Specification tests in semiparametric transformation models

NICK KLOODT and NATALIE NEUMEYER
Department of Mathematics, University of Hamburg
July 10, 2018

Abstract
We consider semiparametric transformation models, where after pre-estimation of a parametric transformation of the response the data are modeled by means of nonparametric regression. We suggest subsequent procedures for testing lack-of-fit of the regression function and for significance of covariables, which – in contrast to procedures from the literature – are asymptotically not influenced by the pre-estimation of the transformation. The test statistics are asymptotically pivotal, have the same asymptotic distribution as in regression models without transformation, and standard wild bootstrap can be applied to the transformed data to conduct the tests.

Key words: Box-Cox transformation, lack-of-fit test, nonparametric regression, post model-selection inference, significance of covariables, U-statistics, wild bootstrap, Yeo-Johnson transformation

1 Introduction
Assume we have observed independent data \((X_i, Y_i), i = 1, \ldots, n\), and after a transformation of the response a regression model shall be fitted. The aim of the data transformation typically is to obtain a simpler model, e.g. a homoscedastic instead of a heteroscedastic model. If the transformation is chosen from a parametric class \(\{\Lambda_\theta \mid \theta \in \Theta\}\) (e.g. Box-Cox power transformations, see Box and Cox, 1964, or their modification suggested by Yeo and Johnson, 2000) one typically assumes the existence of a unique ‘true parameter’ \(\theta_0 \in \Theta\) such that the simpler model holds for the transformed data \((X_i, \Lambda_{\theta_0}(Y_i)), i = 1, \ldots, n\). We will assume a homoscedastic model

\[ \Lambda_{\theta_0}(Y_i) = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \]

where \(m\) denotes the regression function and \(\varepsilon_i\) some unobservable centered error, independent from the covariates \(X_i\). Data-dependent choices of the transformation parameter as considered by

*Financial support by the DFG (Research Unit FOR 1735 Structural Inference in Statistics: Adaptation and Efficiency) is gratefully acknowledged.
Linton, Sperlich and Van Keilegom (2008), among others. However, leave subsequent inference to
be based on \((X_i, \Lambda_{\hat{\theta}}(Y_i)), i = 1, \ldots, n\), where \(\hat{\theta}\) depends on the whole sample \((X_1, Y_1), \ldots, (X_n, Y_n)\).

It would be desirable to be able to apply standard procedures to the transformed data. Yet, the
random transformation may influence the performance of inference procedures severely. For the
famous Box-Cox transformations this phenomenon has been a major discussion topic (see, e.g.
Bickel and Doksum, 1981, or Hinkley and Runger, 1984), but is often ignored in practice. Note
that estimating the transformation is a problem of model selection. The influence of the random
selection procedure on post-model-selection inference is a topic of current high interest, see, e.g.
Berk et al. (2013), Efron (2014) and Lee et al. (2016). Naturally inferential procedures that are
not influenced by the randomness of the model selection (here estimation of the transformation
parameter) have the advantage of ready applicability.

For the model at hand we will consider two typical testing problems in nonparametric regression
models, namely testing for a parametric class of the regression function (lack-of-fit) and testing for
significance of covariables. We will present test statistics that are asymptotically not influenced by
the randomness of the data transformation.

Concerning the first testing problem, recently Colling and Van Keilegom (2016, 2017) suggested
lack-of-fit tests for the regression function \(m\) in a transformation model. The tests are based on
ideas from Van Keilegom, González-Manteiga and Sánchez Sellero (2008), on the one hand, and
from Bierens (1982), Stute (1997) and Escanciano (2006), on the other hand (in models without
transformation). Colling and Van Keilegom (2016, 2017) derive the asymptotic distribution of
the test statistics and show that the estimation of the transformation parameter alters the limit
distribution in both cases. Even when bootstrap is conducted to apply the tests the bootstrap
versions of the original tests (without transformation) cannot be applied to the transformed data,
but the bootstrap procedures have to be adapted for the transformation model as well. In particular,
for each bootstrap replication a new transformation estimation has to be performed. This can be
computationally quite demanding as the estimation is based on nonlinear optimization.

Concerning the second testing problem, Allison, Huškova and Meintanis (2017) consider testing
for significance of covariables in semiparametric transformation models based on ideas from Bierens
(1982) and Hlávka, Hušková, Kirch and Meintanis (2017). They derive the asymptotic distribution
of the test statistic which again is influenced by the transformation estimation. The bootstrap
procedure is adapted to the unknown transformation as well.

The changes in the asymptotic distributions due to the transformation estimation and the nec-
essary modification of standard procedures seems to be rather inconvenient for applications. They
are contrary to the expectation that with the transformation one obtains a simple model, for that
standard inference procedures can be applied. On account of this we will suggest testing proce-
dures that are asymptotically not influenced by the random transformation. Therefore after the
data transformation standard procedures or their standard bootstrap versions can be applied. For
testing for a parametric class of the regression functions we will generalize Härdle and Mammen’s
(1993) test (see also Alcalá, Cristóbal and González-Manteiga, 1999) as well as Zheng’s (1996) test.
For testing significance of covariables we will generalize Lavergne, Maistre and Patilea’s (2015)
procedure.

In section 2 we will define the semiparametric transformation model and briefly discuss the estimation of the transformation parameter. In section 3 we suggest two lack-of-fit tests for the regression function after transformation, while in section 4 we consider testing for significance of covariables. In both settings we prove asymptotic normality under the null hypothesis and local alternatives. In section 5 we discuss wild bootstrap versions and we compare the suggested tests with those from the literature in a simulation study. Appendix A contains the assumptions and appendix B the proofs.

2 Estimation of the transformation

Throughout we assume we have independent realisations \((X_1, Y_1), \ldots, (X_n, Y_n)\) of the model

\[
\Lambda_{\theta_0}(Y) = m(X) + \varepsilon,
\]

where \(Y\) and \(\varepsilon\) are \(\mathbb{R}\)-valued and \(X\) is \(\mathbb{R}^d\)-valued random variable. Moreover, \(\varepsilon\) and \(X\) are independent and \(E[\varepsilon] = 0, \text{Var}(\varepsilon) = \sigma^2 \in (0, \infty)\). The transformation belongs to a class \(\{\Lambda_{\theta} | \theta \in \theta\}\) and the true parameter \(\theta_0\) is unknown. The regression function \(m\) is estimated nonparametrically.

There are several possibilities to estimate the transformation parameter. For our approaches the only property the estimator \(\hat{\theta}\) has to fulfill is root-\(n\)-consistency, i.e.

\[
\hat{\theta} = \theta_0 + O_P(n^{-1/2}).
\]  

This is fulfilled for the profile likelihood and the minimum distance estimator considered in Linton et al. (2008). See Colling and Van Keilegom (2016) for a detailed description of the profile likelihood estimator that we use in our simulations, and for regularity assumptions to obtain (2.2). In a work in progress Benjamin Colling and Ingrid Van Keilegom consider an alternative estimator for \(\theta\) that also fulfills (2.2) and seems to perform better in simulations than the profile likelihood estimator.

For transformation parameter estimation in other semiparametric models, see, e.g. Horowitz (1996) or Linton, Chen, Wang and Härdle (1997). Some parametric classes for transformations are considered by Box and Cox (1964), Zellner and Revankar (1969), Bickel and Doksum (1981) and Yeo and Johnson (2000), among others.

3 Lack-of-fit testing

For model (2.1) we consider tests for the hypothesis

\[
H_0 : \quad m \in \{m(\cdot, \beta) : \beta \in B\}
\]

for some compact parameter space \(B\). The function \(m(\cdot, \beta)\) is known apart from the true regression parameter \(\beta_0\) with \(m(\cdot) = m(\cdot, \beta_0)\). Testing lack-of-fit (in models without transformation) is a classical topic in statistics. A very thorough review on related literature is given in González-Manteiga and Crujeiras (2013). The most commonly used approaches (that also have been very
influential in terms of development of related statistics in different contexts) are arguably those by Härdle and Mammen (1993) (based on $L^2$ distance), Zheng (1996) (based on $U$-statistics) and Stute (1997) (based on a marked empirical process). We take the two first approaches in a model with unknown transformation, while the latter one was considered by Colling and Van Keilegom (2017).

To this end we consider least squares estimators for the regression parameter based on the transformed data $(X, \Lambda_\theta(Y))$ for each $\theta \in \Theta$,

$$\hat{\beta}_\theta = \arg\min_{\beta \in B} \sum_{i=1}^n (\Lambda_\theta(Y_i) - m(X_i, \beta))^2.$$ 

We further define local alternatives

$$m(x) = m(x, \bar{\beta}) + c_n \Delta_n(x)$$

with parameter $\bar{\beta} = \arg\min_{\beta \in B} \int (m(x) - m(x, \beta))^2 f(x) \, dx$, rate $c_n \sim n^{-\frac{1}{2}} h^{-\frac{d}{4}}$ and $\Delta_n(x)$ uniformly bounded in $n$ and $x$. Note that $\bar{\beta}$ may depend on $n$. These local alternatives contain as special case the null hypothesis with $\Delta_n \equiv 0$ and $\bar{\beta} = \beta_0$. Now we define Härdle and Mammen’s (1993) test statistic

$$T_n(\hat{\theta}) = nh^d \int \left( \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)(\Lambda_\theta(Y_i) - m(X_i, \hat{\beta}_\theta)) \right)^2 \, dx$$

and Zheng’s (1996) test statistic

$$V_n(\hat{\theta}) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_h(X_i - X_j)(\Lambda_\theta(Y_i) - m(X_i, \hat{\beta}_\theta))(\Lambda_\theta(Y_j) - m(X_j, \hat{\beta}_\theta))$$

both applied to the transformed data $(X, \Lambda_\theta(Y))$. Here $K_h(\cdot) = K(\cdot/h)/h^d$, where $K : \mathbb{R}^d \rightarrow \mathbb{R}$ denotes a kernel function and $h = h_n$ a sequence of bandwidths fulfilling assumptions [A3] and [A4] in appendix A. Note that in contrast to Härdle and Mammen (1993) we avoid the kernel density estimator in the denominator. As test statistics we consider $T_n(\hat{\theta})$ and $V_n(\hat{\theta})$ for some estimator $\hat{\theta}$ that fulfills (2.2), whereas the asymptotic distributions of $T_n(\theta_0)$ and $V_n(\theta_0)$ are given by Härdle and Mammen (1993) and Zheng (1996), respectively (in a model without transformation).

**Theorem 3.1** Under the assumptions [A1]–[A9] in appendix A, we have $T_n(\hat{\theta}) - T_n(\theta_0) = o_P(1)$ and $V_n(\hat{\theta}) - V_n(\theta_0) = o_P((nh^{d/2})^{-1})$.

The proof is given in appendix B. The theorem shows that the asymptotic distribution is not influenced by the estimation of the transformation. The reason is essentially the faster convergence of the transformation parameter estimator compared to the convergence rate of the test statistics. From Härdle and Mammen (1993) and Zheng (1996) now directly follows the next result.

**Corollary 3.2** Under the assumptions of theorem 3.1 we have for $n \rightarrow \infty$

$$\frac{T_n(\hat{\theta}) - b_n - \mu_n}{\sqrt{V_n}} \overset{D}{\rightarrow} \mathcal{N}(0, 1)$$
\[ \frac{nh^{d/2}V_n(\hat{\theta}) - \mu_n}{\sqrt{\Sigma}} \xrightarrow{d} \mathcal{N}(0,1) \]

with \( b_h = h^{-d/2}\sigma^2 \int K^2(u) du \), \( \mu_n = E[\Delta_n(X_1)^2 f_X(X_1)] \), \( V = 2\sigma^4 \int f_X(x)^2 dx \int (K \ast K)^2(x) dx \), where \((K \ast K)(x) = \int K(x-u)K(u) du\), and \( \Sigma = 2\sigma^4 \int f_X(x)^2 dx \int K^2(u) du \).

Asymptotic level \( \alpha \)-tests for the null hypothesis \( H_0 \) can be constructed from the corollary. To this end, note that with the methods used in the proof of Theorem 3.1 it is easy to show that consistent estimators (under the null) for \( \sigma^2 \) and \( \Sigma \) are given by \( \hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (\Lambda_\theta(Y_i) - m(X_i, \hat{\theta}))^2 \) and

\[ \hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j\neq i}^n K_h(X_i - X_j)(\Lambda_\theta(Y_i) - m(X_i, \hat{\theta}))^2(\Lambda_\theta(Y_j) - m(X_j, \hat{\theta}))^2, \]

respectively. Further \( \hat{s} = n^{-2} \sum_{i=1}^n \sum_{j=1}^n K_h(X_j - X_i) \) consistently estimates \( \int f_X(x)^2 dx \). Note that Zheng (1996) uses the estimator \( \hat{\Sigma}(\theta_0) \) in a heteroscedastic model. In our model, instead of \( \hat{\Sigma}(\hat{\theta}) \), the estimator \( \hat{\Sigma} = 2\hat{\sigma}^4 \hat{s} \int K^2(y) dy \) can be applied as well. We demonstrate the finite sample behavior of the asymptotic tests as well as of wild bootstrap versions in section 5.

The correlation further shows that both tests can detect local alternatives of rate \( (nh^{-d/2})^{-1/2} \). The approaches considered by Colling and Van Keilegom (2016, 2017) can detect faster local alternatives of \( n^{-1/2} \)-rate, but the asymptotic distributions depend in a complicated way on the estimation of the transformation. Further, the estimation of the transformation needs also to be taken into account in the bootstrap procedure.

Other test statistics could be considered as well, e.g. the empirical characteristic function approach by Husková and Meintanis (2009). We conjecture that with this approach the asymptotic distribution will depend on the transformation parameter estimate.

### 4 Testing significance of covariables

Under model (2.1) let \( X = (W, Z) \), where the entries of \( X \) are in such order that the hypothesis of significance of \( Z \) is of interest, i.e.

\[ H_0 : E[\Lambda_{\theta_0}(Y)|W, Z] = E[\Lambda_{\theta_0}(Y)|W]. \]

Note that \( m(X) = m(W, Z) = E[\Lambda_{\theta_0}(Y)|W, Z] \) and throughout we use the definition \( r(W) = E[\Lambda_{\theta_0}(Y)|W] \). Further we will denote the density of \( W \) by \( f_W \) and the conditional density of \( W \), given \( Z = z \), by \( f_{W|Z}(\cdot|z) \).

Literature overviews for testing for the simplifying hypothesis \( H_0 \) (in models without transformation) are given by González-Manteiga and Crujeiras (2013) and Lavergne et al. (2015). Lavergne et al.’s (2015) test is similar to those by Fan and Li (1996) and Lavergne and Vuong (2000) based on U-statistics, but does not involve smoothing with respect to \( Z \) and therefore converges at a faster rate, independent from the dimension of \( Z \).
Assume that $W$ is $p$- and $Z$ $q$-dimensional. Like Lavergne et al. (2015) we consider the local alternative

$$m(W, Z) = r(W) + \delta_n d(W, Z)$$

for some fixed integrable function $d : \mathbb{R}^{p+q} \to \mathbb{R}$ with $E[d(W, Z)] = 0$, but concentrate on the rate $\delta_n \sim n^{-\frac{2}{7}h^{-\frac{2}{7}}}$. The local alternative contains the null hypothesis for $d = 0$.

We define

$$I_n(\theta) = \frac{h^p}{n^{3}} \sum_{i,j,k,l} (\Lambda_\theta(Y_i) - \Lambda_\theta(Y_k))(\Lambda_\theta(Y_j) - \Lambda_\theta(Y_l)) L_g(W_i - W_k)L_g(W_j - W_l)$$

$$\times K_h(W_i - W_j)\psi(Z_i - Z_j).$$

As before, $K_h(\cdot) = K(\cdot/h)/h^p$ and $L_g(\cdot) = L(\cdot/g)/g^p$, where $K$ and $L$ denote bounded and symmetric kernel functions of $p$ variables with compact supports together with two bandwidths $g = g_n$ and $h = h_n$. Let $\psi$ be a bounded, symmetric function with almost everywhere positive Fourier transform. The notation $\sum_{\neq}^{}$ stands for summation over pairwise distinct indices.

Using a transformation parameter estimator $\hat{\theta}$ as in (2.2), we consider $I_n(\hat{\theta})$ as test statistic, which is asymptotically equivalent to $I_n(\theta_0)$ by the following theorem.

**Theorem 4.1** Under the assumptions of theorem 4.1 in appendix A, we have $I_n(\hat{\theta}) - I_n(\theta_0) = o_p(1)$.

The proof of the theorem is given in appendix B, while the following corollary is a direct consequence of Lavergne et al. (2015).

**Corollary 4.2** Under the assumptions of theorem 4.1 we have for $n \to \infty$

$$\frac{I_n(\hat{\theta}) - \mu}{\omega} \xrightarrow{d} N(0, 1)$$

with

$$\mu = E \left[ \int d(w, Z_1)d(w, Z_2)f_W^2(w)f_{W|Z}(w|Z_1)f_{W|Z}(w|Z_2)\psi(Z_1 - Z_2) \right]$$

$$\omega^2 = 2\sigma^4 E \left[ f_W^4(w)f_{W|Z}(w|Z_1)f_{W|Z}(w|Z_2)\psi(Z_1 - Z_2) \right] \int K^2(u) du.$$

The asymptotic variance can be consistently estimated by

$$\hat{\omega} = \frac{2h^p}{n^{3}} \sum_{i,j,k,l,k',l'} (\Lambda_\hat{\theta}(Y_i) - \Lambda_\hat{\theta}(Y_k))(\Lambda_\hat{\theta}(Y_j) - \Lambda_\hat{\theta}(Y_l))(\Lambda_\hat{\theta}(Y_i) - \Lambda_\hat{\theta}(Y_k'))(\Lambda_\hat{\theta}(Y_j) - \Lambda_\hat{\theta}(Y_l'))$$

$$\times L_g(W_i - W_k)L_g(W_i - W_k')L_g(W_j - W_l)L_g(W_j - W_l')(K_h(W_i - W_j))^2 \psi^2(Z_i, Z_j)$$

in order to apply asymptotic tests. In the next section we consider a wild bootstrap version of the test.

In contrast to the empirical characteristic function approach by Allison et al. (2017) the asymptotic distribution of the test we suggest is not influenced by the pre-selection of the transformation.
That means that the test by Lavergne et al. (2015) (either using the asymptotic normality or a standard wild bootstrap) can simply be applied to the randomly transformed data.

Further approaches for testing significance could be followed, e.g. the marked empirical process approach by Delgado and González-Manteiga (2001). However, we conjecture that for this approach the estimation of the transformation has to be taken into account in the asymptotic distribution and the bootstrap procedure.

5 Bootstrap versions and simulations

Evaluating the performance of our procedures involves studying their behaviour for small sample sizes. Therefore in this section we perform some simulations in order to compare them to the other approaches already mentioned. In both of the following sections we try to mimic the simulation settings of, on the one hand, Colling and Van Keilegom (2017) and, on the other hand, Allison et al. (2017). As in these papers we use the language R (R Core Team, 2013) for our simulations. Throughout this chapter we consider the Yeo-Johnson transformation (see Yeo and Johnson, 2000)

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

5.1 Lack-of-Fit Testing

Let \( m_\theta(x) = E[\Lambda_\theta(Y)|X = x] \) denote the expectation of \( \Lambda_\theta(Y) \) conditioned on \( X = x \). Like Colling and Van Keilegom (2017) we use the profile-likelihood estimator developed by Linton et al. (2008) for our simulations, that is

\[
\hat{\theta} = \arg \max_\theta \sum_{i=1}^n \left( \log \hat{f}_{\varepsilon(\theta)}(\Lambda_\theta(Y_i) - \hat{m}_\theta(X_i)) + \log \Lambda'_\theta(Y_i) \right),
\]

where \( \hat{f}_{\varepsilon(\theta)} \) is an estimator for the density of \( \varepsilon(\theta) = \Lambda_\theta(Y) - m_\theta(X) \) and \( \hat{m}_\theta \) denotes an estimator for \( m_\theta \). Here we implement \( \hat{f}_{\varepsilon(\theta)} \) as an ordinary kernel estimator with Epanechnikov kernel and a bandwidth following the normal reference rule. Further, we use a local linear estimator \( \hat{m}_\theta \) (see Fan and Gijbels, 1996, for a detailed description) with a bandwidth obtained by cross validation. Note that the bandwidths also may depend on \( \theta \).

Härdle and Mammen (1993) already mentioned in their article that due to the slow convergence rate of the negligible terms the asymptotic distribution may be inappropriate for describing the test statistic’s behaviour or obtaining critical values. Hence, they suggested a wild bootstrap procedure that is based on the ideas of Wu (1986).

We will call this approach ‘standard wild bootstrap’ (swb) in what follows. To describe it in our context denote the transformed data as \( (X_i, Z_i), i = 1, \ldots, n \), with \( Z_i = \Lambda_\hat{\theta}(Y_i) \), and define
nonparametric estimates of the residuals \( \hat{\varepsilon}_i = Z_i - \hat{m}_\theta(X_i) \), where \( \hat{m} \) denotes the Nadaraya-Watson estimator for \( \hat{m} \) (see Nadaraya, 1964, or Watson, 1964). Further generate independent random variables \( \varepsilon_i^* \) with
\[
E^*[\varepsilon_i^*] = 0, \quad E^*[(\varepsilon_i^*)^2] = \hat{\varepsilon}_i^2, \quad \text{and} \quad E^*[(\varepsilon_i^*)^3] = \hat{\varepsilon}_i^3,
\]
where \( E^* \) denotes conditional expectation given the sample \((X_i, Y_i), i = 1, \ldots, n\). Now use the new sample \((X_i, Z_i^*) = m(X_i, \hat{\theta}) + \varepsilon_i^* \) as bootstrap observations. The bootstrap versions of the test statistics are now defined as
\[
T_n^* = n h^2 \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(Z_i^* - m(X_i, \hat{\theta}^*)) \right)^2 dx
\]
\[
V_n^* = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(Z_i^* - m(X_i, \hat{\theta}^*))(Z_j^* - m(X_j, \hat{\theta}^*)),
\]
where \( \hat{\theta}^* \) is evaluated from \((X_i, Z_i^*), i = 1, \ldots, n\). Note that the bootstrap, in contrast to the one presented by Colling and Van Keilegom (2017), does not take account of the estimation of the transformation parameter. This makes our procedure much faster and easier to implement. To show that our testing method also results in good rejection probabilities, we compare in the following the empirical rejection probabilities of our approach with the ones obtained by Colling and Van Keilegom (2017). From now on we consider the model \( \Lambda_0(Y) = m(X) + \varepsilon \) with the null hypothesis
\[
m \in \{m(\cdot, \beta) : \beta \in B\} = \{x \mapsto \beta_1 + \beta_2 x : \beta_1, \beta_2 \in \mathbb{R} \}
\]
for the true regression parameter \( \beta_0 = (3, 5)^t \). In order to examine the performance of the test under several alternatives we add the same deviation functions \( \Delta \) as Colling and Van Keilegom (2017) namely \( \Delta(x) = 2x^2, \Delta(x) = 3x^2, \Delta(x) = 4x^2, \Delta(x) = 5x^2, \Delta(x) = 2 \exp(x), \Delta(x) = 3 \exp(x), \Delta(x) = 4 \exp(x), \Delta(x) = 5 \exp(x), \Delta(x) = 0.25 \sin(2\pi x), \Delta(x) = 0.5 \sin(2\pi x), \Delta(x) = 0.75 \sin(2\pi x) \) and \( \Delta(x) = \sin(2\pi x) \). Here \( X \) and \( \varepsilon \) follow a uniform distribution on \([0, 1]\) and a standard normal distribution truncated on \([-3, 3]\), respectively.

First, tables [1] and [5] show the (empirical) rejection probabilities of the tests \( T_n(\hat{\theta}) \) and \( V_n(\hat{\theta}) \) developed in section 3 this paper. The simulations are conducted for a sample size of \( n = 200 \) with \( B = 1000 \) bootstrap repetitions in each of 500 simulation runs. Further, the transformation parameter is chosen to be equal to \( \theta_0 = 0, \theta_0 = 0.5 \) and \( \theta_0 = 1 \). The significance level is chosen to be \( \alpha = 0.10 \). We show rejection probabilities based on the standard wild bootstrap (swb) and based on the asymptotic distribution (asym) as given in Corollary 3.2. Even though Colling and Van Keilegom (2017) considered residuals that were truncated at \( \pm 3 \) it does not seem to us as if such a truncation is necessary in our context as the rejection probabilities are very similar (see table [4]). Nevertheless we use truncated residuals in order to keep the models optimally comparable. In the last column we write the rejection probabilities of the test based on the test statistic \( W_1^2 \) from Colling and Van Keilegom (2017) as this test seems to perform best in their article.

Although the estimation of the transformation parameter does not have any influence on the asymptotic distribution of the test statistics we also carry out simulations with a bootstrap procedure
similar to the one described by Colling and Van Keilegom (2017). We will call it ‘transformation wild bootstrap’ (twb) in what follows. To this end define $$\varepsilon_i^* := \zeta_i^* + a_n \xi_i^*$$, where the $$\xi_i^*$$ are standard normally distributed random variables independent of $$(X_i, Y_i)$$, $$i = 1, \ldots, n$$, and the $$\zeta_i^*$$ are drawn with replacement from the nonparametrically estimated residuals. Like Colling and Van Keilegom we choose $$a_n = 0.1$$. Then the bootstrap sample $$(X_i^*, Y_i^*)$$ is obtained by $$X_i^* = X_i$$ and $$Y_i^* = \Lambda_{\hat{\theta}^{-1}}(m(X_i^*, \hat{\beta}) + \varepsilon_i^*)$$. The bootstrap test statistics are then defined as

$$T_n^*\left(\hat{\theta}^*\right) = nh^{-2} \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(\Lambda_{\hat{\theta}^*}(Y_i^*) - m(X_i, \hat{\beta}_{\hat{\theta}^*})) \right)^2 dx$$

$$V_n^*\left(\hat{\theta}^*\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_h(X_i - X_j)(\Lambda_{\hat{\theta}^*}(Y_i^*) - m(X_i, \hat{\beta}_{\hat{\theta}^*}))(\Lambda_{\hat{\theta}^*}(Y_j^*) - m(X_j, \hat{\beta}_{\hat{\theta}^*}))$$

where $$\hat{\theta}^*$$ is the transformation parameter estimator built from the bootstrap sample. The resulting rejection probabilities for $$T_n(\hat{\theta})$$ and $$V_n(\hat{\theta})$$ are also given in tables 1–3 and are denoted by twb.

If we neglect for a moment the columns denoted by twb, tables 1–3 show a few patterns. Probably the first thing that attracts attention is that while using the bootstrap procedure swb, under the null hypothesis the rejection probabilities of the tests treated in this paper lie slightly above the significance level of $$\alpha = 0.10$$. Second, for all considered alternatives $$V_n(\hat{\theta})$$ is at least as good as $$T_n(\hat{\theta})$$ or Colling and Van Keilegom’s test $$W_1^2$$ and mostly even better. Third the opposite is true when looking at the rejection probabilities with the asymptotic critical values (asym). It is also noteworthy that $$T_n(\hat{\theta})$$ with asymptotical critical values (although not recommended by Härdle and Mammen, 1993, in models without transformation) in many cases even outperforms the test provided by Colling and Van Keilegom (2017). Now considering the columns denoted by twb we see that the rejection probabilities under the null hypothesis do not longer exceed the significance level. However, it does not really seem to make sense to use the transformation wild bootstrap, since under the alternatives $$V_n(\hat{\theta})$$ performs much better with the standard wild bootstrap. Further $$T_n(\hat{\theta})$$ is even outperformed by its counterpart using asymptotic critical values. Particularly, the results do not justify the immense increase of the computational costs. Another problem of the transformation wild bootstrap is the instability of the resulting procedure due to the additional estimation of the transformation parameter for every bootstrap sample, because when adding the bootstrap residuals to the estimated regression function in order to calculate $$Y_i^* = \Lambda_{\hat{\theta}}^{-1}(m(X_i^*, \hat{\beta}) + \varepsilon_i^*)$$ it might happen that the argument of the inverse transformation function leaves its domain of definition. This has to be taken account for in the algorithm.

TABLES 1–3 HERE

5.2 Testing Significance of Covarivables

In the second part of our simulation study we examine the small sample size behavior of the test developed in section 4 in order to compare it to the test provided by Allison et al. (2017). The estimation of the transformation parameter is conducted as in the last subsection. Further, $$(W, Z)$$
follows a uniform distribution on $[0, 1]^2$, while the residuals are standard normally distributed (not truncated). In our test statistic $I_n(\hat{\theta})$ we use the Epanechnikov kernel for $K$ and $L$ with bandwidths $h$ and $g$ selected by cross validation, and the standard normal density $\psi$. We use a standard wild bootstrap procedure applied to the transformed data. Note that the bootstrap regression function only depends on $W$ in order to mimic the null hypothesis. While using the approach of Allison et al. (2017) there arises the need to take account of the estimation of the transformation parameter in the bootstrap procedure (transformation wild bootstrap). An exact description of the method used there is omitted here, but the idea is similar to the one described in subsection 5.1. Again we carry out simulations for $I_n(\hat{\theta})$ on the one hand with the standard wild bootstrap (swb) and on the other hand with the transformation wild bootstrap (twb). The simulations are conducted for the sample sizes $n = 75$ and $n = 100$ at the significance level 0.05. We use $B = 1000$ bootstrap data for the standard wild bootstrap, but only $B = 250$ for the transformation wild bootstrap which is computationally demanding. In contrast to Allison et al. (2017) we do not use the warp-speed method of Giacomini, Politis and White (2013). The results are based on 500 simulation runs. The models considered are

\[
\Lambda_{\theta_0}(Y) = W + \varepsilon \\
\Lambda_{\theta_0}(Y) = W + Z + \varepsilon \\
\Lambda_{\theta_0}(Y) = 3 + 2W + 0.25\sin(2\pi Z)\varepsilon \\
\Lambda_{\theta_0}(Y) = 1 + W + \sin(5Z) + \varepsilon \\
\Lambda_{\theta_0}(Y) = 3 + 2W + 5Z^2 \varepsilon \\
\Lambda_{\theta_0}(Y) = 1 + W + Z^2 + \varepsilon \\
\Lambda_{\theta_0}(Y) = 1 + W + \exp(Z^2) + \varepsilon
\]

for true transformation parameter $\theta_0 = 1$. The results depicted in table 5 show that for $I_n(\hat{\theta})$ the transformation wild bootstrap has no advantage over the standard wild bootstrap. We also compare our test with the one based on the test statistic $\tilde{\Psi}_n$ from Allison et al. (2017) (for the choice of $\gamma = 0.01$ and with transformation wild bootstrap) as this test is recommended in their paper. There are cases where $I_n(\hat{\theta})$ has larger power and other cases were it has lower power than its competitor.

TABLE 5 HERE

A Assumptions

A.1 Assumptions for Theorem 3.1

(A1) On its compact support $X$ the density $f_X$ of the covariate $X$ is continuously differentiable.

(A2) $\varepsilon$ is a nondegenerated random variable with $E[\varepsilon^4] < \infty$. 

10
Denote with

(A.2) Assumptions for Theorem 4.1

(A3) \( K \) is a symmetric, nonnegative and continuously differentiable kernel with compact support.

(A4) The bandwidth \( h \) fulfills \( h \to 0 \) and \( nh^d \to \infty \).

(A5) \( \Delta_n \) is continuously differentiable. \( \Delta_n \) and its derivative are bounded uniformly in \( x \) and \( n \), and \( c_n \sim n^{-\frac{1}{2}}h^{-\frac{d}{2}} \).

(A6) \( \hat{\theta} \) is an estimator of \( \theta_0 \) fulfilling \( \hat{\theta} - \theta_0 = O_P(n^{-\frac{1}{2}}) \).

(A7) For all \( y \) the transformation \( \Lambda_\theta(y) \) is two times continuously differentiable with respect to \( \theta \). We denote the gradient and Hessian matrix with \( \hat{\Lambda}_\theta(y) \) and \( \hat{\Lambda}_\theta(y) \), respectively. Let

\[
\sup_{x \in X} E[\|\hat{\Lambda}_\theta(Y)\|^2 | X = x] < \infty
\]

and let there exist an \( \alpha > 0 \) with

\[
E \left[ \sup_{\theta' : \|\theta' - \theta_0\| \leq \alpha} \|\hat{\Lambda}_{\theta'}(Y)\| \right] < \infty.
\]

(A8) The map \( x \mapsto E[\Lambda_\theta(Y)^4 | X = x] \) is continuously differentiable with

\[
E[E[\Lambda_\theta(Y)^4 | X]^2] < \infty.
\]

(A9) All partial derivatives of \( m(x, \beta) \) of order 0, \ldots, 3 with respect to \( x \) and \( \beta \) exist and are continuous for all \( (x, \beta) \). The set \( B \) of regression parameters \( \beta \) is compact and \( \bar{\beta} = \arg\min_{\beta \in B} E[(m(X) - m(X, \beta))^2] \) is (for all \( n \) sufficiently large) a uniquely determined interior value of \( B \) with \( \|\bar{\beta} - \beta_0\| \to 0 \) for some \( \beta_0 \in B \) as \( n \to \infty \). Moreover,

\[
\inf_{\|\beta - \beta_0\| > \delta} E[(m(X, \beta) - m(X, \beta_0))^2] > 0
\]

holds for all \( \delta > 0 \) and \( n \) sufficiently large. The matrix

\[
\Omega(\beta_0) = \left( E \left[ \frac{\partial}{\partial \beta_i} m(X, \beta) \bigg| \beta = \beta_0 \right] \frac{\partial}{\partial \beta_j} m(X, \beta) \bigg| \beta = \beta_0 \right)_{i,j=1,\ldots,q}
\]

is regular.

A.2 Assumptions for Theorem 4.1

Denote with \( \mathcal{U}^p \) the set of all integrable and uniformly continuous functions of \( p \) variables.

(B1) The covariate \( W \) has compact support \( \mathcal{W} \) and density \( f_W \). On their support the functions \( f_W \) and \( r(\cdot)f_W(\cdot) \) lie in \( \mathcal{U}^p \) and are twice continuously differentiable.

(B2) \( E[|d(W, Z)|] < \infty \), \( E[d(W, Z)] = 0 \) and \( \delta_n \sim n^{-1/2}h^{-p/4} \).

(B3) For any \( z \) in the support of \( Z \) the random variable \( W \) admits a conditional density denoted by \( f_W | Z(\cdot|z) \). Further \( E[(\Lambda_{\theta_0}(Y) - r(W))^2 | W] = \|f_W(\cdot) \) and \( E[(\Lambda_{\theta_0}(Y) - r(W))^4 | W] = \|f_W(\cdot) \) belong to \( \mathcal{U}^p \).
Let $K$ and $L$ be symmetric kernel functions of order 2 with compact support and let $L$ be of bounded variation.

Let $\sigma^2(w,z) = E[(\Lambda_{\theta_0}(Y) - r(W))^2|W = w, Z = z]$, then for all $z$ in the support of $Z$ the function $\sigma^2(\cdot,z) f_W(w) f_{W|Z}(\cdot|z) \in U^p$ has an integrable Fourier transform. Further, $E[\sigma^4(W,X)f_W^2(W)f_{W|Z}(W|X)] < \infty$.

$\sigma^2(\cdot,z)$ belongs to $U^p, d(\cdot,x)f_W(\cdot)f_{W|Z}(\cdot|x)$ is integrable and squared integrable for any $x$ in the support of $X$. Further, $E[d^2(W,X)f_W^2(W)f_{W|Z}(W|X)] < \infty$.

$\psi$ is a bounded function with positive and integrable Fourier transform.

The bandwidths $g$ and $h$ fulfill
\[ g \to 0, \quad h \to 0, \quad \frac{ng^4}{\log n} \to \infty, \quad nh^p \to \infty, \quad nh^2 g^4 \to 0, \quad \frac{h}{g} \to 0. \]

$E[\Lambda_{\theta_0}(Y)^8] < \infty$.

$\hat{\theta}$ is an estimator of $\theta_0$ fulfilling $\hat{\theta} - \theta_0 = O_P(n^{-\frac{1}{2}})$.

For all $y$ the transformation $\Lambda_{\theta}(y)$ is two times continuously differentiable with respect to $\theta$. Further,
\[ \sup_{w \in W} E[\|\hat{\Lambda}_{\theta_0}(Y)\|^2|W = w] < \infty \]
holds and there exists an $\alpha > 0$ with
\[ E \left[ \sup_{\theta':\|\theta' - \theta_0\| \leq \alpha} \|\hat{\Lambda}_{\theta'}(Y)\| \right] < \infty. \]

**Remark A.1** Note that $E[\varepsilon^8] < \infty$ is implied by \[(B9)\] Instead of $nh^2 g^4 \to 0$ Lavergne et al. (2015) use the assumption $nh^2 g^{2s} \to 0$ for some $s \geq 2$ with some additional differentiability conditions on $f_W$ and $r(\cdot)f_W(\cdot)$ and a higher order kernel for $L$. The stronger version arises here due to boundary problems. The reason for considering a density with a compact support is to allow for example the profile likelihood estimator of Colling and Van Keilegom (2015) and other estimators for $\theta_0$, that require a compact support of the covariate. By straightforward calculations it can be seen that assumption \[(B8)\] can be fulfilled if and only if $p \leq 5$.

## B Proofs

**B.1 Proof of Theorem 3.1**

Proof for $T_n$.

Consider the decomposition $T_n(\hat{\theta}) = T_n(\theta_0) + I + II + III + IV + V$, where
\[ I = nh^2 \int \left( \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)(\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) \right)^2 dx \]
\[ II = nh^{\frac{d}{2}} \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(m(X_i, \hat{\beta}_0) - m(X_i, \hat{\beta}_0)) \right)^2 dx \]
\[ III = 2nh^{\frac{d}{2}} \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(\Lambda_{\theta_0}(Y_i) - m(X_i, \hat{\beta}_0)) \right) \times \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) \right) dx \]
\[ IV = 2nh^{\frac{d}{2}} \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) \right) \times \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(m(X_i, \hat{\beta}_0) - m(X_i, \hat{\beta}_0)) \right) dx \]
\[ V = 2nh^{\frac{d}{2}} \int \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) \right) \times \left( \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(m(X_i, \hat{\beta}_0) - m(X_i, \hat{\beta}_0)) \right) dx. \]

In what follows we show asymptotic negligibility of the terms \( I, II, III, IV \) and \( V \).

To treat \( I \) we will use a Taylor expansion of \( \Lambda_{\hat{\theta}}(Y) \) around \( \Lambda_{\theta_0}(Y) \). To this end note that due to assumption \([A6]\) we can assume \( \|\hat{\theta} - \theta_0\| \leq \alpha \) with \( \alpha \) from assumption \([A7]\). Let now \( \theta_i^* \) denote suitable parameters between \( \hat{\theta} \) and \( \theta_0 \) such that

\[
I = \frac{h^{\frac{d}{2}}}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int K_h(x - X_i)(\hat{\Lambda}_{\theta_0}(Y_i) - \Lambda_{\theta_0}(Y_i))(\hat{\theta} - \theta_0) \times (\hat{\Lambda}_{\hat{\theta}}(Y_j) - \Lambda_{\hat{\theta}}(Y_j))(\hat{\theta} - \theta_0)) dx \\
\leq \frac{h^{\frac{d}{2}}\|\hat{\theta} - \theta_0\|^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\hat{\Lambda}_{\theta_0}(Y_i)\|\|\Lambda_{\theta_0}(Y_j)\| \int K_h(x - X_i)K_h(x - X_j) dx \tag{B.1} \\
+ \frac{2h^{\frac{d}{2}}\|\hat{\theta} - \theta_0\|^3}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\hat{\Lambda}_{\theta_0}(Y_i)\|\|\Lambda_{\theta_0}(Y_j)\| \int K_h(x - X_i)K_h(x - X_j) dx \tag{B.2} \\
+ \frac{h^{\frac{d}{2}}\|\hat{\theta} - \theta_0\|^4}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\hat{\Lambda}_{\theta_0}(Y_i)\|\|\Lambda_{\theta_0}(Y_j)\| \int K_h(x - X_i)K_h(x - X_j) dx. \tag{B.3} 
\]

For the expectation of the summands of term \( B.1 \) we have for \( i = j \)

\[
E \left[ \|\hat{\Lambda}_{\theta_0}(Y_i)\|^2 \int K_h(x - X_i)^2 dx \right] \leq C h^{-d} E[\|\hat{\Lambda}_{\theta_0}(Y_i)\|^2] = O(h^{-d})
\]

and for \( i \neq j \)

\[
E \left[ \|\hat{\Lambda}_{\theta_0}(Y_i)\|\|\hat{\Lambda}_{\theta_0}(Y_j)\| \int K_h(x - X_i)K_h(x - X_j) dx \right] \\
= \int (E[\|\hat{\Lambda}_{\theta_0}(Y_i)\|\|K_h(x - X_i)\|^2 dx)
\]

\[ = 0 \]

13
\[
\int \left( E \left[ \left| \hat{\Lambda}_{\theta_0}(Y_1) \right| \mid X_1 \right] K_h(x - X_1) \right)^2 \, dx \leq C \int \left( E \left[ K_h(x - X_1) \right] \right)^2 \, dx = O(1)
\]
where the second last equality follows by assumption \([A7]\) for some constant \(C\). Thus term \([B.1]\) is, due to assumption \([A6]\) of order \(O_P(h^{d/2}n^{-2})(O_P((h^{-d}n) + O_P(n^2)) = o_P(1)\).

For term \([B.2]\) one obtains
\[
O_P(h^{d/2}n^{-5/2}) \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| \hat{\Lambda}_{\theta_0}(Y_i) \right\| \left\| \hat{\Lambda}_{\theta_j}(Y_j) \right\| \int K(x)K_h(hx + X_i - X_j) \, dx
\]
\[
\leq O_P(h^{-d/2}n^{-1/2}) \left( \frac{1}{n} \sum_{i=1}^{n} \left\| \hat{\Lambda}_{\theta_0}(Y_i) \right\| \right) \left( \frac{1}{n} \sum_{j=1}^{n} \sup_{\|\theta - \theta_0\| \leq \alpha} \left\| \hat{\Lambda}_{\theta}(Y_j) \right\| \right) = o_P(1)
\]
by assumption \([A7]\). The last term \([B.3]\) is treated similarly and we obtain \(I = o_P(1)\).

Term \(II\) is treated completely analogously with a Taylor expansion of \(m(X_i, \hat{\beta}_0)\) around \(m(X_i, \hat{\beta}_0)\) and using \(\hat{\beta}_0 - \hat{\beta}_0 = O_P(n^{-1/2})\). For the latter equality note that \(\hat{\beta}_0 - \hat{\beta} = O_P(n^{-1/2})\) follows from the proof of Lemma 4 in Colling and Van Keilegom (2016) (taking into account that our local alternative has a slower rate), whereas \(\hat{\beta}_0 - \hat{\beta} = O_P(n^{-1/2})\) follows from standard arguments in the model \(\Lambda_{\theta_0}(Y) = m(X) + \varepsilon\) with known transformation.

With the Cauchy-Schwarz inequality \(V = o_P(1)\) directly follows from \(I = o_P(1)\) and \(II = o_P(1)\).

We further decompose the term \(III = A_n + B_n + C_n\) into the terms
\[
A_n = 2nh^d \int \left( \frac{1}{n} \sum_{i=1}^{n} \left( K_h(x - X_i)(m(X_i, \hat{\beta}) - m(X_i, \hat{\beta}_0)) \right) \right) \times \left( \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j)(\Lambda_{\theta}(Y_j) - \Lambda_{\theta_0}(Y_j)) \right) \, dx
\]
\[
B_n = 2nh^d \int \left( \frac{1}{n} \sum_{i=1}^{n} \left( K_h(x - X_i)(\Lambda_{\theta_0}(Y_i) - m(X_i)) \right) \right) \times \left( \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j)(\Lambda_{\theta}(Y_j) - \Lambda_{\theta_0}(Y_j)) \right) \, dx
\]
\[
C_n = 2nh^d \int \left( \frac{1}{n} \sum_{i=1}^{n} \left( K_h(x - X_i)(m(X_i) - m(X_i, \hat{\beta})) \right) \right) \times \left( \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j)(\Lambda_{\theta}(Y_j) - \Lambda_{\theta_0}(Y_j)) \right) \, dx.
\]

Now \(A_n\) is very similar to \(V\) and can be treated in the same way noting that \(\hat{\beta} - \hat{\beta}_0 = O_P(n^{-1/2})\).

With Taylor expansion we obtain for some \(\theta_j^*\) between \(\hat{\theta}\) and \(\theta_0\),
\[
B_n = 2\frac{h^d}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int K_h(x - X_i)\varepsilon_i K_h(x - X_j) \, dx \hat{\Lambda}_{\theta_0}(Y_j)(\hat{\theta} - \theta_0)
\]
\[
+ 2\frac{h^d}{n} \sum_{i=1}^{n} \int K_h(x - X_i)^2\varepsilon_i \hat{\Lambda}_{\theta_0}(Y_i) \, dx (\hat{\theta} - \theta_0)
\]
\[ + 2 \frac{h^\frac{3}{2}}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int K_h(x - X_i)K_h(x - X_j)\varepsilon_i(\hat{\theta} - \theta_0)^t \hat{\Lambda}_{\theta_j}(Y_j)(\hat{\theta} - \theta_0) \, dx. \]

The expectation of the absolute value of the integral in the second term can be bounded by

\[ E \left[ E[|\varepsilon_i||\hat{\Lambda}_{\theta_0}(Y_i)||X_i] \int K_h(x - X_i)^2 \, dx \right] \leq \sigma E \left[ \sqrt{ E[|\hat{\Lambda}_{\theta_0}(Y_i)||X_i]} \int K_h(x - X_i)^2 \, dx \right] = O(h^d) \]

by assumption \((A7)\). Thus the second term in the decomposition of \(B_n\) is of order \(O_P(h^{-d/2}n^{-1/2}) = o_P(1)\). The third term can be bounded by

\[ 2h^\frac{3}{2}\|\hat{\theta} - \theta_0\|^2 \sum_{j=1}^{n} \|\hat{\Lambda}_{\theta_j}(Y_j)\| \int K_h(x - X_j) \, dx \sup_x \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i K_h(x - X_i) \right| = o_P(h^{d/2}) \]

(consider the last factor as local constant estimator in a regression model with zero regression function). For the first term in the decomposition of \(B_n\) consider the double sum without the factor \(\hat{\theta} - \theta_0 = O_P(n^{-1/2})\). Because \(\varepsilon\) is centered and independent of \(X\), the expectation is zero and the variance is easy to derive. With calculations similar to before it can be shown that the variance is of order \(O(nh^d)\) such that for the first term in \(B_n\) one has \(O_P((nh^d)^{1/2}n^{-1/2}) = o_P(1)\).

Inserting \(m(X_i) - m(X_i, \hat{\beta}) = c_n \Delta_n(X_i)\) and the Taylor expansion for \(\Lambda_\hat{\beta}(Y_j) - \Lambda_{\theta_0}(Y_j)\) one sees that \(C_n\) has the rate \(nh^{d/2}c_nO_P(||\hat{\theta} - \theta_0||) = O_P(h^{d/4}) = o_P(1)\).

The remaining term \(IV\) is treated analogously to before inserting Taylor expansions of \(\Lambda_\hat{\beta}(Y_i)\) around \(\Lambda_{\theta_0}(Y_i)\) and of \(m(X_i, \hat{\beta})\) around \(m(X_i, \hat{\beta}_{\theta_0})\).

\[ \square \]

**Proof for \(V_n\).**

We decompose \(nh^\frac{d}{2}V_n(\hat{\theta}) = nh^\frac{d}{2}V_n(\theta_0) + \tilde{I} + \tilde{II} + \tilde{III} + \tilde{IV} + \tilde{V},\) where

\[ \tilde{I} = \frac{h^\frac{3}{2}}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(\Lambda_{\hat{\beta}}(Y_i) - \Lambda_{\theta_0}(Y_i))(\Lambda_{\hat{\beta}}(Y_j) - \Lambda_{\theta_0}(Y_j)) \]

\[ \tilde{II} = \frac{h^\frac{3}{2}}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(m(X_i, \hat{\beta}_{\theta_0}) - m(X_i, \hat{\beta})) (m(X_j, \hat{\beta}_{\theta_0}) - m(X_j, \hat{\beta})) \]

\[ \tilde{III} = \frac{2h^\frac{3}{2}}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(\Lambda_{\theta_0}(Y_i) - m(X_i, \hat{\beta}_{\theta_0}))(\Lambda_{\hat{\beta}}(Y_j) - \Lambda_{\theta_0}(Y_j)) \]

\[ \tilde{IV} = \frac{2h^\frac{3}{2}}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(\Lambda_{\theta_0}(Y_i) - m(X_i, \hat{\beta}_{\theta_0}))(m(X_j, \hat{\beta}_{\theta_0}) - m(X_j, \hat{\beta})) \]

\[ \tilde{V} = \frac{2h^\frac{3}{2}}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(X_i - X_j)(\Lambda_{\hat{\beta}}(Y_i) - \Lambda_{\theta_0}(Y_i))(m(X_j, \hat{\beta}_{\theta_0}) - m(X_j, \hat{\beta})) \]

15
and we have to show that $\tilde{I}, \ldots, \tilde{V}$ are of order $o_P(1)$. Note that rewriting the terms $I, \ldots, V$ in the decomposition of $T_n(\tilde{\theta})$ one sees that those have the very same structure as the terms $\tilde{I}, \ldots, \tilde{V}$.

The only difference is that for the latter terms the double sums are without diagonal terms and the kernel $K_h(x - X_j)$ in $\tilde{I}, \ldots, \tilde{V}$ corresponds to $\int K_h(x - X_i)K_h(x - X_j)\,dx = (K * K)_h(X_i - X_j)$ in $I, \ldots, V$. The derivations are thus analogous and omitted for the sake of brevity.

$\square$

### B.2 Proof of Theorem 4.1

Using a Taylor expansion for $\Lambda_{\hat{\theta}}(Y)$ around $\Lambda_{\theta_0}(Y)$ one obtains the decomposition $I_n(\hat{\theta}) = I_n(\theta_0) + I + II + III + IV + V$, where

\[
I = \frac{h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \Lambda_{\hat{\theta}}(\hat{\theta}) L_{i,k} L_{j,l} K_{i,j} \psi_{i,j},
\]

\[
II = \frac{h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \Lambda_{\hat{\theta}}(\hat{\theta}) L_{i,k} L_{j,l} K_{i,j} \psi_{i,j},
\]

\[
III = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \left( \Lambda_{\theta_0}(Y_i) - \Lambda_{\theta_0}(Y_k) \right) a_{j,l}(\hat{\theta}) L_{i,k} L_{j,l} K_{i,j} \psi_{i,j},
\]

\[
IV = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \left( \Lambda_{\theta_0}(Y_i) - \Lambda_{\theta_0}(Y_k) \right) (\hat{\theta}) L_{i,k} L_{j,l} K_{i,j} \psi_{i,j},
\]

\[
V = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} a_{i,k}(\hat{\theta}) a_{j,l}(\hat{\theta}) L_{i,k} L_{j,l} K_{i,j} \psi_{i,j},
\]

with $a_{i,k} = (\hat{\Lambda}_{\theta_0}(Y_i) - \hat{\Lambda}_{\theta_0}(Y_k))$ and $A^*_{i,k} = (\hat{\Lambda}_{\theta^*_i}(Y_i) - \hat{\Lambda}_{\theta^*_k}(Y_k))$ for suitable $\theta^*_i$ between $\theta_0$ and $\hat{\theta}$.

Further we use the notations

\[
L_{i,k} = L\left(\frac{W_i - W_k}{g}\right), K_{i,j} = K\left(\frac{W_i - W_j}{h}\right), \psi_{i,j} = \psi(X_i - X_j).
\]

Now for the first term one obtains

\[
|I| \leq \|\hat{\theta} - \theta_0\|^2 \frac{h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \frac{\partial^4}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \|a_{i,k}\| \|a_{j,l}\| \left| L\left(\frac{W_i - W_k}{g}\right) L\left(\frac{W_j - W_i}{g}\right) K\left(\frac{W_i - W_j}{h}\right) \psi(X_i - X_j) \right|
\]

with

\[
\frac{1}{g^{2p} h^p} E\left[\|a_{1,3}\| \|a_{2,4}\| L\left(\frac{W_1 - W_3}{g}\right) L\left(\frac{W_2 - W_4}{g}\right) K\left(\frac{W_1 - W_3}{h}\right) \psi(X_1 - X_2)\right]\n\]

\[
\leq C \frac{1}{g^{2p} h^p} E\left[\left| L\left(\frac{W_1 - W_3}{g}\right) L\left(\frac{W_2 - W_4}{g}\right) K\left(\frac{W_1 - W_3}{h}\right) \right| \right]
\]

\[
E\left[\|\hat{\Lambda}_{\theta_0}(Y_1) - \hat{\Lambda}_{\theta_0}(Y_3)\| \|\hat{\Lambda}_{\theta_0}(Y_2) - \hat{\Lambda}_{\theta_0}(Y_4)\| \|X_1, W_1, \ldots, X_4, W_4\|\right]
\]

16
\[
\leq \frac{C^2}{g^{2p}h^p} E \left[ \left| L \left( \frac{W_1 - W_3}{g} \right) L \left( \frac{W_2 - W_4}{g} \right) K \left( \frac{W_1 - W_3}{h} \right) \right] \right] = O(1)
\]
for some constant \( C \), where for the first inequality we use assumption \([B7]\) and for the second inequality we use assumption \([B11]\). The rate \( O(1) \) follows by standard kernel arguments. Applying \([B10]\) we obtain \( f = O_P(h^2) = o_P(1) \).

Due to \([B10]\) we can assume that \( \|\hat{\theta} - \theta\| \leq \alpha \) from \([B11]\). Then \(|II|\) can be bounded by

\[
\frac{Cn\|\hat{\theta} - \theta_0\|^4}{h^2 g^{2p}} \sum_{i,j,k,l} \| A_{i,k}^n \| \| A_{j,l}^n \| |L_{i,k}L_{j,l}|
\]

\[
\leq \frac{Cn\|\hat{\theta} - \theta_0\|^4}{h^2 g^{2p}} \left( \frac{1}{n^2} \sum_{i \neq k} \| A_{i,k}^n \| |L_{i,k}| \right)^2
\]

\[
\leq 4Cn\|\hat{\theta} - \theta_0\|^4 \left( \frac{1}{n} \sum_{i=1}^n \sup_{\|\theta - \theta_0\| \leq \alpha} \| \hat{L}_g(Y_i) \| \sup_w \frac{1}{n} \sum_{k=1}^n \frac{1}{g^p} |L \left( \frac{W_i - w}{g} \right) | \right)^2
\]

\[
= O_P \left( \frac{1}{nh^2} \right) = o_P(1).
\]

Here, the uniform convergence of a kernel density estimator with kernel \(|L|\) and bandwidth \( g \) is used.

The treatments of \( IV \) and \( V \) are similar and omitted for the sake of brevity.

For \( III \) we need a further decomposition that is obtained by using the model equation (2.1) and the structure of the local alternative. We have \( III = A_n - B_n + C_n + D_n \) with

\[
A_n = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \hat{\varepsilon}_i a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} (\hat{\theta} - \theta_0)
\]

\[
B_n = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \hat{\varepsilon}_k a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} (\hat{\theta} - \theta_0)
\]

\[
C_n = \frac{2h^2}{n^3 g^{2p} h^p} \delta_n \sum_{i,j,k,l} \hat{\varepsilon}_i (d(W_i, Z_i) - d(W_k, Z_k)) a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} (\hat{\theta} - \theta_0)
\]

\[
D_n = \frac{2h^2}{n^3 g^{2p} h^p} \sum_{i,j,k,l} \hat{\varepsilon}_i (r(W_i) - r(W_k)) a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} (\hat{\theta} - \theta_0).
\]

To treat \( A_n \) define the function \( \varphi(x, y) = g^{-2p} h^{-p} E[a_{1,2} L_{3,4} L_{1,2} K_{3,1} \psi_{3,1} | X_3 = x, Y_3 = y] \) and consider \( A_n = A_n^{(1)} + A_n^{(2)} \) with

\[
A_n^{(1)} = O(h^2) \sum_{i=1}^n \hat{\varepsilon}_i \varphi(X_i, Y_i) (\hat{\theta} - \theta_0) = O(h^2) O_P(n^{1/2}) O_P(n^{-1/2}) = o_P(1)
\]

because the random variables \( \varepsilon_i \varphi(X_i, Y_i) \), \( i = 1, \ldots, n \), are iid and centered, while the function \( \varphi \) is bounded. Further, one obtains

\[
A_n^{(2)} = O(h^2) \frac{1}{n^3} \sum_{i,j,k,l} \hat{\varepsilon}_i (a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} - E[a_{j,l} L_{i,k} L_{j,l} K_{i,j} \psi_{i,j} | X_i, Y_i]) (\hat{\theta} - \theta_0) = o_P(1)
\]
with some tedious calculation of the second moment of the vector components of the (centred) sum (without the factor $(\hat{\theta} - \theta_0)$).

For $C_n$ note that $E[g^{-2p}h^{-p} | (d(W_i, Z_i) - d(W_k, Z_k))a_{i,l}L_{i,k}L_{j,l}K_{i,j}\psi_{i,j}] < \infty$ and thus

$$C_n = O(h^{p/2}k_n)O_P(1)O_P(n^{-1/2}) = O_P(h^{p/4}) = o_P(1).$$

It remains to show $D_n = o_P(1)$, which can be done by calculating expectation and variance of the sum (without the factor $(\hat{\theta} - \theta_0)$). As, in contrast to $B_n$, the summands are not centered, to this end one needs to make use of assumptions [B1] and [B4]. Consider, e.g., for $i \neq k$

$$E[(r(W_i) - r(W_k))g^{-p}L_{i,k}|W_i = w] = \int (r(w) - r(v))f_W(v)\frac{1}{g^p}L\left(\frac{w-v}{g}\right)dv$$

$$= \int (r(w) - r(w-ug))f_W(w-ug)L(u)du$$

$$= O(g^2) + O(g)I\{w \in B_g\} = O(g^2)$$

using Taylor’s expansions for $f_W$ and $rf_W$ and the order of the kernel $L$. Here $B_g$ denotes the $g$-boundary of the support of $W$ and one has $P(W \in B_g) = O(g)$. Thus the expectation of the sum (without the factor $(\hat{\theta} - \theta_0)$) is of order $h^{p/2}ng^2$ with $h^{p/2}ng^2n^{-1/2} = o(1)$ by assumption [B8]. The calculation of the variance is more tedious, but analogous to the calculations in Lavergne et al. (2015) and Lavergne and Vuong (2000).

References

Alcalá, J.T., Cristóbal, J.A. and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. Statist. Prob. Lett. 42, 39–46.

Allison, J.S., Hušková, M. and Meintanis, S.G. (2017). Testing the adequacy of semiparametric transformation models. TEST, to appear. DOI 10.1007/s11749-017-0544-4

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. Ann. Statist. 41, 802–837.

Bierens, H.J. (1982). Consistent model specification tests. J. Econometr. 20, 105–134.

Bickel, P. and Doksum, K.A. (1981). An analysis of transformations revisited. J. Amer. Statist. Assoc. 76, 296–311.

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

Carroll, R. J. and Ruppert, D. (1988). Transformation and Weighting in Regression. Monographs on Statistics and Applied Probability. Chapman & Hall, New York.

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. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* 29, 1469–1507.

Efron, B. (2014). Estimation and accuracy after model selection. *J. Amer. Statist. Assoc.* 109, 991–1007.

Fan, J. and Gijbels, I (1996). *Local Polynomial Modelling and its Applications.* Monographs on Statistics and Applied Probability. Chapman & Hall, New York.

Fan, Y. and Li, Q (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

Escanciano, J.C. (2006). A consistent test for regression models using projections. *Econometric Theory* 22, 1030–1051.

Giacomini, R., Politis, D. N., White, H. (2013). A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators. *Econometric Theory* 29, 567–589.

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

Hinkley, D.V. and Runger, G. (1984). The analysis of transformed data. With discussion. *J. Amer. Statist. Assoc.* 79, 302–320.

Hlávka, Z., Hušková, M., Kirch, C. and Meintanis, S.G. (2017). Fourier-type tests involving martingale difference processes *Econometric Reviews* 36, 468–492.

Horowitz, J.L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 102–137.

Horowitz, J.L. (2009). *Semiparametric and nonparametric methods in econometrics.* Springer Series in Statistics. Springer, New York.

Hušková, M. and Meintanis, S.G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. *Kybernetika* 45, 960–971.

Lavergne, P., Maistre, S. and Patilea, V. (2015). A significance test for covariates in nonparametric regression *Electron. J. Statist.* 9, 643–678.

Lavergne, P. and Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory* 16, 576–601.

Lee, J.D., Sun D.L., Sun, Y. and Taylor, J.E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* 44, 907–927.
Linton, O., Chen, R., Wang, N. and Härdle, W. (1997). An analysis of transformations for additive nonparametric regression. *J. Amer. Statist. Assoc.* **92**, 1512–1521.

Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation on a semiparametric transformation model. *Ann. Statist.* **36**, 686–718.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and Its Applications* **9**, 141–142.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL [http://www.R-project.org/](http://www.R-project.org/)

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613–641.

Van Keilegom, I., González-Manteiga, W. and Sánchez Sellero, C. (2008). Goodness of fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17**, 401–415.

Watson, G.S. (1964). Smooth Regression Analysis. *The Indian Journal of Statistics* **26**, 359–372.

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261–1350.

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

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* **75**, 263–289.

Zellner, A. and Revankar, N.S. (1969). Generalized production functions. *Rev. Economic Studies* **36**, 241–250.
Table 1: Rejection probabilities at $\theta_0 = 0$ for the lack-of-fit tests $T_n(\hat{\theta})$, $V_n(\hat{\theta})$ with standard wild bootstrap (swb), transformation wild bootstrap (twb) and using asymptotic critical values (asym), and for $W_2^1$ (twb) with nominal level 0.1.

| Alternative $T_n(\hat{\theta})$ | $V_n(\hat{\theta})$ | $W_2^1$ |
|---------------------------------|----------------------|---------|
|                                | swb | twb | asym | swb | twb | asym | twb |
| 0                              | 0.148  | 0.100  | 0.054 | 0.134  | 0.096  | 0.040  | 0.102 |
| $2X^2$                         | 0.422  | 0.206  | 0.292 | 0.510  | 0.398  | 0.254  | 0.466 |
| $3X^2$                         | 0.716  | 0.604  | 0.612 | 0.784  | 0.692  | 0.558  | 0.754 |
| $4X^2$                         | 0.876  | 0.802  | 0.814 | 0.914  | 0.862  | 0.786  | 0.892 |
| $5X^2$                         | 0.966  | 0.930  | 0.938 | 0.984  | 0.958  | 0.928  | 0.944 |
| $2 \exp(X)$                    | 0.394  | 0.270  | 0.270 | 0.460  | 0.316  | 0.250  | 0.332 |
| $3 \exp(X)$                    | 0.610  | 0.380  | 0.518 | 0.700  | 0.472  | 0.470  | 0.550 |
| $4 \exp(X)$                    | 0.750  | 0.540  | 0.688 | 0.814  | 0.650  | 0.660  | 0.696 |
| $5 \exp(X)$                    | 0.876  | 0.664  | 0.850 | 0.918  | 0.788  | 0.818  | 0.782 |
| $0.25 \sin(2\pi X)$           | 0.324  | 0.214  | 0.148 | 0.324  | 0.244  | 0.104  | 0.210 |
| $0.5 \sin(2\pi X)$            | 0.768  | 0.642  | 0.584 | 0.796  | 0.720  | 0.484  | 0.590 |
| $0.75 \sin(2\pi X)$           | 0.960  | 0.954  | 0.924 | 0.982  | 0.970  | 0.892  | 0.924 |
| $1 \sin(2\pi X)$              | 1.000  | 0.998  | 0.992 | 1.000  | 1.000  | 0.988  | 0.998 |

Table 2: Rejection probabilities at $\theta_0 = 0.5$ for the lack-of-fit tests $T_n(\hat{\theta})$, $V_n(\hat{\theta})$ with standard wild bootstrap (swb), transformation wild bootstrap (twb) and using asymptotic critical values (asym), and for $W_2^1$ (twb) with nominal level 0.1.

| Alternative $T_n(\hat{\theta})$ | $V_n(\hat{\theta})$ | $W_2^1$ |
|---------------------------------|----------------------|---------|
|                                | swb | twb | asym | swb | twb | asym | twb |
| 0                              | 0.114  | 0.098  | 0.054 | 0.136  | 0.108  | 0.036  | 0.072 |
| $2X^2$                         | 0.464  | 0.278  | 0.320 | 0.514  | 0.378  | 0.266  | 0.412 |
| $3X^2$                         | 0.698  | 0.494  | 0.606 | 0.784  | 0.658  | 0.546  | 0.720 |
| $4X^2$                         | 0.894  | 0.748  | 0.848 | 0.940  | 0.852  | 0.812  | 0.882 |
| $5X^2$                         | 0.972  | 0.872  | 0.950 | 0.976  | 0.932  | 0.932  | 0.952 |
| $2 \exp(X)$                    | 0.368  | 0.114  | 0.266 | 0.448  | 0.258  | 0.218  | 0.228 |
| $3 \exp(X)$                    | 0.598  | 0.158  | 0.478 | 0.668  | 0.452  | 0.408  | 0.362 |
| $4 \exp(X)$                    | 0.746  | 0.252  | 0.680 | 0.818  | 0.652  | 0.626  | 0.524 |
| $5 \exp(X)$                    | 0.880  | 0.262  | 0.846 | 0.926  | 0.726  | 0.808  | 0.634 |
| $0.25 \sin(2\pi X)$           | 0.294  | 0.180  | 0.126 | 0.324  | 0.252  | 0.086  | 0.138 |
| $0.5 \sin(2\pi X)$            | 0.752  | 0.522  | 0.560 | 0.772  | 0.734  | 0.474  | 0.450 |
| $0.75 \sin(2\pi X)$           | 0.966  | 0.892  | 0.932 | 0.976  | 0.978  | 0.896  | 0.828 |
| $1 \sin(2\pi X)$              | 1.000  | 0.976  | 0.994 | 1.000  | 1.000  | 0.988  | 0.940 |
Table 3: Rejection probabilities at $\theta_0 = 1$ for the lack-of-fit tests $T_n(\hat{\theta})$, $V_n(\hat{\theta})$ with standard wild bootstrap (swb), transformation wild bootstrap (twb) and using asymptotic critical values (asym), and for $W_1^2$ (twb) with nominal level 0.1.

| Alternative | $T_n(\hat{\theta})$ | $V_n(\hat{\theta})$ | $W_1^2$ |
|-------------|---------------------|---------------------|---------|
|             | swb     | twb     | asym    | swb     | twb     | asym    | swb     | twb     |
| 0           | 0.156   | 0.104   | 0.048   | 0.142   | 0.092   | 0.028   | 0.050   |
| $2X^2$      | 0.386   | 0.338   | 0.264   | 0.470   | 0.422   | 0.210   | 0.386   |
| $3X^2$      | 0.668   | 0.530   | 0.580   | 0.766   | 0.654   | 0.538   | 0.678   |
| $4X^2$      | 0.876   | 0.758   | 0.812   | 0.936   | 0.862   | 0.722   | 0.868   |
| $5X^2$      | 0.954   | 0.900   | 0.940   | 0.986   | 0.944   | 0.932   | 0.932   |
| $2 \exp(X)$ | 0.368   | 0.192   | 0.240   | 0.406   | 0.286   | 0.198   | 0.178   |
| $3 \exp(X)$ | 0.548   | 0.284   | 0.472   | 0.648   | 0.476   | 0.424   | 0.308   |
| $4 \exp(X)$ | 0.768   | 0.412   | 0.706   | 0.808   | 0.618   | 0.660   | 0.446   |
| $5 \exp(X)$ | 0.882   | 0.486   | 0.838   | 0.934   | 0.796   | 0.822   | 0.560   |
| $0.25 \sin(2\pi X)$ | 0.316   | 0.204   | 0.134   | 0.346   | 0.278   | 0.098   | 0.106   |
| $0.5 \sin(2\pi X)$  | 0.720   | 0.554   | 0.576   | 0.772   | 0.702   | 0.488   | 0.340   |
| $0.75 \sin(2\pi X)$ | 0.968   | 0.882   | 0.928   | 0.978   | 0.972   | 0.880   | 0.692   |
| $1 \sin(2\pi X)$    | 1.000   | 0.970   | 0.998   | 1.000   | 1.000   | 0.994   | 0.878   |

Table 4: Rejection probabilities with standard wild bootstrap (swb) and using asymptotic critical values (asym) at $\theta_0 = 0.5$ for the lack-of-fit tests $T_n(\hat{\theta})$ and $V_n(\hat{\theta})$ with and without truncated error distributions.
Table 5: Rejection probabilities for the test of significance with test statistic $I_n(\hat{\theta})$ with standard wild bootstrap (swb) and transformation wild bootstrap (twb) and with test statistic $\Psi_n$ from Allison et al. with transformation wild bootstrap. The transformation parameter is $\theta_0 = 1$ and the nominal level is 0.05.

| Model                        | $n = 75$ |          | $n = 100$ |          |
|------------------------------|----------|----------|-----------|----------|
|                              | twb      | swb      | $\Psi_n$  | twb      | swb      | $\Psi_n$  |
| $A_{\theta_0}(Y) = W + \varepsilon$ | 0.024    | 0.014    | 0.05      | 0.034    | 0.042    | 0.04      |
| $A_{\theta_0}(Y) = W + Z + \varepsilon$ | 0.146    | 0.136    | 0.49      | 0.250    | 0.282    | 0.60      |
| $A_{\theta_0}(Y) = 3 + 2W + 0.25 \sin(2\pi Z) + \varepsilon$ | 0.048    | 0.240    | 0.10      | 0.112    | 0.358    | 0.11      |
| $A_{\theta_0}(Y) = 1 + W + \sin(5Z) + \varepsilon$ | 0.546    | 0.650    | 0.77      | 0.726    | 0.842    | 0.88      |
| $A_{\theta_0}(Y) = 3 + 2W + 5Z^2 + \varepsilon$ | 0.762    | 0.990    | 0.41      | 0.902    | 0.998    | 0.51      |
| $A_{\theta_0}(Y) = 1 + W + Z^2 + \varepsilon$ | 0.078    | 0.168    | 0.34      | 0.170    | 0.274    | 0.40      |
| $A_{\theta_0}(Y) = 1 + W + \exp(Z^2) + \varepsilon$ | 0.248    | 0.422    | 0.35      | 0.356    | 0.622    | 0.44      |