South African Statistical Journal
2018, Vol. 52, No. 1, 1–13

INDEPENDENCE TESTS IN SEMIPARAMETRIC TRANSFORMATION MODELS

Marie Hušková
Department of Statistics, Charles University, Prague, Czech Republic

Simos G. Meintanis [1]
Unit for Business Mathematics and Informatics,
North-West University, Potchefstroom, South Africa

Natalie Neumeyer
Department of Mathematics, University of Hamburg, Hamburg, Germany

Charl Pretorius [2]
Unit for Business Mathematics and Informatics,
North-West University, Potchefstroom, South Africa
e-mail: *charl.pretorius@nwu.ac.za*

Consider an observed response Y which, following a certain transformation $\mathcal{Y}_{\boldsymbol{\vartheta}} := \mathcal{T}_{\boldsymbol{\vartheta}}(Y)$, can be expressed by a homoskedastic nonparametric regression model referenced by a vector $\boldsymbol{X}$ of regressors. If this transformation model is indeed valid then conditionally on $\boldsymbol{X}$, the values of $\mathcal{Y}_{\boldsymbol{\vartheta}}$ may be viewed as being just location shifts of the regression error, for some value of the transformation parameter $\boldsymbol{\vartheta}$. We propose tests for the validity of this model, and establish the limiting distribution of the test statistics under the null hypothesis and under alternatives. Since the null distribution is complicated we also suggest a certain resampling procedure in order to approximate the critical values of the tests, and subsequently use this type of resampling in a Monte Carlo study of the finite-sample properties of the new tests. In estimating the model we rely on the methods proposed by Neumeyer, Noh and Van Keilegom (2016) for the aforementioned transformation model. Our tests however deviate from the tests suggested by Neumeyer et al. (2016) in that we employ an analogue of the test suggested by Hlávka, Hušková and Meintanis (2011) involving characteristic functions, rather than distribution functions.

Key words: Bootstrap test, Independence, Nonparametric regression, Transformation model.

1. Introduction

Consider the semi-parametric transformation model

$$\mathcal{T}_{\boldsymbol{\vartheta}}(Y) = m(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon, \tag{1}$$

where $\mathcal{T}_{\boldsymbol{\vartheta}}(\cdot)$ belongs to a fixed parametric family of monotone transformations, $m(\cdot)$ and $\sigma(\cdot)$ are unknown smooth functions, and where the error ε is supposed to be independent of the p-dimensional vector of covariates $\boldsymbol{X}$, satisfying $\mathbb{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = 1$.

[1] On sabbatical leave from the National and Kapodistrian University of Athens.
[2] Corresponding author.
MSC2010 subject classifications. 62G08, 62G09, 62G10.

Such transformations have been in use mostly with linear regression in the past. This is a special case of the more general model in (1) in which, given $(\boldsymbol{X}, Y)$, we are searching for a transformation $\mathcal{T}_{\boldsymbol{\vartheta}}(Y)$ that efficiently explains dependence of the transformed response $\mathcal{Y}_{\boldsymbol{\vartheta}} := \mathcal{T}_{\boldsymbol{\vartheta}}(Y)$ on covariates $\boldsymbol{X}$ by means of the linear model

$$\mathcal{Y}_{\boldsymbol{\vartheta}} = \boldsymbol{X}^\top \boldsymbol{\beta} + \varepsilon. \tag{2}$$

It should be mentioned at the outset that model (2) is fully parametric in the sense that both the family of transformations as well as the regression function are fixed, and the only unknowns are the transformation and regression parameters, $\boldsymbol{\vartheta}$ and $\boldsymbol{\beta}$, respectively. In fact the majority of approaches to model (2) were by means of the Box-Cox class of power transformations, mostly in the context of normal errors (Box and Cox, 1964; Chen, Lockhart and Stephens, 2002), or without the assumption of normality (Foster, Tian and Wei, 2001), and more recently even with estimation of the conditional quantile, rather than the conditional mean (Mu and He, 2007); for a review on transformations and related inference procedures the interested reader is referred to Horowitz (2009).

For the non-transformation case, which corresponds to $\mathcal{T}_{\boldsymbol{\vartheta}}(Y) \equiv Y$ in (1), there exist a number of specification tests that validate the location/scale structure manifested in that model; see Einmahl and Van Keilegom (2008), Neumeyer (2009), Hlávka et al. (2011), and the review article by González-Manteiga and Crujeiras (2013). Recently however there exists an increasing interest for the more general transformation model defined by (1), the validity of which entails a location/scale structure following a certain transformation of the response. In the context of the transformation model, aspects that have hitherto occupied researchers include estimation (Linton, Sperlich and Van Keilegom, 2008; Colling, Heuchenne, Samb and Van Keilegom, 2015; Neumeyer et al., 2016), and goodness-of-fit for the regression function (Colling and Van Keilegom, 2016, 2017), and for regressors (Allison, Hušková and Meintanis, 2018), as well as model validity (Neumeyer et al., 2016). Here we are concerned with validity of the transformation $\mathcal{Y}_{\boldsymbol{\vartheta}} = \mathcal{T}_{\boldsymbol{\vartheta}}(Y)$, $\boldsymbol{\vartheta} \in \Theta$, as exemplified in Neumeyer et al. (2016) by the null hypothesis

$$H_0 : \exists\, \boldsymbol{\vartheta}_0 \in \Theta \text{ such that } \frac{\mathcal{Y}_{\boldsymbol{\vartheta}_0} - \mathbb{E}(\mathcal{Y}_{\boldsymbol{\vartheta}_0}|\boldsymbol{X})}{\left[\mathrm{Var}(\mathcal{Y}_{\boldsymbol{\vartheta}_0}|\boldsymbol{X})\right]^{1/2}} \perp \boldsymbol{X}, \tag{3}$$

where $\perp$ denotes stochastic independence, and $\Theta \subseteq \mathbb{R}^q$, $q \geq 1$.

If the null hypothesis H_0 is true then there exists a true value $\boldsymbol{\vartheta}_0$ and a corresponding member in the prechosen parametric class of transformations $\{\mathcal{T}_{\boldsymbol{\vartheta}} | \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}$, such that the transformed response variable $\mathcal{Y}_{\boldsymbol{\vartheta}_0} = \mathcal{T}_{\boldsymbol{\vartheta}_0}(Y)$ follows a nonparametric location-scale model. In other words, we wish to test the appropriateness of this specific parametric family of transformations for the data at hand, and in this sense rejection of H_0 may be interpreted as signalling lack-of-fit for the particular transformation family figuring in the null hypothesis.

Neumeyer et al. (2016) construct a procedure for the validity of the model (1) based on distribution functions. Here we develop a procedure for the same problem based on characteristic functions (CFs) since it is often the case that procedures based on CFs work under weaker assumptions. In addition, CF procedures are more convenient from a computational point of view and readily extendible to the multivariate context, which is not always the case with methods based on distribution functions due to lack of proper order in $\mathbb{R}^p$. Here we treat the homoskedastic model

$$\mathcal{T}_{\boldsymbol{\vartheta}}(Y) = m(\boldsymbol{X}) + \varepsilon, \tag{4}$$

where the error ε has zero mean and $\mathrm{Var}(\varepsilon) = \sigma^2$, for some constant $\sigma^2 > 0$.

In this connection let

$$\varepsilon_{\boldsymbol{\vartheta}} = \mathcal{Y}_{\boldsymbol{\vartheta}} - \mathbb{E}[\mathcal{Y}_{\boldsymbol{\vartheta}}|\boldsymbol{X}]$$

and notice that the null hypothesis in (3) can be equivalently formulated as

$$\exists\, \boldsymbol{\vartheta}_0 \ \text{ such that } \ \varphi_{\boldsymbol{X},\varepsilon_{\boldsymbol{\vartheta}_0}} - \varphi_{\boldsymbol{X}}\varphi_{\varepsilon_{\boldsymbol{\vartheta}_0}} \equiv 0, \tag{5}$$

where $\varphi_{\boldsymbol{X},\varepsilon_{\boldsymbol{\vartheta}}}$ denotes the joint CF of $(\boldsymbol{X}, \varepsilon_{\boldsymbol{\vartheta}})$, and $\varphi_{\boldsymbol{X}}$ and $\varphi_{\varepsilon_{\boldsymbol{\vartheta}}}$ denote the marginal CFs of $\boldsymbol{X}$ and $\varepsilon_{\boldsymbol{\vartheta}}$, respectively.

Let $(Y_j, \boldsymbol{X}_j)$, $j = 1, \dots, n$, denote independent copies of $(Y, \boldsymbol{X}) \in \mathbb{R} \times \mathbb{R}^p$, and assume the existence of estimators $\widehat{m}(\cdot)$ and $\widehat{\boldsymbol{\vartheta}}$ of the unknown nonparametric regression function $m(\cdot)$ and of the transformation parameter $\boldsymbol{\vartheta}_0$, respectively, by means of which we obtain residuals $\widehat{\varepsilon}_j = \widehat{\mathcal{Y}}_j - \widehat{m}(\boldsymbol{X}_j)$, with $\widehat{\mathcal{Y}}_j = \mathcal{T}_{\widehat{\boldsymbol{\vartheta}}}(\boldsymbol{X}_j)$, $j = 1, \cdots, n$.

Following the approach in Hlávka et al. (2011) our test procedure will be based on the criterion

$$\Delta_{n,W} = n \int_{-\infty}^{\infty} \int_{\mathbb{R}^p} |D_n(t_1, \boldsymbol{t}_2)|^2 W(t_1, \boldsymbol{t}_2) dt_1 d\boldsymbol{t}_2, \tag{6}$$

where

$$D_n(t_1, \boldsymbol{t}_2) = \widehat{\varphi}(t_1, \boldsymbol{t}_2) - \widehat{\varphi}_{\boldsymbol{X}}(\boldsymbol{t}_2)\widehat{\varphi}_{\widehat{\varepsilon}}(t_1), \ (t_1, \boldsymbol{t}_2) \in \mathbb{R} \times \mathbb{R}^p \tag{7}$$

is an estimator of the quantity in the left-hand side of the identity figuring in (5), and involves the empirical joint CF

$$\widehat{\varphi}(t_1, \boldsymbol{t}_2) = \frac{1}{n}\sum_{j=1}^{n} \exp\{i\boldsymbol{t}_2^\top \boldsymbol{X}_j + it_1\widehat{\varepsilon}_j\}$$

as an estimator corresponding to the joint CF $\varphi_{\boldsymbol{X},\varepsilon_{\boldsymbol{\vartheta}_0}}$ of $(\boldsymbol{X}, \varepsilon_{\boldsymbol{\vartheta}_0})$, as well as the empirical CFs

$$\widehat{\varphi}_{\boldsymbol{X}}(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} \exp\{i\boldsymbol{t}^\top \boldsymbol{X}_j\}$$

and

$$\widehat{\varphi}_{\widehat{\varepsilon}}(t) = \frac{1}{n}\sum_{j=1}^{n} \exp\{it\widehat{\varepsilon}_j\}$$

as estimators of the marginal CFs $\varphi_{\boldsymbol{X}}$ and $\varphi_{\varepsilon_{\boldsymbol{\vartheta}_0}}$ of $\boldsymbol{X}$ and $\varepsilon_{\boldsymbol{\vartheta}_0}$, respectively.

Clearly, for a non-negative weight function $W(\cdot, \cdot)$ to be specified later on, the test statistic $\Delta_{n,W}$ defined in (6) is a weighted L_2 distance which should be small under the null hypothesis H_0 and large under alternatives, at least for large sample size n. Therefore large values of the test statistics $\Delta_{n,W}$ indicate that the null hypothesis is violated.

The rest of the paper is outlined as follows. In Section 2 the asymptotic distribution of the test statistic under the null hypothesis as well as under alternatives is studied, while in Section 3 we particularize the test statistic with respect to the weight function, and suggest a bootstrap procedure which is suitable in order to approximate the null distribution of the test statistic. Section 4 presents the results of a Monte Carlo exercise that shed light to the finite-sample properties of the method. We finally conclude with discussion of our findings in Section 5. Some technical material is deferred to the Appendix.

2. Some theoretical results

We now move on to formulate theoretical properties of the introduced test statistics. More precisely, we present the limit distribution of our test statistics under both the null as well as alternative hypotheses. Since the assumptions are quite technical they are deferred to the Appendix.

We first introduce the required notation. For $\boldsymbol{\vartheta} \in \Theta$, define

$$\begin{aligned} m_{\boldsymbol{\vartheta}}(\boldsymbol{X}_j) &= E\big(\mathcal{T}_{\boldsymbol{\vartheta}}(Y_j)|\boldsymbol{X}_j\big), \\ \varepsilon_{\boldsymbol{\vartheta} j} &= \mathcal{T}_{\boldsymbol{\vartheta}}(Y_j) - m_{\boldsymbol{\vartheta}}(\boldsymbol{X}_j), \\ \widehat{\varepsilon}_j = \widehat{\varepsilon}_{\widehat{\boldsymbol{\vartheta}},j} &= \mathcal{T}_{\widehat{\boldsymbol{\vartheta}}}(Y_j) - \widehat{m}_{\widehat{\boldsymbol{\vartheta}}}(\boldsymbol{X}_j), \end{aligned} \tag{8}$$

where $\widehat{\boldsymbol{\vartheta}}$ is an estimator of $\boldsymbol{\vartheta}_0$ and $\widehat{f}(\cdot)$ and $\widehat{m}_{\boldsymbol{\vartheta}}(\boldsymbol{x})$ are kernel estimators of the density of $\boldsymbol{X}_j$ and $m_{\boldsymbol{\vartheta}}(\boldsymbol{x})$, respectively, defined by

$$\begin{aligned} \widehat{f}(\boldsymbol{x}) &= \frac{1}{nh^p}\sum_{v=1}^{n} K\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_v}{h}\Big), \quad \boldsymbol{x} = (x_1,\dots,x_p)^\top, \\ \widehat{m}_{\boldsymbol{\vartheta}}(\boldsymbol{x}) &= \frac{1}{\widehat{f}(\boldsymbol{x})}\frac{1}{nh^p}\sum_{v=1}^{n} K\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_v}{h}\Big)\mathcal{T}_{\boldsymbol{\vartheta}}(Y_v), \quad \boldsymbol{x} = (x_1,\dots,x_p)^\top, \end{aligned}$$

where $K(\cdot)$ and $h = h_n$ are a kernel and a bandwidth. It is assumed that $\widehat{\boldsymbol{\vartheta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\vartheta}_0$ that allows an asymptotic representation as shown in assumption (A.7).

Under the null hypothesis and the assumptions formulated in the Appendix, the distribution of $T_{n,W}$ is approximately the same as that of

$$\int_{\mathbb{R}^{p+1}} |Z(t_1,\boldsymbol{t}_2)|^2 W(t_1,\boldsymbol{t}_2)dt_1 d\boldsymbol{t}_2,$$

where $\{Z(t_1,\boldsymbol{t}_2), (t_1,\boldsymbol{t}_2) \in \mathbb{R}\times\mathbb{R}^p\}$ is a Gaussian process with zero mean function and the covariance structure as the process $\{\widetilde{Z}(t_1,\boldsymbol{t}_2), (t_1,\boldsymbol{t}_2) \in \mathbb{R}\times\mathbb{R}^p\}$ defined as

$$\begin{aligned} \widetilde{Z}(t_1,\boldsymbol{t}_2) = {} & \{\cos(t_1\varepsilon_1) - C_{\varepsilon_1}(t_1)\}g_+(\boldsymbol{t}_2^\top\boldsymbol{X}_1) + \{\sin(t_1\varepsilon_1) - S_{\varepsilon_1}(t_1)\}g_-(\boldsymbol{t}_2^\top\boldsymbol{X}_1) \\ & + t_1\varepsilon_1\Big(S_{\varepsilon_1}(t_1)g_+(\boldsymbol{t}_2^\top\boldsymbol{X}_1) + C_{\varepsilon_1}(t_1)g_-(\boldsymbol{t}_2^\top\boldsymbol{X}_1)\Big) \\ & + (\boldsymbol{g}(Y_1,\boldsymbol{X}_1))^\top E\Big(\boldsymbol{H}_{\boldsymbol{\vartheta}_0}(\varepsilon_1,\boldsymbol{X}_1,Y_1;\, t_1,\boldsymbol{t}_2)\Big), \end{aligned} \tag{9}$$

where C_{ε_1} and S_{ε_1} are the real and the imaginary part of the CF of ε_j. Similarly, C_X and S_X denote the real and the imaginary part of the CF of $\boldsymbol{X}_j$. Also, $\boldsymbol{g}(Y_1,\boldsymbol{X}_1)$ is specified in Assumption (A.7) and

$$\begin{aligned} g_+(\boldsymbol{t}_2^T\boldsymbol{X}_1) = {} & \cos(\boldsymbol{t}_2^\top\boldsymbol{X}_1) + \sin(\boldsymbol{t}_2^\top\boldsymbol{X}_1) - C_X(\boldsymbol{t}_2) - S_X(\boldsymbol{t}_2), \\ g_-(\boldsymbol{t}_2^\top\boldsymbol{X}_1) = {} & \cos(\boldsymbol{t}_2^\top\boldsymbol{X}_1) - \sin(\boldsymbol{t}_2^\top\boldsymbol{X}_1) - C_X(\boldsymbol{t}_2) - S_X(\boldsymbol{t}_2), \\ \boldsymbol{H}_{\boldsymbol{\vartheta}}(\varepsilon_1,\boldsymbol{X}_1,Y_1;\, t_1,\boldsymbol{t}_2) = {} & \Big(\frac{\partial\mathcal{T}_{\boldsymbol{\vartheta}}(Y_1)}{\partial\vartheta_1},\dots,\frac{\partial\mathcal{T}_{\boldsymbol{\vartheta}}(Y_1)}{\partial\vartheta_q}\Big)^\top\Big(-(\sin(t_1\varepsilon_1) - S_{\varepsilon_1}(t_1))g_+(\boldsymbol{t}_2^\top\boldsymbol{X}_1) \\ & + (\cos(t_1\varepsilon_1) - C_{\varepsilon_1}(t_1))g_-(\boldsymbol{t}_2^\top\boldsymbol{X}_1)\Big). \end{aligned}$$

The proof is omitted here since it follows along the lines of Theorem 1 in Hlávka et al. (2011) where a test for independence in the model $Y = m(\boldsymbol{X}) + \varepsilon$ is studied. The limit distribution under the null hypothesis is a weighted L_2-type functional of a Gaussian process. Concerning the structure, the first row in (9) corresponds to the situation when both $\boldsymbol{\vartheta}_0$ and ε are known, the second row reflects the influence of the estimator of $m(\cdot)$, while the third row reflects the influence of the estimator of $\boldsymbol{\vartheta}$.

For getting an approximation of the critical value one estimates the unknown quantities and simulates the limit distribution described above with unknown parameters replaced by their estimators. However, the bootstrap described in Section 3.2 is probably more useful.

Concerning the consistency of the newly proposed test, note that if H_0 is not true, there is no parameter in Θ that leads to independence. Assume that there exists a $\boldsymbol{\vartheta}_1$ satisfying (A.9). Then the integral

$$\int_{-\infty}^{\infty}\int_{\mathbb{R}^p} |\varphi_{\boldsymbol{X},\varepsilon_{\boldsymbol{\vartheta}_1}}(\boldsymbol{t}_2, t_1) - \varphi_{\boldsymbol{X}}(\boldsymbol{t}_2)\varphi_{\varepsilon_{\boldsymbol{\vartheta}_1}}(t_1)|^2 W(t_1, \boldsymbol{t}_2) dt_1 d\boldsymbol{t}_2 > 0,$$

and therefore $T_{n,W} \stackrel{P}{\to} \infty$ and hence the test is consistent.

3. Computations and resampling

3.1 Computations

First we discuss some computational aspects of the test statistic. Specifically from (7) straightforward algebra shows that

$$\begin{aligned}|D_n(t_1, \boldsymbol{t}_2)|^2 = &\frac{1}{n^2}\sum_{j,k=1}^{n} \cos(t_1\widehat{\varepsilon}_{jk} + \boldsymbol{t}_2^\top \boldsymbol{X}_{jk}) + \frac{1}{n^4}\sum_{j,k,\ell,m=1}^{n} \cos(t_1\widehat{\varepsilon}_{j\ell} + \boldsymbol{t}_2^\top \boldsymbol{X}_{km}) \\ &- \frac{2}{n^3}\sum_{j,k,\ell=1}^{n} \cos(t_1\widehat{\varepsilon}_{jk} + \boldsymbol{t}_2^\top \boldsymbol{X}_{j\ell}),\end{aligned}$$

where $\boldsymbol{X}_{jk} = \boldsymbol{X}_j - \boldsymbol{X}_k$ and $\widehat{\varepsilon}_{jk} = \widehat{\varepsilon}_j - \widehat{\varepsilon}_k,\ j,k = 1,\cdots,n$. Based on this expression it becomes clear that the test statistic will be simplified if we use the decomposition $W(t_1, \boldsymbol{t}_2) = w_1(t_1)w_2(\boldsymbol{t}_2)$ for the weight function with $\int_{-\infty}^{\infty} w_1(t)dt < \infty$ and $\int_{\mathbb{R}^p} w_2(\boldsymbol{t})d\boldsymbol{t} < \infty$ such that Assumption (A.8) is satisfied. Then the test statistic in (6) takes the form

$$\Delta_{n,W} = \frac{1}{n}\sum_{j,k=1}^{n} I_{1,jk}I_{2,jk} + \frac{1}{n^3}\sum_{j,k=1}^{n} I_{1,jk}\sum_{j,k=1}^{n} I_{2,jk} - \frac{2}{n^2}\sum_{j,k,\ell=1}^{n} I_{1,jk}I_{2,j\ell},$$

where $I_{1,jk} := I_{w_1}(\widehat{\varepsilon}_{jk})$, $I_{2,jk} := I_{w_2}(\boldsymbol{X}_{jk})$, with

$$I_{w_m}(\boldsymbol{x}) = \int \cos(\boldsymbol{t}^\top \boldsymbol{x}) w_m(\boldsymbol{t}) \mathrm{d}\boldsymbol{t}, \quad m = 1, 2,$$

and integration performed over the appropriate domain. In this paper we set $w_1(t) = e^{-at^2}$ and $w_2(\boldsymbol{t}) = e^{-a\|\boldsymbol{t}\|^2}$ and make use of the integral

$$\int \cos(\boldsymbol{t}^\top \boldsymbol{x}) e^{-a\|\boldsymbol{t}\|^2} \mathrm{d}\boldsymbol{t} = \left(\frac{\pi}{a}\right)^{p/2} e^{-\|\boldsymbol{x}\|^2/4a}, \quad a > 0.$$

3.2 Resampling

Recall that the null hypothesis H_0 in (5) corresponds to model (1) in which both the true value of transformation parameter $\boldsymbol{\vartheta}$ as well as the error density are unknown. In this connection, and since, as was noted in Section 2, the asymptotic distribution of the test criterion under the null hypothesis depends on these quantities, among other things, we provide here a resampling scheme which can be used in order to compute critical points and actually carry out the test. The resampling scheme, which was proposed by Neumeyer et al. (2016), involves resampling from the observed $\boldsymbol{X}_j$ and independently constructing the bootstrap errors by smoothing the residuals. The bootstrap model then fulfils the null hypothesis since

$$\mathcal{T}_{\widehat{\boldsymbol{\vartheta}}}(Y_j^*) - \mathbb{E}^*[\mathcal{T}_{\widehat{\boldsymbol{\vartheta}}}(Y_j^*)|\boldsymbol{X}_j^*] := \varepsilon_j^* \perp^* \boldsymbol{X}_j^*,$$

where $\mathbb{E}^*$ denotes conditional expectation and $\perp^*$ the conditional independence given the original sample.

We now describe the resampling procedure. Let a_n be a positive smoothing parameter such that $a_n \to 0$ and $n\, a_n \to \infty$, as $n \to \infty$. Also, denote by $\{\xi_j\}_{j=1}^n$ a sequence of random variables which are drawn independently of any other stochastic quantity involved in the test criterion. The bootstrap procedure is as follows:

1. Draw $\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_n^*$ with replacement from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.
2. Generate i.i.d. random variables $\{\xi_j\}_{j=1}^n$ having a standard normal distribution and let $\varepsilon_j^* = a_n\xi_j + \widehat{\varepsilon}_j,\ j = 1, \ldots, n$, with $\widehat{\varepsilon}_j$ defined in (8).
3. Compute the bootstrap responses $Y_j^* = \mathcal{T}_{\widehat{\boldsymbol{\vartheta}}}^{-1}(\widehat{m}_{\widehat{\boldsymbol{\vartheta}}}(\boldsymbol{X}_j^*) + \varepsilon_j^*),\ j = 1, \ldots, n$.
4. On the basis of the observations $(Y_j^*, \boldsymbol{X}_j^*)$, $j = 1, \ldots, n$, refit the model and obtain the bootstrap residuals $\widehat{\varepsilon}_j^*,\ j = 1, \ldots, n$.
5. Calculate the value of the test statistic, say $\Delta_{n,W}^*$, corresponding to the bootstrap sample $(Y_j^*, \boldsymbol{X}_j^*)$, $j = 1, \ldots, n$.
6. Repeat the previous steps a number of times, say B, and obtain $\{\Delta_{n,W}^{*(b)}\}_{b=1}^B$.
7. Calculate the critical point of a size-α test as the $(1-\alpha)$ level quantile $c_{1-\alpha}^*$ of $\Delta_{n,W}^{*(b)}$, $b = 1, \ldots, B$.
8. Reject the null hypothesis if $\Delta_{n,W} > c_{1-\alpha}^*$, where $\Delta_{n,W}$ is the value of the test statistic based on the original observations $(Y_j, \boldsymbol{X}_j)$, $j = 1, \ldots, n$.

4. Simulations

In this section we present the results of a Monte Carlo exercise that sheds light on the small-sample properties of the new test statistic and compare our test with the classical Kolmogorov–Smirnov and Cramér–von Mises criteria suggested by Neumeyer et al. (2016). We consider the family of transformations

$$\mathcal{Y}_\vartheta = \mathcal{T}_\vartheta(Y) = \begin{cases} \left\{(Y+1)^\vartheta - 1\right\}/\vartheta & \text{if } Y \geq 0,\ \vartheta \neq 0, \\ \log(Y+1) & \text{if } Y \geq 0,\ \vartheta = 0, \\ -\left\{(-Y+1)^{2-\vartheta} - 1\right\}/(2-\vartheta) & \text{if } Y < 0,\ \vartheta \neq 2, \\ -\log(-Y+1) & \text{if } Y < 0,\ \vartheta = 2, \end{cases}$$

proposed by Yeo and Johnson (2000), and generate data from the univariate ($p = 1$) heteroskedastic model

$$\mathcal{T}_{\vartheta_0}(Y_j) = m(X_j) + \sigma(X_j)\varepsilon_j, \quad j = 1, \ldots, n, \tag{10}$$

where $\vartheta_0 = 0$, $m(\cdot)$ is some specified regression model, $\sigma(x) = 1 + \beta(x-1)$, $X \sim \mathrm{U}(0,1)$ and ε is an error term such that $X \perp \varepsilon$. We consider the following three combinations of choices of m and ε:

Model A. $m(x) = 1.5 + 0.25\sin(2\pi x)$ and $\varepsilon \sim \mathrm{N}(0,1)$.

Model B. $m(x) = 1.5 + 0.25\sin(2\pi x)$ and $\varepsilon \sim t_3$, where t_ν denotes the t distribution with $\nu > 0$ degrees of freedom.

Model C. $m(x) = 1 + x + x^2$ and $\varepsilon \sim t_2$.

For a test size of $\alpha = 5\%$, the rejection frequency of the test is recorded for $\beta = 0$ (null hypothesis), and $\beta = 0.5, 0.75, 1.0$ (alternative hypothesis) for sample sizes $n = 100$ and $n = 200$.

The bootstrap resampling scheme of Section 3.2 requires a choice of the smoothing parameter a_n. For all simulations we followed Neumeyer et al. (2016) and chose $a_n = 0.5n^{-1/4}$. Since the bootstrap replications are time consuming we have employed the warp-speed method of Giacomini, Politis and White (2013) in order to calculate critical points of the test criterion. With this method we generate only one bootstrap resample for each Monte Carlo sample and thereby compute the bootstrap test statistic T^* for that resample. Then, for a number M of Monte Carlo replications, the size-α critical point is determined similarly as in step 7 of Section 3.2, by computing the $(1-\alpha)$-level quantile of $T^{*(m)}$, $m = 1, \ldots, M$. For all simulations the number of Monte Carlo replications was set to $M = 2\,000$.

4.1 Estimation of the transformation parameter

To estimate the transformation parameter ϑ_0 in (10) we consider two estimators based on a profile likelihood approach. The first estimator, denoted by $\widehat{\vartheta}_0^{PL}$, is that of Linton et al. (2008) which is based on the assumption of a homoskedastic error structure, i.e. under the model in (4).

While employing a method which is tailored to the null hypothesis of homoskedasticity seems to be the only natural way to go about in estimating ϑ_0, there is also the question of the effect that estimation has on the power of any given test. This issue is of a long standing and persistent interest; for a relatively early reference see for instance Drost, Kallenberg and Oosterhoff (1990). In the context of CF-based goodness-of-fit tests the estimation effect has been reported by Gürtler and Henze (2000), Matsui and Takemura (2005), and Potgieter and Genton (2013), under the standard i.i.d. scenario. Specifically, as it is argued by Drost et al. (1990), it is not always the best estimator under the null that renders the test statistic more powerful, but rather that an estimator which behaves

Table 1. Size and power results for Model A.

						$\Delta_{n,W}$					
Estimator	β	n	Mean $\widehat{\vartheta}$	KS	CvM	$a = 0.025$	$a = 0.05$	$a = 0.1$	$a = 0.25$	$a = 0.5$	$a = 1$
$\widehat{\vartheta}_0^{PL}$	0.00	100	0.014	6.1	4.6	3.3	3.5	3.6	3.6	3.5	3.3
		200	0.018	8.6	6.8	4.6	4.5	5.2	4.9	4.6	5.3
	0.50	100	0.174	23.9	37.3	15.2	20.9	29.2	40.2	44.6	46.1
		200	0.222	50.5	70.8	38.0	51.5	61.9	75.2	81.5	83.5
	0.75	100	0.265	68.0	87.3	63.3	75.9	84.7	88.9	89.6	96.4
		200	0.287	95.8	99.8	95.7	98.0	99.1	99.9	99.8	99.5
	1.00	100	0.323	98.5	99.9	99.5	99.8	99.8	98.8	97.5	95.2
		200	0.310	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
$\widehat{\vartheta}_1^{PL}$	0.00	100	0.018	3.0	2.8	3.0	3.0	3.3	3.8	3.8	4.0
		200	0.022	7.6	6.2	4.1	4.4	5.0	5.2	5.6	6.2
	0.50	100	0.019	21.8	37.0	19.4	26.8	33.1	39.4	40.0	40.1
		200	0.015	45.8	74.1	49.6	60.4	71.7	81.9	82.7	80.3
	0.75	100	0.027	61.0	89.4	73.5	82.5	87.0	85.8	79.0	72.4
		200	0.011	94.3	99.8	98.3	99.3	99.8	99.7	98.8	95.8
	1.00	100	0.093	98.5	100.0	99.3	98.9	97.7	92.8	88.0	81.0
		200	0.078	100.0	100.0	100.0	100.0	99.9	99.3	97.8	94.8

well under the null but is highly non-robust and can even diverge under alternatives may have a favourable impact on the power of the test under such alternatives. In this connection we also report results for the profile likelihood estimator $\widehat{\vartheta}_1^{PL}$ developed by Neumeyer et al. (2016), which allows for a heteroskedastic error structure.

These estimators both involve estimating $m(\cdot)$ nonparametrically, for which we used local linear regression with a Gaussian kernel and a fixed bandwidth of $\frac{1}{2}n^{-1/5}$. The estimators also require the estimation of the density of the regression errors. For this purpose we used a Gaussian kernel with bandwidth $1.06\widehat{\sigma}_\varepsilon n^{-1/5}$, where $\widehat{\sigma}_\varepsilon^2$ denotes the sample variance of the residuals involved (see Silverman, 1986, p. 45).

4.2 Simulation results

The simulation results for the three considered models are shown in Tables 1 to 3. The results for the classical Kolmogorov–Smirnov and Cramér–von Mises tests are given in the columns labelled KS and CvM, respectively. The percentage of rejections of our statistic $\Delta_{n,W}$ is given for six different choices of the tuning parameter a appearing in the test. For each configuration we also report the mean value of the estimated transformation parameter over the 2 000 iterations.

Firstly, all tests seem to respect the nominal size of the test well, except in some cases where the classical tests seem to be significantly oversized. This is especially visible in the case of Model C as seen in Table 3. This size distortion seems to be less severe for the statistic $\Delta_{n,W}$. As expected, the power of all tests increases with the extent of violation of the null hypothesis (i.e. as β increases), and also with increasing sample size.

In terms of power the new test based on $\Delta_{n,W}$ exhibits competitive performance when compared to the classical tests. In the case of Model A, our test performs reasonably well with power that is mostly in line with that of the Cramér–von Mises test. However, for Models B and C which have

Table 2. Size and power results for Model B.

						$\Delta_{n,W}$					
Estimator	β	n	Mean $\widehat{\vartheta}$	KS	CvM	$a = 0.025$	$a = 0.05$	$a = 0.1$	$a = 0.25$	$a = 0.5$	$a = 1$
$\widehat{\vartheta}_0^{PL}$	0.00	100	0.013	6.4	6.6	3.8	4.3	4.8	6.3	7.0	6.1
		200	0.012	4.5	4.6	2.7	3.9	3.6	4.2	4.8	4.2
	0.50	100	0.058	8.8	17.2	12.0	15.9	20.4	23.9	23.5	25.3
		200	0.062	16.1	27.8	20.9	29.1	36.0	42.7	44.9	45.5
	0.75	100	0.092	26.6	49.1	41.9	52.1	63.8	67.4	68.0	61.5
		200	0.087	56.3	84.5	78.0	85.8	92.2	95.0	95.0	91.7
	1.00	100	0.119	78.8	97.2	97.3	98.2	98.1	96.5	92.4	84.3
		200	0.098	98.3	100.0	100.0	100.0	100.0	100.0	100.0	99.7
$\widehat{\vartheta}_1^{PL}$	0.00	100	−0.023	6.5	6.2	4.0	4.8	5.5	6.2	6.8	6.0
		200	−0.022	4.4	4.3	2.9	3.4	4.0	4.0	3.9	3.5
	0.50	100	−0.027	10.3	14.3	11.3	15.1	17.6	20.2	19.4	18.1
		200	−0.030	9.3	18.6	16.8	24.6	31.7	36.0	36.0	33.5
	0.75	100	−0.029	22.1	43.1	43.6	52.1	56.9	57.5	52.0	43.4
		200	−0.045	34.3	69.2	61.9	75.5	84.2	87.8	83.0	76.8
	1.00	100	−0.044	76.0	96.8	97.0	95.8	93.9	86.6	73.6	60.8
		200	−0.080	89.8	100.0	99.7	99.3	98.6	96.0	82.7	87.8

heavier-tailed errors, the new test based on $\Delta_{n,W}$ has significantly higher power than the classical tests. This increase in power is even more striking for Model C, which involves infinite-variance heavy-tailed errors.

Before we discuss the effect of the choice of estimator on the power of the tests, first note that under alternative hypotheses the estimator $\widehat{\vartheta}_0^{PL}$ seems to diverge from the value $\vartheta_0 = 0$. This phenomenon is also seen in the results of Neumeyer et al. (2016, Section 4.2), where the authors explain that, whenever $\beta \neq 0$, the estimator $\widehat{\vartheta}_0^{PL}$ targets a value $\vartheta_1 \neq \vartheta_0$ that maximises the likelihood under the assumption of the homoskedastic model in (4). In such cases the data cannot be considered to come from a homoskedastic model, and it is then possible to detect the violation of the null hypothesis through the dependence between the regressor X and the residuals $\widehat{\varepsilon}_{\widehat{\vartheta}_0^{PL}}$ brought about by the model misspecification. This behaviour is not seen with the estimator $\widehat{\vartheta}_1^{PL}$, which is consistent also under alternative heteroskedastic models. As mentioned in the discussion in the previous section, we show the results obtained using both estimators.

Regarding the effect of the estimator on the power of the tests we observe, depending on the underlying regression function and error distribution, different patterns for the different test statistics, but overall the $\Delta_{n,W}$ appears to be the most estimation-sensitive criterion. Specifically the KS test is adversely affected by the use of the heteroskedasticity-consistent estimator $\widehat{\vartheta}_1^{PL}$, with the size of this effect ranging from minor to significant for different combinations of regression function, error distribution, and intensity of violation of the null hypothesis. On the other hand, for the CvM and CF-based criteria the same effect varies not only with respect to the specific design and type of deviation from the null, but also in its direction, being favourable to the use of the heteroskedasticity-consistent estimator $\widehat{\vartheta}_1^{PL}$ in some cases, but also favouring the estimator $\widehat{\vartheta}_0^{PL}$ (which is inconsistent under heteroskedasticity) in other cases. Clearly our results are just indicative, and this issue deserves further attention.

Table 3. Size and power results for Model C.

Estimator	β	n	Mean $\widehat{\vartheta}$	KS	CvM	$\Delta_{n,W}$ $a=0.025$	$a=0.05$	$a=0.1$	$a=0.25$	$a=0.5$	$a=1$
$\widehat{\vartheta}_0^{PL}$	0.00	100	0.004	8.5	8.0	4.7	5.9	6.3	7.2	7.4	7.3
		200	−0.018	7.5	7.5	4.9	6.0	6.5	6.7	7.0	7.5
	0.50	100	−0.054	10.1	9.5	10.8	12.3	13.1	13.8	11.8	10.9
		200	−0.062	9.0	9.5	14.1	16.3	16.6	17.3	17.4	16.3
	0.75	100	−0.114	14.7	17.0	27.6	30.1	29.2	25.7	22.4	19.1
		200	−0.106	18.1	25.2	47.9	50.1	47.1	42.5	35.8	32.4
	1.00	100	−0.191	47.5	67.5	78.0	68.8	59.8	43.9	35.3	28.6
		200	−0.160	66.8	85.7	96.2	91.6	82.5	67.8	56.3	46.8
$\widehat{\vartheta}_1^{PL}$	0.00	100	−0.023	8.2	7.8	4.8	5.5	5.5	6.1	6.7	6.5
		200	−0.051	7.8	7.0	5.1	5.7	6.0	6.3	6.4	5.7
	0.50	100	−0.019	9.7	10.2	9.2	11.6	12.5	13.8	13.9	13.8
		200	−0.042	9.5	9.8	12.2	13.3	14.4	14.1	15.4	16.0
	0.75	100	−0.025	13.3	15.7	28.9	29.0	29.4	29.7	28.5	25.3
		200	−0.048	14.6	16.9	40.5	45.9	47.9	40.6	37.9	32.5
	1.00	100	−0.069	26.1	45.1	80.4	75.5	66.5	51.7	41.4	33.1
		200	−0.058	41.1	73.6	97.5	95.3	90.2	76.8	62.6	49.6

We close by noting that the value of the tuning parameter a clearly has some effect on the power of the CF-test. There exist several interpretations regarding the value of a and for more information on this the reader is referred to the recent review paper in this journal by Meintanis (2016). Our current results are mixed in this respect, favouring a 'higher' or 'intermediate' value of a in most cases of alternatives involving an error distribution which is not very heavy-tailed. On the other hand, with an error distribution as heavy-tailed as (the infinite-variance) t_2 distribution, it seems that a should be taken somewhat closer to zero. Overall we suggest $a = 0.1$ or $a = 0.25$ as compromise values.

5. Conclusions

New tests for the validity of the homoskedastic transformation model are proposed which are based on the well known factorization property of the joint characteristic function into its corresponding marginals. The asymptotic null distribution is derived and the consistency of the new criteria is shown. A Monte Carlo study is included by means of which a resampling version of the proposed method is compared to earlier methods and shows that the new test, aside from being computationally convenient, compares well and often outperforms its competitors, particularly under heavy tailed error distributions.

6. Appendix

We list some technical assumptions on the basis of which the asymptotic results of Section 2 are derived.

(A.1) $(Y_j, \boldsymbol{X}_j)$, $j = 1, \ldots, n$, are i.i.d. random vectors, where the covariates $\boldsymbol{X}_j$, $j = 1, \ldots, n$, have a compact support $\mathcal{R}_X$ with $\mathcal{R}_X \subset \mathbb{R}^p$.

(A.2) We use a product kernel $K(\boldsymbol{y}) = \prod_{s=1}^{p} k(y_s)$, $\quad \boldsymbol{y} = (y_1, \ldots, y_p)^\top$, with $k(\cdot)$ being symmetric and continuous in $[-1,1]$, and satisfying

$$\int_{-1}^{1} u^r k(u)du = \delta_{r,0}, \quad r = 0,\dots,p, \quad \int_{-1}^{1} u^{p+1} k(u)du \neq 0,$$

where $\delta_{r,s}$ stands for Kronecker's delta.

(A.3) The bandwidth $h = h_n$ satisfies

$$(nh^{2p})^{-1} + nh^{3p+\delta} \to 0 \text{ as } n \to \infty$$

for some $\delta > 1$.

(A.4) It holds that $\mathbb{E}||\boldsymbol{X}_j||^2 < \infty$, and that $\boldsymbol{X}_j$ has the density $f(\cdot)$ satisfying

$$0 < \inf_{\boldsymbol{x}\in\mathcal{R}_X} f(\boldsymbol{x}) \leq \sup_{\boldsymbol{x}\in\mathcal{R}_X} f(\boldsymbol{x}) < \infty,$$
$$\Big| f(\boldsymbol{x}) - L(f,\boldsymbol{x},\boldsymbol{x}-\boldsymbol{y},s_1)\Big| \leq ||\boldsymbol{x}-\boldsymbol{y}||^{s_1+\delta_1} d_1(\boldsymbol{y}),$$

where $L(f,\boldsymbol{x},\boldsymbol{y}-\boldsymbol{x},s_1)$ is the Taylor expansion of the density f of order s_1 at $\boldsymbol{x}$, $\delta_1 > 0$ and $\mathbb{E}|d_1(\boldsymbol{X}_j)|^2 < \infty$, for some $s_1 + 1 \geq p/2$.

(A.5) It is assumed $m_{\boldsymbol{\vartheta}_0}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{R}_X$, satisfies

$$\Big| m_{\boldsymbol{\vartheta}_0}(\boldsymbol{x}) - L_{\boldsymbol{\vartheta}_0}(m,\boldsymbol{y}_0,\boldsymbol{y}_0-\boldsymbol{x},s_2)\Big| \leq ||\boldsymbol{x}-\boldsymbol{y}_0||^{s_2+\delta_2} d_2(\boldsymbol{y}_0),$$

where $L_{\boldsymbol{\vartheta}_0}(m_{\boldsymbol{\vartheta}_0},\boldsymbol{y}_0,\boldsymbol{y}_0-\boldsymbol{x},s_2)$ is the Taylor expansions of regression function $m_{\boldsymbol{\vartheta}_0}$ of order s_2 at $\boldsymbol{y}_0$ and $\mathbb{E}|d_2(\boldsymbol{X}_j)|^2 < \infty$ for some $s_2 \geq p/2$ and $\mathbb{E}m^2_{\boldsymbol{\vartheta}_0}(\boldsymbol{X}_j) < \infty$.

(A.6) $\mathcal{L} = \{\mathcal{T}_{\boldsymbol{\vartheta}};\ \boldsymbol{\vartheta} \in \Theta\}$ is a parametric class of strictly increasing transformations, Θ is a nonempty measurable subset of $\mathbb{R}^q$, and for some $\xi > 0$

$$\sup_{||\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0||\leq\xi} \Big| \mathcal{T}_{\boldsymbol{\vartheta}}(Y_j) - \mathcal{T}_{\boldsymbol{\vartheta}_0}(Y_j) - \sum_{s=1}^{q} (\vartheta_s - \vartheta_{s0}) \frac{\partial \mathcal{T}_{\boldsymbol{\vartheta}}(Y_j)}{\partial \vartheta_s}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \Big| / ||\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0||^{2+\delta_4} \leq d_4(Y_j),$$

where $\mathbb{E}d_4^2(Y_j) < \infty$, $\mathbb{E}(|d_4(Y_j)||\boldsymbol{X}_j) < \infty$, a.s., and for some $\delta_4 > 0$ it holds that

$$\mathbb{E}\Big(\frac{\partial \mathcal{T}_{\boldsymbol{\vartheta}}(Y_j)}{\partial \vartheta_s}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \Big| \boldsymbol{X}_j \Big) = \frac{\partial \mathbb{E}\big(\mathcal{T}_{\boldsymbol{\vartheta}}(Y_j)|\boldsymbol{X}_j\big)}{\partial \vartheta_s}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}, \text{ a.s.,}$$
$$\mathbb{E}\Big(\mathbb{E}\Big(\frac{\partial \mathcal{T}_{\boldsymbol{\vartheta}}(Y_j)}{\partial \vartheta_s}\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \Big| \boldsymbol{X}_j \Big)\Big)^2 < \infty, \quad \mathbb{E}\mathcal{T}^2_{\boldsymbol{\vartheta}}(Y_j) < \infty.$$

(A.7) The estimator $\widehat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}_0$ satisfies

$$\sqrt{n}\Big(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\Big) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \boldsymbol{g}(Y_j,\boldsymbol{X}_j) + o_P(1),$$

where $\boldsymbol{g}(Y_j,\boldsymbol{X}_j)$ has zero mean and finite covariance matrix.

(A.8) The weight function $W(t_1, \boldsymbol{t}_2)$ satisfies the decomposition

$$W(t_1, \boldsymbol{t}_2) = w_1(t_1)w_2(\boldsymbol{t}_2), \quad (t_1, \boldsymbol{t}_2) \in \mathbb{R} \times \mathbb{R}^p,$$
$$\int_{-\infty}^{\infty} t^2 w_1(t)dt < \infty, \quad w_1(t) = w_1(-t), \quad w_1(t) \geq 0 \quad \forall t \in \mathbb{R},$$
$$w_2(\boldsymbol{t}) = w_2(-\boldsymbol{t}), \quad w_2(\boldsymbol{t}) \geq 0 \quad \forall \boldsymbol{t} \in \mathbb{R}^p.$$

(A.9) There exists a $\boldsymbol{\vartheta}_1 \in \Theta$ such that

$$\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_1 = o_P(1).$$

Comments on the assumptions:

- Assumptions (A.1), (A.2), (A.3), and (A.8) are quite standard.
- Assumption (A.4) requires smoothness of the density $f(\cdot)$ of $\boldsymbol{X}$.
- Assumption (A.5) formulates the requirements on $m_{\boldsymbol{\vartheta}_0}(\boldsymbol{x}) = \mathbb{E}(\mathcal{T}_{\boldsymbol{\vartheta}_0}(Y_j)|\boldsymbol{X}_j = \boldsymbol{x})$. Motivation for assumptions (A.4) and (A.5) are from Delgado and González-Manteiga (2001).
- Assumption (A.6) concerns smoothness of $\mathcal{T}_{\boldsymbol{\vartheta}}$ w.r.t. $\boldsymbol{\vartheta}$ in a neighbourhood of $\boldsymbol{\vartheta}_0$. Notice that this assumption implies smoothness of $\widehat{m}_{\boldsymbol{\vartheta}}(\boldsymbol{x})$ and $\widehat{\varepsilon}_{\boldsymbol{\vartheta},j} = \mathcal{T}_{\boldsymbol{\vartheta}}(\boldsymbol{Y}_j) - \widetilde{m}_{\boldsymbol{\vartheta}}(\boldsymbol{X}_j)$ in $\boldsymbol{\vartheta}$, in a neighbourhood of $\boldsymbol{\vartheta}_0$.
- Assumption (A.7) requires that a $\sqrt{n}$-estimator of $\boldsymbol{\vartheta}_0$ with an asymptotic representation is available. Such estimators are proposed and studied, e.g. in Breiman and Friedman (1985), Horowitz (2009), Linton et al. (2008). They are either based on a modified least squares method or on profile likelihood estimators or on mean square distance from independence.

References

Allison, J., Hušková, M., and Meintanis, S. (2018). Testing the adequacy of semiparametric transformation models. *TEST*, **27**, 70–94.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Stat. Soc. B.*, **26**, 211–252.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Am. Stat. Assoc.*, **80**, 580–598.

Chen, G., Lockhart, R. A., and Stephens, M. A. (2002). Box–Cox transformations in linear models: large sample theory and tests of normality. *Can. J. Stat.*, **30**, 177–209.

Colling, B., Heuchenne, C., Samb, R., and Van Keilegom, I. (2015). Estimation of the error density in a semiparametric transformation model. *Ann. I. Stat. Math.*, **67**, 1–18.

Colling, B. and Van Keilegom, I. (2016). Goodness-of-fit tests in semiparametric transformation models. *TEST*, **25**, 291–308.

Colling, B. and Van Keilegom, I. (2017). Goodness-of-fit tests in semiparametric transformation models using the integrated regression function. *J. Multivariate Anal.*, **160**, 10–30.

Delgado, M. A. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Stat.*, 1469–1507.

Drost, F., Kallenberg, W., and Oosterhoff, J. (1990). The power of edf tests of fit under non-robust estimation of nuisance parameters. *Stat. Decis.*, **8**, 167–182.

Einmahl, J. H. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *J. Econometrics*, **143**, 88–102.

Foster, A., Tian, L., and Wei, L. (2001). Estimation for the Box-Cox transformation model without assuming parametric error distribution. *J. Am. Stat. Assoc.*, **96**, 1097–1101.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Economet. Theor.*, **29**, 567.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST*, **22**, 361–411.

Gürtler, N. and Henze, N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. I. Stat. Math.*, **52**, 267–286.

Hlávka, Z., Hušková, M., and Meintanis, S. G. (2011). Tests for independence in non-parametric heteroscedastic regression models. *J. Multivariate Anal.*, **102**, 816–827.

Horowitz, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, volume 12. Springer, New York.

Linton, O., Sperlich, S., and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Stat.*, 686–718.

Matsui, M. and Takemura, A. (2005). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann. I. Stat. Math.*, **57**, 183–199.

Meintanis, S. (2016). A review of testing procedures based on the empirical characteristic function. *S. Afr. Stat. J.*, **50**, 1–14.

Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *J. Am. Stat. Assoc.*, **102**, 269–279.

Neumeyer, N. (2009). Testing independence in nonparametric regression. *J. Multivariate Anal.*, **100**, 1551–1566.

Neumeyer, N., Noh, H., and Van Keilegom, I. (2016). Heteroscedastic semiparametric transformation models: estimation and testing for validity. *Stat. Sinica*, **26**, 925–954.

Potgieter, C. and Genton, M. (2013). Characteristic function based semiparametric inference for skew-symmetric models. *Scand. J. Stat.*, **40**, 471–490.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. Chapman & Hall, New York.

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954–959.

Manuscript received 2018-02-27, revised 2018-03-19, accepted 2018-03-20.