A note on martingale transforms for model checks

Peter Behl, Holger Dette, Mathias Vetter
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 234 32 14559

May 5, 2010

Abstract

Martingale transforms are a well known tool to derive asymptotically distribution free tests for statistics based on empirical processes. Since its introduction by Khmaladze (1981) they have been frequently applied to many testing problems. In this paper martingale transforms for empirical processes are discussed from a non standard perspective with a specific focus on the case where the null hypothesis is not satisfied. For the sake of a transparent presentation we restrict our investigations to the problem of checking model assumptions in regression models, but the conclusions are generally valid. We show the weak convergence of empirical processes under fixed alternatives and introduce a new version of the martingale transform such that the transformed limiting process is a Brownian motion in scaled time, even if the null hypothesis is not satisfied.

Keywords: martingale transform, marked empirical process, weak convergence under fixed alternatives, model checks, nonparametric regression

1 Introduction

The problem of testing for the parametric form of a regression or the parametric form of the distribution of the given sample has a long history in statistics [see Durbin (1974), Loynes (1980) for some early references among others]. Several authors have proposed to use marked empirical or partial sum processes for testing model assumptions in nonparametric regression models [see

Stute (1997), Stute et al. (1998), An and Cheng (1991), Khmaladze and Koul (2004)]. It is well known that these processes are not asymptotically distribution free. A powerful tool to obtain asymptotically distribution free modifications of these processes is the martingale transform which was proposed in a far reaching publication by Khmaladze (1981). Since its introduction this transform has frequently been applied to tests based on empirical processes for various testing problems [see Khmaladze (1993), Khmaladze and Koul (2004), Stute et al. (1998), Koenker and Xiao (2002, 2006), Delgado et al. (2005), Koul and Yi (2006) and Dette and Hetzler (2009) among others]. Most papers consider the asymptotic distribution of the process under the null hypothesis and use a linear transform such that the limiting process is asymptotically distribution free.
The present paper is devoted to the investigation of martingale transforms in the case where the null hypothesis is not satisfied. In this case a standardized (in particular appropriately centered) version of the empirical process converges still to a Gaussian process, but the usual martingale transform does not lead to an asymptotically distribution free process. In the present paper we propose a new transform which has this property even under fixed alternatives. For the sake of brevity, we restrict ourselves to the problem of checking model assumptions regarding the conditional expectation in a nonparametric regression model. However, most of the conclusions and ideas are easily generalized to any other testing problem which has been discussed in the literature so far.

To be precise let $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ denote a sample of independent, identically distributed, bivariate observations where X has a continuous distribution function F. Define

(1.1) $$m(x) := E[Y|X = x] , \quad \sigma^2(x) := Var[Y|X = x]$$

as the conditional expectation and variance, respectively. Let $\Theta \subset \mathbb{R}^p$ denote a set of parameters and consider a class of parametric models

(1.2) $$\mathcal{M} = \{m(\cdot, \theta) : \theta \in \Theta\}.$$

In order to construct a test for the hypothesis of a parametric form of the regression function, i.e.

(1.3) $$H_0 : m \in \mathcal{M} \quad \text{versus} \quad H_1 : m \notin \mathcal{M},$$

Stute (1997) proposed to use the marked empirical process

(1.4) $$R_n(x) = n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} [Y_i - m(X_i, \theta_n)],$$

where θ_n denotes an appropriate estimator of the parameter vector θ under the assumption $m \in \mathcal{M}$. In the following we will study several properties of this process in the case where the null hypothesis is not true. In Section 2 we will present a result on the weak convergence of the process $\{R_n(x)\}_x$ under fixed alternatives, i.e. $m \notin \mathcal{M}$. In Section 3 we will introduce a generalization of the martingale transform proposed by Stute et al. (1998), which yields a martingale even under a fixed alternative. The proposed transform is a two-step procedure and reduces to the transform proposed by Stute et al. (1998) if the null hypothesis is satisfied. In concrete applications these transforms have to be estimated from the data and have to be applied to the corresponding empirical processes [see e.g. Koul and Song (2010) and Dette and Hetzler (2009)].

2 Weak convergence under fixed alternatives

Throughout this paper, we will denote by θ_0 a parameter corresponding to a "best" approximation of the unknown regression function m by the parametric class $\mathcal{M}$, where the specific metric depends on the method of estimation. For example, if least squares estimation is used (i.e. if θ_n denotes the least squares estimator of θ), we have

$$\theta_0 = \arg\min_\theta D(\theta) = \arg\min_\theta E[(m(X) - m(X,\theta))^2]. \tag{2.1}$$

Similarly, if maximum likelihood estimation is used, let $f(\cdot,\cdot)$ and $f(\cdot,\cdot,\theta)$ denote the density of (X,Y) in the general model and under the null hypothesis $m \in \mathcal{M}$, respectively. As a "best" approximation we then consider the minimizer of the Kullback-Leibler-distance, i.e.

$$\theta_{0_{ML}} = \arg\min_\theta D_{KL}(\theta) = \arg\min_\theta E\left[\log \frac{f(X,Y)}{f(X,Y,\theta)}\right]. \tag{2.2}$$

In order to investigate the asymptotic properties of the process defined in (1.4) we assume that the following assumptions are satisfied:

(A1) The estimate θ_n has a stochastic expansion of the form

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2}\sum_{i=1}^n l(X_i, Y_i, \theta_0) + o_P(1), \tag{2.3}$$

where $l(x,y,\theta_0)$ is a vector-valued square integrable function such that $E[l(x,y,\theta_0)l^T(x,y,\theta_0)]$ exists and $E[l(X,Y,\theta_0)] = 0$ is satisfied.

(A2) $m(x,\theta)$ is continuously differentiable at each interior point θ of Θ.

(A3) Let $g(x,\theta) := \frac{\partial m(x,\theta)}{\partial \theta} = (g_1(x,\theta), ..., g_p(x,\theta))^T$. Then there exists a function $M(x)$ which is integrable with respect to F such that

$$|g_i(x,\theta)| \le M(x) \quad \text{for all} \quad \theta \in \Theta \quad \text{and} \quad 1 \le i \le p.$$

(A4) $[m(x) - m(x,\theta_0)]^2 \le H(x)$ for all x and $\theta \in \Theta$ where H is integrable with respect to F.

(A5) $E[Y^2] < \infty$.

In Example 2.4 and Example 2.5 we will show the existence of a stochastic expansion of the form (2.3) for the least squares estimator θ_0 and the maximum likelihood estimator $\theta_{0_{ML}}$, respectively. Furthermore we will denote by

$$\Delta(t) = m(t) - m(t, \theta_0) \tag{2.4}$$

the distance between the regression function m and its approximation in the parametric class and by

$$R(x) := \int_{-\infty}^{x} (m(t) - m(t, \theta_0)) dF(t) = \int_{-\infty}^{x} \Delta(t) dF(t) \tag{2.5}$$

the integrated difference between the functions m and $m(\cdot, \theta_0)$. Then the following result specifies the asymptotic behavior of the centered marked empirical process

$$R_n^1(x) := R_n(x) - n^{1/2} R(x) \ , \qquad -\infty \le x \le \infty, \tag{2.6}$$

where we extend R_n^1 continuously on $\bar{\mathbb{R}} = [-\infty, \infty]$ using the definitions $R_n^1(-\infty) = 0$ and $R_n^{(1)}(\infty) = n^{-1/2} \sum_{i=1}^{n} (Y_i - m(X_i, \theta_n)) - \int_{-\infty}^{\infty} \Delta(t) dF(t)$.

Theorem. 2.1 *If the assumptions (A1)–(A5) are satisfied then as $n \to \infty$ we have*

$$\{R_n^1(x)\}_{x \in \bar{\mathbb{R}}} \xrightarrow{\mathcal{D}} \{R_\infty(x)\}_{x \in \bar{\mathbb{R}}},$$

in $D[-\infty, \infty]$ where R_∞ denotes a centered Gaussian process whose covariance kernel coincides with the covariance kernel of the process

$$\{\mathbb{1}_{\{X \le x\}}(Y - m(X)) - G^T(x, \theta_0)\ell(X, Y, \theta_0) + \mathbb{1}_{\{X \le x\}}\Delta(X)\}_{x \in \bar{\mathbb{R}}}, \tag{2.7}$$

and the vector G is defined by

$$G(x, \theta) = \int_{-\infty}^{x} g(u, \theta) dF(u).$$

Remark. 2.2 A straightforward calculation shows that for $s \le t$ the covariance kernel of the process (3.4) is given by

(2.8)
$$\begin{aligned} K(s,t) \ = \ & \int_{-\infty}^{s} \sigma^2(u) dF(u) + G^T(s, \theta_0) L(\theta_0) G(t, \theta_0) \\ & - G^T(s, \theta_0) E\left[\mathbb{1}_{\{X \le t\}}[Y - m(X)] l(X, Y, \theta_0)\right] - G^T(t, \theta_0) E\left[\mathbb{1}_{\{X \le s\}}[Y - m(X)] l(X, Y, \theta_0)\right] \\ & - G^T(s, \theta_0) E\left[\mathbb{1}_{\{X \le t\}}\Delta(X) l(X, Y, \theta_0)\right] - G^T(t, \theta_0) E\left[\mathbb{1}_{\{X \le s\}}\Delta(X) l(X, Y, \theta_0)\right] \\ & + E\left[\mathbb{1}_{\{X \le s\}}\Delta^2(X)\right] - E\left[\mathbb{1}_{\{X \le s\}}\Delta(X)\right] E\left[\mathbb{1}_{\{X \le t\}}\Delta(X)\right] \end{aligned}$$

where $L(\theta_0) := E[l(x, y, \theta_0)^T l(x, y, \theta_0)]$. Note that under the null hypothesis (1.3) we have $\Delta \equiv 0$ and in this case the kernel coincides with the kernel derived by Stute (1997).

Remark. 2.3 Let W and B denote a Brownian motion and a Brownian bridge, respectively, which are mutually independent. Then a straightforward calculation shows that the limiting process in Theorem 2.1 has a representation of the form

$$R_\infty(t) = (W \circ \psi)(t) + \int_{-\infty}^{t} \Delta(u) d(B \circ F)(u) - G^T(t, \theta_0)V \tag{2.9}$$

where

$$\psi(t) = \int_{-\infty}^{t} \sigma^2(u) dF(u), \tag{2.10}$$

and V denotes a centered normally distributed random vector with covariance matrix $L(\theta_0)$. This result follows easily by a straightforward calculation. For example, an application of Ito's rule [see e.g. Goldstein and McCabe (1993)] yields

$$\begin{aligned} Cov\Big[&\int_{-\infty}^{s} \Delta(u) d(B \circ F)(u), \int_{-\infty}^{t} \Delta(u) d(B \circ F)(u)\Big] \\ &= \int_{-\infty}^{s\wedge t} \Delta^2(u) dF(u) - 2\int_{-\infty}^{s} \Delta(u) dF(u) \int_{-\infty}^{t} \Delta(u) dF(u) + \int_{-\infty}^{s} \Delta(u) dF(u) \int_{-\infty}^{t} \Delta(u) dF(u) \\ &= \int_{-\infty}^{s\wedge t} \Delta^2(u) dF(u) - \int_{-\infty}^{s} \Delta(u) dF(u) \int_{-\infty}^{t} \Delta(u) dF(u) \end{aligned}$$

which gives the last two terms in (2.8).

Proof of Theorem 2.1. We make use of the decomposition

$$\begin{aligned} R_n^1(x) &= n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} [Y_i - m(X_i, \theta_n)] - n^{1/2} R(x) \\ &= n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} [Y_i - m(X_i)] - n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} [m(X_i, \theta_n) - m(X_i, \theta_0)] \\ &\quad + \left[n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} \Delta(X_i) - n^{1/2} R(x) \right] \\ &= A_n^1(x) - B_n^1(x) + C_n^1(x), \end{aligned} \tag{2.11}$$

where the last identity defines the quantities $A_n^1(x)$, $B_n^1(x)$ and $C_n^1(x)$ in an obvious manner. For the second term $B_n^1(x)$, a Taylor expansion and an application of the assumptions (A1) and (A3) yield

$$B_n^1(x) = n^{-1/2} \sum_{i=1}^{n} l^T(X_i, Y_i, \theta_0) G(x, \theta_0) + o_P(1), \tag{2.12}$$

uniformly with respect to $x \in \mathbb{R}$. Because $R_n(x)$ is a (rescaled) sum of independent identically distributed random variables with mean $R(x)$, the finite-dimensional distributions of the process

$\{R_n^1(x)\}_{x\in\mathbb{R}}$ are asymptotically normal with mean 0 and covariance kernel defined by (2.8). In order to prove tightness, note that similar arguments as in Stute (1997) show that it suffices to assume that the underlying distribution F of X is the uniform distribution on $[0,1]$. Furthermore note that under assumptions (A1)–(A3) it follows from Stute (1997) that the processes $\{A_n^1(x)\}_{x\in\bar{\mathbb{R}}}$ and $\{B_n^1(x)\}_{x\in\bar{\mathbb{R}}}$ are tight, and it remains to prove tightness of the process $\{C_n^1(x)\}_{x\in\bar{\mathbb{R}}}$. We do this by proving the inequality

$$h(u,u_1,u_2) = E[(C_n^1(u) - C_n^1(u_1))^2(C_n^1(u_2) - C_n^1(u))^2] \le (H(u_2) - H(u_1))^2 \tag{2.13}$$

for $0 \le u_1 \le u \le u_2 \le 1$ and some continuous non-decreasing function H. Then tightness of $\{C_n^1(x)\}_{x\in\mathbb{R}}$ follows from Theorem 15.6 in Billingsley (1968). For a proof of the inequality (2.13) we use Lemma 5.1 in Stute (1997) and obtain

$$h(u,u_1,u_2) \;=\; \frac{1}{n^2}E\Big[\Big(\sum_{i=1}^n \alpha_i\Big)^2\Big(\sum_{i=1}^n \beta_i\Big)^2\Big] \le \frac{1}{n}E[\alpha_1^2\beta_1^2] + 3E[\alpha_1^2]E[\beta_1^2], \tag{2.14}$$

where the random variables α_i and β_i are defined by

$$\begin{aligned}
\alpha_i &= \mathbb{1}_{\{u_1\le U_i\le u\}}\Delta(U_i) - [R(u) - R(u_1)],\\
\beta_i &= \mathbb{1}_{\{u< U_i\le u_2\}}\Delta(U_i) - [R(u_2) - R(u)],
\end{aligned}$$

respectively. Using the notations (2.5) and $\mu(u) := \int_0^u \Delta^2(t)dF(t)$, we obtain $(0 \le u \le 1)$ for the first term on the right hand side of (2.14)

$$\begin{aligned}
E[\alpha_1^2\beta_1^2] &= E\big[\big(-\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta(U_i)[R(u_2)-R(u)]\\
&\quad -\mathbb{1}_{\{u\le U_i\le u_2\}}\Delta(U_i)[R(u)-R(u_1)] + [R(u)-R(u_1)][R(u_2)-R(u)]\big)^2\big]\\
&= E\Big[\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta^2(U_i)[R(u_2)-R(u)]^2\\
&\quad +\mathbb{1}_{\{u\le U_i\le u_2\}}\Delta^2(U_i)[R(u)-R(u_1)]^2 + [R(u)-R(u_1)]^2[R(u_2)-R(u)]^2\\
&\quad -2\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta(U_i)[R(u_2)-R(u)]^2[R(u)-R(u_1)]\\
&\quad -2\mathbb{1}_{\{u\le U_i\le u_2\}}\Delta(U_i)[R(u)-R(u_1)]^2[R(u_2)-R(u)]\Big]\\
&= [\mu(u)-\mu(u_1)][R(u_2)-R(u)]^2\\
&\quad +[\mu(u_2)-\mu(u)][R(u)-R(u_1)]^2 - 3[R(u)-R(u_1)]^2[R(u_2)-R(u)]^2.
\end{aligned}$$

For the calculation of the second term in (2.14) note that

$$\begin{aligned}
E[\alpha_1^2] &= E\Big[\big(\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta(U_i) - [R(u)-R(u_1)]\big)^2\Big]\\
&= E[\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta^2(U_i)]\\
&\quad -2E[\mathbb{1}_{\{u_1\le U_i\le u\}}\Delta(U_i)][R(u)-R(u_1)] + [R(u)-R(u_1)]^2\\
&= \mu(u) - \mu(u_1) - [R(u)-R(u_1)]^2,
\end{aligned}$$

and similarly $E[\beta_1^2] = \mu(u_2) - \mu(u) - [R(u_2) - R(u)]^2$. This yields

$$\begin{aligned}
h(u, u_1, u_2) \;&\le\; 3[\mu(u) - \mu(u_1)][\mu(u_2) - \mu(u)] + (3 - \frac{3}{n})[R(u) - R(u_1)]^2[R(u_2) - R(u)]^2 \\
&-\; (3 - \frac{1}{n})\Big\{[\mu(u) - \mu(u_1)][R(u_2) - R(u)]^2 + [\mu(u_2) - \mu(u)][R(u) - R(u_1)]^2\Big\} \\
&\le 3[\mu(u) - \mu(u_1)][\mu(u_2) - \mu(u)] - (3 - \frac{1}{n})[\mu(u_2) - \mu(u)][R(u) - R(u_1)]^2 \\
&\quad -(-\frac{1}{n} + \frac{3}{n})[\mu(u) - \mu(u_1)][R(u_2) - R(u)]^2 \\
&\le 3[\mu(u) - \mu(u_1)][\mu(u_2) - \mu(u)] \le 3[\mu(u_2) - \mu(u_1)]^2,
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality in the second step in order to obtain $[R(u) - R(u_1)]^2 \le \mu(u) - \mu(u_1)$. Therefore the assertion (2.13) follows with $H(u) = \sqrt{3}\mu(u)$, which completes the proof of Theorem 2.1. □

Example. 2.4 Consider the distance function $D(\theta)$ defined in (2.1) and the least squares estimate

$$\theta_n = \arg\min_\theta D_n(\theta), \tag{2.15}$$

where

$$D_n(\theta) = n^{-1}\sum_{i=1}^n [(Y_i - m(X_i, \theta)]^2.$$

Let D_n' and D_n'' denote the gradient and the Hessian matrix of the function D_n, respectively. In addition to the previous chapters, assume that

(B1) θ_0 is a unique minimizer of $D(\theta)$ on Θ and an interior point of Θ.

(B2) $m(x, \theta)$ is twice continuously differentiable with respect to θ at each interior point $\theta \in \Theta$.

(B3) The Hessian matrix D_n'' is invertible at each interior point $\theta \in \Theta$ and the matrix D'' is invertible at θ_0.

Under assumptions (A4) and (B1), White (1981) proved that θ_n is a strongly consistent estimator for θ_0. Therefore, together with assumptions (B2) and (B3) a Taylor expansion and an application of Slutsky's theorem yield

$$\sqrt{n}(\theta_n - \theta_0) = -\sqrt{n}D_n''(\theta_0)^{-1}D_n'(\theta_0) + o_P(1).$$

Note that the random variable

$$-D_n'(\theta_0) = 2n^{-1}\sum_{i=1}^n (Y_i - m(X_i, \theta_0))\frac{\partial}{\partial\theta}m(X_i, \theta)\big|_{\theta=\theta_0}$$

has expectation 0, because θ_0 minimizes $D(\theta)$, which implies

$$\frac{\partial}{\partial\theta}D(\theta)\big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta}E[(m(X)-m(X,\theta)^2]\big|_{\theta=\theta_0} = -2E\Big[\frac{\partial}{\partial\theta}m(X,\theta)\big|_{\theta=\theta_0}(Y-m(X,\theta_0))\Big] \;=\; 0$$

by successive conditioning. On the other hand the law of large numbers implies

$$D_n''(\theta_0) \xrightarrow{P} D''(\theta_0) = 2\int_{-\infty}^{\infty}\Big[\Big(\frac{\partial}{\partial\theta}m(x,\theta)\big|_{\theta=\theta_0}\Big)\Big(\frac{\partial}{\partial\theta}m(x,\theta)\big|_{\theta=\theta_0}\Big)^T$$
$$-\Delta(x)\frac{\partial^2}{\partial\theta^2}m(x,\theta)\big|_{\theta=\theta_0}\Big]dF(x),$$

and it follows that

$$\begin{aligned}\sqrt{n}(\theta_n-\theta_0) \;&=\; -\sqrt{n}D''(\theta_0)^{-1}D_n'(\theta_0)+o_P(1)\\ &=\; 2D''(\theta_0)^{-1}n^{-1/2}\sum_{i=1}^{n}(Y_i-m(X_i,\theta_0))\frac{\partial}{\partial\theta}m(X_i,\theta)\big|_{\theta=\theta_0}+o_P(1).\end{aligned}$$

This yields the representation (2.3) with

$$l(X,Y,\theta_0) = 2D''(\theta_0)^{-1}(Y-m(X,\theta_0))\frac{\partial}{\partial\theta}m(X,\theta)\big|_{\theta=\theta_0}.$$

Note that under the null hypothesis the matrix $D''(\theta_0)$ equals

$$E[\frac{\partial}{\partial\theta}m(X,\theta)\big|_{\theta=\theta_0}(\frac{\partial}{\partial\theta}m(X,\theta)\big|_{\theta=\theta_0})^T],$$

and this identity also holds under a fixed alternative if the model $m(x,\theta)$ is a linear model. However, in general the two matrices are different under the alternative $H_1 : m \notin \mathcal{M}$.

Example. 2.5 Let D_{KL} denote the Kullback-Leibler-Distance as defined in (2.2). Let $f(\cdot,\cdot,\theta)$ and $f(\cdot,\cdot)$ be the density of (X,Y) under the null hypothesis $m \in \mathcal{M}$ and the alternative $m \notin \mathcal{M}$ and θ_{ML} denote the maximum likelihood estimate of the parameter θ. Similarly as in the previous example, we assume that

(C1) $E[\log f(X,Y)]$ exists and $\log f(x,y,\theta) \leq \tilde{H}(x)$ for all x and $\theta \in \Theta$, where $\tilde{H}$ is an integrable function with respect to F.

(C2) $\theta_{0_{ML}}$ is a unique minimizer of $D_{KL}(\theta)$ on Θ and an interior point of Θ.

(C3) $\log f(x,y,\theta)$ is twice continuosly differentiable at each interior point $\theta \in \Theta$.

(C4) The Hessian matrix $\frac{\partial^2}{\partial^2\theta}\log f(x,y,\theta)$ is invertible.

Under assumption (C1) and (C2), θ_{ML} is a strongly consistent estimator for $\theta_{0_{ML}}$ [for details see White (1982)] and we get that

$$\sqrt{n}(\theta_{ML} - \theta_{0_{ML}}) = E^{-1}(\theta_{0_{ML}}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i, Y_i, \theta)\big|_{\theta=\theta_{0_{ML}}} + o_P(1),$$

where the matrix $E(\theta_{0_{ML}})$ is defined by

$$E(\theta_{0_{ML}}) = -\int \frac{\partial^2}{\partial^2 \theta} \log f(x, y, \theta)\big|_{\theta=\theta_{0_{ML}}} f(x, y) d(x, y) \in \mathbb{R}^{p \times p}.$$

Note that this matrix simplifies to

$$E\Big[\Big(\frac{\partial}{\partial \theta} \log f(X, Y, \theta)\big|_{\theta=\theta_{0_{ML}}}\Big)\Big(\frac{\partial}{\partial \theta} \log f(X, Y, \theta)\big|_{\theta=\theta_{0_{ML}}}\Big)^T\Big]$$

if the null hypothesis is satisfied, i.e. $f(\cdot, \cdot) = f(\cdot, \cdot, \theta_{0_{ML}})$.

3 Martingale transforms under the alternative

It turns out that there exists a martingale transform with the desired properties for a weighted version of the process $\{R_n^1(x)\}_{x \in \bar{\mathbb{R}}}$. For this purpose we consider the weighted marked empirical process

(3.1) $$\tilde{R}_n^1(x) := n^{-1/2} \sum_{i=1}^{n} \Big(\mathbb{1}_{\{X_i \le x\}} \beta(X_i)[Y_i - m(X_i, \theta_n)] - \tilde{R}(x) \Big),$$

(with an obvious continuous extension at $x = \mp\infty$) where the centering term $\tilde{R}(x)$ is given by

(3.2) $$\tilde{R}(x) = \int_{-\infty}^{x} \beta(t) \Delta(t) dF(t)$$

and $\beta(x)$ is a continuous real-valued weight function such that the following assumptions are satisfied.

(D1) There exist functions $M(x)$, $\tilde{M}(x)$ and $M^*(x)$ which are integrable with respect to F such that

$$|\beta(x) g_i(x, \theta)| \le M(x) \quad \text{for all} \quad \theta \in \Theta \quad \text{and} \quad 1 \le i \le p,$$
$$\beta^2(x) \Delta^2(x) \le M^*(x) \quad \text{for all} \quad \theta \in \Theta$$

and

$$\beta^2(x) \sigma^2(x) \le \tilde{M}(x) \quad \text{for all} \quad \theta \in \Theta.$$

(D2) $\beta(x) > 0$ and $\sigma^2(x) > 0$ for all $x \in \mathbb{R}$ and there exists a positive constant c such that $\beta^2(x)[\sigma^2(x) + \Delta^2(x)] \ge c$ for all $x \in \mathbb{R}$.

Then similar arguments as given in the proof of Theorem 2.1 yield the following result:

Theorem. 3.1 *If the assumptions (A1)–(A5) and (D1)–(D2) are satisfied, then for $n \to \infty$ we have on $D[-\infty,\infty]$*

$$\{\tilde{R}_n^1(x)\}_{x\in\bar{\mathbb{R}}} \xrightarrow{\mathcal{D}} \{\tilde{R}_\infty(x)\}_{x\in\bar{\mathbb{R}}},$$

where

$$\tilde{R}_\infty(x) = \int_{-\infty}^{x} \beta(u)\sigma(u)d(W\circ F)(u) + \int_{-\infty}^{x} \beta(u)\Delta(u)d(B\circ F)(u) - \tilde{G}(x,\theta_0)^T V, \tag{3.3}$$

W and B denote a Brownian motion and a Brownian bridge, respectively, which are mutually independent, V is a centered, normally distributed random variable and $\tilde{G}(x,\theta)$ is defined by

$$\tilde{G}(x,\theta) = \int_{-\infty}^{x} \beta(u)g(u,\theta)dF(u).$$

The covariance kernel of the process $\tilde{R}_\infty$ coincides with the covariance kernel of the process

$$\{\mathbb{1}_{\{X\le x\}}\beta(X)(Y-m(X)) - \tilde{G}^T(x,\theta_0)\ell(X,Y,\theta_0) + \mathbb{1}_{\{X\le x\}}\beta(X)\Delta(X)\}_{x\in\bar{\mathbb{R}}}. \tag{3.4}$$

In the following we will consider a composition of two linear transforms which maps the process $\tilde{R}_\infty$ onto a martingale. Let W^* denote a Brownian motion independent of W, such that $B_t = W_t^* - tW_1^*$. Note that the limiting process $\tilde{R}_\infty$ in Theorem 3.2 can be represented as

$$\begin{aligned}
\tilde{R}_\infty(x) &\stackrel{D}{=} \int_{-\infty}^{x} \beta(u)\sigma(u)d(W\circ F)(u) + \int_{-\infty}^{x} \beta(u)\Delta(u)d(W^*\circ F)(u) \\
&\quad -W_1^* \int_{-\infty}^{x} \beta(u)\Delta(u)dF(u) + V^T \int_{-\infty}^{x} \beta(u)g(u,\theta_0)dF(u) \\
&\stackrel{D}{=} \int_{-\infty}^{x} \beta(u)[\sigma^2(u)+\Delta^2(u)]^{1/2} d(\tilde{W}\circ F)(u) \\
&\quad -W_1^* \int_{-\infty}^{x} \beta(u)\Delta(u)dF(u) + V^T \int_{-\infty}^{x} \beta(u)g(u,\theta_0)dF(u) \\
&= R_\infty^A(x) - R_\infty^B(x) + R_\infty^C(x)
\end{aligned} \tag{3.5}$$

where $\stackrel{D}{=}$ denotes equality in distribution, $\tilde{W}$ is a standard Brownian motion and the quantitites $R_\infty^A(x)$, $R_\infty^B(x)$ and $R_\infty^C(x)$ are defined in an obvious manner. Our first transform is therefore a standard martingale transform in the spirit of Khmaladze (1981), which transforms R_∞^B onto 0. To this end, consider for a process of the form

$$\int_{-\infty}^{x} r(u)d(W\circ F)(u) + V^T \int_{-\infty}^{x} s(u)dF(u)$$

with non-random real and vector-valued functions r and s, respectively, and a random variable V that does not depend on x. Note that a martingale transform can be obtained by defining

$$T(f) := f - \int_{-\infty}^{\cdot} s^T(u)A^{-1}(u)\left[\int_u^\infty \frac{s(v)}{r^2(v)}df(v)\right]dF(u)$$

where

$$A(u) = \int_u^\infty \frac{s(v)s^T(v)}{r^2(v)} dF(v).$$

Thus, in our case for the process R_∞^B in (3.5) the first transform T is given by (here we have $s = \beta\Delta$ and $r = \beta(\sigma^2 + \Delta^2)^{1/2}$)

$$(3.6)\quad T(f) := f - \int_{-\infty}^{\cdot} \Delta(u)\beta(u)A^{-1}(u)\Big[\int_u^\infty \Delta(v)\beta^{-1}(v)\left[\sigma^2(v) + \Delta^2(v)\right]^{-1} df(v)\Big] dF(u),$$

where

$$A(u) = \int_u^\infty \frac{\Delta^2(v)}{\sigma^2(v) + \Delta^2(v)} F(dv).$$

In order to define a second transform which maps the remaining term $T(R_\infty^C)$ onto 0, we introduce the notation

$$(3.7)\qquad l(u) = \beta(u)g(u,\theta_0) - \Delta(u)\beta(u)A^{-1}(u)\int_u^\infty \frac{\Delta(v)g(v,\theta_0)}{\sigma^2(v) + \Delta^2(v)} dF(v)$$

and, using the same procedure as above, we define a mapping S by

$$(3.8)\qquad S(f) := f - \int_{-\infty}^{\cdot} l^T(u)\tilde{A}^{-1}(u)\left[\int_u^\infty l(v)\left[\beta^2(v)[\sigma^2(v) + \Delta^2(v)]\right]^{-1} df(v)\right] dF(u)$$

with

$$\tilde{A}(u) := \int_u^\infty \left[\beta^2(s)[\sigma^2(s) + \Delta^2(s)]\right]^{-1} l(s)l^T(s)dF(s).$$

With these notations we obtain the following result:

Theorem. 3.2 *Let T and S denote the transforms defined in* (3.6) *and* (3.8), *respectively, then*

$$(S \circ T)(\tilde{R}_\infty) \stackrel{\mathcal{D}}{=} \tilde{W} \circ K,$$

where

$$(3.9)\qquad K(t) = \int_{-\infty}^t \beta(u)[\sigma^2(u) + \Delta^2(u)]^{1/2} dF(u)$$

and $\tilde{W}$ denotes a standard Brownian motion. In particular, if $\beta(u) = [\sigma^2(u) + \Delta^2(u)]^{-1/2}$ we have

$$(S \circ T)(\tilde{R}_\infty) \stackrel{\mathcal{D}}{=} \tilde{W} \circ F.$$

Proof. We start to investigate the transformed process $T(\tilde{R}_\infty)$ and discuss the terms in the decomposition (3.5) separately. For the first term we have

$$\begin{aligned} T(R_\infty^A)(x) &= T\left(\int_{-\infty}^x \beta(u)[\sigma^2(u) + \Delta^2(u)]^{1/2} d(\tilde{W} \circ F)(u)\right) \\ &= \int_{-\infty}^x \beta(u)(\sigma^2(u) + \Delta^2(u))^{1/2} d(\tilde{W} \circ F)(u)) \\ &\quad - \int_{-\infty}^x \Delta(u)\beta(u)A^{-1}(u)\int_u^\infty \frac{\Delta(v)}{[\sigma^2(v) + \Delta^2(v)]^{1/2}} d(\tilde{W} \circ F)(v) dF(u) \end{aligned}$$

and a similar calculation as in Stute et al. (1998) shows

$$Cov[T(R_\infty^A)(s), T(R_\infty^A)(t)] = \int_{-\infty}^{s\wedge t} \beta^2(u)[\sigma^2(u) + \Delta^2(u)]F(du)$$

which implies that

$$T(R_\infty^A) \stackrel{D}{=} \tilde{W} \circ K, \tag{3.10}$$

where K is defined in (3.9). For the second term R_∞^B in (3.5) we have

$$T(R_\infty^B) = T\left(W_1^* \int_{-\infty}^{\cdot} \beta(u)\Delta(u)dF(u)\right) = 0. \tag{3.11}$$

Finally, we obtain for the remaining term R_∞^C

$$\begin{aligned} T(R_\infty^C) &= V^T \int_{-\infty}^{\cdot} \beta(u)g(u,\theta_0)dF(u) \\ &\quad - V^T \int_{-\infty}^{\cdot} \Delta(u)\beta(u)A^{-1}(u)\left[\int_u^\infty \frac{\Delta(v)g(v,\theta_0)}{[\sigma^2(v)+\Delta^2(v)]}dF(v)\right]dF(u) \\ &= V^T \int_{-\infty}^{\cdot} l(u)dF(u), \end{aligned} \tag{3.12}$$

where the function l is defined in (3.7). Combining the results in (3.10)–(3.12) it follows that

$$T(\tilde{R}_\infty) \stackrel{D}{=} \tilde{W} \circ K + V^T \int_{-\infty}^{\cdot} l(u)dF(u).$$

Now the second transform defined in (3.8) obviously satisfies

$$S\left(V^T \int_{-\infty}^{\cdot} l(u)dF(u)\right) = 0,$$

and similar arguments as in Stute et al. (1998) show

$$S[T(\tilde{W} \circ K)] \stackrel{D}{=} S[\tilde{W} \circ K] \stackrel{D}{=} \tilde{W} \circ K.$$

Consequently, the composition of both transforms yields

$$(S \circ T)(\tilde{R}_\infty) \stackrel{D}{=} \tilde{W} \circ K$$

which simplifies to $\tilde{W} \circ F$ if $\beta(x) = [\sigma^2(x) + \Delta^2(x)]^{-1/2}$. □

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. The work of Holger Dette and Mathias Vetter has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG). The work of Peter Behl has been funded by a doctoral scholarship of the Hanns-Seidel-Foundation.

References

An, H. Z. and Cheng, B. (1991). A Kolmogorov–Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.*, 59:287–307.

Billingsley, P. (1968). *Convergence of Probability Measures. 1st Ed.* Wiley Series in Probability and Statistics, Chichester.

Delgado, M. A., Hidalgo, J., and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics*, 33:2568–2609.

Dette, H. and Hetzler, B. (2009). Khmaladze transformation of integrated variance processes with applications to goodness-of-fit testing. *Mathematical Methods of Statistics*, 18:97–116.

Durbin, J. (1974). *Distribution Theory for Tests Based on the Sample Distribution Function.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

Goldstein, L. and McCabe, B. (1993). On the moments of certain stochastic integrals. *Statistics and Probability Letters*, 18:65–72.

Khmaladze, E. V. (1981). A 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 Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Annals of Statistics*, 32:995–1034.

Koenker, R. and Xiao, Z. (2002). Inference of the quantile regression process. *Econometrica*, 70:1583–1612.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101:980–990.

Koul, H. L. and Song, W. (2010). Conditional variance model checking. *Journal of Statistical Planning and Inference*, 140:1056–1072.

Koul, H. L. and Yi, T. (2006). Goodness-of-fit testing in interval censoring case 1. *Statistics and Probability Letters*, 76:709–718.

Loynes, R. M. (1980). The empirical distribution function of residuals from generalised regression. *Annals of Statistics*, 8:285–298.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25:613–641.

Stute, W., Thies, S., and Zhu, L. (1998). Model checks for regression: an innovation process approach. *Annals of Statistics*, 26:1916–1934.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, 76:419–433.

White, H. (1982). Maximum-likelihood-estimation of misspecified models. *Econometrica*, 50:1–25.