Specification testing in semi-parametric transformation models

Nick Kloodt¹ · Natalie Neumeyer¹ · Ingrid Van Keilegom²

Received: 10 January 2020 / Accepted: 19 January 2021 / Published online: 25 February 2021

© The Author(s) 2021, corrected publication 2021

Abstract

In transformation regression models, the response is transformed before fitting a regression model to covariates and transformed response. We assume such a model where the errors are independent from the covariates and the regression function is modeled nonparametrically. We suggest a test for goodness-of-fit of a parametric transformation class based on a distance between a nonparametric transformation estimator and the parametric class. We present asymptotic theory under the null hypothesis of validity of the semi-parametric model and under local alternatives. A bootstrap algorithm is suggested in order to apply the test. We also consider relevant hypotheses to distinguish between large and small distances of the parametric transformation class to the ‘true’ transformation.

Keywords Bootstrap · Goodness-of-fit test · Nonparametric regression · Nonparametric transformation estimator · Parametric transformation class · U-statistics

Mathematics Subject Classification 62G10

1 Introduction

It is very common in applications to transform data before investigation of functional dependence of variables by regression models. The aim of the transformation is to obtain a simpler model, e.g. with a specific structure of the regression function,

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11749-021-00756-0>.

✉ Nick Kloodt
Nick.Kloodt@uni-hamburg.de

¹ Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany

² Research Centre for Operations Research and Statistics (ORSTAT), KU Leuven, Leuven, Belgium

or a homoscedastic instead of a heteroscedastic model. Typically, flexible parametric classes of transformations are considered from which a suitable one is selected data-dependently. A classical example is the class of Box-Cox power transformations (see Box and Cox (1964)). For purely parametric transformation models, see Carroll and Ruppert (1988) and references therein. Powell (1991) and Mu and He (2007) consider transformation quantile regression models. Nonparametric estimation of the transformation in the context of parametric regression models has been considered by Horowitz (1996) and Chen (2002), among others. Horowitz (2009) reviews estimation in transformation models with parametric regression in the cases where either the transformation or the error distribution or both are modeled nonparametrically. Linton et al. (2008) suggest a profile likelihood estimator for a parametric class of transformations, while the error distribution is estimated nonparametrically and the regression function semi-parametrically. Heuchenne et al. (2015) suggest an estimator of the error distribution in the same model. Neumeyer et al. (2016) consider profile likelihood estimation in heteroscedastic semi-parametric transformation regression models, i.e. the mean and variance function are modeled nonparametrically, while the transformation function is chosen from a parametric class. A completely nonparametric (homoscedastic) model is considered by Chiappori et al. (2015). Lewbel et al. (2015) provide a test for the validity of such a model. The approach of Chiappori et al. (2015) is modified and corrected by Colling and Van Keilegom (2019). The version of the nonparametric transformation estimator considered in the latter paper is then applied by Colling and Van Keilegom (2020) to suggest a new estimator of the transformation parameter if it is assumed that the transformation belongs to a parametric class.

In general, asymptotic theory for nonparametric transformation estimators is sophisticated and parametric transformation estimators show much better performance if the parametric model is true. A parametric transformation will thus lead to better estimates of the regression function. Moreover, parametric transformations are easier to interpret and allow for subsequent inference in the transformation model. For the latter purpose note that for transformation models with parametric transformation, lack-of-fit tests for the regression function as well as tests for significance for covariate components have been suggested by Colling and Van Keilegom (2016), Colling and Van Keilegom (2017), Allison et al. (2018) and Kloodt and Neumeyer (2020). Those tests cannot straightforwardly be generalized to nonparametric transformation models because known estimators in that model do not allow for uniform rates of convergence over the whole real line, see Chiappori et al. (2015) and Colling and Van Keilegom (2019).

However, before applying a transformation model with parametric transformation, it would be appropriate to test the goodness-of-fit of the parametric transformation class. In the context of parametric quantile regression, Mu and He (2007) suggest such a goodness-of-fit test. In the context of nonparametric mean regression Neumeyer et al. (2016), develop a goodness-of-fit test for the parametric transformation class based on an empirical independence process of pairs of residuals and covariates. The latter approach was modified by Hušková et al. (2018), who applied empirical characteristic functions. In a linear regression model with transformation of the response, Szydłowski (2020) suggests a goodness-of-fit test for the parametric transformation class that is

based on a distance between the nonparametric transformation estimator considered by Chen (2002) and the parametric class. We will follow a similar approach but consider a nonparametric regression model. The aim of the transformations we consider is to induce independence between errors and covariates. The null hypothesis is that the unknown transformation belongs to a parametric class. Note that when applied to the special case of a class of transformations that contains as only element the identity, our test provides indication on whether a classical homoscedastic regression model (without transformation) is appropriate or whether first the response should be transformed. Our test statistic is based on a minimum distance between a nonparametric transformation and the parametric transformations. We present the asymptotic distribution of the test statistic under the null hypothesis of a parametric transformation and under local alternatives of $n^{-1/2}$ -rate. Under the null hypothesis, the limit distribution is that of a degenerate U-statistic. With a flexible parametric class applying an appropriate transformation can reduce the dependence enormously, even if the ‘true’ transformation does not belong to the class. Thus, for the first time in the context of transformation goodness-of-fit tests, we consider testing for so-called precise or relevant hypotheses. Here, the null hypothesis is that the distance between the true transformation and the parametric class is large. If this hypothesis is rejected, then the model with the parametric transformation fits well enough to be considered for further inference. Under the new null hypothesis, the test statistic is asymptotically normally distributed. The term “precise hypotheses” refers to Berger and Delampady (1987). Dette et al. (2020) considered precise hypotheses in the context of comparing mean functions in the context of functional time series. Note that the idea of precise hypotheses is related to that of equivalence tests, which originate from the field of pharmacokinetics (see Lakens (2017)). Throughout, we assume that the nonparametric transformation estimator fulfills an asymptotic linear expansion. It is then shown that the estimator considered by Colling and Van Keilegom (2019) fulfills this expansion and thus can be used for evaluating the test statistic.

The remainder of the paper is organized as follows. In Sect. 2, we present the model and the test statistic. Asymptotic distributions under the null hypothesis of a parametric transformation class and under local alternatives are presented in Sect. 3, which also contains a consistency result and asymptotic results under relevant hypotheses. Section 4 presents a bootstrap algorithm and a simulation study. Section 1 of the supplementary material contains assumptions for bootstrap results, while Section 2 there treats a specific nonparametric transformation estimator and shows that it fulfills the required conditions. The proofs of the main results are given in Section 3 and a rigorous treatment of bootstrap asymptotics is given in Section 4 of the supplement.

2 The model and test statistic

Assume we have observed (X_i, Y_i) , $i = 1, \dots, n$, which are independent with the same distribution as (X, Y) that fulfill the transformation regression model

$$h(Y) = g(X) + \varepsilon, \quad (1)$$

where $E[\varepsilon] = 0$ holds and ε is independent of the covariate X , which is $\mathbb{R}^{d_X}$ -valued, while Y is univariate. The regression function g will be modelled nonparametrically. The transformation $h : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing. Throughout we assume that given the joint distribution of (X, Y) and some identification conditions, there exists a unique transformation h such that this model is fulfilled. It then follows that the other model components are identified via $g(x) = E[h(Y)|X = x]$ and $\varepsilon = h(Y) - g(X)$. See Chiappori et al. (2015) for conditions under which the identifiability of h holds. In particular, conditions are required to fix location and scale, and we will assume throughout that

$$h(0) = 0 \quad \text{and} \quad h(1) = 1. \quad (2)$$

Now let $\{\Lambda_\theta : \theta \in \Theta\}$ be a class of strictly increasing parametric transformation functions $\Lambda_\theta : \mathbb{R} \rightarrow \mathbb{R}$, where $\Theta \subseteq \mathbb{R}^{d_\theta}$ is a finite-dimensional parameter space. Our purpose is to test whether a semi-parametric transformation model holds, i.e.

$$\Lambda_{\theta_0}(Y) = \tilde{g}(X) + \tilde{\varepsilon},$$

for some parameter $\theta_0 \in \Theta$, where $\tilde{\varepsilon}$ and X are independent. Due to the assumed uniqueness of the transformation h one obtains $h = h_0$ under validity of the semi-parametric model, where

$$h_0(\cdot) = \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0)}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)}.$$

Thus, we can write the null hypothesis as

$$H_0 : h \in \left\{ \frac{\Lambda_\theta(\cdot) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} : \theta \in \Theta \right\} \quad (3)$$

which thanks to (2) can be formulated equivalently as

$$H_0 : h \in \left\{ \frac{\Lambda_\theta(\cdot) - c_2}{c_1} : \theta \in \Theta, c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R} \right\}. \quad (4)$$

Our test statistics will be based on the following L^2 -distance

$$d(\Lambda_\theta, h) = \min_{c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}} E[w(Y)\{h(Y)c_1 + c_2 - \Lambda_\theta(Y)\}^2], \quad (5)$$

where w is a positive weight function with compact support $\mathcal{Y}_w$. Its empirical counterpart is

$$d_n(\Lambda_\theta, \hat{h}) := \min_{c_1 \in C_1, c_2 \in C_2} \frac{1}{n} \sum_{j=1}^n w(Y_j) \{\hat{h}(Y_j)c_1 + c_2 - \Lambda_\theta(Y_j)\}^2,$$

where $\hat{h}$ denotes a nonparametric estimator of the true transformation h as discussed below, and $C_1 \subset \mathbb{R}^+$, $C_2 \subset \mathbb{R}$ are compact sets. Assumption (A6) assures that the sets are large enough to contain the true values. Let $\gamma := (c_1, c_2, \theta)$ and $\mathcal{Y} := C_1 \times C_2 \times \Theta$. The test statistic is defined as

$$T_n = n \min_{\theta \in \Theta} d_n(\Lambda_\theta, \hat{h}) = \min_{\gamma = (c_1, c_2, \theta) \in \mathcal{Y}} \sum_{j=1}^n w(Y_j) \{\hat{h}(Y_j) c_1 + c_2 - \Lambda_\theta(Y_j)\}^2 \quad (6)$$

and the null hypothesis should be rejected for large values of the test statistic. If the null hypothesis holds, the minimizing parameters c_1, c_2 in Eq. (5) can be written as $c_1 = \Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)$ and $c_2 = \Lambda_{\theta_0}(0)$ for some $\theta_0 \in \Theta$. Hence, an alternative test statistic

$$\bar{T}_n = \min_{\theta \in \Theta} \sum_{j=1}^n w(Y_j) [\hat{h}(Y_j) \{\Lambda_\theta(1) - \Lambda_\theta(0)\} + \Lambda_\theta(0) - \Lambda_\theta(Y_j)]^2 \quad (7)$$

can be considered as well.

We will derive the asymptotic distributions under the null hypothesis and local and fixed alternatives in Section 3 and suggest a bootstrap version of the tests in Section 4.

Remark 2.1 Colling and Van Keilegom (2019) consider the estimators

$$\hat{\theta} := \arg \min_{\theta \in \Theta} d_n(\Lambda_\theta, \hat{h})$$

and

$$\tilde{\theta} := \arg \min_{\theta \in \Theta} n^{-1} \sum_{j=1}^n w(Y_j) [\hat{h}(Y_j) \{\Lambda_\theta(1) - \Lambda_\theta(0)\} + \Lambda_\theta(0) - \Lambda_\theta(Y_j)]^2$$

for the parametric transformation (assuming H_0) corresponding to T_n and $\bar{T}_n$. They observe that $\hat{\theta}$ outperforms $\tilde{\theta}$ in simulations.

Nonparametric estimation of the transformation h has been considered by Chiappori et al. (2015) and Colling and Van Keilegom (2019). For our main asymptotic results, we need that $\hat{h}$ has a linear expansion, not only under the null hypothesis, but also under fixed alternatives and the local alternatives as defined in the next section. The linear expansion should have the form

$$\hat{h}(y) - h(y) = \frac{1}{n} \sum_{i=1}^n \psi(Z_i, \mathcal{T}(y)) + o_P(n^{-1/2}) \text{ uniformly in } y \in \mathcal{Y}_w. \quad (8)$$

Here, ψ needs to fulfil condition (A8) in Section 3, and we use the definitions ($i = 1, \dots, n$)

$$Z_i = (U_i, X_i), \quad U_i = \mathcal{T}(Y_i), \quad \mathcal{T}(y) = \frac{F_Y(y) - F_Y(0)}{F_Y(1) - F_Y(0)}, \quad (9)$$

where F_Y denotes the distribution of Y and is assumed to be strictly increasing on the support of Y . To ensure that $\mathcal{T}$ is well-defined, the values 0 and 1 are w.l.o.g. assumed to belong to the support of Y , but can be replaced by arbitrary values $a < b \in \mathbb{R}$ (in the support of Y). The expansion (8) could also be formulated with a linear term $n^{-1} \sum_{i=1}^n \tilde{\psi}(X_i, Y_i, y)$. In Section 2 of the supplement, we reproduce the definition of the estimator $\hat{h}$ that was suggested by Colling and Van Keilegom (2019) as modification of the estimator by Chiappori et al. (2015). We give regularity assumptions under which the desired expansion holds, see Lemma 1. Other nonparametric estimators of the transformation that fulfill the expansion could be applied as well.

3 Asymptotic results

In this section, we will derive the asymptotic distribution under the null hypothesis and under local and fixed alternatives. For the formulation of the local alternatives, consider the null hypothesis as given in (4), i.e. $h(\cdot)c_1 + c_2 = \Lambda_{\theta_0}(\cdot)$ for some $\theta_0 \in \Theta$, $c_1 \in \mathbb{R}^+$, $c_2 \in \mathbb{R}$, and instead assume

$$H_{1,n} : h(\cdot)c_1 + c_2 = \Lambda_{\theta_0}(\cdot) + n^{-1/2}r(\cdot) \text{ for some } \theta_0 \in \Theta, c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R} \text{ and some function } r.$$

Due to the identifiability conditions (2), one obtains $c_2 = \Lambda_{\theta_0}(0) + n^{-1/2}r(0)$ and $c_1 = \Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-1/2}(r(1) - r(0))$. Assumption (A5) yields boundedness of r , so that we rewrite the local alternative as

$$\begin{aligned} h(\cdot) &= \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0) + n^{-1/2}(r(\cdot) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-1/2}(r(1) - r(0))} \\ &= h_0(\cdot) + n^{-1/2}r_0(\cdot) + o(n^{-1/2}), \end{aligned} \quad (10)$$

where $h_0(\cdot) = (\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0))/(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))$ and

$$r_0(\cdot) = \frac{r(\cdot) - r(0) - h_0(\cdot)(r(1) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)}.$$

Note that the null hypothesis H_0 is included in the local alternative $H_{1,n}$ by considering $r \equiv 0$ which gives $h = h_0$. We assume the following data generating model under the local alternative $H_{1,n}$. Let the regression function g , the errors ε_i and the covariates X_i be independent of n and define $Y_i = h^{-1}(g(X_i) + \varepsilon_i)$ ($i = 1, \dots, n$), which under local alternatives depends on n through the transformation h . Throughout we use the notation ($i = 1, \dots, n$)

$$S_i = h(Y_i) = g(X_i) + \varepsilon_i. \quad (11)$$

Further, recall the definition of U_i in (9). Note that the distribution of U_i does not depend on n , even under local alternatives, because $F_Y(Y_i)$ is uniformly distributed

on $[0, 1]$, while $F_Y(0) = P(Y_i \leq 0) = P(h(Y_i) \leq h(0)) = P(S_i \leq 0)$ due to (2), and similarly $F_Y(1) = P(S_i \leq 1)$.

To formulate our main result, we need some more notations. With ψ from (8), Z_i from (9) and S_i from (11) define ($i = 1, \dots, n$)

$$\dot{\Lambda}_\theta(y) = \left(\frac{\partial}{\partial \theta_k} \Lambda_\theta(y) \right)_{k=1, \dots, d_\Theta} \quad (12)$$

$$R(s) = (s, 1, -\dot{\Lambda}_{\theta_0}(h_0^{-1}(s)))^t \quad (12)$$

$$R_f(s) = (\dot{\Lambda}_\theta(1)^t - \dot{\Lambda}_\theta(0)^t, \dot{\Lambda}_\theta(0)^t, \dot{\Lambda}_{\theta_0}(h_0^{-1}(s))^t)(s, 1, -1)^t \quad (13)$$

$$\Gamma_0 = E[w(h_0^{-1}(S_1))R(S_1)R(S_1)^t] \quad (14)$$

$$\Gamma_{0,f} = E[w(h_0^{-1}(S_1))R_f(S_1)R_f(S_1)^t] \quad (15)$$

$$\varphi(z) = E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)R(S_2) \mid Z_1 = z] \quad (16)$$

$$\varphi_f(z) = E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)R_f(S_2) \mid Z_1 = z] \quad (17)$$

$$\begin{aligned} \zeta(z_1, z_2) = & E \left[w(h_0^{-1}(S_3)) \{ \psi(Z_1, U_3) - \varphi(Z_1)^t \Gamma_0^{-1} R(S_3) \} \right. \\ & \left. \times \{ \psi(Z_2, U_3) - \varphi(Z_2)^t \Gamma_0^{-1} R(S_3) \} \mid Z_1 = z_1, Z_2 = z_2 \right] \end{aligned} \quad (18)$$

$$\begin{aligned} \zeta_f(z_1, z_2) = & E \left[w(h_0^{-1}(S_3)) \{ \psi(Z_1, U_3) - \varphi_f(Z_1)^t \Gamma_{0,f}^{-1} R_f(S_3) \} \right. \\ & \left. \times \{ \psi(Z_2, U_3) - \varphi_f(Z_2)^t \Gamma_{0,f}^{-1} R_f(S_3) \} \mid Z_1 = z_1, Z_2 = z_2 \right] \end{aligned} \quad (19)$$

$$\bar{r}(s) = r_0(h_0^{-1}(s)) - E[w(h_0^{-1}(S_1))r_0(h_0^{-1}(S_1))R(S_1)]^t \Gamma_0^{-1} R(s) \quad (20)$$

$$\bar{r}_f(s) = r_0(h_0^{-1}(s)) - E[w(h_0^{-1}(S_1))r_0(h_0^{-1}(S_1))R_f(S_1)]^t \Gamma_{0,f}^{-1} R_f(s) \quad (21)$$

$$\tilde{\zeta}(z_1) = 2E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)\bar{r}(S_2) \mid Z_1 = z] \quad (22)$$

$$\tilde{\zeta}_f(z_1) = 2E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)\bar{r}_f(S_2) \mid Z_1 = z] \quad (23)$$

and let P_Z and F^Z denote the law and distribution function, respectively, of Z_i . The quantities which are marked with an “ f ”, referring to “fixed” parameters $c_1 = \Lambda_{\hat{\theta}}(1) - \Lambda_{\hat{\theta}}(0)$ and $c_2 = \Lambda_{\hat{\theta}}(0)$, will be used to describe the asymptotic behaviour of the test statistic $\bar{T}_n$.

With these notations, the assumptions for the asymptotic results can be formulated. To this end, let $\mathcal{Y}$ denote the support of Y (which depends on n under local alternatives). Further, F_S denotes the distribution function of S_1 as in (11) and $\mathcal{T}_S$ denotes the transformation $s \mapsto (F_S(s) - F_S(0))/(F_S(1) - F_S(0))$. The following assumptions are used.

- (A1) The sets C_1 , C_2 and Θ are compact.
- (A2) The weight function w is continuous with a compact support $\mathcal{Y}_w \subset \mathcal{Y}$.
- (A3) The map $(y, \theta) \mapsto \Lambda_\theta(y)$ is twice continuously differentiable on $\mathcal{Y}_w$ with respect to θ and the (partial) derivatives are continuous in $(y, \theta) \in \mathcal{Y}_w \times \Theta$.
- (A4) There exists a unique strictly increasing and continuous transformation h such that model (1) holds with X independent of ε .

- (A5) The function h_0 defined in (10) is strictly increasing and continuously differentiable and r is continuous on $\mathcal{Y}_w$. F_Y is strictly increasing on the support of Y .
- (A6) Minimizing the functions $M : \mathcal{Y} \rightarrow \mathbb{R}, \gamma = (c_1, c_2, \theta) \mapsto E[w(Y)(h_0(Y)c_1 + c_2 - \Lambda_\theta(Y))^2]$ and $\bar{M} : \Theta \rightarrow \mathbb{R}, \theta \mapsto E[w(Y)(h_0(Y)(\Lambda_\theta(1) - \Lambda_\theta(1)) + \Lambda_\theta(0) - \Lambda_\theta(Y))^2]$ leads to unique solutions $\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0)$ and θ_0 in the interior of $\mathcal{Y}$ and Θ , respectively. For all $\theta \neq \tilde{\theta}$ it is $\sup_{y \in \text{supp}(w)} \left| \frac{\Lambda_\theta(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_\theta(1) - \Lambda_{\tilde{\theta}}(0)} - \frac{\Lambda_{\tilde{\theta}}(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)} \right| > 0$.
- (A7) The Hessian matrices $\Gamma_0 := \text{Hess } M(\gamma_0)$ and $\Gamma_{0,f} := \text{Hess } \bar{M}(\theta_0)$ are positive definite.
- (A8) The transformation estimator $\hat{h}$ fulfils (8) for some function ψ . For some $\mathcal{U}_0$ (independent of n under local alternatives) with $\mathcal{T}_S(h(\mathcal{Y}_w)) \subset \mathcal{U}_0$ the function class $\{z \mapsto \psi(z, t) : t \in \mathcal{U}_0\}$ is Donsker with respect to P^Z and $E[\psi(Z_1, t)] = 0$ for all $t \in \mathcal{U}_0$. The fourth moments $E[w(h_0^{-1}(S_1))\psi(Z_1, U_1)^4]$ and $E[w(h_0^{-1}(S_1))\psi(Z_2, U_1)^4]$ are finite.

When considering a fixed alternative H_1 or the relevant hypothesis H'_0 below, (A6) and (A8) are replaced by the following Assumptions (A6') and (A8') (assumption (A8') is only relevant for H'_0). Note that h is a fixed function then, not depending on n .

- (A6') Minimizing the functions $M : \mathcal{Y} \rightarrow \mathbb{R}, \gamma = (c_1, c_2, \theta) \mapsto E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2]$ and $\bar{M} : \Theta \rightarrow \mathbb{R}, \theta \mapsto E[w(Y)(h(Y)(\Lambda_\theta(1) - \Lambda_\theta(1)) + \Lambda_\theta(0) - \Lambda_\theta(Y))^2]$ leads to unique solutions $\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0)$ and θ_0 in the interior of $\mathcal{Y}$ and Θ , respectively. For all $\theta \neq \tilde{\theta}$, it is $\sup_{y \in \text{supp}(w)} \left| \frac{\Lambda_\theta(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_\theta(1) - \Lambda_{\tilde{\theta}}(0)} - \frac{\Lambda_{\tilde{\theta}}(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)} \right| > 0$.
- (A8') The transformation estimator $\hat{h}$ fulfils (8) for some function ψ . For some $\mathcal{U}_0 \supset \mathcal{T}_S(h(\mathcal{Y}_w))$, the function class $\{z \mapsto \psi(z, t) : t \in \mathcal{U}_0\}$ is Donsker with respect to P^Z and $E[\psi(Z_1, t)] = 0$ for all $t \in \mathcal{U}_0$. Further, one has $E[\psi(Z_1, U_2)^2] < \infty$.

Remark 3.1 Assumptions concerning compactness of the parameter spaces, differentiability of model components and uniqueness of the minimizer γ_0 are standard assumptions in the context of goodness of fit tests. Moreover, it can be shown that the definitions of Γ_0 and $\Gamma_{0,f}$ in (A7) coincide with those in Eqs. (14) and (15), respectively. Assumption (A8) controls the asymptotic behaviour of $\hat{h} - h$ and thus the rate of local alternatives which can be detected. The Donsker and boundedness conditions are needed to obtain uniform convergence rates of $\hat{h} - h$ and some negligible remainders in the proof. Assumption (A8') is the counterpart of Assumption (A8) for precise hypotheses as considered in (24).

Theorem 3.2 Assume (A1)–(A8). Let $(\lambda_k)_{k \in \{1,2,\dots\}}$ and $(\lambda_{k,f})_{k \in \{1,2,\dots\}}$ be the eigenvalues of the operators

$$K\rho(z_1) := \int \rho(z_2)\zeta(z_1, z_2) dF_Z(z_2) \quad \text{and} \quad K_f\rho(z_1) := \int \rho(z_2)\zeta_f(z_1, z_2) dF_Z(z_2),$$

respectively, with corresponding eigenfunctions $(\rho_k)_{k \in \{1, 2, \dots\}}$ and $(\rho_{k,f})_{k \in \{1, 2, \dots\}}$, which are each orthonormal in the L^2 -space corresponding to the distribution F_Z . Let $(W_k)_{k \in \{1, 2, \dots\}}$ be independent and standard normally distributed random variables and let W_0 be centred normally distributed with variance $E[\tilde{\zeta}(Z_1)^2]$ such that for all $K \in \mathbb{N}$ the random vector $(W_0, W_1, \dots, W_K)^t$ follows a multivariate normal distribution with $\text{Cov}(W_0, W_k) = E[\tilde{\zeta}(Z_1)\rho_k(Z_1)]$ for all $k = 1, \dots, K$. Let $W_{0,f}$ and $(W_{k,f})_{k \in \{1, 2, \dots\}}$ be defined similarly with $E[W_{0,f}^2] = E[\tilde{\zeta}_f(Z_1)^2]$ and $\text{Cov}(W_0, W_k) = E[\tilde{\zeta}_f(Z_1)\rho_k(Z_1)]$ for all $k \in \mathbb{N}_0$. Then, under the local alternative $H_{1,n}$, T_n converges in distribution to

$$(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2 \left(\sum_{k=1}^{\infty} \lambda_k W_k^2 + W_0 + E \left[w(h_0^{-1}(S_1)) \bar{r}(S_1)^2 \right] \right)$$

and $\bar{T}_n$ converges in distribution to

$$(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2 \left(\sum_{k=1}^{\infty} \lambda_{k,f} W_{k,f}^2 + W_{0,f} + E \left[w(h_0^{-1}(S_1)) \bar{r}_f(S_1)^2 \right] \right).$$

In particular, under H_0 (i.e. for $r \equiv 0$), T_n and $\bar{T}_n$ converge in distribution to

$$T = (\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2 \sum_{k=1}^{\infty} \lambda_k W_k^2 \quad \text{and} \quad \bar{T} = (\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2 \sum_{k=1}^{\infty} \lambda_{k,f} W_{k,f}^2,$$

respectively.

The proof is given in Section 3 of the supplementary material. Asymptotic level- α tests should reject H_0 if T_n or $\bar{T}_n$ are larger than the $(1 - \alpha)$ -quantile of the distribution of T or $\bar{T}$, respectively. As the distributions of T and $\bar{T}$ depend in a complicated way on unknown quantities, we will propose a bootstrap procedure in Section 4. Although most results hold similarly for T_n and $\bar{T}_n$, for ease of presentation, we will mainly focus on results for T_n in the remainder.

Remark 3.3

1. Note that $\zeta(z_1, z_2) = E[I(z_1)I(z_2)]$ with

$$I(z) := w(h_0^{-1}(S_1))^{1/2} \left(\psi(z, U_1) - \varphi(z)^t \Gamma_0^{-1} R(S_1) \right)$$

and ψ from (8). Thus, the operator K defined in Theorem 3.2 is positive semi-definite.

2. The appearance of W_0 under the local alternative results from asymptotic theory for degenerate U-statistics. Related phenomena occur in the case of quadratic forms. Similar to the proof of Theorem 3.1, consider some $z_n + cn^{-1/2}$, where $n^{1/2}z_n$ converges to a centred normally distributed random variable, say z , and

we have $c = 0$ under H_0 . Moreover, consider a quadratic form $z_n^T A_n z_n$, where A_n is a positive definite matrix and $n^{-1} A_n$ converges to a matrix A . Then, under H_0 , $z_0^T A_n z_0 = z_n^T A_n z_n$ converges to $z^T A z$, which has a χ^2 distribution. However, under $H_{1,n}$, we have

$$z_0^T A_n z_0 = z_n^T A_n z_n + 2c^T n^{-1/2} A_n z_n + c^T n^{-1} A_n c,$$

where the first term on the right-hand side is as before. However, the second term converges to $2c^T A z$, which is normally distributed and corresponds to W_0 in our context. The last term converges to a constant $c^T A c$, corresponding to the constant summand in the limit in Theorem 3.1. Note that the limit of $z_0^T A_n z_0$ cannot be negative due to the positive definiteness of A_n .

Next, we consider fixed alternatives of a transformation h that do not belong to the parametric class, i. e.

$$H_1 : d(\Lambda_\theta, h) > 0 \quad \text{for all } \theta \in \Theta.$$

Theorem 3.4 *Assume (A1)–(A4), (A6') and let $\hat{h}$ estimate h uniformly consistently on compact sets. Then, under H_1 , $\lim_{n \rightarrow \infty} P(T_n > q) = 1$ and $\lim_{n \rightarrow \infty} P(\bar{T}_n > q) = 1$ for all $q \in \mathbb{R}$, that is, the proposed tests are consistent.*

The proof is given in Section 3 of the supplement.

The transformation model with a parametric transformation class might be useful in applications even if the model does not hold exactly. With a good choice of θ applying the transformation Λ_θ can reduce the dependence between covariates and errors enormously. Estimating an appropriate θ is much easier than estimating the transformation h nonparametrically. Consequently, one might prefer the semi-parametric transformation model over a completely nonparametric one. It is then of interest how far away we are from the true model. Therefore, in the following, we consider testing precise hypotheses (relevant hypotheses)

$$H'_0 : \min_{\theta \in \Theta} d(\Lambda_\theta, h) \geq \eta \quad \text{and} \quad H'_1 : \min_{\theta \in \Theta} d(\Lambda_\theta, h) < \eta. \quad (24)$$

If a suitable test rejects H'_0 for some small η (fixed beforehand by the experimenter), the model is considered “good enough” to work with, even if it does not hold exactly. To test those hypotheses, we will use the same test statistic as before, but we have to standardize differently. Assume H'_0 , then h is a transformation which does not belong to the parametric class, i.e. the former fixed alternative H_1 holds. Let

$$M(\gamma) = M(c_1, c_2, \theta) = E\{w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2\},$$

and let

$$\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0) := \arg \min_{(c_1, c_2, \theta) \in \mathcal{T}} M(c_1, c_2, \theta).$$

If $\mathbb{R}_+$ and $\mathbb{R}$ are replaced by C_1 and C_2 in the definition of d in (5) one has $\min_{c_1 \in C_1, c_2 \in C_2} M(\gamma) = d(\Lambda_\theta, h)$ for all $\theta \in \Theta$. Assume that

$$\Gamma' = E \left[w(Y_1) \begin{pmatrix} h(Y_1)^2 & h(Y_1) & -h(Y_1)\dot{\Lambda}_{\theta_0}(Y_1) \\ h(Y_1) & 1 & -\dot{\Lambda}_{\theta_0}(Y_1) \\ -h(Y_1)\dot{\Lambda}_{\theta_0}(Y_1)^t & -\dot{\Lambda}_{\theta_0}(Y_1)^t & \Gamma'_{3,3} \end{pmatrix} \right] \quad (25)$$

is positive definite, where $\Gamma'_{3,3} = \dot{\Lambda}_{\theta_0}(Y_1)^t \dot{\Lambda}_{\theta_0}(Y_1) - \ddot{\Lambda}_{\theta_0}(Y_1) \tilde{R}_1$ with

$$\ddot{\Lambda}_\theta(y) = \left(\frac{\partial^2}{\partial \theta_k \partial \theta_\ell} \Lambda_\theta(y) \right)_{k, \ell=1, \dots, d_\Theta}$$

and $\tilde{R}_i = h(Y_i)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_i)$ ($i = 1, \dots, n$).

Theorem 3.5 Assume (A1)–(A4), (A6'), (A8'), let (A7) hold with γ_0 from (A6') and let Γ' be positive definite. Then,

$$n^{1/2}(T_n/n - M(\gamma_0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

with $\sigma^2 = \text{Var}(w(Y_1)\tilde{R}_1^2 + \delta(Z_1))$, where $\delta(Z_1) = 2c_{1,0}E[w(Y_2)\psi(Z_1, U_2)\tilde{R}_2 \mid Z_1]$.

The proof is given in Section 3 of the supplementary material. It is conjectured, that a similar result can be derived for $\bar{T}_n$, although the corresponding Hessian matrix might become more complex.

A consistent asymptotic level- α -test rejects H_0' if $(T_n - n\eta)/(n\hat{\sigma}^2)^{1/2} < u_\alpha$, where u_α is the α -quantile of the standard normal distribution and $\hat{\sigma}^2$ is a consistent estimator of σ^2 . Further research is required on suitable estimators of σ^2 . Let $\hat{\gamma} = (\hat{c}_1, \hat{c}_2, \hat{\theta})^t$ be the minimizer in Eq. (6). For some intermediate sequences $(m_n)_{n \in \mathbb{N}}$, $(q_n)_{n \in \mathbb{N}}$ with $q_n = \lfloor n/m_n \rfloor - 1$ we considered

$$\begin{aligned} \hat{\sigma}^2 := & \frac{1}{q_n} \sum_{s=1}^{q_n} \left(\frac{2\hat{c}_1\sqrt{m_n}}{n} \sum_{k=1}^n w(Y_k) (\hat{h}^{(s)}(Y_k) - \hat{h}(Y_k)) (\hat{h}(Y_k)\hat{c}_1 + \hat{c}_2 - \Lambda_{\hat{\theta}}(Y_k)) \right. \\ & + \frac{1}{\sqrt{m_n}} \sum_{j=(s-1)m_n+1}^{sm_n} \left(w(Y_j) (\hat{h}(Y_j)\hat{c}_1 + \hat{c}_2 - \Lambda_{\hat{\theta}}(Y_j))^2 \right. \\ & \left. \left. - \frac{1}{n} \sum_{i=1}^n w(Y_i) (\hat{h}(Y_i)\hat{c}_1 + \hat{c}_2 - \Lambda_{\hat{\theta}}(Y_i))^2 \right) \right)^2 \end{aligned}$$

as an estimator of σ^2 , where $\hat{h}^{(s)}$ denotes the nonparametric estimator of h depending on the subsample $(Y_{(s-1)m_n+1}, X_{(s-1)m_n+1}), \dots, (Y_{sm_n}, X_{sm_n})$, $s = 1, \dots, q_n$, but suitable choices for m_n are still unclear. Alternatively, a self-normalization approach as in Shao (2010), Shao and Zhang (2010) or Dette et al. (2020) can be applied. For this purpose, let $s \in (0, 1)$ and let $\hat{h}_s$ and $\hat{\gamma}_s = (\hat{c}_{s,1}, \hat{c}_{s,2}, \hat{\theta}_s)^t$ be defined as $\hat{h}$ and $\hat{\gamma}$, but

based on the subsample $(Y_1, X_1), \dots, (Y_{[ns]}, X_{[ns]})$. Moreover, let $K \in \mathbb{N}$, $0 < t_1 < \dots < t_K < 1$ and let ν be a probability measure on $(0, 1)$ with $\nu(\{t_1, \dots, t_K\}) = 1$. Define

$$V_n := \int_0^1 \left(\sum_{k=1}^{[ns]} w(Y_k) (\hat{h}_s(Y_k) \hat{c}_{1,s} + \hat{c}_{2,s} - \Lambda_{\hat{\theta}_s}(Y_k))^2 - s \sum_{k=1}^n w(Y_k) (\hat{h}(Y_k) \hat{c}_1 + \hat{c}_2 - \Lambda_{\hat{\theta}}(Y_k))^2 \right)^2 \nu(ds)$$

as well as

$$\tilde{T}_n := \frac{T_n - nM(\gamma_0)}{\sqrt{V_n}}.$$

In Section 5 of the supplementary material, it is shown that $\tilde{T}_n \xrightarrow{\mathcal{D}} \tilde{T}$ for some random variable $\tilde{T}$ and that the distribution of $\tilde{T}$ does not depend on any unknown parameters. Hence, its quantiles can be simulated and $\tilde{T}_n$ can be used to test for the hypotheses H'_0 and H'_1 .

Remark 3.6 Note that not rejecting the null hypothesis H_0 does not mean that the null hypothesis is valid. Consequently, alternative approaches like for example increasing the level to accept more transformation functions instead of testing for precise hypotheses as in (24) in general do not necessarily result in evidence for applying a transformation model.

4 A bootstrap version and simulations

Although Theorem 3.2 shows how the test statistic behaves asymptotically under H_0 , it is hard to extract any information about how to choose appropriate critical values of a test that rejects H_0 for large values of T_n . The main reasons for this are that first for any function ζ the eigenvalues of the operator defined in Theorem 3.2 are unknown that second this function is unknown and has to be estimated as well, and that third even ψ (which would be needed to estimate ζ) mostly is unknown and rather complex (see e.g. Section 2 of the supplement). Therefore, approximating the α -quantile, say q_α , of the distribution of T in Theorem 3.2 in a direct way is difficult and instead we suggest a smooth bootstrap algorithm to approximate q_α .

Algorithm 4.1 Let $(Y_1, X_1), \dots, (Y_n, X_n)$ denote the observed data, define

$$h_\theta(y) = \frac{\Lambda_\theta(y) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} \quad \text{and} \quad g_\theta(x) = E[h_\theta(Y)|X = x]$$

and let $\hat{g}$ be a consistent estimator of g_{θ_0} , where θ_0 is defined as in (A6) under the null hypothesis and as in (A6') under the alternative. Let κ and ℓ be smooth Lebesgue

densities on $\mathbb{R}^{d_X}$ and $\mathbb{R}$, respectively, where ℓ is strictly positive, κ has bounded support and $\kappa(0) > 0$. Let $(a_n)_n$ and $(b_n)_n$ be positive sequences with $a_n \rightarrow 0$, $b_n \rightarrow 0$, $na_n \rightarrow \infty$, $nb_n^{d_X} \rightarrow \infty$. Denote by $m \in \mathbb{N}$ the sample size of the bootstrap sample.

- (1) Calculate $\hat{\gamma} = (\hat{c}_1, \hat{c}_2, \hat{\theta})^t = \arg \min_{\gamma \in \mathcal{Y}} \sum_{i=1}^n w(Y_i) (\hat{h}(Y_i) c_1 + c_2 - \Lambda_{\theta}(Y_i))^2$. Estimate the parametric residuals $\varepsilon_i(\theta_0) = h_{\theta_0}(Y_i) - g_{\theta_0}(X_i)$ by $\hat{\varepsilon}_i = h_{\hat{\theta}}(Y_i) - \hat{g}(X_i)$ and denote centered versions by $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - n^{-1} \sum_{j=1}^n \hat{\varepsilon}_j$, $i = 1, \dots, n$.
- (2) Generate X_j^* , $j = 1, \dots, m$, independently (given the original data) from the density

$$f_{X^*}(x) = \frac{1}{nb_n^{d_X}} \sum_{i=1}^n \kappa\left(\frac{x - X_i}{b_n}\right)$$

(which is a kernel density estimator of f_X with kernel κ and bandwidth b_n). For $j = 1, \dots, m$ define bootstrap observations as

$$Y_j^* = (h^*)^{-1}(\hat{g}(X_j^*) + \varepsilon_j^*) \quad \text{for } h^*(\cdot) = \frac{\Lambda_{\hat{\theta}}(\cdot) - \Lambda_{\hat{\theta}}(0)}{\Lambda_{\hat{\theta}}(1) - \Lambda_{\hat{\theta}}(0)}, \quad (26)$$

where ε_j^* is generated independently (given the original data) from the density

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{a_n} \ell\left(\frac{\tilde{\varepsilon}_i - \cdot}{a_n}\right)$$

(which is a kernel density estimator of the density of $\varepsilon(\theta_0)$ with kernel ℓ and bandwidth a_n).

- (3) Calculate the bootstrap estimate $\hat{h}^*$ for h^* from (Y_j^*, X_j^*) , $j = 1, \dots, m$.
- (4) Calculate the bootstrap statistic $T_{n,m}^* = \min_{(c_1, c_2, \theta) \in \mathcal{Y}} \sum_{j=1}^m w(Y_j^*) (\hat{h}^*(Y_j^*) c_1 + c_2 - \Lambda_{\theta}(Y_j^*))^2$.
- (5) Let $B \in \mathbb{N}$. Repeat steps (2)–(4) B times to obtain the bootstrap statistics $T_{n,m,1}^*, \dots, T_{n,m,B}^*$. Let q_{α}^* denote the quantile of $T_{n,m}^*$ conditional on (Y_i, X_i) , $i = 1, \dots, n$. Estimate q_{α}^* by

$$\hat{q}_{\alpha}^* = \min \left\{ z \in \{T_{n,m,1}^*, \dots, T_{n,m,B}^*\} : \frac{1}{B} \sum_{k=1}^B I_{\{T_{n,m,k}^* \leq z\}} \geq \alpha \right\}.$$

Remark 4.2

1. The reason for resampling the bootstrap data X_j^* , $j = 1, \dots, m$, nonparametrically consists in the need to mimic the original transformation estimator and its asymptotic behaviour with the bootstrap estimator conditional on the data. Therefore, to proceed in the proof as in Colling and Van Keilegom (2019), it is necessary

to smooth the distribution of X^* . The properties $nb_n^{d_X} \rightarrow \infty$ and $\kappa(0) > 0$ ensure that conditional on the original data $(Y_1, X_1), \dots, (Y_n, X_n)$ the support of X^* contains that of v (from assumption **(B7)** in Section 2 of the supplement) with probability converging to one. Thus, v can be used for calculating $\hat{h}^*$ as well.

2. To proceed as in Algorithm 4.1, it may be necessary to modify h^* so that $S_j^* = \hat{g}(X_j^*) + \varepsilon_j^*$ belongs to the domain of $(h^*)^{-1}$ for all $j = 1, \dots, m$. As long as these modifications do not have any influence on $h^*(y)$ for $y \in \mathcal{Y}_w$, the influence on the $\hat{h}^*$ and $T_{n,m}$ should be asymptotically negligible (which can be proven for the estimator by Colling and Van Keilegom (2019)).

The bootstrap algorithm should fulfil two properties: On the one hand, under the null hypothesis, the algorithm has to provide, conditionally on the original data, consistent estimates of the quantiles of T_n , or rather its asymptotic distribution from Theorem 3.2. On the other hand, to be consistent under H_1 the bootstrap quantiles have to stabilize or at least converge to infinity with a rate less than that of T_n . To formalize this, let $(\Omega, \mathcal{A}, P)$ denote the underlying probability space. Assume that $(\Omega, \mathcal{A})$ can be written as $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ for some measurable spaces $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$. Further, assume that P is characterized as the product of a probability measure P_1 on $(\Omega_1, \mathcal{A}_1)$ and a Markov kernel

$$P_2^1 : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, 1],$$

that is $P = P_1 \otimes P_2^1$. While randomness with respect to the original data is modelled by P_1 , randomness with respect to the bootstrap data and conditional on the original data is modelled by P_2^1 . Moreover, assume

$$P_2^1(\omega, A) = P(\Omega_1 \times A | (Y_1(\omega), X_1(\omega)), \dots, (Y_n(\omega), X_n(\omega)))$$

for all $\omega \in \Omega_1, A \in \mathcal{A}_2$.

With these notations, the assumptions **(A8*)** and **(A9*)** from Section 1 of the supplementary material can be formulated.

Theorem 4.3 *Let q_α^* denote the bootstrap quantile from Algorithm 4.1.*

1. *Assume $H_0, (\mathbf{A1})-(\mathbf{A8}), (\mathbf{A8}^*), (\mathbf{A9}^*)$. Then, q_α^* fulfils*

$$P_1\left(\omega \in \Omega_1 : \limsup_{m \rightarrow \infty} |q_\alpha^* - q_\alpha| > \delta\right) = o(1)$$

for all $\delta > 0$. Hence, $P(T_n > q_\alpha^) = \alpha + o(1)$ under the null hypothesis.*

2. *Assume $H_1, (\mathbf{A1})-(\mathbf{A4}), (\mathbf{A6}'), (\mathbf{A8}^*)$. Then, q_α^* fulfils*

$$P_1\left(\omega \in \Omega_1 : T_n > \limsup_{m \rightarrow \infty} q_\alpha^*\right) = 1 + o(1),$$

so that $P(T_n > q_\alpha^) = 1 + o(1)$ under the alternative.*

The proof is given in the supplement. Since only $\hat{\theta}$ is used to generate the bootstrap observations in Algorithm 4.1, it is conjectured that Theorem 4.3 can be generalized to the usage of $\bar{T}_n$ from (7) in Algorithm 4.1.

Simulations

Throughout this section, $g(X) = 4X - 1$, $X \sim \mathcal{U}([0, 1])$ and $\varepsilon \sim \mathcal{N}(0, 1)$ are chosen. Moreover, the null hypothesis of h belonging to the Yeo and Johnson (2000) transformations

$$\Lambda_{\theta}(Y) = \begin{cases} \frac{(Y+1)^{\theta}-1}{\theta}, & \text{if } Y \geq 0, \theta \neq 0 \\ \log(Y+1), & \text{if } Y \geq 0, \theta = 0 \\ -\frac{(1-Y)^{2-\theta}-1}{2-\theta}, & \text{if } Y < 0, \theta \neq 2 \\ -\log(1-Y), & \text{if } Y < 0, \theta = 2. \end{cases}$$

with parameter $\theta \in \Theta_0 = [0, 2]$ is tested. Under H_0 , we generate data using the transformation $h = (\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0)) / (\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))$ to match the identification constraints $h(0) = 0, h(1) = 1$. Under the alternative, we choose transformations h with an inverse given by the following convex combination,

$$h^{-1}(Y) = \frac{(1-c)(\Lambda_{\theta_0}^{-1}(Y) - \Lambda_{\theta_0}^{-1}(0)) + c(r(Y) - r(0))}{(1-c)(\Lambda_{\theta_0}^{-1}(1) - \Lambda_{\theta_0}^{-1}(0)) + c(r(1) - r(0))} \quad (27)$$

for some $\theta_0 \in [0, 2]$, some strictly increasing function r and some $c \in [0, 1]$. In general, it is not clear if a growing factor c leads to a growing distance (5). Indeed, the opposite might be the case, if r is somehow close to the class of transformation functions considered in the null hypothesis. Simulations were conducted for $r_1(Y) = 5\Phi(Y)$, $r_2(Y) = \exp(Y)$ and $r_3(Y) = Y^3$, where Φ denotes the cumulative distribution function of a standard normal distribution, and $c = 0, 0.2, 0.4, 0.6, 0.8, 1$. The prefactor in the definition of r_1 is introduced because the values of Φ are rather small compared to the values of Λ_{θ} , that is, even when using the presented convex combination in (27), Λ_{θ_0} (except for $c = 1$) would dominate the “alternative part” r of the transformation function without this factor. Note that r_2 and Λ_0 only differ with respect to a different standardization. Therefore, if h is defined via (27) with $r = r_2$ the resulting function is for $c = 1$ close to the null hypothesis case.

For calculating the test statistic the weighting function w was set equal to one. The nonparametric estimator of h was calculated as in Colling and Van Keilegom (2019) (see Section 2 of the supplement for details) with the Epanechnikov kernel $K(y) = \frac{3}{4}(1-y^2)^2 I_{[-1,1]}(y)$ and a normal reference rule bandwidth (see for example Silverman (1986))

$$h_u = \left(\frac{40\sqrt{\pi}}{n} \right)^{\frac{1}{5}} \hat{\sigma}_u, \quad h_x = \left(\frac{40\sqrt{\pi}}{n} \right)^{\frac{1}{5}} \hat{\sigma}_x,$$

where $\hat{\sigma}_u^2$ and $\hat{\sigma}_x^2$ are estimators of the variance of $U = \mathcal{T}(Y)$ and X , respectively. The number of evaluation points N_x for the nonparametric estimator of h was set equal to 100 (see Section 2 of the supplement for details). The integral in (S3) was computed by applying the function *integrate* implemented in *R*. In each simulation run, $n = 100$ independent and identically distributed random pairs $(Y_1, X_1), \dots, (Y_n, X_n)$ were generated as described before, and 250 bootstrap quantiles, which are based on $m = 100$ bootstrap observations $(Y_1^*, X_1^*), \dots, (Y_m^*, X_m^*)$, were calculated as in Algorithm 4.1 using κ the $U([-1, 1])$ -density, ℓ the standard normal density and $a_n = b_n = 0.1$. To obtain more precise estimators of the rejection probabilities under the null hypothesis, 800 simulation runs were performed for each choice of θ_0 under the null hypothesis, whereas in the remaining alternative cases 200 runs were conducted. Among other things the nonparametric estimation of h , the integration in (S3), the optimization with respect to θ and the number of bootstrap repetitions cause the simulations to be quite computationally demanding. Hence, an interface for C++ as well as parallelization were used to conduct the simulations.

The main results of the simulation study are presented in Table 1. There, the rejection probabilities of the settings with $h = (\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0))/(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))$ under the null hypothesis, and h as in (27) under the alternative with $r \in \{r_1, r_2, r_3\}$, $c \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ and $\theta_0 \in \{0, 0.5, 1, 2\}$ are listed. The significance level was set equal to 0.05 and 0.10. Note that the test sticks to the level or is even a bit conservative. Under the alternatives, the rejection probabilities not only differ between different choices of r , but also between different transformation parameters θ_0 that are inserted in (27). While the test shows high power for some alternatives, there are also cases, where the rejection probabilities are extremely small. There are certain reasons that explain these observations. First, the class of Yeo-Johnson transforms seems to be quite general and second the testing approach itself is rather flexible due to the minimization with respect to γ . Having a look at the definition of the test statistic in (6), it attains small values if the true transformation function can be approximated by a linear transformation of $\Lambda_{\tilde{\theta}}$ for some appropriate $\tilde{\theta} \in [0, 2]$. In the following, this issue will be explored further by analysing some graphics. All of the three figures that occur in the following have the same structure and consist of four panels. The upper left panel shows the true transformation function with inverse function (27). Since Y depends on the transformation function, the values of $S = h(Y) = g(X) + \varepsilon$, which are displayed on the vertical axis, are fixed for a comparison of different transformation functions. Due to the choice of $g(X) = 4X - 1$ and $X \sim \mathcal{U}([0, 1])$ the vertical axis reaches from -1 to 3 , which would be the support of $h(Y)$ if the error is neglected. In the upper right panel, the parametric estimator of this function is displayed. Both of these functions are then plotted against each other in the lower left panel by pairing values with the same S component. Finally, the function $Y \mapsto \Lambda_{\theta_0}(Y(\Lambda_{\theta_0}^{-1}(1) - \Lambda_{\theta_0}^{-1}(0)) + \Lambda_{\theta_0}^{-1}(0))$, which represents the part of h corresponding to the null hypothesis, is plotted against the true transformation function in the last panel.

In the lower left panel, one can see if the true transformation function can be approximated by a linear transform of $\Lambda_{\tilde{\theta}}$ for some $\tilde{\theta} \in [0, 2]$, which is an indicator for rejecting or not rejecting the null hypothesis as was pointed out before. As already mentioned, the rejection probabilities not only differ between different deviation functions r , but also within these settings. For example, when considering $r = r_1$ with

Table 1 Rejection probabilities at $\theta_0 \in \{0, 0.5, 1, 2\}$ and $r \in \{r_1, r_2, r_3\}$ for the test statistic T_η

Level α		$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$		$\theta_0 = 2$	
		0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10
r_1	Null hyp.	0.01000	0.04000	0.03125	0.08750	0.03125	0.07750	0.01625	0.05625
	$c = 0.2$	0.000	0.010	0.075	0.105	0.010	0.015	0.000	0.020
	$c = 0.4$	0.000	0.000	0.020	0.045	0.000	0.015	0.120	0.200
	$c = 0.6$	0.100	0.155	0.035	0.050	0.085	0.150	0.415	0.545
	$c = 0.8$	0.685	0.765	0.110	0.210	0.505	0.645	0.785	0.890
r_2	$c = 1$	0.965	0.990	0.925	0.975	0.975	0.985	0.985	0.990
	$c = 0.2$	0.010	0.035	0.030	0.045	0.515	0.640	0.885	0.965
	$c = 0.4$	0.015	0.040	0.000	0.005	0.060	0.135	0.870	0.980
	$c = 0.6$	0.035	0.085	0.000	0.005	0.005	0.005	0.625	0.815
	$c = 0.8$	0.020	0.040	0.010	0.040	0.000	0.005	0.185	0.325
r_3	$c = 1$	0.020	0.065	0.030	0.090	0.025	0.095	0.050	0.105
	$c = 0.2$	0.330	0.505	0.730	0.855	0.810	0.905	0.930	0.995
	$c = 0.4$	0.730	0.865	0.815	0.945	0.875	0.970	0.915	0.990
	$c = 0.6$	0.880	0.940	0.895	0.960	0.950	0.995	0.940	0.990
	$c = 0.8$	0.895	0.965	0.925	0.975	0.935	0.990	0.915	0.980
	$c = 1$	0.980	0.990	0.960	0.990	0.939	0.990	0.940	0.985

Fig. 1 Some transformation functions for $\theta_0 = 0.5$, $c = 0.6$ and $r = r_1$

Fig. 2 Some transformation functions for $\theta_0 = 2$, $c = 0.6$ and $r = r_1$

$c = 0.6$ the rejection probabilities for $\theta_0 = 0.5$ amount to 0.035 for $\alpha = 0.05$ and to 0.050 for $\alpha = 0.10$, while for $\theta_0 = 2$, they are 0.415 and 0.545. Figures 1 and 2 explain why the rejection probabilities differ that much. While for $\theta_0 = 0.5$ the transformation function can be approximated quite well by transforming $\Lambda_{1.06}$ linearly, the best approximation for $\theta_0 = 2$ is given by $\Lambda_{1.94}$ and seems to be relatively bad. The best approximation for $c = 1$ can be reached for θ around 1.4. In contrast to that considering $\theta_0 = 2$ and $r = r_3$ results in a completely different picture. As can be

Fig. 3 Some transformation functions for $\theta_0 = 2$, $c = 0.2$ and $r = r_3$

seen in Fig. 3 even for $c = 0.2$ the resulting h differs so much from the null hypothesis that it cannot be linearly transformed into a Yeo-Johnson transform (see the lower left subgraphic). Consequently, the rejection probabilities are rather high.

A way to overcome this problem can consist in applying the modified test statistic $\tilde{T}_n$ from (7). Although Colling and Van Keilegom (2020) showed that the estimator $\hat{\theta}$ seems to outperform $\tilde{\theta}$ from Remark 2.1 in simulations, fixing c_1, c_2 beforehand might lead due to the reduced flexibility of the minimization procedure to higher rejection probabilities when using $\tilde{T}_n$ instead of T_n . Table 2 contains rejection probabilities which are based on the bootstrap version of $\tilde{T}_n$. The same simulation setting and procedures as before have been used. Indeed, some of the rejection probabilities have increased compared to Table 1. For example, the rejection probabilities for $r = r_1, \theta_0 = 0.5$ and $c = 0.6$ amount to 0.115 and 0.17 instead of 0.0035 and 0.05 in Table 1. Nevertheless, this cannot be generalized since the rejection probabilities when using $\tilde{T}_n$ are sometimes below those for T_n , e.g. for $\theta_0 = 0$ and $r = r_1$ or $\theta_0 = 2$ and $r = r_2$.

Under some alternatives the rejection probabilities are even smaller than the level. This behaviour indicates that from the presented test's perspective, these models seem to fulfil the null hypothesis more convincingly than the null hypothesis models themselves. The reason for this is shown in Fig. 4 for the setting $\theta_0 = 1, c = 0.4$ and $r = r_1$. There, the relationship between the nonparametric estimator of the transformation function and the true transformation function is shown. While the diagonal line represents the identity, the nonparametric estimator seems to flatten the edges of the transformation function. In contrast to this, using $r = r_1$ in (27) steepens the edges so that both effects neutralize each other. Similar effects cause low rejection probabilities for $r = r_2$, although the reasoning is slightly more sophisticated and is also associated with the boundedness of the parameter space $\Theta_0 = [0, 2]$.

Table 2 Rejection probabilities at $\theta_0 \in \{0, 0.5, 1, 2\}$ and $r \in \{r_1, r_2, r_3\}$ for the test statistic $\bar{T}_n$

Level α		$\theta_0 = 0$			$\theta_0 = 0.5$			$\theta_0 = 1$			$\theta_0 = 2$		
		0.05	0.10	0.10	0.05	0.10	0.10	0.05	0.10	0.10	0.05	0.10	0.10
r_1	Null hyp.	0.03375	0.08250	0.08375	0.03875	0.08375	0.08375	0.04250	0.07500	0.07500	0.06000	0.10375	0.10375
	$c = 0.2$	0.025	0.040	0.115	0.035	0.115	0.115	0.055	0.090	0.090	0.230	0.285	0.285
	$c = 0.4$	0.010	0.025	0.155	0.070	0.155	0.155	0.155	0.245	0.245	0.525	0.595	0.595
	$c = 0.6$	0.055	0.130	0.170	0.115	0.170	0.170	0.505	0.585	0.585	0.780	0.840	0.840
	$c = 0.8$	0.235	0.485	0.610	0.515	0.610	0.610	0.840	0.885	0.885	0.925	0.940	0.940
r_2	$c = 1$	0.990	0.995	0.995	0.990	0.995	0.995	0.995	1.000	1.000	0.980	0.985	0.985
	$c = 0.2$	0.035	0.040	0.045	0.035	0.045	0.045	0.185	0.405	0.405	0.795	0.950	0.950
	$c = 0.4$	0.035	0.060	0.060	0.020	0.060	0.060	0.030	0.055	0.055	0.715	0.920	0.920
	$c = 0.6$	0.035	0.065	0.055	0.020	0.055	0.055	0.015	0.050	0.050	0.470	0.705	0.705
	$c = 0.8$	0.050	0.085	0.045	0.035	0.045	0.045	0.030	0.060	0.060	0.105	0.245	0.245
r_3	$c = 1$	0.030	0.060	0.095	0.045	0.095	0.095	0.080	0.130	0.130	0.055	0.125	0.125
	$c = 0.2$	0.230	0.395	0.765	0.560	0.765	0.765	0.725	0.880	0.880	0.900	1.000	1.000
	$c = 0.4$	0.625	0.780	0.915	0.760	0.915	0.915	0.810	0.945	0.945	0.780	0.970	0.970
	$c = 0.6$	0.750	0.935	0.975	0.865	0.975	0.975	0.845	0.985	0.985	0.780	0.985	0.985
	$c = 0.8$	0.810	0.980	0.965	0.825	0.965	0.965	0.805	0.960	0.960	0.810	0.990	0.990
	$c = 1$	0.725	0.970	0.970	0.725	0.970	0.970	0.680	0.970	0.970	0.750	0.990	0.990

Fig. 4 Transformation function for $\theta_0 = 1$, $c = 0.4$ and $r = r_1$ on the horizontal axis and its nonparametric estimator on the vertical axis. The identity is displayed in red

Table 3 Rejection probabilities at $\theta_0 = 1$ and $\theta_0 = 2$ for $r = r_1$

Param.	Alternative Level	Original framework		Modified weighting	
		$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$\theta_0 = 1$	Null hyp.	0.03125	0.07750	0.02875	0.07875
	$c = 0.2$	0.010	0.015	0.040	0.100
	$c = 0.4$	0.000	0.015	0.205	0.320
	$c = 0.6$	0.085	0.150	0.590	0.715
	$c = 0.8$	0.505	0.645	0.950	0.980
	$c = 1$	0.975	0.985	1.000	1.000
$\theta_0 = 2$	Null hyp.	0.01625	0.05625	0.05500	0.10375
	$c = 0.2$	0.000	0.020	0.225	0.350
	$c = 0.4$	0.120	0.200	0.575	0.710
	$c = 0.6$	0.415	0.545	0.910	0.965
	$c = 0.8$	0.785	0.890	0.990	1.000
	$c = 1$	0.985	0.990	0.995	1.000

One possible solution could consist in adjusting the weight function w such that the boundary of the support of Y does no longer belong to the support of w . In Table 3, the rejection probabilities for a modified weighting approach are presented. There, the weight function was chosen such that the smallest five percent and the largest five percent of observations were omitted to avoid the flattening effect of the nonparametric estimation. Indeed, the resulting rejection probabilities under the alternatives increase and lie above those under the null hypotheses.

At last, simulations for precise hypotheses as in (24) were conducted. For sake of brevity, only rejection probabilities resulting from the self normalized test statistic are

Table 4 Rejection probabilities at $\theta_0 = 0.5$ and $\theta_0 = 1$ for precise hypotheses and $r = r_1, r_2, r_3, r_4$

Level α		$\theta_0 = 0$			$\theta_0 = 0.5$			$\theta_0 = 1$			$\theta_0 = 2$		
		Mean (T)		Mean (T)	Mean (T)		Mean (T)	Mean (T)		Mean (T)	Mean (T)		Mean (T)
		0.05	0.10		0.05	0.10		0.05	0.10		0.05	0.10	
r_1	$c = 0$	0.095	0.185	6.564	0.995	1.000	0.886	0.965	0.990	1.446	0.680	0.830	2.629
	$c = 0.2$	0.700	0.800	2.443	0.965	0.975	1.081	0.905	0.970	1.540	0.205	0.310	5.588
	$c = 0.4$	0.790	0.895	2.259	0.930	0.975	1.401	0.385	0.550	3.930	0.000	0.015	14.360
	$c = 0.6$	0.215	0.390	4.638	0.660	0.815	2.581	0.010	0.060	9.477	0.000	0.000	22.146
	$c = 0.8$	0.085	0.185	6.336	0.030	0.050	9.239	0.000	0.000	19.834	0.000	0.000	28.955
r_2	$c = 1$	0.000	0.000	30.996	0.000	0.000	32.081	0.000	0.000	30.907	0.000	0.000	31.710
	$c = 0.2$	0.090	0.130	7.121	0.985	0.995	0.995	0.180	0.320	5.050	0.000	0.000	71.936
	$c = 0.4$	0.035	0.080	7.611	0.905	0.980	1.331	0.785	0.880	2.266	0.000	0.000	38.728
	$c = 0.6$	0.030	0.075	8.080	0.580	0.755	2.899	0.790	0.900	2.089	0.010	0.035	11.850
	$c = 0.8$	0.025	0.045	8.778	0.125	0.245	5.810	0.260	0.380	4.939	0.485	0.610	3.576
r_3	$c = 1$	0.005	0.015	9.254	0.005	0.015	9.550	0.015	0.045	8.756	0.005	0.030	8.785
	$c = 0.2$	0.030	0.060	8.158	0.080	0.155	8.108	0.000	0.000	25.054	0.000	0.000	197.373
	$c = 0.4$	0.000	0.000	18.715	0.000	0.000	29.072	0.000	0.000	56.932	0.000	0.000	243.476
	$c = 0.6$	0.000	0.000	51.102	0.000	0.000	75.638	0.000	0.000	106.332	0.000	0.000	246.010
	$c = 0.8$	0.000	0.000	118.049	0.000	0.000	135.888	0.000	0.000	166.425	0.000	0.000	259.728
r_4	$c = 1$	0.000	0.000	248.025	0.000	0.000	262.134	0.000	0.000	254.237	0.000	0.000	236.894
	$c = 0.2$	0.475	0.635	3.436	0.995	1.000	0.942	0.960	0.980	1.265	0.685	0.820	2.695
	$c = 0.4$	0.815	0.915	1.920	0.995	1.000	0.972	0.880	0.940	1.793	0.420	0.570	4.062
	$c = 0.6$	0.580	0.780	2.943	0.975	0.995	1.124	0.590	0.775	2.773	0.135	0.235	6.390
	$c = 0.8$	0.155	0.310	5.137	0.800	0.895	2.185	0.120	0.295	5.528	0.060	0.100	8.490
	$c = 1$	0.010	0.015	10.722	0.015	0.015	10.519	0.000	0.020	10.680	0.015	0.030	10.211

presented since this approach seems to outperform that based on the estimator $\hat{\sigma}^2$ from Section 3 by far in the simulated settings. Since only a fraction of the data is used to calculate V_n , the sample size was increased to $n = 500$. The settings and techniques remain the same as before. The probability measure ν was set to

$$\nu = \frac{1}{10}\delta_{0.6} + \frac{2}{10}\delta_{0.7} + \frac{3}{10}\delta_{0.8} + \frac{4}{10}\delta_{0.9}$$

to put a higher weight on those parts of V_n where more data points are used. Furthermore, the threshold was chosen to be $\eta = 0.02$, which roughly corresponds to plugging the logit-function $r_4(y) := \frac{5\exp(y)}{1+\exp(y)}$ and $c = 1$ into Eq. (27) and calculating $\min_{\theta \in \Theta} d(\Lambda_\theta, h)$. Hence, we expect the test to reject the null hypothesis H'_0 if $T_n < n\eta = 10$ holds.

A detailed analysis would go beyond the scope of this manuscript, so that only some rejection probabilities are given in Table 4. Moreover, the mean values of the test statistic T_n are listed to link the rejection probabilities to the distance between the expected value of the test statistic and the threshold of $\eta = 0.02$. First, the smaller the value of T_n is the more likely the test seems to reject the null hypothesis H'_0 . Further, the test holds the level, but is slightly conservative. Alternatives seem to be detected for mean values of T_n around or below eight. Nevertheless, the power of the test is quite high in scenarios with small expected values of the test statistic, which often corresponds to transformations functions which are close to the parametric class. For $\theta = 0.5$ and $\theta = 1$ the rejection probabilities are in these cases above 0.90 and sometimes even close to one. Although the influence of simulation parameters such as the sample size n or the probability measure ν has not been examined, the results indicate that using the self normalized test statistic can be a good way to test for the precise hypotheses H'_0 and H'_1 .

Acknowledgements Natalie Neumeyer acknowledges financial support by the DFG (Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency*). Ingrid Van Keilegom acknowledges financial support by the European Research Council (2016–2021, Horizon 2020/ERC Grant Agreement No. 694409). Moreover, we would like to thank the associate editor and two anonymous referees for their very helpful suggestions and comments on the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Allison JS, Hušková M, Meintanis SG (2018) Testing the adequacy of semiparametric transformation models. *TEST* 27(1):70–94

- Berger JO, Delampady M (1987) Testing precise hypotheses. *Stat Sci* 2(3):317–335
- Box GEP, Cox DR (1964) An analysis of transformations. *J R Stat Soc B* 26(2):211–252
- Carroll RJ, Ruppert D (1988) Transformation and weighting in regression. CRC Press, Boca Raton
- Chen S (2002) Rank estimation of transformation models. *Econometrica* 70:1683–1697
- Chiappori P-A, Komunjer I, Kristensen D (2015) Nonparametric identification and estimation of transformation. *J Economet* 188(1):22–39
- Colling B, Van Keilegom I (2016) Goodness-of-fit tests in semiparametric transformation models. *TEST* 25(2):291–308
- Colling B, Van Keilegom I (2017) Goodness-of-fit tests in semiparametric transformation models using the integrated regression function. *J Multivar Anal* 160:10–30
- Colling B, Van Keilegom I (2019) Estimation of fully nonparametric transformation models. *Bernoulli* 25:3762–3795
- Colling B, Van Keilegom I (2020) Estimation of a semiparametric transformation model: a novel approach based on least squares minimization. *Electron J Stat* 14:769–800
- Dette H, Kokot K, Volgushev S (2020) Testing relevant hypotheses in functional time series via self-normalization. *J R Stat Soc* 82:629–660
- Heuchenne C, Samb R, Van Keilegom I (2015) Estimating the error distribution in semiparametric transformation models. *Electron J Stat* 9:2391–2419
- Hoeffding W (1948) A class of statistics with asymptotically normal distribution. *Ann Math Stat* 19(3):293–325
- Horowitz JL (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64(1):103–137
- Horowitz JL (2009) Semiparametric and nonparametric methods in econometrics. Springer, Berlin
- Hušková M, Meintanis SG, Neumeyer N, Pretorius C (2018) Independence tests in semiparametric transformation models. *S Afr J Stat* 52:1–13
- Kloodt N, Neumeyer N (2020) Specification tests in semiparametric transformation models—a multiplier bootstrap approach. *Comput Stat Data Anal* 145
- Lakens D (2017) Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc Psychol Personal Sci* 8(4):355–362
- Lee AJ (1990) U-Statistics: theory and practice. Dekker, New York City
- Lewbel A, Lu X, Su L (2015) Specification testing for transformation models with an application to generalized accelerated failure-time models. *J Economet* 184(1):81–96
- Linton O, Sperlich S, Van Keilegom I (2008) Estimation of a semiparametric transformation model. *Ann Stat* 36(2):686–718
- Mu Y, He X (2007) Power transformation toward a linear regression quantile. *J Am Stat Assoc* 102(477):269–279
- Neumeyer N, Noh H, Van Keilegom I (2016) Heteroscedastic semiparametric transformation models: estimation and testing for validity. *Stat Sin* 26:925–954
- Powell J (1991) Estimation of monotonic regression models under quantile restrictions. In: Barnett WA, Powell J, Tauchen GE (eds) Nonparametric and semiparametric methods in econometrics and statistics: proceedings of the 5th international symposium on economic theory and econometrics, pp 357–384
- Shao X (2010) A self-normalized approach to confidence interval construction in time series. *J R Stat Soc B* 72(3):343–366
- Shao X, Zhang X (2010) Testing for change points in time series. *J Am Stat Assoc* 105(491):1228–1240
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Szydlowski A (2020) Testing a parametric transformation model versus a nonparametric alternative. *Economet Theory*. <https://doi.org/10.1017/S0266466619000355>
- Wellner JA (2005) Empirical processes: theory and applications. <https://www.stat.washington.edu/jaw/RESEARCH/TALKS/Delft/emp-proc-delft-big.pdf>
- Witting H, Müller-Funk U (1995) Mathematical statistics II—Asymptotic statistics: parametric models and nonparametric functionals. B. G Teubner
- Yeo I-K, Johnson RA (2000) A new family of power transformations to improve normality or symmetry. *Biometrika* 98(4):954–959