I N S T I T U T   D E   S T A T I S T I Q U E

B I O S T A T I S T I Q U E   E T

S C I E N C E S   A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/17

Goodness-of-fit tests in semiparametric transformation models

COLLING B. AND I. VAN KEILEGOM

Goodness-of-fit tests in semiparametric transformation models

Benjamin Colling [*] Ingrid Van Keilegom [*,§]

April 24, 2014

Abstract

Consider a semiparametric transformation model of the form $\Lambda_\theta(Y) = m(X) + \varepsilon$, where Y is a univariate dependent variable, X is a d-dimensional covariate, and ε is independent of X and has mean zero. We assume that $\{\Lambda_\theta : \theta \in \Theta\}$ is a parametric family of strictly increasing functions, while m is an unknown regression function. The goal of the paper is to develop tests for the null hypothesis that $m(\cdot)$ belongs to a certain parametric family of regression functions. We propose a Kolmogorov-Smirnov and a Cramér-von Mises type test statistic, which measure the distance between the distribution of ε estimated under the null hypothesis and the distribution of ε without making use of this null hypothesis. The estimated distributions are based on a profile likelihood estimator of θ and a local polynomial estimator of $m(\cdot)$. The limiting distributions of these two test statistics are established under the null hypothesis and under a local alternative. We use a bootstrap procedure to approximate the critical values of the test statistics under the null hypothesis. Finally, a simulation study is carried out to illustrate the performance of our testing procedures, and we apply our tests to data on the scattering of sunlight in the atmosphere.

Key Words: Bootstrap; Goodness-of-fit; Local polynomial smoothing; Profile likelihood; Semiparametric regression; Transformation model.

[*]Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium. Research supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), and by the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

[§]Research also supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650.

1 Introduction

Consider the following semiparametric transformation model :

$$\Lambda_\theta(Y) = m(X) + \varepsilon \;, \tag{1.1}$$

where $\Lambda_\theta(\cdot)$ belongs to a parametric family of strictly increasing functions and the function $m(\cdot)$ is unknown. We assume that X is a d-dimensional covariate, Y is a univariate response variable, the error term ε is independent of X, and $E(\varepsilon) = 0$. Let θ belong to a finite dimensional compact subset Θ of $\mathbb{R}^k$, and denote the true but unknown values of θ and $m(\cdot)$ by θ_0 and $m_0(\cdot)$.

The motivation for considering this model comes from the rich literature on parametric transformations in regression, starting from the seminal paper by Box and Cox (1964). They proposed a parametric family of power transformations that includes as special cases the logarithm and the identity. They suggested that when this power transformation is applied to the response in a linear regression model, the regression function of the new model might have an additive structure, and the new error might be approximately normal and homoscedastic. Other transformations have been proposed in the literature, like for example, the Zellner and Revankar (1969) transform and the Bickel and Doksum (1981) transform. See also the book by Carroll and Ruppert (1988) and the review paper by Sakia (1992) for more details and references on this topic.

Whereas the above references restrict attention to models in which the regression function (as well as the transformation) is parametric, we will focus in this paper on model (1.1), which assumes that the regression function is nonparametric. The estimation of this (semiparametric) transformation model has been studied by Linton, Sperlich and Van Keilegom (2008). They proposed two different estimators of the transformation parameter θ and developed the asymptotic properties of these estimators. Moreover, Colling, Heuchenne, Samb and Van Keilegom (2013) and Heuchenne, Samb and Van Keilegom (2014) studied nonparametric estimators of the density and of the distribution function of the error term ε under this model. Other papers that have studied the estimation of this model include Vanhems and Van Keilegom (2013), who suppose that some of the regressors are endogenous as a result of e.g. omitted variables, measurement error or simultaneous equations. We also like to mention the work by Horowitz (1996), who worked with a nonparametric transformation Λ and a parametric regression function m, and the papers by Horowitz (2001) and Jacho-Chavez, Lewbel and Linton (2008), who suppose that both Λ and m are nonparametric.

All the above papers focus on the problem of estimation of a transformation model (that can be of parametric, semiparametric or nonparametric nature). As far as we know, no paper has considered so far the problem of testing in a transformation model. Several aspects of the model can be tested, like the form of the transformation, the form of the regression function, the homoscedasticity of the error term, or the separability of the regression model. In this paper, we like to test the hypothesis

$$H_0 : m \in \mathcal{M}, \tag{1.2}$$

where $\mathcal{M} = \{m_\beta : \beta \in \mathcal{B}\}$ is some parametric class of regression functions and $\mathcal{B} \subset \mathbb{R}^q$. We will construct two test statistics, which measure a certain distance between the distribution function of ε estimated in a semiparametric way and the distribution function of ε estimated under the null hypothesis. We will show that the two distributions are equal if and only if the null hypothesis H_0 is true.

The idea of testing the form of the regression function by comparing two estimators of the distribution of the error term, was introduced for the first time by Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). Their test was developed for a nonparametric location-scale model without transforming the response variable. In the present paper we will see how their ideas and methodology can be carried over to a transformation model. For the same location-scale model, a similar testing approach was also used (among others) by Pardo-Fernández, Van Keilegom and González-Manteiga (2007) for testing the equality of regression curves, and by Dette, Neumeyer and Van Keilegom (2007) for testing the form of the variance function. All these papers build further on the work of Akritas and Van Keilegom (2001), who studied the asymptotic properties of a nonparametric estimator of the error distribution in a location-scale model without transforming the response variable.

Instead of using the idea based on the comparison of error distributions, other approaches could be used as well. We refer to the nice review paper by González-Manteiga and Crujeiras (2013) for a recent overview of developments on goodness-of-fit tests for regression models. Among the possible alternative testing procedures are the tests in the spirit of the seminal papers by Härdle and Mammen (1993) and Stute (1997). They will be considered in forthcoming papers.

The paper is organized as follows. In the next section we explain in detail the testing procedure. Section 3 contains the main asymptotic results concerning the proposed test statistics. In Section 4 we explain how the critical values of these test statistics can be obtained using a bootstrap procedure, and a simulation study is carried out to illustrate the performance of our tests. Section 5 is devoted to the application of our testing procedures to data on the scattering of sunlight in the atmosphere, and in Section 6 we give some general conclusions. Finally, the Appendix contains the technical assumptions and the proofs of the main results.

2 The proposed test

2.1 Notations and definitions

We suppose that we have randomly drawn an *iid* sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from model (1.1), where we denote the components of X_i by $(X_{i1}, \ldots, X_{id})$ for $i = 1, \ldots, n$. We denote by F_X and F_ε the distribution functions of X and ε respectively. The probability density functions of X and ε will be denoted respectively by f_X and f_ε. Moreover, we assume that X has compact support $\chi \subset \mathbb{R}^d$, we define the regression function

$$m(x, \theta) = E[\Lambda_\theta(Y)|X = x] \ ,$$

and let $\sigma^2 = V(\varepsilon)$. Note that $m(x, \theta_0) = m(x)$. We also denote

$$\frac{\partial}{\partial x} f_X(x) = \left(\frac{\partial}{\partial x_1} f_X(x), \ldots, \frac{\partial}{\partial x_d} f_X(x) \right)^t \ ,$$

which is a $(d \times 1)$-vector where $x = (x_1, \ldots, x_d)^t$, and let

$$\dot{\Lambda}_\theta(y) = \left(\frac{\partial}{\partial \theta_1} \Lambda_\theta(y), \ldots, \frac{\partial}{\partial \theta_k} \Lambda_\theta(y) \right)^t$$

be a $(k \times 1)$-vector where $\theta = (\theta_1, \ldots, \theta_k)^t$. Similar notations will be used for other functions. For any function φ, we define $\varphi'(u) = \partial \varphi / \partial u$. Finally, let $\varepsilon(\theta) = \Lambda_\theta(Y) - m(X, \theta)$ and let $F_{\varepsilon(\theta)}$ and $f_{\varepsilon(\theta)}$ be the distribution and the density function of $\varepsilon(\theta)$, respectively.

2.2 Estimation of the model

We start by estimating the parameter θ. Linton, Sperlich and Van Keilegom (2008) proposed two estimation methods for the unknown true parameter vector θ_0 : a profile likelihood method and a mean squared distance from independence method. Here, we will use the profile likelihood estimator, since it was shown in the latter paper that it outperforms the other estimator. Note however that our model and estimation method are slightly different from what Linton, Sperlich and Van Keilegom (2008) did : we assume that $m(\cdot)$ is completely unspecified (whereas they assume an additive or multiplicative structure on $m(\cdot)$), and we will use local polynomial smoothing (instead of kernel smoothing based on higher order kernels). This has however no impact on how the profile likelihood estimator of θ is constructed.

The idea of the profile likelihood method is to calculate the log-likelihood function of Y given X and to replace all unknown expressions by nonparametric estimators. The log-likelihood function of Y given X is given by :

$$\sum_{i=1}^{n} \left\{ \log f_{\varepsilon(\theta_0)}(\Lambda_{\theta_0}(Y_i) - m(X_i, \theta_0)) + \log \Lambda'_{\theta_0}(Y_i) \right\} .$$

In this expression, $\Lambda_{\theta_0}(\cdot)$ and $\Lambda'_{\theta_0}(\cdot)$ are known (except for the parameter θ_0), unlike $m(\cdot, \theta_0)$ and $f_{\varepsilon(\theta_0)}(\cdot)$. We will replace these two quantities by nonparametric estimators. First, for an arbitrary point $x = (x_1, \ldots, x_d)^t$ in the support χ of X, we start by estimating the regression function $m(x, \theta)$ by a local polynomial estimator of degree p (like in Neumeyer and Van Keilegom (2010)), i.e. $\widehat{m}(x, \theta) = \widehat{b}_0(\theta)$ where $\widehat{b}_0(\theta)$ is the first component of the vector $\widehat{b}(\theta)$, which is the solution of the following local minimization problem :

$$\min_b \sum_{i=1}^{n} (\Lambda_\theta(Y_i) - P_i(b, x, p))^2 K_1\left(\frac{X_i - x}{h}\right) ,$$

where $P_i(b, x, p)$ is a polynomial of order p built up with all products of $0 \le l \le p$ factors of the form $X_{ij} - x_j$ for $j = 1, \ldots, d$. Moreover, $h = (h_1, \ldots, h_d)^t$ is a d-dimensional bandwidth vector and for $u = (u_1, \ldots, u_d)^t$, $K_1(u)$ is a d-dimensional product kernel of the form $K_1(u) = \prod_{j=1}^d k_1(u_j)$ where k_1 is a univariate kernel. We introduce also the following notation : $K_{1h}(u) = \prod_{j=1}^d k_1(u_j/h_j)/h_j$. Second, we estimate the error density function $f_{\varepsilon(\theta)}(y)$ by the classical kernel estimator of a density function :

$$\widehat{f}_{\varepsilon(\theta)}(y) = \frac{1}{ng} \sum_{i=1}^{n} k_2\left(\frac{y - \widehat{\varepsilon}_i(\theta)}{g}\right) ,$$

where $\widehat{\varepsilon}_i(\theta) = \Lambda_\theta(Y_i) - \widehat{m}(X_i, \theta)$, k_2 is a kernel (which can be different from k_1) and g is a bandwidth. We define $k_{2g}(u) = k_2(u/g)/g$. The profile likelihood estimator of θ is now defined by :

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \left\{ \log \widehat{f}_{\varepsilon(\theta)}(\Lambda_\theta(Y_i) - \widehat{m}(X_i, \theta)) + \log \Lambda'_\theta(Y_i) \right\} .$$

The asymptotic properties of this estimator have been established by Linton, Sperlich and Van Keilegom (2008). In their Theorem 4.1, they prove the following asymptotic representation for $\widehat{\theta} - \theta_0$:

$$\widehat{\theta} - \theta_0 = -n^{-1}\Gamma^{-1}\sum_{i=1}^{n}\xi(\theta_0, X_i, Y_i) + o_P(n^{-1/2}) \ ,$$

where

$$\xi(\theta, X, Y) = \frac{1}{f_{\varepsilon(\theta)}(\varepsilon(\theta))}[f'_{\varepsilon(\theta)}(\varepsilon(\theta))(\dot{\Lambda}_\theta(Y) - \dot{m}(X,\theta)) + \dot{f}_{\varepsilon(\theta)}(\varepsilon(\theta))] + \frac{\dot{\Lambda}'_\theta(Y)}{\Lambda'_\theta(Y)} \ ,$$

and

$$\Gamma = \frac{\partial}{\partial\theta}E[\xi(\theta, X, Y)]\bigg|_{\theta=\theta_0} \ .$$

We will also denote

$$g(X, Y) = \Gamma^{-1}\xi(\theta_0, X, Y) \ . \tag{2.1}$$

However, as mentioned before, our model and estimation method are slightly different from those considered in Linton, Sperlich and Van Keilegom (2008). In the Appendix, we will show that their Theorem 4.1 continues to hold true in our case, under appropriate regularity conditions. Finally, let (for reasons of simplicity of notation)

$$\widehat{m}(x) = \widehat{m}(x, \widehat{\theta}) \ .$$

2.3 The test statistics

The main idea of the test statistics is to compare the distribution function of the error term $\varepsilon = \Lambda_{\theta_0}(Y) - m(X)$ estimated in a semiparametric way with the distribution function of ε estimated under H_0. That this leads to a valid testing procedure, is shown in the next theorem.

Theorem 2.1. *Let m be a continuous function. Then, H_0 is valid if and only if the random variables*

$$\Lambda_{\theta_0}(Y) - m(X) \quad \text{and} \quad \Lambda_{\theta_0}(Y) - m_{\widetilde{\beta}_0}(X)$$

have the same distribution, where $\widetilde{\beta}_0 = \arg\min_{\beta\in\mathcal{B}} E[(m(X) - m_\beta(X))^2]$.

The proof is given in the Appendix. Clearly, when H_0 is true, then $\widetilde{\beta}_0 = \beta_0$ where β_0 is the true value of β under H_0. We remind that $F_\varepsilon(y) = F_{\varepsilon(\theta_0)}(y) = P(\Lambda_{\theta_0}(Y) - m(X) \leq y)$, and we define $F_{\varepsilon_0}(y) = P(\Lambda_{\theta_0}(Y) - m_{\widetilde{\beta}_0}(X) \leq y)$. Next, we explain how to estimate $F_\varepsilon(\cdot)$ and $F_{\varepsilon_0}(\cdot)$ in order to construct the test statistics. First, define

$$\widehat{F}_\varepsilon(y) = n^{-1} \sum_{i=1}^{n} I(\widehat{\varepsilon}_i \leq y) \ , \tag{2.2}$$

where $\widehat{\varepsilon}_i = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}(X_i)$ are the semiparametric residuals. Second, we estimate $m_{\widetilde{\beta}_0}(x)$ by the least squares method for nonlinear regression, i.e. we estimate $m_{\widetilde{\beta}_0}(x)$ by $m_{\widehat{\beta}}(x)$, where $\widehat{\beta}$ is a minimizer over $\beta \in \mathcal{B}$ of the expression

$$S_n(\beta) = n^{-1} \sum_{i=1}^{n} (\Lambda_{\widehat{\theta}}(Y_i) - m_\beta(X_i))^2 \ . \tag{2.3}$$

Next, we follow the idea of Härdle and Mammen (1993) and we smooth the function $m_{\widehat{\beta}}(x)$ by a local polynomial estimator of degree p, i.e. we define $\widehat{m}_{\widehat{\beta}}(x) = \widehat{c}_0$, where $\widehat{c}_0$ is the first component of the vector $\widehat{c}$, which is the solution of the following local minimization problem :

$$\min_c \sum_{i=1}^{n} (m_{\widehat{\beta}}(X_i) - P_i(c, x, p))^2 K_1\left(\frac{X_i - x}{h}\right) ,$$

where $P_i(c, x, p)$ is a polynomial of order p built up with all products of $0 \leq l \leq p$ factors of the form $X_{ij} - x_j$ for $j = 1, \ldots, d$. Note that we use here the same d-dimensional kernel K_1, the same d-dimensional bandwidth h and the same order p of the local polynomial as in the local polynomial estimator of the regression function $m(x)$. This is to ensure that these two estimators have the same asymptotic bias under H_0. Hence, we obtain the following estimator of the distribution function of ε under H_0 :

$$\widehat{F}_{\varepsilon_0}(y) = n^{-1} \sum_{i=1}^{n} I(\widehat{\varepsilon}_{i0} \leq y) \ , \tag{2.4}$$

where $\widehat{\varepsilon}_{i0} = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}_{\widehat{\beta}}(X_i)$ are the residuals estimated under H_0. The test statistics that we will use are Kolmogorov-Smirnov and Cramér-von Mises type statistics defined by

$$T_{KS} = n^{1/2} \sup_{y \in \mathbb{R}} |\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y)| \quad \text{and} \quad T_{CM} = n \int (\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y))^2 \, d\widehat{F}_\varepsilon(y) \ .$$

Next, to study the power of the test statistics, we consider the following local alternative hypothesis :

$$H_{1n} : m(x) = m_{\beta_0}(x) + n^{-1/2} r(x) \text{ for all } x$$

for some fixed function $r \neq 0$. Note that the local alternative H_{1n} only affects the regression function $m(x)$ and not the error distribution.

3 Asymptotic results

Before stating the main results of this paper, we need to introduce the following notations :

$$\Omega = \left\{ E\left[\frac{\partial m_{\beta_0}(X)}{\partial \beta_r} \left(\frac{\partial m_{\beta_0}(X)}{\partial \beta_s} \right)^t \right] \right\}_{r,s=1,\ldots,q},$$

$$\eta_\beta(x,y) = \Omega^{-1} \frac{\partial m_\beta(x)}{\partial \beta} (\Lambda_{\theta_0}(y) - m_\beta(x)),$$

where

$$\frac{\partial m_\beta(x)}{\partial \beta} = \left(\frac{\partial m_\beta(x)}{\partial \beta_1}, \ldots, \frac{\partial m_\beta(x)}{\partial \beta_q} \right)^t$$

is a $(q \times 1)$-vector and $\beta = (\beta_1, \ldots, \beta_q)^t$. The regularity conditions under which the results of this section are valid, can be found in the Appendix.

3.1 Results under H_0

First, under H_0, the following theorem states an asymptotic representation for $\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y)$ and gives the limiting distribution of the process $n^{1/2}(\widehat{F}_\varepsilon(\cdot) - \widehat{F}_{\varepsilon_0}(\cdot))$.

Theorem 3.1. *Assume (A1)-(A9) and suppose that H_0 holds.*

(i) Then,

$$\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y) = f_\varepsilon(y)\ n^{-1} \sum_{i=1}^n H(X_i, Y_i, \theta_0, \beta_0) + R_n(y)\ ,$$

where

$$\begin{aligned} H(X,Y,\theta,\beta) \;=\; & \Lambda_\theta(Y) - m(X) - \int \left(\frac{\partial m_\beta(x)}{\partial \beta} \right)^t dF_X(x)\ \eta_\beta(X,Y) - E[(\dot{\Lambda}_\theta(Y))^t] g(X,Y) \\ & + \int \left(\frac{\partial m_\beta(x)}{\partial \beta} \right)^t dF_X(x)\ \Omega^{-1} E\left[\frac{\partial m_\beta(X)}{\partial \beta} (\dot{\Lambda}_\theta(Y))^t \right] g(X,Y)\ , \end{aligned}$$

where $\sup_y |R_n(y)| = o_P(n^{-1/2})$ *and* $g(X,Y)$ *is defined in (2.1).*

(ii) Moreover, the process $n^{1/2}(\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y))$ *($-\infty < y < +\infty$) converges weakly to* $f_\varepsilon(y)W$*, where* W *is a zero mean normal random variable with variance*

$$V(W) = E[H^2(X, Y, \theta_0, \beta_0)] \ .$$

This theorem states that the difference between the two empirical distribution functions factorizes in the error density function and a certain sum of *iid* terms, plus negligeable terms. Note that the second term in this asymptotic expansion is due to the estimation of β and the third and the last terms are due to the estimation of θ. If β and θ would be known, then $V(W)$ would simply be equal to σ^2.

As a consequence, we obtain the following corollary, which gives the limiting distribution of the Kolmogorov-Smirnov and Cramér-von Mises statistics under H_0.

Corollary 3.1. *Assume (A1)-(A9). Then, under* H_0*,*

$$T_{KS} \xrightarrow{d} \sup_{y \in \mathbb{R}} |f_\varepsilon(y)| \, |W| \qquad \text{and} \qquad T_{CM} \xrightarrow{d} \int f_\varepsilon^2(y) \, dF_\varepsilon(y) \, W^2 \ .$$

3.2 Results under H_{1n}

First, we define $S_0(\beta) = \sigma^2 + E[(m_\beta(X) - m_{\beta_0}(X))^2]$, and

$$\widetilde{S}_{0n}(\beta) = \sigma^2 + E[(m_\beta(X) - m(X))^2] \ , \tag{3.1}$$

and let $\widetilde{\beta}_{0n}$ be a minimizer over $\beta \in \mathcal{B}$ of $\widetilde{S}_{0n}(\beta)$, which depends on n under H_{1n}. Similarly to Section 3.1, but now under H_{1n}, the following theorem states an asymptotic representation for $\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y)$ and gives the limiting distribution of the process $n^{1/2}(\widehat{F}_\varepsilon(\cdot) - \widehat{F}_{\varepsilon_0}(\cdot))$.

Theorem 3.2. *Assume (A1)-(A10) and suppose that* H_{1n} *holds.*

(i) Then,

$$\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y) = f_\varepsilon(y) \ n^{-1} \sum_{i=1}^n H(X_i, Y_i, \theta_0, \widetilde{\beta}_{0n}) + n^{-1/2} f_\varepsilon(y) b + R_n(y) \ ,$$

where $\sup_y |R_n(y)| = o_P(n^{-1/2})$*,* $H(X, Y, \theta, \beta)$ *is defined in Theorem 3.1 and*

$$b = -\int \left(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \right)^t dF_X(x) \ \Omega^{-1} \int r(x) \ \frac{\partial m_{\beta_0}(x)}{\partial \beta} \, dF_X(x) + \int r(x) \, dF_X(x) \, .$$

(ii) Moreover, the process $n^{1/2}(\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y))$ *($-\infty < y < +\infty$) converges weakly to* $f_\varepsilon(y)(W + b)$*, where* W *is the same normal random variable as in Theorem 3.1(ii).*

Note that the bias term $f_\varepsilon(y)b$ equals zero under H_0, i.e. when $r \equiv 0$. Finally, the following corollary states the limiting distribution of the two test statistics under the local alternative H_{1n}.

Corollary 3.2. *Assume (A1)-(A10). Then, under* H_{1n}*,*

$$T_{KS} \xrightarrow{d} \sup_{y \in \mathbb{R}} |f_\varepsilon(y)| \, |W + b| \qquad \textit{and} \qquad T_{CM} \xrightarrow{d} \int f_\varepsilon^2(y) \, dF_\varepsilon(y) \, (W + b)^2 \ .$$

One advantage of our approach is that our tests can detect alternatives at the rate $n^{-1/2}$, which is faster than the rate $n^{-1/2}h^{-d/4}$ obtained by other approaches in the literature. Note however that there are situations in which the random variable W defined in Theorem 3.1 is non degenerate, and the bias term b in Theorem 3.2 is equal to zero. Take e.g. the case where X is uniform on $[-1, 1]$, $r(x) = x$ and we are interested in testing $H_0 : m(x) = \beta x^2$ for all x. Then, it is easily seen that $b = 0$, whereas in general $H(X, Y, \theta_0, \beta_0)$ will be a.s. different from zero. Although this example shows that cases can be constructed where the tests have no power under the local alternative H_{1n}, there are very many cases where the test is consistent under H_{1n}. Similar features have been found in other papers, see e.g. Van Keilegom, González-Manteiga and Sánchez-Sellero (2008) and Pardo-Fernández, Van Keilegom and González-Manteiga (2007), among others. Also, remind that our test is consistent in the sense of Theorem 2.1.

In order to apply the result of Corollary 3.1 in practice, we need to estimate the limiting distribution of T_{KS} and T_{CM} by plugging in estimators of f_ε, $\dot{f}_\varepsilon$, f'_ε, $\dot{m}$ and f_X. Although this is in principle possible, it is not an easy task, as it requires the introduction of new bandwidths. Therefore, we prefer to approximate the distribution of the test statistics under H_0 by using a bootstrap procedure. This will be described in detail in the next section.

4 Simulations

In this section, we carry out simulations to evaluate the performance of our proposed tests for small samples. The simulated model is $\Lambda_\theta(Y_i) = 3 + \beta X_i + c(X_i) + \varepsilon_i$, where Λ_θ is the

Box-Cox (1964) transformation

$$\Lambda_\theta(y) = \begin{cases} \frac{y^\theta - 1}{\theta}, & \theta \neq 0 \\ \log(y), & \theta = 0. \end{cases}$$

Moreover, $X_1, \ldots, X_n$ are independent, of dimension $d = 1$ and uniformly distibuted on [0,1], and $\varepsilon_1, \ldots, \varepsilon_n$ are independent standard normal random variables truncated on [-3,3]. We consider the following null hypothesis :

$$H_0 : m(x) = 3 + \beta x \quad \text{for all } x \ .$$

We perform simulations for three different values of the parameter θ : $\theta_0 = 0$ which corresponds to a logarithmic transformation, $\theta_0 = 0.5$ which corresponds to a square root transformation and $\theta_0 = 1$ which corresponds to the identity. The true value of the parameter β is $\beta_0 = 2$. The term $c(x)$ represents different deviations from the null hypothesis and we consider here $c(x) = 5x^2$, $c(x) = 7.5x^2$, $c(x) = 10x^2$, $c(x) = 0.5\exp(x)$, $c(x) = \exp(x)$, $c(x) = 2\exp(x)$, $c(x) = 3\exp(x)$, $c(x) = 0.25\sin(2\pi x)$, $c(x) = 0.5\sin(2\pi x)$, $c(x) = \sin(2\pi x)$ and $c(x) = 1.5\sin(2\pi x)$ for sample size $n = 200$.

Next, we use the Epanechnikov kernel $k_1(x) = k_2(x) = \frac{3}{4}(1 - x^2)\, 1_{\{|x| \leq 1\}}$ for both the estimator of the regression function and the density function. For the estimation of θ, h, g and β, we proceed as follows. We maximize the following function with respect to θ for some optimal values of h and g :

$$l_\theta(h, g) = \sum_{i=1}^{n} \left\{ \log \widehat{f}_{\varepsilon(\theta)}(\Lambda_\theta(Y_i) - \widehat{m}(X_i, \theta, h)) + \log \Lambda'_\theta(Y_i) \right\} ,$$

where $\widehat{m}(x, \theta, h)$ denotes $\widehat{m}(x, \theta)$ constructed with a bandwidth h. For each value of θ, let $h^*(\theta)$ be the bandwidth obtained by least squares cross-validation :

$$h^*(\theta) = \arg\min_h \sum_{i=1}^{n} (\Lambda_\theta(Y_i) - \widehat{m}_{-i,\theta}(X_i))^2 \ ,$$

where

$$\widehat{m}_{-i,\theta}(X_i) = \frac{\sum_{j=1, j\neq i}^{n} \Lambda_\theta(Y_j) k_1\left(\frac{X_j - X_i}{h}\right)}{\sum_{j=1, j\neq i}^{n} k_1\left(\frac{X_j - X_i}{h}\right)} \ .$$

Moreover, g can be chosen by a classical bandwidth selection rule for kernel density estimation. For simplicity, we choose here the normal reference rule, i.e. $\widehat{g}(\theta) = (40\sqrt{\pi})^{1/5} n^{-1/5} \widehat{\sigma}_{\widehat{\varepsilon}(\theta, h^*(\theta))}$, where $\widehat{\sigma}_{\widehat{\varepsilon}(\theta,h^*(\theta))}$ is the classical estimator of the standard deviation of the error term $\widehat{\varepsilon}(\theta, h^*(\theta)) = \Lambda_\theta(Y) - \widehat{m}(X, \theta, h^*(\theta))$. Consequently, the optimal value of θ, $\widehat{\theta} = \arg\max_\theta l_\theta(h^*(\theta), \widehat{g}(\theta))$, is obtained iteratively with the function *optimize* in R. Finally, to estimate β, we minimize the following expression :

$$\widehat{\beta} = \arg\min_\beta \sum_{i=1}^n (\Lambda_{\widehat{\theta}}(Y_i) - m_\beta(X_i))^2 \ .$$

The critical values of the test statistics T_{KS} and T_{CM} will be approximated by means of the following bootstrap procedure. First, we standardize the residuals $\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n$ in order to have mean zero. Let $\check{F}_\varepsilon$ be the empirical distribution of these standardized residuals. Next, we smooth the estimator $\check{F}_\varepsilon$ by using a small bandwidth equal to 0.1, and we denote this new estimator by $\widetilde{F}_\varepsilon$. From this smoothed distribution, bootstrap samples of the errors will be drawn. Note that we have to work with a smoothed distribution in the bootstrap procedure, because the asymptotic representation of $\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y)$ given in Theorem 3.1 involves the density $f_\varepsilon(y)$ (see e.g. Silverman and Young (1987) or Neumeyer (2009) for similar bootstrap procedures). The bootstrap procedure can now be described as follows. For fixed B and for $b = 1, \dots, B$:

1. Let $\varepsilon^*_{1b}, \dots, \varepsilon^*_{nb}$ be independent random errors drawn from $\widetilde{F}_\varepsilon$, and let $X^*_{ib} = X_i$ ($i = 1, \dots, n$).

2. Define new responses $Y^*_{ib} = \Lambda^{-1}_{\widehat{\theta}}(m_{\widehat{\beta}}(X^*_{ib}) + \varepsilon^*_{ib})$, $i = 1, \dots, n$, obtained under the null hypothesis.

3. Let $T^*_{KS,b}$ and $T^*_{CM,b}$ be the test statistics obtained from the bootstrap sample (X^*_{ib}, Y^*_{ib}), $i = 1, \dots, n$.

Then, the $[(1-\alpha)B]$-th order statistic of $T^*_{KS,1}, \dots, T^*_{KS,B}$ approximates the $(1-\alpha)$-th quantile of the distribution of T_{KS}, and similarly for T_{CM}. We refer to Neumeyer (2009) for the consistency of this bootstrap procedure in the case where the response is not transformed. In our simulations, we take $B = 250$.

Tables 1 and 2 show respectively the percentage of rejection under the null hypothesis and under the different deviations $c(x)$ we have introduced above. These percentages of rejection are obtained with the test statistics T_{KS} and T_{CM} for 500 samples. The nominal level is 5%.

	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
$c(x)$	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	4.2	5.8	5.0	7.0	6.4	8.4

Table 1: Percentage of rejection under the null hypothesis (nominal level 5%) for samples of size $n = 200$.

	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
$c(x)$	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
$5x^2$	96.6	96.6	52.2	55.6	11.0	10.0
$7.5x^2$	98.4	98.6	90.2	89.4	67.8	68.4
$10x^2$	99.0	98.8	92.6	93.2	79.6	78.8
$0.5\exp(x)$	31.8	44.2	16.4	19.8	11.0	14.2
$\exp(x)$	89.8	94.6	38.6	42.8	27.2	29.4
$2\exp(x)$	97.4	97.2	70.0	71.8	47.0	47.8
$3\exp(x)$	99.6	99.6	72.8	72.8	55.4	55.4
$0.25\sin(2\pi x)$	11.8	16.2	12.2	14.4	11.4	13.8
$0.5\sin(2\pi x)$	25.2	29.8	17.2	19.6	14.6	16.8
$\sin(2\pi x)$	51.4	52.8	23.2	24.4	15.0	16.2
$1.5\sin(2\pi x)$	76.4	75.8	26.0	24.6	20.6	22.2

Table 2: Percentage of rejection under the alternative hypothesis (nominal level 5%) for samples of size $n = 200$.

We see that the different estimations of the nominal level under H_0 are globally good and we note that the results for T_{KS} are sligthly better. Indeed, the percentages of rejection given by T_{CM} are a little bit too high. Next, under the alternative, the power is largest for $\theta_0 = 0$, followed by $\theta_0 = 0.5$ and then $\theta_0 = 1$. This result seems logical, because Linton, Sperlich and Van Keilegom (2008) showed that the mean squared error of the profile likelihood estimator of θ is largest for $\theta_0 = 1$, followed by $\theta_0 = 0.5$ and then $\theta_0 = 0$. Moreover, we see that the percentages of rejection are slightly larger for T_{CM} than for T_{KS}, which is in line with what happens under H_0.

5 Application

We apply our testing procedure to a data set composed of 355 observations resulting from an experiment on the scattering of sunlight in the atmosphere (see Bellver (1987)). The data can be found in Cleveland (1993). The response Y is the scattering angle at which the polarization of sunlight vanishes, called the Babinet point. Note that the response is positive, which justifies the use of a Box-Cox transformation. Moreover, the covariate X is the cube root of a measure of particulate concentration in the atmosphere and we standardize it.

This data set has already been analyzed, but without transformation of the response variable, in different articles, like in Hart (1997), in Zhang (2003) and in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). A test for linearity of the underlying regression function was realized in Hart (1997), while different tests for lth degree polynomial regression ($l = 1, 2, 3, 4$) were realized in Zhang (2003) and in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008), both with their own testing procedure.

Here, a Box-Cox transformation of the response variable is considered and we check the goodness-of-fit of the lth degree polynomial regression ($l = 1, 2, 3, 4$) by using the Kolmogorov-Smirnov and the Cramér-von Mises test statistics defined in this paper. The distributions and p-values of these two test statistics are approximated by bootstrap on the basis of 1000 replicates. The results are given in Table 3.

	$l=1$	$l=2$	$l=3$	$l=4$
T_{KS}	0.333	0.734	0.737	0.475
T_{CM}	0.208	0.748	0.666	0.549

Table 3: p-values of the lth degree polynomial fit for the sunlight data.

First, note that the profile likelihood estimator of θ is equal to $\widehat{\theta} = 1.9428$. This implies that we transform the response variable Y by taking approximatively its square. Table 3 indicates that there is no evidence against a polynomial fit of order $l = 2, 3, 4$, similarly as in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). Moreover, there is also no evidence against a linear fit, which is different from the conclusions in Hart (1997), Zhang (2003) and Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). This can be explained by the transformation realized on the response variable Y, which has an important impact on the regression function.

6 Conclusions and future research

In this paper, we constructed a test for the parametric form of the regression function in a semiparametric transformation model. The transformation of the dependent variable in this model was supposed to belong to some parametric family of strictly increasing functions. We defined a Kolmogorov-Smirnov and a Cramér-von Mises test statistic, where the main idea was to compare the distribution of the error term estimated in a semiparametric way to the one estimated under H_0. We established the limiting distribution of these two test statistics under the null hypothesis and under a local alternative. We evaluated the performance of our test by means of some simulations and we applied our method on a real data set.

It would be interesting to extend the paper of Pardo-Fernández, Van Keilegom and González-Manteiga (2007) by constructing a test for the equality of regression curves in the case of semiparametric transformation models. Another possibility of future research could be the extension of this paper to the case of censored data. Finally, the extension of the methods of Härdle and Mammen (1993) and Stute (1997) to the context of transformation models would be an useful alternative for the tests developed in this paper, and it would then be informative to know under which model conditions which test behaves best.

7 Appendix: Proofs

The Appendix is structured as follows. We start in Subsection 7.1 by introducing a number of notations and by stating the assumptions under which the main results of this paper are valid. Then, in Subsections 7.2, 7.3 and 7.4, we prove certain results regarding, respectively, the estimation of θ, β and $m(\cdot)$. Finally, in Subsection 7.5, the main results of this paper are shown.

7.1 Notations and technical assumptions

For $0 < \alpha < \delta/2$, where δ is defined as in condition $(A2)$ (see below), let $C_1^{d+\alpha}(\chi)$, be the set of d-times differentiable functions $f : \chi \to \mathbb{R}$ such that :

$$||f||_{d+\alpha} := \max_{j. \leq d} \sup_{x \in \chi} |D^j f(x)| + \max_{j. = d} \sup_{x, x' \in \chi} \frac{|D^j f(x) - D^j f(x')|}{||x - x'||^\alpha} \leq 1$$

where $j = (j_1, \ldots, j_d)$, $j. = \sum_{i=1}^d j_i$, $D^j = \frac{\partial^{j.}}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}$ and $||.||$ is the Euclidean norm on $\mathbb{R}^d$.

The main results of the asymptotic theory require the following regularity conditions on the kernels, the bandwidths, the distributions of X and ε, the transformation Λ_θ and the functions $m_\beta(x)$, $m(x)$ and $r(x)$:

(A1) The functions k_j $(j = 1, 2)$ are symmetric, have support [-1,1], $\int k_1(u)\,du = 1$, $\int u^k k_2(u)\,du = 0$ for $k = 1, \ldots, q_2 - 1$ and $\int u^{q_2} k_2(u)\,du \neq 0$ for some $q_2 \geq 4$. Moreover, k_1 is d-times continuously differentiable, $k_1^{(l)}(\pm 1) = 0$ for $l = 0, \ldots, d-1$ and k_2 is twice continuously differentiable.

(A2) h_l (for $l = 1, \ldots, d$) satisfies $h_l/h \to c_l$ for some $0 < c_l < \infty$ and the bandwidths h and g satisfy $nh \to \infty$, $nh^{2p+2} \to 0$ for some $p \geq 3$, $nh^{3d+\delta} \to \infty$ for some $\delta > 0$, $ng \prod_{i=1}^d h_i \to \infty$, $ng^6(\ln g^{-1})^{-2} \to \infty$ and $ng^{2q_2} \to 0$ when $n \to \infty$, where q_2 is defined in condition $(A1)$.

(A3) (i) The support χ of the covariate X is a compact subset of $\mathbb{R}^d$.

(ii) The distribution function F_X is $2d + 1$-times continuously differentiable.

(iii) $\inf_x f_X(x) > 0$, $\sup_x f_X(x) < \infty$ and $\sup_x ||\frac{\partial}{\partial x} f_X(x)|| < \infty$.

(A4) (i) The error term $\varepsilon = \Lambda_{\theta_0}(Y) - m(X)$ has finite fourth moment and is independent of X.

(ii) The distribution function $F_{\varepsilon(\theta)}(y)$ is three times continuously differentiable with respect to y and θ, and

$$\sup_{\theta,y} \left|\left| \frac{\partial^{i+j}}{\partial y^i \partial \theta^j} F_{\varepsilon(\theta)}(y) \right|\right| < \infty \qquad \text{and} \qquad E\left[\sup_{\substack{y \\ ||\theta - \theta_0|| \leq \alpha}} \left|\left| \frac{\partial^2}{\partial y \partial \theta} F_{\varepsilon(\theta)}(y|X) \right|\right| \right] < \infty$$

for all i and j such that $0 \leq i + j \leq 2$, and for some $\alpha > 0$.

(A5) (i) The transformation $\Lambda_\theta(y)$ is three times continuously differentiable with respect to both y and θ, and there exists $\alpha > 0$ such that :

$$E\left[\sup_{\theta' : ||\theta' - \theta|| \leq \alpha} \left|\left| \frac{\partial^{i+j}}{\partial y^i \partial \theta^j} \Lambda_{\theta'}(Y) \right|\right| \right] < \infty$$

for all $\theta \in \Theta$ and all i and j such that $0 \leq i + j \leq 3$.

(ii) $\sup_{x \in \chi} ||E[\dot{\Lambda}_{\theta_0}^4(Y)|X = x]|| < \infty$.

(iii) $\sup_{\theta,x} ||E[\dot{\Lambda}_\theta(Y)|X = x]|| < \infty$.

(iv) The density function of $(\dot{\Lambda}_\theta(Y), X)$ exists and is continuous for all $\theta \in \Theta$.

(A6) (i) $\mathcal{B}$ is a compact subset of $\mathbb{R}^q$ and β_0 is an interior point of $\mathcal{B}$.

(ii) All partial derivatives of $m_\beta(x)$ with respect to the components of x and β of order 0, 1, 2 and 3 exist and are continuous in (x, β) for all x and β.

(iii) For all $\varepsilon > 0$:

$$\inf_{||\beta-\beta_0||>\varepsilon} E[(m_\beta(X) - m_{\beta_0}(X))^2] > 0 \ .$$

(iv) Ω is non singular.

(A7) The functions $m(x)$ and $\frac{\partial}{\partial\theta} m(x, \theta) := \dot{m}(x)$ are $p + 2$ times continuously differentiable with respect to the components of x on $\chi \times N(\theta_0)$, where $N(\theta_0)$ is a neighbourhood of θ_0 and all derivatives up to order $p + 2$ are bounded, uniformly in (x, θ) in $\chi \times N(\theta_0)$.

(A8) (i) For all $\eta > 0$, there exists $\varepsilon(\eta) > 0$ such that

$$\inf_{||\theta-\theta_0||>\eta} ||E(\xi(\theta, X, Y))|| \geq \varepsilon(\eta) > 0 \ .$$

(ii) The matrix Γ is of full rank.

(A9) $\Lambda_{\theta_0}(\alpha) = a$ and $\Lambda_{\theta_0}(\beta) = b$ for some $\alpha < \beta$ and $a < b$, and the set $\{x \in \chi : \frac{\partial}{\partial x} m(x) \neq 0\}$ has nonempty interior.

(A10) $E(r^2(X)) < \infty$ and $r(x)$ is twice continuously differentiable for all x.

Note that condition (A9) is needed for identifying the model (see Vanhems and Van Keilegom (2013)). Moreover, conditions (A6) and (A10) come from Van Keilegom, González-Manteiga and Sánchez Sellero (2008), conditions (A4)(ii), (A5)(i) and (A8) come from Linton, Sperlich and Van Keilegom (2008), condition (A3)(ii) come from Neumeyer and Van Keilegom (2010) and conditions (A4)(i), (A5)(ii) and (A5)(iv) come from Heuchenne, Samb and Van Keilegom (2014). Finally, note that conditions (A1) and (A2), which are assumptions on the different kernels and bandwidths and condition (A7) come partially from Linton, Sperlich and Van Keilegom (2008) and partially from Neumeyer and Van Keilegom (2010).

7.2 Results regarding the estimation of θ

First, we need to define the notion of 'equivalent kernel' introduced by Fan and Gijbels (1996). We omit here the details but the main idea is that the estimator $\widehat{m}(x)$ can be written in the form of a Nadaraya-Watson type estimator :

$$\widehat{m}(x) = \sum_{i=1}^{n} W_0^n\left(\frac{x - X_i}{h}, x\right)\Lambda_{\widehat{\theta}}(Y_i) \ ,$$

where the weights $W_0^n(\cdot)$ depend on x and satisfy

$$\sum_{i=1}^{n} W_0^n\left(\frac{x - X_i}{h}, x\right) = 1 \quad \text{and} \quad W_0^n(u, x) = \frac{1}{n\prod_{j=1}^{d} h_j}\frac{1}{f_X(x)}K_1^*(u)(1 + o_P(1)) \qquad (7.1)$$

uniformly in $u \in [-1, 1]^d$ and $x \in \chi$. Here, K_1^* is the so-called equivalent kernel and is a product kernel $K_1^*(u_1, \ldots, u_d) = \prod_{j=1}^{d} k_1^*(u_j)$ where k_1^* is a univariate kernel that satisfies $\int u^{p'} k_1^*(u)\, du = 1_{\{p'=0\}}$ $(0 \le p' \le p)$. We will denote $K_{1h}^*(u) = \prod_{j=1}^{d} k_1^*(u_j/h_j)/h_j$.

Next, we prove that Lemmas A.1, A.2 and A.3 in Linton, Sperlich and Van Keilegom (2008) remain valid when the regression function $m(x)$ is estimated by a local polynomial estimator of degree p and when a fully nonparametric estimator of $m(x)$ is used. This ensures that we can apply the asymptotic representation of $\widehat{\theta} - \theta_0$ given in Theorem 4.1 in Linton, Sperlich and Van Keilegom (2008) to our context.

Lemma 7.1. *Assume (A1)-(A9). Then, for all* $y \in \mathbb{R}$,

$$\widehat{f}_{\varepsilon(\theta_0)}(y) - f_{\varepsilon(\theta_0)}(y) = n^{-1} f_\varepsilon'(y) \sum_{i=1}^{n} \varepsilon_i + n^{-1} \sum_{i=1}^{n} k_{2g}(\varepsilon_i - y) - f_\varepsilon(y) + \widehat{r}(y) \ ,$$

where $\sup_y |\widehat{r}(y)| = o_P(n^{-1/2})$.

Proof. First, we have by a Taylor expansion :

$$\begin{aligned}
&\widehat{f}_{\varepsilon(\theta_0)}(y) - f_{\varepsilon(\theta_0)}(y) \\
&= \frac{1}{ng}\sum_{i=1}^{n} k_{2g}'(\varepsilon_i - y)(\widehat{\varepsilon}_i(\theta_0) - \varepsilon_i(\theta_0)) + \frac{1}{n}\sum_{i=1}^{n} k_{2g}(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \\
&= -\frac{1}{ng}\sum_{i=1}^{n} k_{2g}'(\varepsilon_i - y)(\widehat{m}(X_i, \theta_0) - m(X_i, \theta_0)) + \frac{1}{n}\sum_{i=1}^{n} k_{2g}(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \ .
\end{aligned}$$

Next, using the equivalent kernels introduced above, we have

$$\begin{aligned}\widehat{f}_{\varepsilon(\theta_0)}(y) - f_{\varepsilon(\theta_0)}(y) &= \Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \varphi_{nij}(\Lambda_{\theta_0}(Y_j) - m(X_i)\Big)(1 + o_P(1)) \\ &\quad + \frac{1}{n}\sum_{i=1}^n k_{2g}(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \ ,\end{aligned}$$

where

$$\varphi_{nij} = -\frac{1}{g}k'_{2g}(\varepsilon_i - y)\frac{1}{f_X(X_i)}\Big\{\prod_{k=1}^d \frac{1}{h_k}k_1^*\Big(\frac{X_{ik} - X_{jk}}{h_k}\Big)\Big\} \ .$$

Since $X_i - X_j = O(h)$, using condition (A2), we have that

$$\widehat{f}_{\varepsilon(\theta_0)}(y) - f_{\varepsilon(\theta_0)}(y) = \Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \varphi_{nij}\varepsilon_j\Big)(1 + o_P(1)) + \frac{1}{n}\sum_{i=1}^n k_{2g}(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \ .$$

As $ng\prod_{i=1}^d h_i \to \infty$ by condition (A2), we can apply Theorem 3.1 in Powell, Stock and Stoker (1989) and the first term on the right hand side above is consequently equal to :

$$\frac{1}{2}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{i<j}\psi_{nij} + o_P(n^{-1/2}) = \frac{1}{2}E(\psi_{nij}) + \frac{1}{n}\sum_{j=1}^n [E(\psi_{nij}|X_j, \varepsilon_j) - E(\psi_{nij})] + o_P(n^{-1/2}) \ ,$$

where $\psi_{nij} = \varphi_{nij}\varepsilon_j + \varphi_{nji}\varepsilon_i$. By independence of X and ε, we have

$$E[\varphi_{nij}\varepsilon_j|X_j, \varepsilon_j] = \varepsilon_j E\Big[-\frac{1}{g}k'_{2g}(\varepsilon_i - y)\Big]E\Big[\frac{1}{f_X(X_i)}\Big\{\prod_{k=1}^d \frac{1}{h_k}k_1^*\Big(\frac{X_{ik} - X_{jk}}{h_k}\Big)\Big\}\Big] \ .$$

By a Taylor expansion, the first expectation is equal to

$$-\frac{1}{g^2}\int k'_2\Big(\frac{t-y}{g}\Big)f_\varepsilon(t)\, dt = \frac{1}{g}\int k_2\Big(\frac{t-y}{g}\Big)f'_\varepsilon(t)\, dt = f'_\varepsilon(y) + O(g^{q_2}) \ ,$$

and the second expectation is equal to 1. Next, $E[\varphi_{nji}\varepsilon_i|X_j, \varepsilon_j] = 0$ since $E(\varepsilon|X) = 0$. Consequently,

$$\frac{1}{n}\sum_{j=1}^n E(\psi_{nij}|X_j, \varepsilon_j) = n^{-1}f'_\varepsilon(y)\sum_{j=1}^n \varepsilon_j + o_P(n^{-1/2})$$

and $E(\psi_{nij}) = 0$. This concludes the proof. □

Note that we can adapt in exactly the same way Lemmas A.2 and A.3 in Linton, Sperlich and Van Keilegom (2008), and so we omit here the details. These lemmas give an asymptotic representation for the estimators $\dot{\widehat{f}}_{\varepsilon(\theta_0)}(y)$ and $\widehat{f}'_{\varepsilon(\theta_0)}(y)$ respectively. These three asymptotic representations are the key ingredients for proving the asymptotic normality of the estimator $\widehat{\theta}$, given below.

Proposition 7.1. *Assume (A1)-(A9). Then,*

$$\widehat{\theta} - \theta_0 = -n^{-1} \sum_{i=1}^{n} g(X_i, Y_i) + o_P(n^{-1/2}) \ ,$$

where $g(X, Y)$ is defined in (2.1). Moreover,

$$n^{1/2}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, V(g(X, Y))) \ .$$

We do not give the proof of this result, as it is very similar to the proof of Theorem 4.1 in Linton, Sperlich and Van Keilegom (2008). The main difference lies in the proofs of Lemmas A.1-A.3, which we have proved above.

7.3 Results regarding the estimation of β

The following three lemmas state the consistency of $\widehat{\beta}$, the consistency of $\widetilde{\beta}_{0n}$, defined in Section 3.2, and an asymptotic representation for $\widehat{\beta} - \beta_0$, all under H_{1n}. Note that to obtain these results under H_0, it suffices to take $r \equiv 0$ and to remove condition (A10).

Lemma 7.2. *Assume (A1)-(A10). Then, under H_{1n},*

$$\widehat{\beta} - \beta_0 \xrightarrow{P} 0 \ .$$

Proof. We verify the two conditions of Theorem 5.7 in Van der Vaart (1998) with $M_n(\beta) = -S_n(\beta)$ and $M(\beta) = -S_0(\beta)$ as in the proof of Lemma 4.1 in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). The second condition, i.e. $\sup_{||\beta-\beta_0||>\varepsilon} -S_0(\beta) < -S_0(\beta_0)$, is equivalent to condition (A6)(iii). To verify the first condition, we have to prove that $\sup_{\beta\in\mathcal{B}} |S_n(\beta) - S_0(\beta)| \xrightarrow{P} 0$. We have :

$$\begin{aligned} S_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n}(\Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i) + \Lambda_{\theta_0}(Y_i) - m_{\beta_0}(X_i) + m_{\beta_0}(X_i) - m_{\beta}(X_i))^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(\Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i))^2 + \frac{1}{n}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i) - m_{\beta_0}(X_i))^2 \end{aligned}$$

$$+\frac{1}{n}\sum_{i=1}^{n}(m_{\beta_0}(X_i)-m_\beta(X_i))^2+\frac{2}{n}\sum_{i=1}^{n}(\Lambda_{\widehat{\theta}}(Y_i)-\Lambda_{\theta_0}(Y_i))(m_{\beta_0}(X_i)-m_\beta(X_i))$$
$$+\frac{2}{n}\sum_{i=1}^{n}(\Lambda_{\widehat{\theta}}(Y_i)-\Lambda_{\theta_0}(Y_i))(\Lambda_{\theta_0}(Y_i)-m_{\beta_0}(X_i))$$
$$+\frac{2}{n}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i)-m_{\beta_0}(X_i))(m_{\beta_0}(X_i)-m_\beta(X_i)) \ .$$

From Proposition 7.1, we know that $\widehat{\theta}-\theta_0=O_P(n^{-1/2})$. By a Taylor expansion, $\Lambda_{\widehat{\theta}}(Y_i)-\Lambda_{\theta_0}(Y_i)=(\dot{\Lambda}_\xi(Y_i))^t(\widehat{\theta}-\theta_0)$ for some ξ between $\widehat{\theta}$ and θ_0. Consequently, using conditions (A5)(i) and (A6)(ii), the first, fourth and fifth terms on the right hand side above are $o_P(1)$, uniformly in β. Next, using Theorem 2 in Jennrich (1969), we have uniformly in β :

$$\frac{1}{n}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i)-m_{\beta_0}(X_i))^2\xrightarrow{P}E[(\Lambda_{\theta_0}(Y)-m_{\beta_0}(X))^2]=\sigma^2 \ ,$$

$$\frac{1}{n}\sum_{i=1}^{n}(m_{\beta_0}(X_i)-m_\beta(X_i))^2\xrightarrow{P}E[(m_{\beta_0}(X)-m_\beta(X))^2] \ ,$$

$$\frac{2}{n}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i)-m_{\beta_0}(X_i))(m_{\beta_0}(X_i)-m_\beta(X_i))\xrightarrow{P}2E[(\Lambda_{\theta_0}(Y)-m_{\beta_0}(X))(m_{\beta_0}(X)-m_\beta(X))] \ .$$

The uniformity in β is given by Theorem 2 in Jennrich (1969). As ε is independent of X and $E(\varepsilon)=0$, this last expression is equal to 0. Consequently, $S_n(\beta)$ converges in probability to $\sigma^2+E[(m_{\beta_0}(X)-m_\beta(X))^2]=S_0(\beta)$. □

Lemma 7.3. *Assume (A1)-(A10). Then, under H_{1n},*

$$\widetilde{\beta}_{0n}-\beta_0\xrightarrow{P}0 \ .$$

Proof. Similarly as in the proof of Lemma 7.2, we will use Theorem 5.7 in Van der Vaart (1998), this time with $M_n(\beta)=-\widetilde{S}_{0n}(\beta)$ and $M(\beta)=-S_0(\beta)$. The proof for the second condition of Theorem 5.7 in Van der Vaart (1998) is exactly the same as in Lemma 7.2. Moreover,

$$\begin{aligned}|\widetilde{S}_{0n}(\beta)-S_0(\beta)| &= |\sigma^2+E[(m_\beta(X)-m(X))^2]-\sigma^2-E[(m_\beta(X)-m_{\beta_0}(X))^2]| \\ &= |E[(m(X)-m_{\beta_0}(X))(m(X)+m_{\beta_0}(X)-2m_\beta(X))]| \\ &= |n^{-1/2}E[r(X)(m(X)+m_{\beta_0}(X)-2m_\beta(X))]|\end{aligned}$$

and this converges to zero, because

$$E[r(X)(m(X)+m_{\beta_0}(X)-2m_\beta(X))] \le \{E[r^2(X)]\}^{1/2}\{E[(m(X)+m_{\beta_0}(X)-2m_\beta(X))^2]\}^{1/2} < \infty$$

by conditions (A10) and (A6)(ii). Consequently, the first condition of Theorem 5.7 in Van der Vaart (1998) is also satisfied and this concludes the proof. □

Lemma 7.4. *Assume (A1)-(A10). Then, under* H_{1n}*,*

$$\begin{aligned}\widehat{\beta}-\beta_0 &= -\Omega^{-1}E\Big[\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}(\dot{\Lambda}_{\theta_0}(Y))^t\Big]n^{-1}\sum_{i=1}^n g(X_i,Y_i)\\ &\quad +n^{-1}\sum_{i=1}^n \eta_{\widetilde{\beta}_{0n}}(X_i,Y_i)+\Omega^{-1}n^{-1/2}\int r(x)\frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial\beta}\,dF_X(x)+o_P(n^{-1/2})\ ,\end{aligned}$$

where $g(X,Y)$ *is defined in (2.1).*

Proof. The proof is somewhat similar to the proof of Lemma 4.2 in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008), but we have to take care of the estimation of the parameter θ_0, whereas in the latter paper the data are not transformed. The proof will be divided in two parts : first we will search an asymptotic expression for $\widehat{\beta}-\widetilde{\beta}_{0n}$ and second we will search an asymptotic expression for $\widetilde{\beta}_{0n}-\beta_0$. By a Taylor expansion, we obtain :

$$\widehat{\beta}-\widetilde{\beta}_{0n} = \left(\frac{\partial^2 S_n(\beta_{1n})}{\partial\beta\partial\beta^t}\right)^{-1}\left(\frac{\partial S_n(\widehat{\beta})}{\partial\beta}-\frac{\partial S_n(\widetilde{\beta}_{0n})}{\partial\beta}\right),$$

where β_{1n} is some value between $\widehat{\beta}$ and $\widetilde{\beta}_{0n}$. By definition of $\widehat{\beta}$, the expression $\frac{\partial S_n(\widehat{\beta})}{\partial\beta}$ is equal to 0. Then,

$$\widehat{\beta}-\widetilde{\beta}_{0n} = -\left(\frac{\partial^2 S_n(\beta_{1n})}{\partial\beta\partial\beta^t}\right)^{-1}\frac{\partial S_n(\widetilde{\beta}_{0n})}{\partial\beta}\ . \tag{7.2}$$

First note that

$$\frac{\partial S_n(\widetilde{\beta}_{0n})}{\partial\beta} = -2n^{-1}\sum_{i=1}^n(\Lambda_{\widehat{\theta}}(Y_i)-m_{\widetilde{\beta}_{0n}}(X_i))\frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta}\ . \tag{7.3}$$

Next, using that $\widehat{\beta}-\widetilde{\beta}_{0n}=o_P(1)$ by Lemmas 7.2 and 7.3 and thus $\beta_{1n}=\widetilde{\beta}_{0n}+o_P(1)=\beta_0+o_P(1)$, we have by condition (A6)(ii) that :

$$\begin{aligned}&\frac{\partial^2 S_n(\beta_{1n})}{\partial\beta\partial\beta^t}\\ &= 2n^{-1}\sum_{i=1}^n\frac{\partial m_{\beta_{1n}}(X_i)}{\partial\beta}\left(\frac{\partial m_{\beta_{1n}}(X_i)}{\partial\beta}\right)^t-2n^{-1}\sum_{i=1}^n(\Lambda_{\widehat{\theta}}(Y_i)-m_{\beta_{1n}}(X_i))\frac{\partial^2 m_{\beta_{1n}}(X_i)}{\partial\beta\partial\beta^t}\end{aligned}$$

$$= 2n^{-1}\sum_{i=1}^n \frac{\partial m_{\beta_0}(X_i)}{\partial\beta}\left(\frac{\partial m_{\beta_0}(X_i)}{\partial\beta}\right)^t - 2n^{-1}\sum_{i=1}^n(\Lambda_{\widehat{\theta}}(Y_i) - m_{\beta_0}(X_i))\frac{\partial^2 m_{\beta_0}(X_i)}{\partial\beta\partial\beta^t} + o_P(1) \ .$$

The first term on the right hand side above is equal to $2E[\frac{\partial m_{\beta_0}(X)}{\partial\beta}(\frac{\partial m_{\beta_0}(X)}{\partial\beta})^t] + o_P(1) = 2\Omega + o_P(1)$. As ε is independent of X and using a Taylor expansion with a certain θ_{1n} between $\widehat{\theta}$ and θ_0, the second term on the right hand side is equal to

$$\begin{aligned}
&-2n^{-1}\sum_{i=1}^n(\Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i))\left(\frac{\partial^2 m_{\beta_0}(X_i)}{\partial\beta\partial\beta^t}\right) - 2E\left[(\Lambda_{\theta_0}(Y) - m_{\beta_0}(X))\left(\frac{\partial^2 m_{\beta_0}(X)}{\partial\beta\partial\beta^t}\right)\right] + o_P(1)\\
&= -2n^{-1}\sum_{i=1}^n(\dot{\Lambda}_{\theta_{1n}}(Y_i))^t(\widehat{\theta} - \theta_0)\left(\frac{\partial^2 m_{\beta_0}(X_i)}{\partial\beta\partial\beta^t}\right) + o_P(1)\\
&= o_P(1) \ .
\end{aligned}$$

Note that to obtain the last equality, we used Proposition 7.1, and conditions (A2), (A5)(i) and (A6)(ii). Hence,

$$\frac{\partial^2 S_n(\beta_{1n})}{\partial\beta\partial\beta^t} = 2\Omega + o_P(1) \ . \tag{7.4}$$

Inserting (7.3) and (7.4) in (7.2), we obtain :

$$\begin{aligned}
\widehat{\beta} - \widetilde{\beta}_{0n} \ &= \ (2\Omega + o_P(1))^{-1}\left(2n^{-1}\sum_{i=1}^n(\Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i))\frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta}\right.\\
&\qquad \left. +2n^{-1}\sum_{i=1}^n(\Lambda_{\theta_0}(Y_i) - m_{\widetilde{\beta}_{0n}}(X_i))\frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta}\right) \ .
\end{aligned}$$

By a Taylor expansion and as the factor between large parentheses is equal to $O_P(n^{-1/2})$, we have

$$\begin{aligned}
\widehat{\beta} - \widetilde{\beta}_{0n} \ &= \ \Omega^{-1}n^{-1}\sum_{i=1}^n \frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta}(\dot{\Lambda}_{\theta_0}(Y_i))^t(\widehat{\theta} - \theta_0)\\
&\qquad +\Omega^{-1}n^{-1}\sum_{i=1}^n(\Lambda_{\theta_0}(Y_i) - m_{\widetilde{\beta}_{0n}}(X_i))\frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta} + o_P(n^{-1/2})\\
&= \ \Omega^{-1}n^{-1}\sum_{i=1}^n \frac{\partial m_{\widetilde{\beta}_{0n}}(X_i)}{\partial\beta}(\dot{\Lambda}_{\theta_0}(Y_i))^t(\widehat{\theta} - \theta_0) + n^{-1}\sum_{i=1}^n \eta_{\widetilde{\beta}_{0n}}(X_i, Y_i) + o_P(n^{-1/2}) \ .
\end{aligned}$$

Hence, using (2.1) and Proposition 7.1,

$$\widehat{\beta} - \widetilde{\beta}_{0n} = -\Omega^{-1}E\left[\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}(\dot{\Lambda}_{\theta_0}(Y))^t\right]n^{-1}\sum_{i=1}^n g(X_i, Y_i) + n^{-1}\sum_{i=1}^n \eta_{\widetilde{\beta}_{0n}}(X_i, Y_i) + o_P(n^{-1/2}) \ . \tag{7.5}$$

By a very similar reasoning, we obtain the following asymptotic expression for $\widetilde{\beta}_{0n} - \beta_0$:

$$\widetilde{\beta}_{0n} - \beta_0 = \Omega^{-1} n^{-1/2} \int r(x) \frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial \beta} \, dF_X(x) + o_P(n^{-1/2}) \ . \tag{7.6}$$

Finally, we get the result by combining (7.5) and (7.6). □

7.4 Results regarding the estimation of $m(\cdot)$

We state now some results concerning $\widehat{m}(x)$ that will be useful to establish the limiting distributions of the test statistics. The three following propositions are generalizations of Propositions 5.1-5.3 in Heuchenne, Samb and Van Keilegom (2014) when the dimension of the variable X is equal to $d \geq 1$:

Proposition 7.2. *Assume* $(A1) - (A9)$*. Then*

$$\sup_{x \in \mathcal{X}} |\widehat{m}(x) - m(x)| = O_P((nh^d)^{-1/2} (\log n)^{1/2}) + O(h^{p+1}) \ .$$

Proof. First, we have

$$\widehat{m}(x) - m(x) \ = \ \sum_{i=1}^n W_0^n \left(\frac{x - X_i}{h}, x \right) (\Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) + \widehat{m}(x, \theta_0) - m(x) \ .$$

By a Taylor expansion, for some value ξ between $\widehat{\theta}$ and θ_0, the absolute value of the first term on the right hand side is equal to

$$\left| \sum_{i=1}^n W_0^n \left(\frac{x - X_i}{h}, x \right) (\dot{\Lambda}_\xi(Y_i))^t (\widehat{\theta} - \theta_0) \right| = (|E(\dot{\Lambda}_\xi(Y)|X = x)| + o_P(1)) O_P(n^{-1/2})$$

and this is $O_P(n^{-1/2})$ because $\sup_{\theta,x} |E(\dot{\Lambda}_\theta(Y)|X = x)|$ is finite by condition (A5)(iii) and $\widehat{\theta} - \theta_0 = O_P(n^{-1/2})$ by Proposition 7.1. Concerning the second term on the right hand side, i.e. $\widehat{m}(x, \theta_0) - m(x)$, we see that the transformation is fixed, and hence we can use Theorem 6 in Masry (1996) with $j = 0$, which gives $\sup_{x \in \mathcal{X}} |\widehat{m}(x, \theta_0) - m(x)| = O((nh^d)^{-1/2} (\log n)^{1/2}) + O(h^{p+1})$. This concludes the proof. □

Proposition 7.3. *Assume* $(A1) - (A9)$*. Then for all vectors* $j = (j_1, \ldots, j_d)$ *such that* $j_\cdot = \sum_{i=1}^d j_i \leq d$*,*

$$\sup_{x \in \mathcal{X}} |D^j \widehat{m}(x) - D^j m(x)| = O_P((nh^{d+2j_\cdot})^{-1/2} (\log n)^{1/2}) + O(h^{p+1-j_\cdot}) \ .$$

where $D^j = \frac{\partial^{j_\cdot}}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}$.

Proposition 7.4. *Assume (A1) − (A9). Then,*

$$\sup_{x,x' \in \mathcal{X}} \frac{|D^d\widehat{m}(x) - D^d m(x) - D^d\widehat{m}(x') + D^d m(x')|}{||x - x'||^\alpha} = O_P((nh^{3d+2\alpha})^{-1/2}(\log n)^{1/2}) + O(h^{p+1-d-\alpha}) .$$

The proofs of Propositions 7.3 and 7.4 are similar to the one of Proposition 7.2 and are therefore omitted. For Proposition 7.4, we use Lemma A1 in Neumeyer and Van Keilegom (2010) in the proof. We also need the following result concerning $\widehat{m}(x) - m(x)$:

Proposition 7.5. *Assume (A1)-(A9). Then,*

$$\int (\widehat{m}(x) - m(x))\, dF_X(x) = n^{-1} \sum_{i=1}^n \Big\{ \Lambda_{\theta_0}(Y_i) - m(X_i) - E((\dot{\Lambda}_{\theta_0}(Y))^t) g(X_i, Y_i) \Big\} + o_P(n^{-1/2}) .$$

Proof. In order to keep the presentation simple, we restrict attention to the case where $d = 1$. The general case $d \geq 1$ is technically a bit harder, but follows the same idea. In the proof of Proposition 7.2, we have shown that

$$\widehat{m}(x) - m(x) = \sum_{i=1}^n W_0^n\Big(\frac{x - X_i}{h}, x\Big)(\dot{\Lambda}_\xi(Y_i))^t(\widehat{\theta} - \theta_0) + \sum_{i=1}^n W_0^n\Big(\frac{x - X_i}{h}, x\Big)(\Lambda_{\theta_0}(Y_i) - m(x)) ,$$

where ξ is some value between $\widehat{\theta}$ and θ_0. Using (7.1), we have that

$$\begin{aligned} &\int (\widehat{m}(x) - m(x))\, dF_X(x) \\ &= \sum_{i=1}^n \int \frac{1}{nh} \frac{1}{f_X(x)} k_1^*\Big(\frac{x - X_i}{h}\Big)(\Lambda_{\theta_0}(Y_i) - m(x))\, dF_X(x)(1 + o_P(1)) \\ &\quad + \sum_{i=1}^n \int \frac{1}{nh} \frac{1}{f_X(x)} k_1^*\Big(\frac{x - X_i}{h}\Big)(\dot{\Lambda}_\xi(Y_i))^t(\widehat{\theta} - \theta_0)\, dF_X(x)(1 + o_P(1)) . \end{aligned} \tag{7.7}$$

The first term on the right hand side of (7.7) is equal to

$$\Big[n^{-1} \sum_{i=1}^n \Lambda_{\theta_0}(Y_i) \int k_{1h}^*(x - X_i)\, dx - n^{-1} \sum_{i=1}^n \int k_{1h}^*(x - X_i) m(x)\, dx \Big](1 + o_P(1)) .$$

By a Taylor expansion of order $p + 1$, with a certain ξ_i between X_i and x, this last term is equal to :

$$\Big[n^{-1}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i)-m(X_i))\int k_{1h}^*(x-X_i)\,dx$$
$$-n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{p}\frac{1}{l!}m^{(l)}(X_i)\int k_{1h}^*(x-X_i)(x-X_i)^l\,dx$$
$$-n^{-1}\sum_{i=1}^{n}\frac{1}{(p+1)!}\int k_{1h}^*(x-X_i)(x-X_i)^{p+1}m^{(p+1)}(\xi_i)\,dx\Big](1+o_P(1))\ .$$

Using the properties of k_1^*, the first term on the right hand side of (7.7) is equal to :

$$n^{-1}\sum_{i=1}^{n}(\Lambda_{\theta_0}(Y_i)-m(X_i))+o_P(n^{-1/2})\ .$$

Next, since $\int k_1^*((x-X_i)/h)/h\,dx=1$, using Proposition 7.1, the second term on the right hand side of (7.7) is equal to

$$n^{-1}\sum_{i=1}^{n}(\dot\Lambda_{\theta_0}(Y_i))^t\Big(-n^{-1}\sum_{j=1}^{n}g(X_j,Y_j)\Big)+o_P(n^{-1/2})=-n^{-1}\sum_{j=1}^{n}E((\dot\Lambda_{\theta_0}(Y))^t)g(X_j,Y_j)+o_P(n^{-1/2})\ ,$$

because $-n^{-1}\sum_{j=1}^{n}g(X_j,Y_j)=O_P(n^{-1/2})$ by Proposition 7.1 and $n^{-1}\sum_{i=1}^{n}\dot\Lambda_{\theta_0}(Y_i)-E(\dot\Lambda_{\theta_0}(Y))=O_P(n^{-1/2})$ by the central limit theorem. This concludes the proof. □

7.5 Proofs of the main results

We start this subsection with the proof that H_0 is true if and only if $F_\varepsilon\equiv F_{\varepsilon_0}$.

Proof of Theorem 2.1. If H_0 is true, then $F_\varepsilon(y)=P(\Lambda_{\theta_0}(Y)-m(X)\le y)=P(\Lambda_{\theta_0}(Y)-m_{\beta_0}(X)\le y)=F_{\varepsilon_0}(y)$. Conversely, if $F_\varepsilon(y)=F_{\varepsilon_0}(y)$ for all y, then $E(\varepsilon)=E(\varepsilon_0)$ and $V(\varepsilon)=V(\varepsilon_0)$. Moreover, $E(\varepsilon_0)=E(m(X)-m_{\beta_0}(X))+E(\varepsilon)$ and thus

$$E(m(X)-m_{\beta_0}(X))=0\ . \qquad (7.8)$$

Similarly, as ε is independent of X, we have that $V(\varepsilon_0)=V(m(X)-m_{\beta_0}(X))+V(\varepsilon)$ and consequently

$$V(m(X)-m_{\beta_0}(X))=0 \qquad (7.9)$$

Finally, (7.8) and (7.9) imply that $m(X) - m_{\beta_0}(X) = 0$ a.s. As m is a continuous function, $m(X) = m_{\beta_0}(X)$. $\square$

We continue with two technical lemmas that are needed for the proof of Theorem 3.1.

Lemma 7.5. *Assume (A1)-(A9). Then,*

$$Var_n\left[1(\Lambda_{\widehat{\theta}}(Y) \leq t + \widehat{m}(X)) - 1(\Lambda_{\theta_0}(Y) \leq t + m(X))\right] = o_P(1) \ ,$$

where $Var_n(.)$ is the conditional variance given (X_i, Y_i), $i = 1, \dots, n$.

The proof of this lemma can be found in Proposition 5.4 in Heuchenne, Samb and Van Keilegom (2014). We also need the following technical result, the proof of which is based on different results in Van der Vaart and Wellner (1996) :

Lemma 7.6. *Assume (A1)-(A9). Then,*

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} \Big\{ & I(\Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}(X_i) \leq t) - I(\Lambda_{\theta_0}(Y_i) - m(X_i) \leq t) \\ & -P(\Lambda_{\widehat{\theta}}(Y) - \widehat{m}(X) \leq t) + P(\Lambda_{\theta_0}(Y) - m(X) \leq t) \Big\} = o_P(n^{-1/2}) \end{aligned}$$

uniformly in $t \in \mathbb{R}$, where $P(\Lambda_{\widehat{\theta}}(Y) - \widehat{m}(X) \leq t)$ is the probability with respect to the joint distribution of (X, Y), conditional on the value of $\widehat{\theta}$ and $\widehat{m}(\cdot)$.

Proof. The proof is a combination of the proof of Lemma 1 in Heuchenne, Samb and Van Keilegom (2014) and of the proof of Lemma A.3 in Neumeyer and Van Keilegom (2010). The latter two papers are based on the work of Akritas and Van Keilegom (2001), who studied the estimation of the error distribution in a nonparametric location-scale regression model, whereas Heuchenne, Samb and Van Keilegom (2014) considered a semiparametric transformation model with a one-dimensional covariate and Neumeyer and Van Keilegom (2010) considered a heteroscedastic regression model with a d-dimensional variable X. For this reason, we will restrict to explaining here the main steps of the proof. To start, note that $\Lambda_{\widehat{\theta}}(Y) - \widehat{m}(X) = \Lambda_{\widehat{\theta}}(Y) - m(X) - h_n(X)$ where $h_n(X) = \widehat{m}(X) - m(X)$. We define :

$$\begin{aligned} \mathcal{F}_1 \ = \ & \Big\{ (x,y) \longrightarrow 1(\Lambda_\theta(y) \leq t + m(x) + h(x)) \ , \ \Lambda_\theta : \mathbb{R} \to \mathbb{R} \ \text{ strictly increasing} \ , \\ & \theta \in \Theta \subset \mathbb{R}^k, \ t \in \mathbb{R}, \ h \in C_1^{d+\alpha}(\chi) \Big\} \ . \end{aligned}$$

By Propositions 7.2, 7.3 and 7.4 we have that $P(h_n \in C_1^{d+\alpha}(\chi)) \to 1$ as $n \to \infty$, which justifies the fact that we let h be an element of $C_1^{d+\alpha}(\chi)$. First, we want to prove that the class $\mathcal{F}_1$ is Donsker. Using Theorem 2.5.6 in Van der Vaart and Wellner (1996), it suffices to prove that

$$\int_0^\infty \sqrt{\log N_{[]}(\widetilde{\varepsilon}, \mathcal{F}_1, L_2(P))}\, d\widetilde{\varepsilon} < \infty \ ,$$

where $N_{[]}(\widetilde{\varepsilon}, \mathcal{F}_1, L_2(P))$ is the $\widetilde{\varepsilon}$-bracketing number of the class $\mathcal{F}_1$, P is the probability measure corresponding to the joint distribution of (X, Y) and $L_2(P)$ is the L_2-norm. Corollary 2.7.2 in Van der Vaart and Wellner (1996) implies that $N_{[]}(\widetilde{\varepsilon}^2, C_1^{d+\alpha}(\chi), L_2(P)) \leq \exp(C\widetilde{\varepsilon}^{-2d/(d+\alpha)})$, where C is a certain positive constant. Using the same arguments as in Lemma 1 in Heuchenne, Samb and Van Keilegom (2014), it follows that the $\widetilde{\varepsilon}-$bracketing number of the class $\mathcal{F}_1$ is at most $O(\widetilde{\varepsilon}^{-2(k+1)} \exp(C\widetilde{\varepsilon}^{-2d/(d+\alpha)}))$ and hence the class $\mathcal{F}_1$ is Donsker. Next, using Theorem 2.5.6 and Theorem 2.10.6 in Van der Vaart and Wellner (1996), we deduce that the class

$$\begin{aligned}
\mathcal{F} \ = \ \Big\{ & (x, y) \longrightarrow 1(\Lambda_\theta(y) \leq t + m(x) + h(x)) - 1(\Lambda_{\theta_0}(y) \leq t + m(x)) \\
& - P(\Lambda_\theta(Y) \leq t + m(X) + h(X)) + P(\Lambda_{\theta_0}(Y) \leq t + m(X)) \ , \\
& \Lambda_\theta : \mathbb{R} \to \mathbb{R} \ \text{ strictly increasing}, \ \theta \in \Theta \subset \mathbb{R}^k, \ t \in \mathbb{R}, \ h \in C_1^{d+\alpha}(\chi) \Big\}
\end{aligned}$$

is also Donsker. Finally, Corollary 2.3.12 in Van der Vaart and Wellner (1996) implies for each $\delta > 0$ that

$$\lim_{\gamma \searrow 0} \limsup_{n \to \infty} P\Bigg(\sup_{f \in \mathcal{F}, Var(f) < \gamma} n^{-1/2} \Bigg| \sum_{i=1}^n f(X_i) \Bigg| > \delta \Bigg) = 0 \ .$$

By restricting the supremum inside this probability to the elements in $\mathcal{F}$ corresponding to $h \equiv h_n$ as defined above, the result follows. □

We are now ready to give the proofs of the main asymptotic results of the paper.

Proof of Theorem 3.1. The proof will be divided in two parts : first we will develop an asymptotic representation for $\widehat{F}_\varepsilon(y) - F_\varepsilon(y)$ and second we will obtain an asymptotic representation for $F_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y)$. Using the definition of the semiparametric transformation model, we have that :

$$\begin{aligned} P(\widehat{\varepsilon} \le y) &= P(\Lambda_{\widehat{\theta}}(Y) \le y + \widehat{m}(X, \widehat{\theta})) \\ &= P(\varepsilon(\widehat{\theta}) \le y + \widehat{m}(X, \widehat{\theta}) - m(X, \widehat{\theta})) \\ &= \int F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x)\, dF_X(x) \ . \end{aligned}$$

Next, using Lemma 7.6 and the last expression above, we have that

$$\begin{aligned} &\widehat{F}_\varepsilon(y) - F_\varepsilon(y) \\ &= n^{-1}\sum_{i=1}^n I(\varepsilon_i \le y) - F_\varepsilon(y) + \int \Big[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) - F_\varepsilon(y) \Big]\, dF_X(x) + R_{n1}(y), \end{aligned}$$

where $\sup_y |R_{n1}(y)| = o_P(n^{-1/2})$. By conditions (A2), (A4), (A5) and Propositions 7.1 and 7.2, we obtain from a Taylor expansion the following asymptotic representation for $\widehat{F}_\varepsilon(y) - F_\varepsilon(y)$:

$$\begin{aligned} &\widehat{F}_\varepsilon(y) - F_\varepsilon(y) \\ &= n^{-1}\sum_{i=1}^n I(\varepsilon_i \le y) - F_\varepsilon(y) + f_\varepsilon(y) \int (\widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta}))\, dF_X(x) \qquad (7.10) \\ &\quad + \int \Big[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) - F_\varepsilon(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \Big]\, dF_X(x) + R_{n1}(y) \ . \end{aligned}$$

Next, we need a similar representation for $\widehat{F}_{\varepsilon_0}(y) - F_\varepsilon(y)$. We can easily prove the equivalence of Propositions 7.2 to 7.5 when the function $\widehat{m}(x)$ is replaced by $\widehat{m}_{\widehat{\beta}}(x)$ (using condition (A6)). Then, Lemma 7.6 can be adapted when $\widehat{m}(x)$ is replaced by $\widehat{m}_{\widehat{\beta}}(x)$. Consequently, in exactly the same way as the above calculations for $\widehat{F}_\varepsilon(y) - F_\varepsilon(y)$, we obtain the following asymptotic representation :

$$\begin{aligned} &\widehat{F}_{\varepsilon_0}(y) - F_\varepsilon(y) \\ &= n^{-1}\sum_{i=1}^n I(\varepsilon_i \le y) - F_\varepsilon(y) + f_\varepsilon(y) \int (\widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta}))\, dF_X(x) \qquad (7.11) \\ &\quad + \int \Big[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) - F_\varepsilon(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \Big]\, dF_X(x) + R_{n2}(y) \ , \end{aligned}$$

where $\sup_y |R_{n2}(y)| = o_P(n^{-1/2})$. Next, by a Taylor expansion with some value ξ between $\widehat{\theta}$

and θ_0, we obtain :

$$\int \Big[F_{\varepsilon(\widehat{\theta})}(y+\widehat{m}(x,\widehat{\theta})-m(x,\widehat{\theta})|x) - F_{\varepsilon}(y+\widehat{m}(x,\widehat{\theta})-m(x,\widehat{\theta})|x)$$

$$-F_{\varepsilon(\widehat{\theta})}(y+\widehat{m}_{\widehat{\beta}}(x,\widehat{\theta})-m(x,\widehat{\theta})|x) + F_{\varepsilon}(y+\widehat{m}_{\widehat{\beta}}(x,\widehat{\theta})-m(x,\widehat{\theta})|x)\Big] \, dF_X(x)$$

$$= \int \Big[\Big(\dot{F}_{\varepsilon(\theta)}(z|x)\Big|_{\substack{\theta=\xi \\ z=y+\widehat{m}(x,\widehat{\theta})-m(x,\widehat{\theta})}} \Big)^t - \Big(\dot{F}_{\varepsilon(\theta)}(z|x)\Big|_{\substack{\theta=\xi \\ z=y+\widehat{m}_{\widehat{\beta}}(x,\widehat{\theta})-m(x,\widehat{\theta})}} \Big)^t \Big] (\widehat{\theta}-\theta_0) \, dF_X(x).$$

By another Taylor expansion, this last expression is equal to

$$\int \Big[\Big(\frac{\partial}{\partial z} \dot{F}_{\varepsilon(\theta)}(z|x)\Big|_{\substack{\theta=\xi \\ z=y-m(x,\widehat{\theta})+\eta}} \Big)^t (\widehat{\theta}-\theta_0)(\widehat{m}(x,\widehat{\theta})-\widehat{m}_{\widehat{\beta}}(x,\widehat{\theta})) \Big] \, dF_X(x) \qquad (7.12)$$

with η some value between $\widehat{m}(x,\widehat{\theta})$ and $\widehat{m}_{\widehat{\beta}}(x,\widehat{\theta})$. Consequently, using assumption (A4)(ii), Lemma 7.2 and Propositions 7.1 and 7.2, we conclude that expression (7.12) is $o_P(n^{-1/2})$. This shows that

$$\widehat{F}_{\varepsilon}(y) - \widehat{F}_{\varepsilon_0}(y) = f_{\varepsilon}(y) \int [\widehat{m}(x) - \widehat{m}_{\widehat{\beta}}(x)] \, dF_X(x) + R_n(y) \ ,$$

where $\sup_y |R_n(y)| = o_P(n^{-1/2})$.

Next, in a very similar way as in the proof of Proposition 7.5, we can prove that :

$$\int [\widehat{m}_{\widehat{\beta}}(x) - m(x)] \, dF_X(x) = n^{-1} \sum_{i=1}^n (m_{\widehat{\beta}}(X_i) - m(X_i)) + o_P(n^{-1/2}) \ .$$

As we are under H_0, $m(X_i) = m_{\beta_0}(X_i)$, and using a Taylor expansion and Lemma 7.2, we obtain :

$$f_{\varepsilon}(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)] \, dF_X(x) = f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n \Big(\frac{\partial m_{\beta_0}(X_i)}{\partial \beta} \Big)^t (\widehat{\beta} - \beta_0) + o_P(n^{-1/2}) \ . \quad (7.13)$$

Now, using Lemma 7.4 and Proposition 7.5, we have that

$$\begin{aligned} \widehat{F}_{\varepsilon}(y) - \widehat{F}_{\varepsilon_0}(y) &= f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n \Big(\Lambda_{\theta_0}(Y_i) - m(X_i) - E((\dot{\Lambda}_{\theta_0}(Y))^t) g(X_i, Y_i) \\ &\quad - \int \Big(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \Big)^t dF_X(x) \eta_{\beta_0}(X_i, Y_i) \\ &\quad + \int \Big(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \Big)^t dF_X(x) \Omega^{-1} E\Big[\frac{\partial m_{\beta_0}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \Big] g(X_i, Y_i) \Big) + R_n(y) \ , \end{aligned}$$

where $\sup_y |R_n(y)| = o_P(n^{-1/2})$.

We will now prove (ii). If we look at (i), we see that the factor that multiplies $f_\varepsilon(y)$ is a sum of *iid* terms and consequently, by the central limit theorem, the process $n^{1/2}(\widehat{F}_\varepsilon(\cdot) - \widehat{F}_{\varepsilon_0}(\cdot))$ $(-\infty < y < +\infty)$ converges to $f_\varepsilon(\cdot)W$ where W is a zero-mean normal random variable. That $E(W) = 0$ follows from the fact that $E(\varepsilon) = 0$, $E(\eta_{\beta_0}(X,Y)) = 0$ and $E(g(X,Y)) = 0$. The latter property holds since $g(X,Y) = \Gamma^{-1}\xi(\theta_0, X, Y)$ and $\xi(\theta_0, X, Y)$ is the derivative of the likelihood. This concludes the proof. □

Proof of Corollary 3.1. The proof follows along the same lines as the proof of Corollary 3.2 in Van Keilegom, González-Manteiga and Sánchez-Sellero (2008). We refer to their proof for more details. □

Proof of Theorem 3.2. In a very similar way as in the proof of Theorem 3.1(i), but now under H_{1n}, we can show that :

$$\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y) = f_\varepsilon(y) \int [\widehat{m}(x) - m(x)]\, dF_X(x) - f_\varepsilon(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)]\, dF_X(x) + R_n(y)$$

and

$$f_\varepsilon(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)]\, dF_X(x) = f_\varepsilon(y) n^{-1} \sum_{i=1}^n (m_{\widehat{\beta}}(X_i) - m(X_i)) + o_P(n^{-1/2}) \ . \qquad (7.14)$$

Next, under $H_{1n} : m(x) = m_{\beta_0}(x) + n^{-1/2} r(x)$, expression (7.14) becomes :

$$\begin{aligned}
&f_\varepsilon(y) n^{-1} \sum_{i=1}^n (m_{\widehat{\beta}}(X_i) - m_{\beta_0}(X_i)) - n^{-1/2} f_\varepsilon(y) \int r(x)\, dF_X(x) + o_P(n^{-1/2}) \\
&= f_\varepsilon(y) \int \Big(\frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial \beta} \Big)^t dF_X(x) \Big(- \Omega^{-1} n^{-1} \sum_{i=1}^n E\Big[\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \Big] g(X_i, Y_i) \\
&\quad + n^{-1} \sum_{i=1}^n \eta_{\widetilde{\beta}_{0n}}(X_i, Y_i) + \Omega^{-1} n^{-1/2} \int r(x) \frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial \beta}\, dF_X(x) \Big) \\
&\quad - n^{-1/2} f_\varepsilon(y) \int r(x)\, dF_X(x) + o_P(n^{-1/2}) \ , \qquad (7.15)
\end{aligned}$$

using Lemma 7.2 – 7.4 and a Taylor expansion. Using Proposition 7.5 and (7.15), this gives

$$\widehat{F}_\varepsilon(y) - \widehat{F}_{\varepsilon_0}(y) = f_\varepsilon(y) n^{-1} \sum_{i=1}^n H(X_i, Y_i, \theta_0, \widetilde{\beta}_{0n}) + n^{-1/2} f_\varepsilon(y) b + R_n(y) \ , \qquad (7.16)$$

where $\sup_y |R_n(y)| = o_P(n^{-1/2})$.

We will now prove (ii). It follows from (7.16) that if we can prove that the limiting distribution under H_{1n} of the expression

$$n^{-1/2}\sum_{i=1}^{n} H(X_i, Y_i, \theta_0, \widetilde{\beta}_{0n}) \tag{7.17}$$

is the same as the limiting distribution under H_0 of the expression

$$n^{-1/2}\sum_{i=1}^{n} H(X_i, Y_i, \theta_0, \beta_0)\ , \tag{7.18}$$

then we can conclude that the process $n^{1/2}(\widehat{F}_\varepsilon(\cdot) - \widehat{F}_{\varepsilon_0}(\cdot))$ converges weakly to $f_\varepsilon(\cdot)(W+b)$, where W is the same normal distribution as in Theorem 3.1(ii). First, we clearly see that these two expressions are sums of *iid* terms and hence they converge to a normal distribution. We have to prove that the means and variances of (7.17) and (7.18) are asymptotically the same. We proved in Theorem (3.1)(ii) that the mean of (7.18) under H_0 is equal to 0. Moreover, as $\widetilde{\beta}_{0n}$ is by definition a minimizer over β of $E[(m(X)-m_\beta(X))^2]$, we have that $E[(m(X)-m_{\widetilde{\beta}_{0n}}(X))\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}] = 0$. Consequently, under H_{1n} :

$$E[\eta_{\widetilde{\beta}_{0n}}(X,Y)] = E[E[\eta_{\widetilde{\beta}_{0n}}(X,Y)|X]] = \Omega^{-1}\ E\left[\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}(m(X) - m_{\widetilde{\beta}_{0n}}(X))\right] = 0\ .$$

Hence, under H_{1n}, the mean of (7.17) is also equal to 0. Next, under H_{1n}, the variance of $\eta_{\widetilde{\beta}_{0n}}(X,Y)$ is equal to :

$$\begin{aligned}
&\Omega^{-1}E\left[\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)(\Lambda_{\theta_0}(Y) - m_{\widetilde{\beta}_{0n}}(X))^2\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)^t\right]\Omega^{-1}\\
&= \Omega^{-1}E\left[\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)\left\{(\Lambda_{\theta_0}(Y) - m(X))^2 + (m(X) - m_{\widetilde{\beta}_{0n}}(X))^2\right\}\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)^t\right]\Omega^{-1}\ ,
\end{aligned}$$

since ε is independent of X and $E(\varepsilon) = 0$. Then, using the fact that $m(X) = m_{\beta_0}(X) + n^{-1/2}r(X)$ under H_{1n}, a Taylor expansion and Lemma 7.3, the last expression on the right hand side is equal to :

$$\begin{aligned}
V[\eta_{\widetilde{\beta}_{0n}}(X,Y)] &= \Omega^{-1}E\Bigg[\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)\Big\{(\Lambda_{\theta_0}(Y) - m(X))^2\\
&\quad + (m_{\beta_0}(X) - m_{\widetilde{\beta}_{0n}}(X) + n^{-1/2}r(X))^2\Big\}\left(\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial\beta}\right)^t\Bigg]\Omega^{-1}\\
&= \Omega^{-1}E\left[\left(\frac{\partial m_{\beta_0}(X)}{\partial\beta}\right)(\Lambda_{\theta_0}(Y) - m(X))^2\left(\frac{\partial m_{\beta_0}(X)}{\partial\beta}\right)^t\right]\Omega^{-1} + o(1)\ .
\end{aligned}$$

This last expression is equal to $V[\eta_{\beta_0}(X,Y)]+o(1)$ under H_0 because under the null hypothesis $m(X)=m_{\beta_0}(X)$. □

Proof of Corollary 3.2. The proof follows from Theorem 3.2 and using similar techniques as in the proof of Corollary 3.1. □

References

Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **28**, 549-568.

Bellver, C. (1987). Influence of particulate pollution on the positions of neutral points in the sky in Seville (Spain). *Atmospheric Environment*, **21**, 699-702

Bickel, P.J. and Doksum, K. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, **76**, 296-311.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society - Series B*, **26**, 211-252.

Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New-York.

Cleveland, W.S. (1993). *Visualizing Data.* Hobart Press, Summit.

Colling, B., Heuchenne, C., Samb, R. and Van Keilegom, I. (2013). Estimation of the error density in a semiparametric transformation model. *Annals of the Institute of Statistical Mathematics* (under revision).

Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society - Series B*, **69**, 903-917.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, London.

González-Manteiga, W. and Crujeiras, R. (2013). An updated review of goodness-of-fit tests for regression models (with discussion). *TEST*, **22**, 361-447.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926-1947.

Hart, J.D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests.* Springer, New-York.

Heuchenne, C., Samb, R. and Van Keilegom, I. (2014). Estimating the residual distribution in semiparametric transformation models (submitted).

Horowitz, J.L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103-137.

Horowitz, J.L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, **69**, 499-513.

Jacho-Chavez, D., Lewbel, A. and Linton, O. (2008). Identification and nonparametric estimation of a transformation additively separable model. Technical report.

Jennrich, R.I. (1969). Asymptotic properties of non linear least squares estimators. *Annals of Mathematical Statistics*, **40**, 633-643.

Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Annals of Statistics*, **36**, 686-718.

Masry, E. (1996). Multivariate local polynomial regression for time series : uniform strong consistency and rates. *Journal of Time Series Analysis*, **17**, 571-599.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics*, **36**, 204-228.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101**, 1067-1078.

Pardo-Fernández, J.C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, **17**, 1115-1137.

Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **57**, 1403-1430.

Sakia, R.M. (1992). The Box-Cox transformation technique : a review. *The Statistician*, **41**, 169-178.

Silverman, B.W. and Young, G.A. (1987). The bootstrap : to smooth or not to smooth. *Biometrika*, **74**, 469-479.

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

Van der Vaart, A.W. (1998). *Asymptotics Statistics.* Cambridge University Press, Cambridge.

Van der Vaart, A.W. and Wellner, J.A (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

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.

Vanhems, A. and Van Keilegom, I. (2013). Semiparametric transformation model with endogeneity : a control function approach (submitted).

Zellner, A. and Revankar, N.S. (1969). Generalized production functions. *Review of Economic Studies*, **36**, 241-250.

Zhang, C.M. (2003). Adaptative tests of regression functions via multiscale generalized likelihood ratio. *Canadian Journal of Statistics*, **31**, 151-171.