdot)$. Using now the relationship $\widehat{V}_{jn_j}(\cdot) = n^{1/2}[\widehat{F}_{jn_j}\{F_j^{-1}(\cdot)\} - Q_{1j}(\cdot)]$, the linearity of $\Gamma_j(\cdot)$ and (17), routine calculations yield that, under H_0 , $\widehat{V}_{jn_j}(\cdot) - \Gamma_j\{\widehat{V}_{jn_j}(\cdot)\} = \widehat{W}_{jn_j}(\cdot)$. Using (2) it also follows from there that

$$\widehat{W}_{jn_j}(\cdot) = V_{jn_j}(\cdot) - \Gamma_j\{V_{jn_j}(\cdot)\} + o_p(1),$$

because $\Gamma_j\{g_j(\cdot)' \xi_{jn_j}\} = g_j(\cdot)' \xi_{jn_j}$ by (17). Thus, as $V_{jn_j}(\cdot)$ converges weakly to a standard Brownian bridge $V(\cdot)$, $\widehat{W}_{jn_j}(\cdot)$ converges weakly to $V(\cdot) - \Gamma_j\{V(\cdot)\}$, which is a standard Brownian motion in $[0, 1]$ (see Khmaladze, 1981).

Now define $\widetilde{W}_{jn_j}(\cdot) := \widetilde{W}_{jn_j}\{F_j^{-1}(\cdot)\}$. To prove that $\widetilde{W}_{jn_j}(\cdot) = \widehat{W}_{jn_j}(\cdot) + o_p(1)$, we follow a similar approach to that used in Theorem 4 of Bai (2003), though now some additional terms turn up because the estimated functions $\widehat{f}_j(\cdot)$ also appear

in the integrals to be computed. Observe that for every $t_0 \in (0, 1)$,

$$\begin{aligned}
& \sup_{t \in [0, t_0]} \left| \widetilde{W}_{jn_j}(t) - \widehat{W}_{jn_j}(t) \right| \leq \\
& \sup_{x \in (-\infty, F_j^{-1}(t_0))} n_j^{1/2} \left| \int_{-\infty}^x \widehat{q}_{3-j}(y)' \widehat{C}_{3-j}(y)^{-1} \left\{ \int_y^\infty \widehat{q}_{3-j}(z) d\widehat{F}_{jn_j}(z) \right\} \right. \\
& \quad \left. \{ \widehat{f}_{3-j}(y) - f_{3-j}(y) \} dy \right| + \\
& \sup_{x \in (-\infty, F_j^{-1}(t_0))} n_j^{1/2} \left| \int_{-\infty}^x \widehat{q}_{3-j}(y)' \{ \widehat{C}_{3-j}(y)^{-1} - \overline{C}_{3-j}(y)^{-1} \} \right. \\
& \quad \left. \{ \int_y^\infty \widehat{q}_{3-j}(z) d\widehat{F}_{jn_j}(z) \} f_{3-j}(y) dy \right| + \\
& \sup_{x \in (-\infty, F_j^{-1}(t_0))} n_j^{1/2} \left| \int_{-\infty}^x [\widehat{q}_{3-j}(y)' \overline{C}_{3-j}(y)^{-1} \{ \int_y^\infty \widehat{q}_{3-j}(z) d\widehat{F}_{jn_j}(z) \} - \right. \\
& \quad \left. q_{3-j}(y)' C_{3-j}(y)^{-1} \{ \int_y^\infty q_{3-j}(z) d\widehat{F}_{jn_j}(z) \}] f_{3-j}(y) dy \right| \\
& \equiv (I) + (II) + (III), \text{ say,}
\end{aligned}$$

where we define $\overline{C}_j(y) := \int_y^\infty \widehat{q}_j(w) \widehat{q}_j(w)' f_j(w) dw$. We prove below that (I), (II) and (III) are all $o_p(1)$. Hence, under H_0 , $\widetilde{W}_{jn_j}(\cdot)$ also converges weakly to a Brownian motion $W^{(1)}(\cdot)$ in the space $D[0, t_0]$. Thus, $\overline{K}_{n_1, n_2, x_0}^{(j)} = \widehat{F}_{jn_j}(x_0)^{-1/2} \sup_{t \in [0, F_j(x_0)]} \left| \widetilde{W}_{jn_j}(t) \right|$ converges in distribution to $F_j(x_0)^{-1/2} \sup_{t \in [0, F_j(x_0)]} |W^{(1)}(t)| = \sup_{t \in [0, 1]} |W(t)|$, where $W(t) := F_j(x_0)^{-1/2} W^{(1)}\{F_j(x_0)t\}$ is a Brownian motion in the space $D[0, 1]$. On the other hand, reasoning as in Theorem 1, it follows that $\overline{C}_{n_1, n_2, x_0}^{(j)} = \widehat{F}_{jn_j}(x_0)^{-2} \int I(x \leq x_0) \{ \widetilde{W}_{jn_j}(x) \}^2 dF_j(x) + o_p(1)$; thus $\overline{C}_{n_1, n_2, x_0}^{(j)}$ converges in distribution to $F_j(x_0)^{-2} \int_0^{F_j(x_0)} \{W^{(1)}(t)\}^2 dt = \int_0^1 \{W(t)\}^2 dt$.

Therefore, to derive the result under H_0 it only remains to prove that (I), (II) and (III) are $o_p(1)$. First, we prove that (III) is $o_p(1)$. To simplify notation, hereafter we drop the argument s and denote $\widehat{q}_j \equiv \widehat{q}_j\{F_j^{-1}(s)\}$, $\overline{C}_j \equiv \overline{C}_j\{F_j^{-1}(s)\}$, $q_j \equiv q_j\{F_j^{-1}(s)\}$, $C_j \equiv C_j\{F_j^{-1}(s)\}$. Note that the first column in \overline{C}_j is $\int_s^1 \widehat{q}_j\{F_j^{-1}(r)\} dr$; hence $\overline{C}_j^{-1} \overline{C}_j = \mathbf{I}_3$ implies that $\overline{C}_j^{-1} \int_s^1 \widehat{q}_j\{F_j^{-1}(r)\} dr = (1, 0, 0)'$. Thus, $\int_0^t \widehat{q}_j' \overline{C}_j^{-1} [\int_s^1 \widehat{q}_j\{F_j^{-1}(r)\} dr] ds = t$; on the other hand, (17) for $l = 1$ can be rewritten as $\int_0^t q_j' C_j^{-1} [\int_s^1 q_j\{F_j^{-1}(r)\} dr] ds = t$. These two equalities

imply that (III) can also be expressed as

$$\begin{aligned}
& \sup_{t \in [0, t_0]} \left| \int_0^t \left[\widehat{q}_{3-j} \overline{C}_{3-j}^{-1} \int_s^1 \widehat{q}_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) - \right. \right. \\
& \quad \left. \left. q'_{3-j} C_{3-j}^{-1} \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right] ds \right| \leq \\
& \sup_{t \in [0, t_0]} \int_0^t \left| \widehat{q}_{3-j} \overline{C}_{3-j}^{-1} \int_s^1 (\widehat{q}_{3-j} \{F_j^{-1}(r)\} - q_{3-j} \{F_j^{-1}(r)\}) d\widehat{V}_{jn_j}(r) \right| ds + \\
& \sup_{t \in [0, t_0]} \int_0^t \left| \widehat{q}_{3-j} (\overline{C}_{3-j}^{-1} - C_{3-j}^{-1}) \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right| ds + \\
& \sup_{t \in [0, t_0]} \int_0^t \left| (\widehat{q}_{3-j} - q_{3-j})' C_{3-j}^{-1} \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right| ds.
\end{aligned} \tag{18}$$

Now observe that, for all s in $(0, t_0)$, from (2) we derive that

$$\begin{aligned}
& \left| \widehat{q}_{3-j} \overline{C}_{3-j}^{-1} \int_s^1 (\widehat{q}_{3-j} \{F_j^{-1}(r)\} - q_{3-j} \{F_j^{-1}(r)\}) d\widehat{V}_{jn_j}(r) \right| \leq \|\widehat{q}_{3-j}\| \times \\
& \left\| \overline{C}_{3-j} \{F_j^{-1}(t_0)\}^{-1} \right\| \left(\left\| \int_s^1 [\widehat{q}_{3-j} \{F_j^{-1}(r)\} - q_{3-j} \{F_j^{-1}(r)\}] dV_{jn_j}(r) \right\| + \right. \\
& \quad \left. \left[\int_s^1 \|\widehat{q}_j \{F_j^{-1}(r)\} - q_j \{F_j^{-1}(r)\}\|^2 dr \right]^{1/2} \right) \\
& \left(\left[\int_s^1 \|\dot{g}_j(r)\|^2 dr \right]^{1/2} O_p(1) + o_p(1) \right) = \|\widehat{q}_{3-j}\| O_p(1) o_p(1),
\end{aligned} \tag{19}$$

where the last equality follows from (4) and (3). Similarly,

$$\begin{aligned}
& \left| \widehat{q}_{3-j} (\overline{C}_{3-j}^{-1} - C_{3-j}^{-1}) \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right| \leq \\
& \|\widehat{q}_{3-j}\| \left\| \overline{C}_{3-j}^{-1} - C_{3-j}^{-1} \right\| \left\| \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right\| \leq \\
& \|\widehat{q}_{3-j}\| \left\| \overline{C}_{3-j}^{-1} - C_{3-j}^{-1} \right\| \left[\left\| \int_s^1 q_{3-j} \{F_j^{-1}(r)\} dV_{jn_j}(r) \right\| + O_p(1) \right] \\
& = \|\widehat{q}_{3-j}\| o_p(1) O_p(1),
\end{aligned} \tag{20}$$

where the second inequality follows using (2) and reasoning as in (19), and the last equality follows because by the functional central limit theorem $\int_s^1 q_{3-j} \{F_j^{-1}(r)\} dV_{jn_j}(r) = O_p(1)$ and, using the same argument as in Bai (2003, p. 548), it follows that $\left\| \overline{C}_{3-j}^{-1} - C_{3-j}^{-1} \right\| = o_p(1)$, uniformly in s . Finally,

$$\begin{aligned}
& \left| (\widehat{q}_{3-j} - q_{3-j})' C_{3-j}^{-1} \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right| \leq \\
& \|\widehat{q}_{3-j} - q_{3-j}\| \left\| C_{3-j} \{F_j^{-1}(t_0)\}^{-1} \right\| \left\| \int_s^1 q_{3-j} \{F_j^{-1}(r)\} d\widehat{V}_{jn_j}(r) \right\| \\
& = \|\widehat{q}_{3-j} - q_{3-j}\| O_p(1) O_p(1).
\end{aligned} \tag{21}$$

Thus, from (18), (19), (20) and (21), and using (3) it follows that

$$(III) \leq o_p(1) \left(\int_0^1 \|\widehat{q}_{3-j}\|^2 ds \right)^{1/2} + O_p(1) \left(\int_0^1 \|\widehat{q}_{3-j} - q_{3-j}\|^2 ds \right)^{1/2} = o_p(1).$$

To analyze (I) and (II) observe that, under our assumptions, using the results in Silverman (1978) and Lemma 3 in the same way as in Proposition 4, it follows that $\sup_{y \in \mathbb{R}} \left| \widehat{f}_j(y) - f_j(y) \right| = o_p(1)$, which implies that $\sup_{y \in (-\infty, x_0]} \left| \widehat{C}_j(y)^{-1} - \overline{C}_j(y)^{-1} \right| = o_p(1)$. Thus, using the same arguments as above it also follows that (I) = $o_p(1)$ and (II) = $o_p(1)$. This completes the proof of part **a**.

Under H_1 , note that the probability limit of $n_j^{-1/2} \widetilde{W}_{jn_j}(x)$ is

$$\Xi(x) := F_j(x) - \int_{-\infty}^x q_{3-j}(y)' C_{3-j}(y)^{-1} \left\{ \int_y^{\infty} q_{3-j}(z) f_j(z) dz \right\} f_{3-j}(y) dy.$$

Assume that $\Xi(x) = 0$ for every $x \in \mathbb{R}$. Then $\dot{\Xi}(x) = 0$, i.e.,

$$f_j(x) - q_{3-j}(x)' \Upsilon(x) f_{3-j}(x) = 0, \quad (22)$$

where $\Upsilon(x) := C_{3-j}(x)^{-1} \left\{ \int_x^{\infty} q_{3-j}(z) f_j(z) dz \right\}$; but using the rules for matrix derivatives it follows that $\dot{\Upsilon}(x) = -C_{3-j}(x)^{-1} \dot{C}_{3-j}(x) C_{3-j}(x)^{-1} \left\{ \int_x^{\infty} q_{3-j}(z) f_j(z) dz \right\} - C_{3-j}(x)^{-1} q_{3-j}(x) f_j(x) = -C_{3-j}(x)^{-1} q_{3-j}(x) \dot{\Xi}(x) = 0$ and hence $\Upsilon(\cdot)$ is constant, say $(\lambda_1, \lambda_2, \lambda_3)'$. Therefore, (22) implies that $f_j(x) = (\lambda_1 + \lambda_3) f_{3-j}(x) + \lambda_2 \dot{f}_{3-j}(x) + \lambda_3 x \dot{f}_{3-j}(x)$. Now, if we integrate the two terms in this equation, and also the two terms premultiplied by x and by x^2 , taking into account that $E(\varepsilon_{ji}) = 0$, $E(\varepsilon_{ji}^2) = 1$ and $E(\varepsilon_{ji}^3) < \infty$, we derive three equations which imply that $\lambda_1 = 1$, $\lambda_2 = \lambda_3 = 0$. This proves that if $\Xi(x) = 0$ for every $x \in \mathbb{R}$, then H_0 holds. Hence, under H_1 there exists x_* such that $\Xi(x_*) \neq 0$; thus, if $x_0 \geq x_*$, $n_j^{-1/2} \overline{K}_{n_1, n_2, x_0}^{(j)}$ converges in probability to $F_j(x_0)^{-1/2} \sup_{x \in (-\infty, x_0]} |\Xi(x)| > 0$ and $n_j^{-1/2} \overline{C}_{n_1, n_2, x_0}^{(j)}$ converges in probability to $F_j(x_0)^{-2} \int_{-\infty}^{x_0} \Xi(x)^2 dF_j(x) > 0$, and part **b** follows from there. ■

Proof of the Corollary: Let $\mathbf{W}(\cdot) := (W_1(\cdot), W_2(\cdot))'$ be a Gaussian process on $D[0, 1] \times D[0, 1]$ with zero mean vector and $\text{Cov}\{\mathbf{W}(s), \mathbf{W}(t)\} = \min(s, t) \mathbf{I}_2$, where \mathbf{I}_2 denotes the identity matrix. As $\widehat{W}_{1n_1}(\cdot)$ and $\widehat{W}_{2n_2}(\cdot)$ are constructed with independent random samples, it follows from the proof of Theorem 3 that under H_0 $\widehat{\mathbf{W}}_n(\cdot) := (\widehat{W}_{1n_1}(\cdot), \widehat{W}_{2n_2}(\cdot))'$ converges weakly to $\mathbf{W}(\cdot)$. Thus, the result

in part **a** follows because $\sup_{t \in [0,1]} \left\| \widehat{\mathbf{W}}_n(t) - \widetilde{\mathbf{W}}_n(t) \right\| = o_p(1)$ where $\widetilde{\mathbf{W}}_n(\cdot) := (\widetilde{W}_{1n_1}(\cdot), \widetilde{W}_{2n_2}(\cdot))'$, and the result in part **b** follows from part **b** of Theorem 3. ■

REFERENCES

- Akritis, M. and I. Van Keilegom (2001) Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics* **28**, 549-567.
- Bai, J. (1994) Weak convergence of sequential empirical processes of residual in ARMA models. *Annals of Statistics* **22**, 2051-2061.
- Bai, J. (2003) Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* **85**, 531-549.
- Bai, J. and S. Ng (2001) A consistent test for conditional symmetry of time series models. *Journal of Econometrics* **103**, 225-258.
- Bickel, P.J. (1982) On adaptive estimation. *Annals of Statistics* **10**, 647-671.
- Delgado, M.A. and J. Mora (2000) A nonparametric test for serial independence of regression errors. *Biometrika* **87**, 228-234.
- González-Manteiga, W. and M.A. Delgado (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* **29**, 1469-1507.
- Hájek, J. and Z. Sidák (1967) *Theory of rank tests*. New York: Academic Press.
- Khmaladze, E.V. (1981) Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and its Applications* **26**, 240-257.
- Khmaladze, E.V. (1993) Goodness of fit problem and scanning innovation martingales. *Annals of Statistics* **21**, 798-829.
- Khmaladze, E.V. and H.L. Koul (2004) Martingale transforms of goodness-of-fit tests in regression models. *Annals of Statistics* **32**, 995-1034.

- Koul, H.L. (1996) Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Annals of Statistics* **24**, 380-404.
- Koul, H.L. and S.N. Lahiri (1994) On bootstrapping M-estimated residual processes in multiple linear regression models. *Journal of Multivariate Analysis* **49**, 255-265.
- Koul, H.L. and P.K. Sen (1985) On a Kolmogorov-Smirnov type aligned test in linear regression. *Statistics and Probability Letters* **3**, 111-115.
- Koul, H.L. and W. Stute (1999) Nonparametric model checks for time series. *Annals of Statistics* **27**, 204-236.
- Loynes, R.M. (1980) The empirical distribution function of residuals from generalized regression. *Annals of Statistics* **8**, 285-298.
- Neumeyer, N. and H. Dette (2003) Nonparametric comparison of regression curves: an empirical process approach. *Annals of Statistics* **31**, 880-920.
- Neumeyer, N., E.R. Nagel and H. Dette (2005a) A note on testing symmetry of the error distribution in linear regression models. *Journal of Nonparametric Statistics*, forthcoming.
- Neumeyer, N., E.R. Nagel and H. Dette (2005b) Bootstrap tests for the error distribution in linear and nonparametric regression models. *Australian and New Zealand Journal of Statistics*, forthcoming.
- Nolan, D. and D. Pollard (1987) U -processes: rates of convergence. *Annals of Statistics* **15**, 780-799.
- Nolan, D. and D. Pollard (1988) Functional limit theorems for U -process. *Annals of Probability* **16**, 1291-1298.

Pierce, D.A. and K.J. Kopecky (1979) Testing goodness-of-fit for the distribution of errors in regression models. *Biometrika* **66**, 1-6.

Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer.

Rao, J.S. and J. Sethuraman (1975) Weak convergence of empirical distribution functions of random variables subject to permutations and scale factors. *Annals of Statistics* **3**, 299-313.

Romano, J.P. (1988) A bootstrap revival of some nonparametric distance tests. *Journal of the American Statistical Association* **83**, 698-708.

Rothman, E.D. and M. Woodroffe (1972), A Cramér-von Mises type statistic for testing symmetry. *Annals of Mathematical Statistics* **43**, 2035-2038.

Shorack, G. and J. Wellner (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.

Silverman, B.W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics* **6**, 177-184.

Stute, W., W. González-Manteiga and M. Presedo-Quindimil (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* **93**, 141-149.

Stute, W. and L-X. Zhu (2002), Model checks for generalized linear models. *Scandinavian Journal of Statistics* **29**, 535-545.