Corrigendum: Detecting heteroskedasticity in nonparametric regression using weighted empirical processes

[*J. R. Statist. Soc. B (2018)* **80**, *951-974*]

Justin Chown[1] and Ursula U. Müller[2]

The second statement of Lemma 2 in the published article is wrong. Consequently, the statements about the limiting distributions of the test statistics in Theorems 1, 2 and 3 are also incorrect.

The claim that the term

$$R_2(t) = \frac{1}{n}\sum_{j=1}^{n} W_j\bigl[F\{t + \hat{r}(X_j) - r(X_j)\} - F(t)\bigr], \quad t \in \mathbb{R},$$

of the approximation of the test statistic is of order $o_P(n^{-1/2})$ and therefore asymptotically negligible is false: Professor Heng Lian has pointed out that the decomposition of R_2 in the proof of Lemma 2 does not lead to the desired rate $o_P(n^{-1/2})$. (The first term in the formula before equation (5.11) is not correctly centred.) Instead, we expect that $R_2(t)$ can be approximated by an extra term as follows. Let $\Delta(x) = \hat{r}(x) - r(x)$. Imposing more smoothness on f (a Hölder condition with exponent ξ), one obtains

$$\sup_{t\in\mathbb{R}}\Bigl|\frac{1}{n}\sum_{j=1}^{n} W_j\bigl[F\{t + \Delta(X_j)\} - F(t) - f(t)\Delta(X_j)\bigr]\Bigr| \le C\frac{1}{n}\sum_{j=1}^{n}|W_j||\Delta(X_j)|^{1+\xi},$$

where C is some constant. For large enough ξ (tied to the convergence rate of Δ), the right-hand side will be of order $o_P(n^{-1/2})$. The new term $n^{-1}\sum_{j=1}^{n} W_j f(t)\Delta(X_j)$ can be dealt with along the lines of Müller, Schick and Wefelmeyer (2009), who derive a uniform expansion for the unweighted residual-based empirical distribution function, i.e. for the case $W_j = 1$. We expect

$$\frac{1}{n}\sum_{j=1}^{n} W_j\Delta(X_j) = \frac{1}{n}\sum_{j=1}^{n} W_j\varepsilon_j + o_P(n^{-1/2}),$$

so the correct expansion should be

$$\sup_{t\in\mathbb{R}}\Bigl|\frac{1}{n}\sum_{j=1}^{n} W_j\{\mathbf{1}(\hat{\varepsilon}_j \le t) - \mathbf{1}(\varepsilon \le t) - f(t)\varepsilon_j\}\Bigr| = o_P(n^{-1/2}). \tag{1}$$

[1] *Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE*

[2] *Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.*

This result holds if the W_j (i.e. $\omega(X_j)$) have finite seconds moments; the Hölder exponent ξ should, as in Müller et al. (2009), be greater than $m/(2s-m)$. It immediately gives expansion (1) with $\hat{W}_j$ in place of W_j and, after some calculations,

$$\sup_{t\in\mathbb{R}} \Big| \frac{1}{n}\sum_{j=1}^{n} \hat{W}_j \mathbf{1}(\hat{\varepsilon}_j \le t) - \frac{1}{n}\sum_{j=1}^{n} W_j\{\mathbf{1}(\varepsilon_j \le t) + f(t)\varepsilon_j\} \Big| = o_P(n^{-1/2}). \tag{2}$$

The extra term involving the error density f will affect the limiting distribution, i.e. the test is not asymptotically distribution free, contrary to what is claimed in the article. More precisely, expansion (2) implies that the weighted residual-based empirical process converges under the null hypothesis weakly in $D[-\infty,\infty]$ to a centred Gaussian process $\mathcal{G}$ with covariance function

$$F(s\wedge t) - F(s)F(t) + f(s)c(t) + f(t)c(s) + f(s)f(t)\sigma_0^2$$

with $c(t) = E\{\varepsilon \mathbf{1}(\varepsilon \le t)\}$. This yields for the test statistic

$$T_n = \sup_{t\in\mathbb{R}} |n^{-1/2} \hat{W}_j \mathbf{1}(\hat{\varepsilon}_j \le t)| \xrightarrow{D} \sup_{t\in R} |\mathcal{G}(t)|.$$

The limit on the right-hand side should replace the limiting distribution involving the standard Brownian bridge in Theorem 1, and also in Theorem 2, which is a version of Theorem 1 with estimated weights. Note that additionally to the assumptions in Theorem 1 the above Hölder condition $\xi > m/(2s-m)$ must be satisfied.

Since the limiting distribution is not a standard distribution, quantiles are not readily available. We therefore recommend bootstrap to generate critical values. The same arguments apply to the complete case statistic in the missing data model. It has the same limiting distribution as the statistic for complete data given above since in both models the weights are standardised. Here also the smoothness condition on f must be satisfied.

Müller, U.U., Schick, A. and Wefelmeyer, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.*, **79**, 957-964.

The authors thank Professor Heng Lian for pointing out the mistake and Professor Anton Schick for helpful comments.