

Estimation and goodness-of-fit tests in semi- and nonparametric transformation models

Benjamin Colling

A thesis submitted to the
Université catholique de Louvain
in partial fulfillment of the requirements for the degree of

DOCTOR OF SCIENCES

Thesis committee:

Ingrid Van Keilegom, Supervisor (Katholieke Universiteit Leuven, Université catholique de Louvain)

Anouar El Ghouch (Université catholique de Louvain)

Cédric Heuchenne (Université de Liège, Université catholique de Louvain)

Natalie Neumeyer (Universität Hamburg)

Olivier Thas (Universiteit Gent)

September 2017

Acknowledgements

First of all, I would like to thank wholeheartedly my supervisor Ingrid Van Keilegom for her availability, her confidence in me, all of her advice and the knowledge that she passed on to me during these six years. Ingrid, you have all the qualities that one would hope for in a supervisor.

I would also like to express my gratitude to my other jury members: Anouar El Ghouch, Cédric Heuchenne, Natalie Neumeyer and Olivier Thas, for their constructive remarks and suggestions that have considerably improved the final version of this thesis. In particular, I would like to thank Anouar, because it was such a pleasure for me to collaborate with you on teaching duties.

Next, I'm grateful to Nadja, Nancy, Sophie and Tatiana for their help on administrative issues and of course for providing cookies every day! In particular, a big thanks goes to Nancy for her availability and her replies to my numerous practical questions concerning my PhD - without forgetting all of the candy I ate in her office!

My thanks also goes to the IT team and to Laurent and Raphaël in particular who have always been there in case of computer problems. I also have a thought for Jean-Pierre, the big supporter of Standard, who had to support four titles of Anderlecht and all my jokes that followed during the six years I have spent at the institute. I also thank Christophe and Pierre for the good daily atmosphere in our hallway.

I would like to thank the entire SMCS team for offering me the opportunity to work on a variety of consultancy jobs. Alain, Catherine, Céline and Nathalie were always there with advice and help, and their experiences taught me a lot. In particular, thanks to Nathalie for organizing a very nice yearly SMCS dinner at her home and for making me discover a cocktail with minced meat, and thanks to Alain for his precious help in deblocking the computers in the researcher's room over and over again, even if one day you took revenge by killing my simulations!

My last six years were excellent because of many colleagues that made the institute a pleasant working place. Thanks to Alain, Alexandre, Anna, Anne, Antoine, Aurélie, Baptiste, Catherine, Cédric, Céline, Cheikh, Florian, Gauthier, Hélène, Hugues, Joris, Laetitia (even if she is not part of the institute), Lieven, Mailis, Majda, Manon, Marco, Mathieu, Mickaël, Nathalie (the old),

Nathalie (the young), Nathan, Nicolas, Oswaldo, Rebecca, Samuel, Sylvie and Vincent for the 10h45 cookie breaks, the lunches, the teaching activities, the conferences (or rather the afternoons and evenings spent at the swimming pool or in the sea and the football matches that we watched from a terrace while drinking rosé, the persons concerned will recognize themselves) and the Friday night drinks.

In particular, a big thanks goes to those who participated in my football prediction leagues or my musical quizzes, to Hélène and Maïlis for being very nice office neighbours and for tolerating my noise, my music and worse, my singing, to Anna, Aurélie, Hélène, Marco, Nathalie and Sylvie for the Davis Cup weekends, to Anouar, Antoine, Cédric, Cheikh, Hélène, Manon and Nathalie for the many badminton games (even if I'm afraid that you dislike me after all of my drop shots), to Alain, Florian, Marco, Nicolas and Vincent for the never-ending sports discussions and the football/beer/Playstation/Mario Kart evenings and to Anna, Aurélie, Marco, Sylvie and Vincent for the weekends at the Belgian coast, the time spent at the water parcs and the summer holidays (thanks to you, I know for sure that I will not play a beach racket game again and that I don't like mokka ice cream)!

I would also like to thank Caroline, Claude, Florence, Gaëtan, Kristel, Laura and Michaël, my great longtime friends on whom I can always count - I wonder how you still put up with me after all this time!

Finally, I would like to thank my parents, my sister, my brother-in-law, my grandparents, my godmother, my godfather, my aunts and my uncles who have always encouraged me and believed in me, without forgetting my adorable goddaughter Lucie who always brings a smile to my face.

I would also like to thank Anna for her support and her love. She always encouraged and helped me, even though I haven't been easy these last months. It was very important for me to have you with me. Thank you also for having translated these acknowledgements in English, it was not an easy task!

Table of contents

1	Transformation models and goodness-of-fit tests	9
1.1	Introduction	9
1.2	Semiparametric transformation models	15
1.2.1	Definition and notations	15
1.2.2	Identification	16
1.2.3	Estimation of the model	19
1.2.4	Estimation of $f_\varepsilon(\cdot)$, $F_\varepsilon(\cdot)$ and related contributions . . .	22
1.3	Nonparametric transformation models	24
1.3.1	Definition, notations and identification	24
1.3.2	Estimation of the model	26
1.4	Goodness-of-fit tests in regression models	28
1.4.1	Introduction	28
1.4.2	Using the error distribution function	29
1.4.3	Using the integrated regression function	32
1.4.4	Other ways to construct a test	35
1.4.4.1	Using the regression function	36
1.4.4.2	Using the characteristic function	37
1.4.4.3	Empirical likelihood ratio tests	38
1.5	Contents of the thesis	42
2	Goodness-of-fit tests in semiparametric transformation models using the error distribution function	45
2.1	Introduction	45
2.2	The proposed test	46
2.2.1	Estimation of the model	46
2.2.2	Test statistics	47
2.3	Asymptotic results	49
2.3.1	Under H_0	49
2.3.2	Under H_{1n}	50
2.3.3	Bootstrap	52
2.4	Simulations	53
2.4.1	Estimation of θ , h , g and β in practice	53

2.4.2	Performance of $\hat{f}_{\hat{\varepsilon}}(y)$	54
2.4.3	Performance of $\hat{F}_{\hat{\varepsilon}}(y)$	60
2.4.4	Performance of the test in dimension 1	68
2.4.5	Performance of the test in dimension 2	71
2.5	Application on sunlight data	73
2.5.1	Previous analyses of the data	73
2.5.2	Results	73
2.6	Conclusions and recommendations	74
2.7	Technical assumptions	75
2.8	Proofs	76
2.8.1	Results regarding the estimation of θ	76
2.8.2	Results regarding the estimation of β	79
2.8.3	Results regarding the estimation of $m(\cdot)$	83
2.8.4	Proofs of the main results	87
3	Goodness-of-fit tests in semiparametric transformation models using the integrated regression function	95
3.1	Introduction	95
3.2	The proposed test	96
3.3	Asymptotic results	102
3.3.1	Under H_0	102
3.3.2	Under H_{1n}	103
3.4	Simulations	105
3.4.1	Comparison of both tests in dimension 1	105
3.4.2	Comparison of both tests in dimension 2	110
3.5	Application on ultrasonic data	112
3.5.1	Previous analyses of the data	112
3.5.2	Results	112
3.6	Conclusions and recommendations	113
3.7	Technical assumptions	114
3.8	Proofs	116
3.8.1	Proofs regarding Conditions (B1) and (B2)	116
3.8.2	Proofs of the main results under H_0	121
3.8.3	Proofs of the main results under H_{1n}	128
4	Estimation of fully nonparametric transformation models	133
4.1	Introduction and motivation	133
4.2	Main idea of the estimation procedure	134
4.3	Identification of $\Gamma(\cdot)$	135
4.4	Definition of the new estimators	136
4.5	Asymptotic results	138
4.6	Simulations	140
4.7	Conclusions and recommendations	148
4.8	Assumptions	149

4.9	Proofs	151
4.9.1	Useful technical results	151
4.9.2	Proofs of the main results	180
5	Discussion	195
5.1	General conclusions	195
5.2	Future research	197
5.2.1	Related extensions	197
5.2.2	Transformation models in other contexts	199

Chapter 1

Transformation models and goodness-of-fit tests

1.1 Introduction

The simple linear regression model is the most commonly used model in statistics when we want to explain the relationship between a dependent variable Y and a vector of explanatory variables denoted X . This model is commonly defined by $Y = X^t\beta + \varepsilon$, where β is a vector of unknown regression parameters, ε is the error term and X^t denotes the transpose of X . However, this model relies on heavy assumptions that are not always satisfied in practice, namely that the structure of this model is additive and linear, the variance of the error term ε is constant and ε is normally distributed. Unfortunately, in practice, the residuals may contain outliers and are often heteroscedastic and/or skewed as we will see in the example below. The violation of one or several of these assumptions could lead to an inefficient estimation of the parameter β and also to a wrong prediction of the response Y .

The book written by Carroll and Ruppert (1988) proposes different analyzes of regression data, based on a lot of examples, when some of these assumptions are violated. An appropriate example of these issues is the relationship between the size of the annual spawning stock of fish, denoted by X , and the production of new catchable fish that can eventually be recruited into the fishery, denoted by Y . We can find in Ricker and Smith (1975) the number of spawners and recruits from 1940 until 1967 for the Skeena River sockeye salmon stock. Note that we will omit the data obtained in 1951 since a rockslide occurred and reduced considerably the number of recruits.

Figure 1.1 shows a scatterplot of recruits and spawners where units are thousands of fish. First, we can see that a linear relationship between spawners and recruits would not be an inappropriate model for some purposes, since the number of recruits tends to increase when the number of spawners increases.

However, applying a linear fit suggests that recruitment could increase indefinitely by allowing more spawners, which is biologically impossible. Hence, in the literature, there are several other models more appropriate to analyze the relation between recruits and spawners. For instance, Ricker (1954) derived the theoretical deterministic model

$$Y = \beta_1 X \exp(-\beta_2 X) \quad \text{where } \beta_2 \geq 0 ,$$

and Beverton and Holt (1957) derived the following alternative model:

$$Y = \frac{1}{\beta_1 + \frac{\beta_2}{X}} \quad \text{where } \beta_1 \geq 0 \text{ and } \beta_2 \geq 0 .$$

In the model of Ricker (1954), Y tends to 0 as X goes to 0, which is expected of a realistic model, and Y also tends to 0 when X tends to ∞ since β_2 is positive. The biological interpretation is the following: if the number of young fish increases, then it also increases competition and predation and the survival rate is affected so that the number of fish reaching maturity decreases. In the model of Beverton and Holt (1957), Y also tends to 0 as X goes to 0 and alternatively to the model of Ricker (1954), Y tends to $1/\beta_1$ when X goes to ∞ . This means that $1/\beta_1$ is the maximum number of recruits that the available space, food and other resources can support.

The grey lines in Figure 1.1 represent an estimation of the model of Ricker (1954), see the graph on the left, and of the model of Beverton and Holt (1957), see the graph on the right, constructed with the least squares method. We obtained $\hat{\beta}_1 = 3.95$ and $\hat{\beta}_2 = 0.0008$ for the model of Ricker (1954), $\hat{\beta}_1 = 0.0004$ and $\hat{\beta}_2 = 0.22$ for the model of Beverton and Holt (1957) and we can observe in Figure 1.1 that these two adjustments are very similar.

However, these two fits are nonlinear fits and the two graphs in Figure 1.1 suggest clearly that the variance of the residuals are not constant and increase in absolute value when the number of spawners increases for both models. The heteroscedasticity of the error term is confirmed by the two graphs on the left in Figures 1.2 and 1.3 that show the residuals against the predicted values obtained with estimations of the models of Ricker (1954) and Beverton and Holt (1957) respectively. Moreover, the two graphs on the right in Figures 1.2 and 1.3 represent the quantiles of the normal distribution against the quantiles of the residuals for the models of Ricker (1954) and Beverton and Holt (1957) respectively, and are commonly called normal probability plots. Since these plots are roughly linear, there is no reason to conclude that the residuals are not normally distributed in both cases even if some left-skewness appears to be present.

In conclusion, using these data, some of the requirements of the linear regression model are not respected. One way to linearize the model in the parameters and to stabilize the variance is to transform the data. The Ricker (1954) model can be linearized by a logarithmic transformation:

$$\log(Y/X) = \alpha_1 + \alpha_2 X,$$

where $\alpha_1 = \log \beta_1$ and $\alpha_2 = -\beta_2$ and the Beverton and Holt (1957) model can be linearized by an inverse transformation:

$$\frac{1}{Y} = \beta_1 + \beta_2 \frac{1}{X}$$

Hence, by obtaining the additivity and the linearity of the effects in the parameters, we could apply the classical estimation procedures of a linear regression model but with a transformed response and sometimes also with a transformed X , like in the Beverton and Holt (1957) model for example. These two transformed models were adjusted with the classical least squares procedure of a linear regression model and we have redone Figures 1.2 and 1.3 but with the transformed data.

We can see in the two graphs on the left in Figures 1.4 and 1.5 that the variance of the residuals seems stabilized in the Ricker (1954) and Beverton and Holt (1957) models after a logarithm and an inverse transformation respectively. The graph on the right in Figure 1.4 indicates that the residuals in the Ricker (1954) model seem slightly better normally approximated after a logarithmic transformation than without transformation even if the slight left-skewness has clearly not disappeared. However, in the Beverton and Holt (1957) model, we can see in the graph on the right in Figure 1.5 that the residuals are worse normally approximated after an inverse transformation. This suggests that we have probably not applied the most appropriate transformation in that case and that the choice of a suitable transformation is an important problem.

Hence, other transformations of Y and/or X could be used to induce constant variance, approximate normality and reduced skewness, and maybe in a better way than the logarithm and inverse transforms proposed in Figures 1.4 and 1.5. However, in this thesis, we will consider the case where only the response Y is transformed. More generally, Box and Cox (1964) introduced the following parametric family of power transformations:

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

and suggested that this power transformation, when it is applied to the response variable Y , might induce additivity of the effects, homoscedasticity and normality of the new error term and reduce skewness with an optimal choice of the parameter θ . Note that the Box and Cox (1964) transformation also includes as special cases the logarithm, the square root, the inverse and the identity.

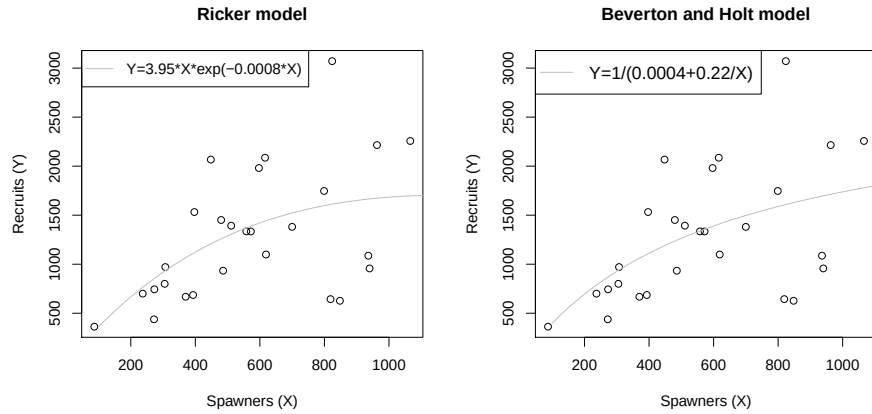


Figure 1.1: Scatterplot of recruits and spawners for the Skeena River sockeye salmon stock where units are thousands of fish. The grey lines represent estimations of the models of Ricker (1954) (on the left) and Beverton and Holt (1957) (on the right) based on these data.

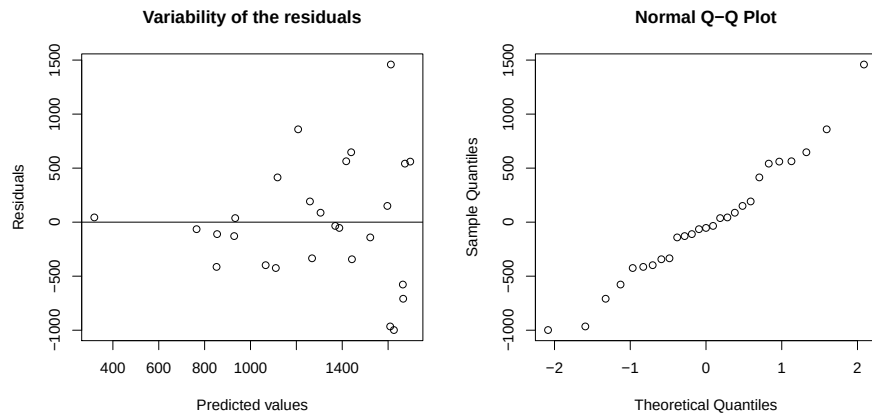


Figure 1.2: On the left: scatterplot of residuals and predicted values for the Skeena River sockeye salmon stock where units are thousands of fish and on the right: normal probability plot of the residuals, both obtained with the Ricker (1954) model.

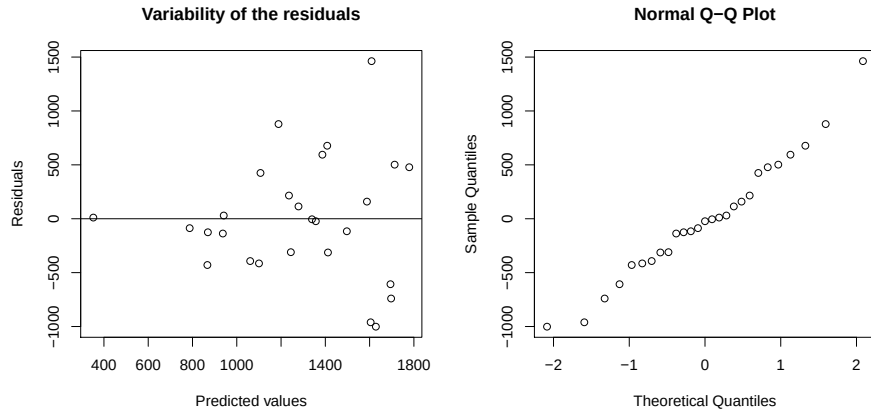


Figure 1.3: On the left: scatterplot of residuals and predicted values for the Skeena River sockeye salmon stock where units are thousands of fish and on the right: normal probability plot of the residuals, both obtained with the Beverton and Holt (1957) model.

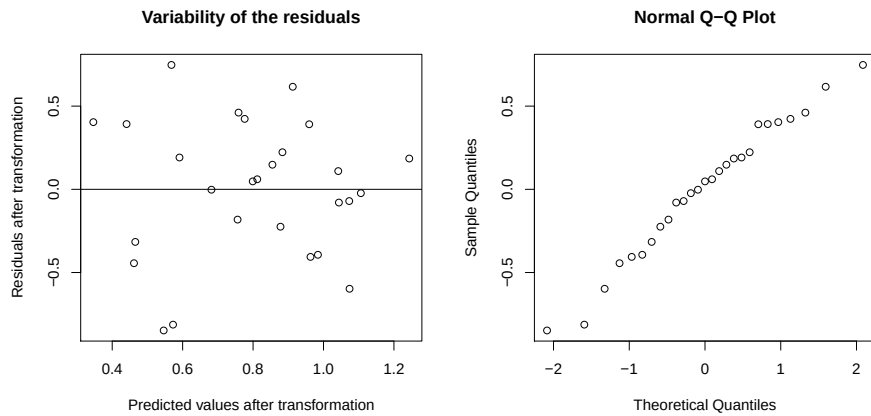


Figure 1.4: On the left: scatterplot of new residuals and predicted values for the Skeena River sockeye salmon stock where units are thousands of fish and on the right: normal probability plot of the new residuals, both obtained after a logarithmic transformation in the Ricker (1954) model.

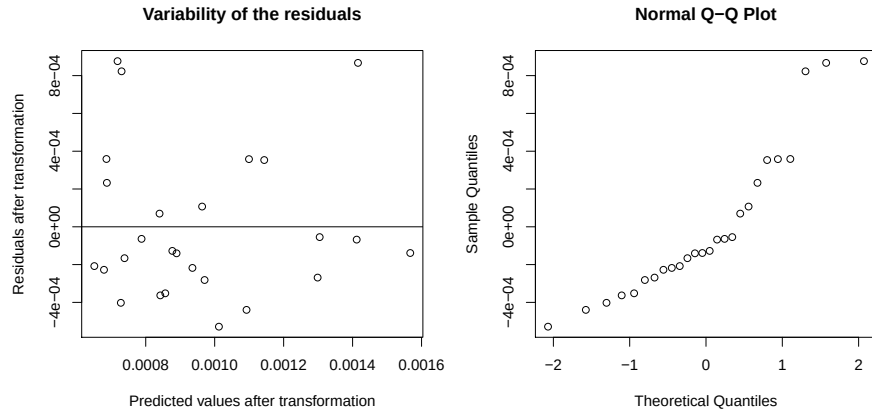


Figure 1.5: On the left: scatterplot of new residuals and predicted values for the Skeena River sockeye salmon stock where units are thousands of fish and on the right: normal probability plot of the new residuals, both obtained after an inverse transformation in the Beverton and Holt (1957) model.

The class of transformations introduced by Box and Cox (1964) has been generalized, see for example the Yeo and Johnson (2000) transform. Other types of transformations have also been introduced in the literature, like for example, the Zellner and Revankar (1969), the John and Draper (1980), the Bickel and Doksum (1981) and the MacKinnon and Magee (1990) transforms among others. We also refer to the review paper by Sakia (1992) for more details and references on this topic.

All the above mentioned papers consider a model where both the transformation and the regression function are parametric. In the literature, we can also find papers where the transformation is nonparametric and the regression function is parametric, e.g. Horowitz (1996) who proposed nonparametric estimators of the transformation and the cumulative distribution function of the error term that are consistent and asymptotically normal. For the same model, Chen (2002) proposed a rank-based estimator for the transformation that has the advantage of not involving nonparametric smoothing.

In this thesis, we will focus on the two other types of existing models, i.e., the models where the transformation is parametric and the regression function is nonparametric, see Chapters 2 and 3 and the models where both the transformation and the regression function are nonparametric, see Chapter 4. These two types of models, as well as the main contributions already made on it in the literature, are introduced in Sections 1.2 and 1.3 respectively.

Moreover, besides the problem of estimation of a transformation model that can be of semiparametric or nonparametric nature, we will also focus on 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 thesis, we will consider the problem of testing the parametric form of the regression function. Hence, a literature review of existing goodness-of-fit tests in regression models is proposed in Section 1.4. Finally, Section 1.5 explains more precisely the contents of each chapter of this thesis.

1.2 Semiparametric transformation models

1.2.1 Definition and notations

In this section, we will focus on a model where the transformation is parametric and the regression function is nonparametric, i.e., we will consider a semiparametric transformation model of the following form:

$$\Lambda_\theta(Y) = m(X) + \varepsilon, \quad (1.2.1)$$

where $m(\cdot)$ is an unknown regression function, $\Lambda_\theta(y)$ is some parametric transformation of the response variable Y , is continuously differentiable with respect to y and belongs to a parametric family of strictly increasing functions and $\theta \in \Theta$ where Θ is a compact subset of $\mathbb{R}^k$. We will denote by θ_0 and $m_0(\cdot)$ the true but unknown values of θ and $m(\cdot)$. Moreover, we assume that X is a d -dimensional covariate, Y is a univariate response variable and the error term ε has zero mean and is independent of X .

We suppose that Y has support $\mathcal{Y}$, X has compact support $\chi \subset \mathbb{R}^d$ and we denote by F_X , f_X , F_ε and f_ε the distribution and the probability density functions of X and ε respectively. We assume that these functions exist and are continuously differentiable. Moreover, let $\sigma^2 = V(\varepsilon) < \infty$ and define the function

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

Note that $m(x, \theta_0) = m(x)$. We also denote

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

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

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

be a $(k \times 1)$ -vector where $\theta = (\theta_1, \dots, \theta_k)^t$. Similar notations will be used for other functions. For any function ζ , we define $\zeta'(z) = \partial \zeta / \partial z$. Finally, let $\varepsilon(\theta) = \Lambda_\theta(Y) - m(X, \theta)$, $\varepsilon(\theta_0) = \varepsilon$ and let $F_{\varepsilon(\theta)}$ and $f_{\varepsilon(\theta)}$ be the distribution

and the density function of $\varepsilon(\theta)$, respectively. Finally, we suppose that we have randomly drawn an *iid* sample $(X_1, Y_1), \dots, (X_n, Y_n)$ from model (1.2.1), where the components of X_i are denoted by $(X_{i1}, \dots, X_{id})$ for $i = 1, \dots, n$. These assumptions and notations will be constantly used in Chapters 2 and 3.

1.2.2 Identification

It is clear that some normalization conditions are needed to identify the semiparametric transformation model (1.2.1). Indeed, if we consider $\Lambda_{\tilde{\theta}_0} \equiv \mu + \lambda \Lambda_{\theta_0}$, $\tilde{m} \equiv t + \lambda m$ and $\tilde{\varepsilon} = \mu - t + \lambda \varepsilon$, then the structure $(\Lambda_{\theta_0}, m, F_\varepsilon)$ is equivalent to the structure $(\Lambda_{\tilde{\theta}_0}, \tilde{m}, F_{\tilde{\varepsilon}})$ for any $\lambda > 0$, $\mu \in \mathbb{R}$ and $t \in \mathbb{R}$. Hence, a location and a scale normalization of Λ_{θ_0} is in particular needed to ensure that $\mu = 0$ and $\lambda = 1$. We need also a condition to ensure that $t = 0$. One possibility to identify the semiparametric transformation model (1.2.1) is to impose the two following conditions:

- (I1) $\Lambda_\theta(\alpha_1) = a_1$ and $\Lambda_\theta(\alpha_2) = a_2$ for all $\theta \in \Theta$ and for some $\alpha_1 < \alpha_2$ and $a_1 < a_2$,
- (I2) There exists $j \in \{1, \dots, d\}$ such that the set $\{x \in \chi : \frac{\partial}{\partial x_j} m(x) \neq 0\}$ has nonempty interior.

Note that the set of conditions (I1) can also be replaced by the following alternative set of conditions:

- (I1') $\Lambda_\theta(\alpha_1) = a_1$ and $\Lambda'_\theta(\alpha_3) = a_3$ for all $\theta \in \Theta$ and for some $a_1, a_3, \alpha_1, \alpha_3$.

The main idea of the two sets of conditions (I1) and (I1') is to fix the location and the scale of the model. Condition $\Lambda_\theta(\alpha_1) = a_1$ fixes the location of model (1.2.1) in both cases and (I1) and (I1') differ in their second condition, i.e., in the way they fix the scale of model (1.2.1).

More precisely, suppose that there exists another transformation $\Lambda_{\tilde{\theta}_0}(\cdot)$ such that:

$$\Lambda_{\tilde{\theta}_0}(Y) = \ell_1 \Lambda_{\theta_0}(Y) + \ell_2 ,$$

$\Lambda_{\tilde{\theta}_0}(\alpha_1) = a_1$, either $\Lambda_{\tilde{\theta}_0}(\alpha_2) = a_2$ or $\Lambda'_{\tilde{\theta}_0}(\alpha_3) = a_3$, and where ℓ_1 and ℓ_2 are some constants. Then,

$$\Lambda_{\tilde{\theta}_0}(\alpha_1) = \ell_1 \Lambda_{\theta_0}(\alpha_1) + \ell_2 = a_1 ,$$

and either

$$\Lambda_{\tilde{\theta}_0}(\alpha_2) = \ell_1 \Lambda_{\theta_0}(\alpha_2) + \ell_2 = a_2 ,$$

or

$$\Lambda'_{\tilde{\theta}_0}(\alpha_3) = \ell_1 \Lambda'_{\theta_0}(\alpha_3) = a_3 ,$$

which implies that $\ell_1 \cdot a_1 + \ell_2 = a_1$ and either $\ell_1 \cdot a_2 + \ell_2 = a_2$ or $\ell_1 \cdot a_3 = a_3$. We conclude that $\ell_1 = 1$ and $\ell_2 = 0$ in both cases, and then $\Lambda_{\theta_0} = \Lambda_{\tilde{\theta}_0}$. Consequently, $\theta_0 = \tilde{\theta}_0$.

Moreover, we have to impose (I2). Indeed, suppose that Condition (I2) is not satisfied, then we cannot find $j \in \{1, \dots, d\}$ such that $\frac{\partial m}{\partial x_j}(x) \neq 0$ for all x in a set of non-empty interior and either $m(x) = c$ for some constant c or $m(x)$ is a piecewise constant function. For simplicity, suppose here that $m(x) = c$ for some constant c , since the idea is the same for the second case. Hence, the semiparametric transformation model becomes $\Lambda_{\theta_0}(Y) = c + \varepsilon$ with ε independent of X . Next, consider another transformation $\Lambda_{\tilde{\theta}_0}(Y)$ of Y . Then,

$$\Lambda_{\tilde{\theta}_0}(Y) = \Lambda_{\tilde{\theta}_0} \Lambda_{\theta_0}^{-1} \Lambda_{\theta_0}(Y) = \Lambda_{\tilde{\theta}_0} \Lambda_{\theta_0}^{-1}(c + \varepsilon) = Z - E(Z) + E(Z) ,$$

where $Z = \tilde{\Lambda}_{\theta_0} \Lambda_{\theta_0}^{-1}(c + \varepsilon)$. Finally,

$$\Lambda_{\tilde{\theta}_0}(Y) = \tilde{c} + \tilde{\varepsilon} ,$$

where $\tilde{c} = E(Z)$, $\tilde{\varepsilon} = Z - E(Z)$, $\tilde{\varepsilon}$ has zero mean and $\tilde{\varepsilon}$ is independent of X . The model is thus not identified if (I2) is not assumed. Hence, Condition (I2) is necessary to identify the semiparametric transformation model (1.2.1).

We will finish this section by proving that the sets of conditions (I1) (or (I1')) and (I2) are sufficient to identify each quantity of that model. To prove the identification of the structure $(\Lambda_{\theta_0}, m, F_\varepsilon)$, we proceed in two steps: we first prove the identification of Λ_{θ_0} and then the identification of m and F_ε . More precisely, consider another structure $(\Lambda_{\tilde{\theta}_0}, \tilde{m}, F_{\tilde{\varepsilon}})$ that has the same properties and that is equivalent to the structure $(\Lambda_{\theta_0}, m, F_\varepsilon)$ and we will prove first that $\theta_0 = \tilde{\theta}_0$ and second that $\tilde{m} = m$ and $F_{\tilde{\varepsilon}} = F_\varepsilon$.

The proof of the first step is inspired from the proofs of Proposition 3.1 in Vanhems and Van Keilegom (2017) and Theorem 1 in Chiappori et al. (2015). Consider $F_{Y|X}$ the conditional distribution function of Y given X . We have

$$\begin{aligned} F_{Y|X}(y, x) &= P(Y \leq y | X = x) \\ &= P(\Lambda_{\theta_0}(Y) \leq \Lambda_{\theta_0}(y) | X = x) \\ &= P(\varepsilon(\theta_0) \leq \Lambda_{\theta_0}(y) - m(X, \theta_0) | X = x) \\ &= F_{\varepsilon(\theta_0)}(\Lambda_{\theta_0}(y) - m(x, \theta_0)) \\ &= F_\varepsilon(\Lambda_{\theta_0}(y) - m(x)) , \end{aligned} \tag{1.2.2}$$

where the second equality was obtained since Λ_{θ_0} is a strictly increasing function and the one but last equality was obtained by independence between X and ε . Next, as we assumed that F_ε and Λ_{θ_0} are continuously differentiable functions, we can compute

$$\frac{\partial}{\partial y} F_{Y|X}(y, x) = \Lambda'_{\theta_0}(y) \times f_\varepsilon(\Lambda_{\theta_0}(y) - m(x)) . \tag{1.2.3}$$

Similarly, using Condition (I2), we can compute for some $j \in \{1, \dots, d\}$

$$\frac{\partial}{\partial x_j} F_{Y|X}(y, x) = -\frac{\partial}{\partial x_j} m(x) \times f_\varepsilon(\Lambda_{\theta_0}(y) - m(x)) . \quad (1.2.4)$$

Dividing (1.2.4) by (1.2.3), we obtain

$$\Lambda'_{\theta_0}(y) = -\frac{\partial}{\partial x_j} m(x) \times s(y, x) , \quad (1.2.5)$$

where $s(y, x) = \frac{\partial}{\partial y} F_{Y|X}(y, x) / \frac{\partial}{\partial x_j} F_{Y|X}(y, x)$. Next, integrating the last expression from α_1 to y , we obtain:

$$\Lambda_{\theta_0}(y) - a_1 = -\frac{\partial}{\partial x_j} m(x) \times S(y, x) , \quad (1.2.6)$$

where $S(y, x) = \int_{\alpha_1}^y s(t, x) dt$ and we have used the fact that $\Lambda_{\theta_0}(\alpha_1) = a_1$ by the first condition in (I1) or in (I1'). From equation (1.2.6) and since $\Lambda_{\theta_0}(\alpha_2) = a_2$ by the second condition in (I1), we obtain

$$-\frac{\partial}{\partial x_j} m(x) = \frac{a_2 - a_1}{S(\alpha_2, x)} .$$

Combining this last result and equation (1.2.6), this gives

$$\Lambda_{\theta_0}(y) = a_1 + (a_2 - a_1) \frac{S(y, x)}{S(\alpha_2, x)} . \quad (1.2.7)$$

Alternatively, if we consider the second condition in (I1') instead of the second condition in (I1), i.e., $\Lambda'_{\theta_0}(\alpha_3) = a_3$ instead of $\Lambda_{\theta_0}(\alpha_2) = a_2$, we deduce from equation (1.2.5) that

$$-\frac{\partial}{\partial x_j} m(x) = \frac{a_3}{s(\alpha_3, x)} .$$

Hence, combining this last result and equation (1.2.6), this gives

$$\Lambda_{\theta_0}(y) = a_1 + a_3 \frac{S(y, x)}{s(\alpha_3, x)} . \quad (1.2.8)$$

Consequently, assuming either Conditions (I1) and (I2) or Conditions (I1') and (I2), we have rewritten the transformation $\Lambda_{\theta_0}(y)$ in terms of the functions $S(y, x)$ and $s(y, x)$ up to additive and multiplicative constants, see (1.2.7) and (1.2.8) respectively. Moreover, these functions $S(y, x)$ and $s(y, x)$ do not depend on the regression function m nor on the error distribution function F_ε since they are constructed using only the conditional distribution function $F_{Y|X}$ of Y given X and its derivatives with respect to y and x_j . Then, for the equivalent structure $(\Lambda_{\tilde{\theta}_0}, \tilde{m}, F_{\tilde{\varepsilon}})$, we can prove exactly in the same way as above that

$$\Lambda_{\tilde{\theta}_0}(y) = a_1 + (a_2 - a_1) \frac{S(y, x)}{S(\alpha_2, x)} ,$$

under (I1) and (I2) and that

$$\Lambda_{\tilde{\theta}_0}(y) = a_1 + a_3 \frac{S(y, x)}{s(\alpha_3, x)} ,$$

under (I1') and (I2). In conclusion, $\Lambda_{\theta_0}(y) = \Lambda_{\tilde{\theta}_0}(y)$ for all $y \in \mathcal{Y}$, which implies that $\theta_0 = \tilde{\theta}_0$ and the component Λ_{θ_0} in the semiparametric transformation model (1.2.1) is identifiable.

Second, recall the definition of the function $m(x)$ given in Section 1.2.1:

$$m(x) = E[\Lambda_{\theta_0}(Y)|X = x] .$$

Similarly, for the equivalent structure $(\Lambda_{\tilde{\theta}_0}, \tilde{m}, F_{\tilde{\varepsilon}})$, we also have

$$\tilde{m}(x) = E[\Lambda_{\tilde{\theta}_0}(Y)|X = x] .$$

Using the identifiability of the transformation Λ_{θ_0} proved just before, we can deduce that $m(x) = \tilde{m}(x)$ for all $x \in \mathcal{X}$. Hence, the regression function m is also identifiable in the semiparametric transformation model (1.2.1).

Finally, using equation (1.2.2), we know that

$$F_{Y|X}(y, x) = F_{\varepsilon}(\Lambda_{\theta_0}(y) - m(x)) .$$

Similarly, for the equivalent structure $(\Lambda_{\tilde{\theta}_0}, \tilde{m}, F_{\tilde{\varepsilon}})$, we also have

$$F_{Y|X}(y, x) = F_{\tilde{\varepsilon}}(\Lambda_{\tilde{\theta}_0}(y) - \tilde{m}(x)) .$$

Using the identifiability of the transformation Λ_{θ_0} and of the regression function $m(x)$, we can deduce that $F_{\varepsilon}(e) = F_{\tilde{\varepsilon}}(e)$ for all $e \in \mathcal{E}$ where $\mathcal{E} \subset \mathbb{R}$ is the support of ε . Consequently, the error distribution function F_{ε} is also identifiable. This concludes the proof of the identifiability of the semiparametric transformation model (1.2.1).

1.2.3 Estimation of the model

Linton et al. (2008) have extensively studied the semiparametric transformation model (1.2.1). Their main objective was to propose different estimators of the transformation parameter θ and to establish the asymptotic properties of these estimators.

They proposed two estimation methods for the unknown true parameter vector θ_0 : a profile likelihood method and a mean squared distance from independence method.

First, the idea of the profile likelihood method is to calculate the log-likelihood function of the vector (X, Y) and to replace all unknown expressions by nonparametric estimators. We consider $\Lambda_{\theta_0}(Y) = m(X, \theta_0) + \varepsilon(\theta_0)$ for

$\theta_0 \in \Theta$. Hence, the cumulative distribution function of Y given X is

$$\begin{aligned} P(Y \leq y|X = x) &= P(\Lambda_{\theta_0}(Y) \leq \Lambda_{\theta_0}(y)|X = x) \\ &= P(m(X, \theta_0) + \varepsilon(\theta_0) \leq \Lambda_{\theta_0}(y)|X = x) \\ &= F_{\varepsilon(\theta_0)}(\Lambda_{\theta_0}(y) - m(x, \theta_0)) . \end{aligned}$$

The first equality was obtained since $\Lambda_{\theta_0}(\cdot)$ is a strictly increasing function and the last equality was obtained by independence between X and $\varepsilon(\theta_0)$. Taking the derivative of the last expression, we obtain the following expression for the probability density function of Y given X :

$$f_{Y|X}(y|x) = f_{\varepsilon(\theta_0)}(\Lambda_{\theta_0}(y) - m(x, \theta_0))\Lambda'_{\theta_0}(y) .$$

Then, for a given $\theta \in \Theta$ and as $f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x)$ where $f_{X,Y}(x, y)$ is the joint density function of X and Y , the log-likelihood function is given by

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

In this expression, $\Lambda_{\theta}(\cdot)$ and $\Lambda'_{\theta}(\cdot)$ are known, unlike $m(\cdot, \theta)$ and $f_{\varepsilon(\theta)}(\cdot)$. These two quantities will be replaced by suitable nonparametric estimators denoted respectively by $\hat{m}(\cdot, \theta)$ and $\hat{f}_{\varepsilon(\theta)}(\cdot)$. In Section 2.2.1, we will define more precisely the estimators $\hat{m}(\cdot, \theta)$ and $\hat{f}_{\varepsilon(\theta)}(\cdot)$ we will consider in this thesis. Moreover, the last term in the log-likelihood function, i.e., $\log f_X(X_i)$, can be omitted since it does not depend on θ and it will consequently not influence the maximization with respect to θ hereafter. Consequently, the profile likelihood estimator of θ is defined by

$$\hat{\theta}_{PL} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \left\{ \log \hat{f}_{\varepsilon(\theta)}(\Lambda_{\theta}(Y_i) - \hat{m}(X_i, \theta)) + \log \Lambda'_{\theta}(Y_i) \right\} . \quad (1.2.9)$$

The asymptotic properties of this estimator have been established by Linton et al. (2008). In their Theorem 4.1, they prove the following asymptotic representation for $\hat{\theta}_{PL} - \theta_0$:

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

and

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

where

$$g(X, Y) = \Xi^{-1}\xi(\theta_0, X, Y) , \quad (1.2.10)$$

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

and

$$\Xi = \frac{\partial}{\partial \theta} E \left[\xi(\theta, X, Y) \right] \Big|_{\theta=\theta_0} .$$

The second approach considered by Linton et al. (2008) was the minimization of the mean squared distance from independence. The construction of this estimator relies on the independence assumption between X and ε . Under this particular assumption, we have

$$F_{X,\varepsilon(\theta_0)}(x, e) = F_X(x)F_{\varepsilon(\theta_0)}(e) ,$$

where $F_{X,\varepsilon(\theta_0)}(x, e)$ is the joint distribution function of X and $\varepsilon(\theta_0)$. The main idea of the mean squared distance from independence estimator is to minimize with respect to θ some criterion function that is constructed on the basis of the difference between a nonparametric estimator of $F_{X,\varepsilon(\theta)}(x, e)$ and a nonparametric estimator of $F_X(x)F_{\varepsilon(\theta)}(e)$.

More precisely, for each $\theta \in \Theta$, Linton et al. (2008) defined the following empirical distribution functions to estimate $F_X(x)$, $F_{\varepsilon(\theta)}(e)$ and $F_{X,\varepsilon(\theta)}(x, e)$ respectively:

$$\begin{aligned} \hat{F}_X(x) &= \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} , \\ \hat{F}_{\varepsilon(\theta)}(e) &= \frac{1}{n} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_i(\theta) \leq e\}} , \end{aligned}$$

and

$$\hat{F}_{X,\varepsilon(\theta)}(x, e) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} 1_{\{\hat{\varepsilon}_i(\theta) \leq e\}} ,$$

where $\hat{\varepsilon}_i(\theta) = \Lambda_\theta(Y_i) - \hat{m}(X_i, \theta)$ and $\hat{m}(x, \theta)$ is an appropriate nonparametric estimator of $m(x, \theta)$. Moreover,

$$1_A(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases} ,$$

and $1_{\{X_i \leq x\}}$ is a component by component indicator for $i = 1, \dots, n$, i.e.,

$$1_{\{X_i \leq x\}} = \prod_{j=1}^d 1_{\{X_{ij} \leq x_j\}} .$$

Next, they also defined the moment function

$$G_{n,\theta}(x, e) = \hat{F}_{X,\varepsilon(\theta)}(x, e) - \hat{F}_X(x) \hat{F}_{\varepsilon(\theta)}(e) ,$$

and the criterion function

$$\|G_{n,\theta}\|_2^2 = \int [G_{n,\theta}(x, e)]^2 d\mu(x, e) ,$$

where μ is some probability measure. For simplicity, we can use the empirical measure $d\mu_n$ of $\{X_i, \widehat{\varepsilon}_i(\theta)\}_{i=1, \dots, n}$ which results in a criterion function

$$\|\widetilde{G}_{n,\theta}\|_2^2 = \frac{1}{n} \sum_{i=1}^n G_{n,\theta}^2(X_i, \widehat{\varepsilon}_i(\theta)) .$$

Finally, the mean squared distance from independence estimator of θ is defined by

$$\widehat{\theta}_{MD} = \arg \min_{\theta \in \Theta} \|\widetilde{G}_{n,\theta}\|_2^2 . \quad (1.2.11)$$

Note that Linton et al. (2008) also established the asymptotic properties of $\widehat{\theta}_{MD}$ and in particular an asymptotic distribution for $n^{1/2}(\widehat{\theta}_{MD} - \theta_0)$. This asymptotic distribution is not the same as the one of $n^{1/2}(\widehat{\theta}_{PL} - \theta_0)$ but we will not introduce it here since we will always use the profile likelihood estimator of θ in this thesis when an estimator of θ will be required.

Indeed, Linton et al. (2008) performed a simulation study in order to compare $\widehat{\theta}_{PL}$ to $\widehat{\theta}_{MD}$ and the profile likelihood estimator clearly outperforms the mean squared distance from independence estimator. However, it is important to notice that the profile likelihood estimator is computationally more expensive than the mean squared distance from independence estimator since we not only have to estimate the regression function $m(\cdot, \theta)$ but also the error density function $f_{\varepsilon(\theta)}(\cdot)$.

1.2.4 Estimation of $f_{\varepsilon}(\cdot)$, $F_{\varepsilon}(\cdot)$ and related contributions

First, we mention the work of Neumeyer and Van Keilegom (2010) who constructed an estimator for the error distribution function in a heteroscedastic nonparametric regression model with multivariate covariates but without transformation of the response. As estimator, they considered the empirical distribution function of residuals, which are obtained from multivariate local polynomial fits of the regression and variance functions, respectively. Moreover, as an application, they constructed in detail a test for additivity of the regression function.

Next, Colling et al. (2015) and Heuchenne et al. (2015) introduced and studied respectively nonparametric estimators for the error density function $f_{\varepsilon}(y)$ and the error distribution function $F_{\varepsilon}(y)$ in a semiparametric transformation model of the form (1.2.1) with $d = 1$.

The estimation approach of Colling et al. (2015) and Heuchenne et al. (2015) is based on a two-step strategy. In a first step, the idea is to obtain suitable estimators $\widehat{\theta}$ and $\widehat{m}(x, \theta)$ of θ_0 and $m(x, \theta)$ respectively and to form the semiparametric regression residuals

$$\widehat{\varepsilon}_i = \widehat{\varepsilon}_i(\widehat{\theta}) = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}(X_i, \widehat{\theta}) .$$

In a second step, the density function f_ε is estimated by the classical kernel estimator of a density function:

$$\hat{f}_\varepsilon(y) = \frac{1}{ng} \sum_{i=1}^n k_2\left(\frac{y - \hat{\varepsilon}_i}{g}\right), \quad (1.2.12)$$

where g is some univariate bandwidth and k_2 is some univariate kernel, and the distribution function F_ε is estimated by the empirical distribution function of the $\hat{\varepsilon}_i$'s:

$$\hat{F}_\varepsilon(y) = n^{-1} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_i \leq y\}}. \quad (1.2.13)$$

Note that Colling et al. (2015) and Heuchenne et al. (2015) used a profile likelihood approach, developed in Linton et al. (2008) and defined in (1.2.9) to estimate θ_0 , whereas for each fixed θ , $m(x, \theta)$ is estimated by means of the Nadaraya (1964) and Watson (1964) method. We refer to Colling et al. (2015) and Heuchenne et al. (2015) for more details about the estimators $\hat{f}_\varepsilon(y)$ and $\hat{F}_\varepsilon(y)$. In particular, Theorems 1 and 2 in Colling et al. (2015) establish respectively an asymptotic representation and the weak convergence of $n^{1/2}(\hat{f}_\varepsilon(\cdot) - f_\varepsilon(\cdot))$ and Theorem 3.1 and Corollary 3.1 in Heuchenne et al. (2015) establish respectively an asymptotic representation and the weak convergence of $n^{1/2}(\hat{F}_\varepsilon(\cdot) - F_\varepsilon(\cdot))$.

Estimating the error density function and the error distribution function may be very useful in various regression problems. First, taking transformations of the data may induce normality and error variance homogeneity in the transformed model. Then, the estimation of the error density and the error distribution in the transformed model may be used for testing these hypotheses. It may also be used for goodness-of-fit tests of a specified error distribution in a parametric or nonparametric regression setting, see for example Akritas and Van Keilegom (2001) and Cheng and Sun (2008). The estimation of these two functions can also be useful for testing the symmetry of the residual distribution, see Ahmad and Li (1997), Dette et al. (2002) and Neumeyer and Dette (2007), among others.

Next, using the estimation of the error density function, Escanciano and Jacho-Chávez (2012) constructed an estimator for the marginal density of the response Y . Another possible application of the estimation of the error density in model (1.2.1) is the forecasting of $\Lambda_{\theta_0}(Y)$ by means of the mode approach, since the mode of the probability density function of $\Lambda_{\theta_0}(Y)$ given $X = x$ is $m(x) + \arg \max_{e \in \mathbb{R}} f_\varepsilon(e)$.

Moreover, the estimation of the error distribution function can also be used for various other testing problems, like tests for the parametric form of the regression or variance function, see Van Keilegom et al. (2008a) and Dette et al. (2007), tests for the equality of regression functions, see Pardo-Fernández et al. (2007), tests for the validity of the model, see Einmahl and Van Keilegom

(2008) and tests for additivity of the regression function, see Neumeyer and Van Keilegom (2010), among others. We will explain in detail in Section 1.4.2 how we can construct such tests using estimators of $F_\varepsilon(\cdot)$.

Note that all the mentioned references in the last three paragraphs considered a model where the response is not transformed. We also refer to Neumeyer et al. (2016) who introduced estimators for the different components of a heteroscedastic transformation model, proved the asymptotic normality of these estimators and constructed a test for the validity of this model. Finally, we would also like to mention the work of Vanhems and Van Keilegom (2017), who studied the estimation of model (1.2.1) and supposed that some of the regressors are endogenous as a result of e.g. omitted variables, measurement error or simultaneous equations.

Finally, we have already seen in Section 1.2.3 that an estimator of the error density is needed to define the profile likelihood estimator of θ_0 and an estimator of the error distribution is needed to define the mean squared distance from independence estimator of θ_0 . Moreover, in Chapter 2, we will use an estimator of the error distribution to construct a test for the parametric form of the regression function in model (1.2.1). Hence, in Section 1.4.2, we will explain in detail how Van Keilegom et al. (2008a) constructed such test using an estimator of the error distribution function but in the case where a model is not transformed.

In conclusion, the error density and the error distribution play a very important role, both for exploratory analyses and for statistical inference. Consequently, a simulation study will be performed in Sections 2.4.2 and 2.4.3 to illustrate respectively the finite sample performance of the estimators of the functions f_ε and F_ε introduced in (1.2.12) and (1.2.13).

1.3 Nonparametric transformation models

1.3.1 Definition, notations and identification

In this section, we will consider a model in which both the transformation and the regression function are nonparametric, i.e., we will consider a nonparametric transformation model of the form

$$\Lambda(Y) = m(X) + \varepsilon, \quad (1.3.1)$$

where $\Lambda(\cdot)$ and $m(\cdot)$ are respectively an unknown transformation and an unknown regression function. Similarly as in Section 1.2, we assume that $\Lambda(\cdot)$ is continuously differentiable and is a strictly increasing function, X is a d -dimensional covariate, Y is a continuous univariate response variable and the error term ε has zero mean and is independent of X . Note that we will use similar notations and assumptions as the ones introduced in Section 1.2.1.

One way to identify model (1.3.1) is to impose one of the following normalization conditions

(J1) either $\Lambda(\alpha_1) = a_1$ and $\Lambda(\alpha_2) = a_2$ for some $\alpha_1 < \alpha_2$ and $a_1 < a_2$,

(J1') or $\Lambda(\alpha_1) = a_1$ and $\Lambda'(\alpha_3) = a_3$ for some $a_1, a_3, \alpha_1, \alpha_3$,

and also the condition

(J2) There exists $j \in \{1, \dots, d\}$ such that the set $\{x \in \chi : \frac{\partial}{\partial x_j} m(x) \neq 0\}$ has nonempty interior.

The main idea of the two sets of conditions (J1) and (J1') is the same as in Section 1.2.2 and is to fix the location and the scale of the model. Condition $\Lambda(\alpha_1) = a_1$ fixes the location of model (1.3.1) in both cases and (J1) and (J1') differ in their second condition, i.e., in the way they fix the scale of model (1.3.1).

Moreover, we can prove that these conditions are sufficient to identify model (1.3.1) simply by replacing the parametric transformation Λ_θ by the nonparametric transformation Λ in the proof given in Section 1.2.2. Indeed, the proof of the identifiability given in Section 1.2.2 was not influenced by the parameter θ and hence by the parametric form of the transformation, which implies that Λ_θ can be replaced by a nonparametric transformation Λ in every step of that proof. More precisely, if we consider another structure $(\tilde{\Lambda}, \tilde{m}, F_{\tilde{\varepsilon}})$ that has the same properties and that is equivalent to the structure $(\Lambda, m, F_\varepsilon)$, we can prove for both structures exactly in the same way as in Section 1.2.2 that

$$\Lambda(y) = a_1 + (a_2 - a_1) \frac{S(y, x)}{S(\alpha_2, x)} = \tilde{\Lambda}(y) ,$$

under (J1) and (J2) and that

$$\Lambda(y) = a_1 + a_3 \frac{S(y, x)}{s(\alpha_3, x)} = \tilde{\Lambda}(y) ,$$

under (J1') and (J2), where the functions $s(y, x)$ and $S(y, x)$ are defined by $s(y, x) = \frac{\partial}{\partial y} F_{Y|X}(y, x) / \frac{\partial}{\partial x_j} F_{Y|X}(y, x)$ and $S(y, x) = \int_{\alpha_1}^y s(t, x) dt$. Hence, $\Lambda(y) = \tilde{\Lambda}(y)$ for all $y \in \mathcal{Y}$ and the component Λ in the nonparametric transformation model (1.3.1) is identifiable. Second, we have

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

and

$$\tilde{m}(x) = E[\tilde{\Lambda}(Y)|X = x] ,$$

which implies that $m(x) = \tilde{m}(x)$ for all $x \in \chi$ using the identifiability of the transformation Λ . Finally, using the same reasoning as in (1.2.2), we obtain

$$F_{Y|X}(y, x) = F_\varepsilon(\Lambda(y) - m(x)) ,$$

and

$$F_{Y|X}(y, x) = F_{\tilde{\varepsilon}}\left(\tilde{\Lambda}(y) - \tilde{m}(x)\right) ,$$

which implies that $F_{\varepsilon}(e) = F_{\tilde{\varepsilon}}(e)$ for all $e \in \mathcal{E}$ where $\mathcal{E} \subset \mathbb{R}$ is the support of ε . Consequently, the regression function m and the error distribution function F_{ε} are also identifiable.

Consequently, considering either a parametric or a nonparametric transformation of the response does not influence the way we choose to identify the corresponding transformation model and it is therefore not surprising at all that the identification conditions (J1), (J1') and (J2) introduced here are very similar to Conditions (I1), (I1') and (I2) introduced in Section 1.2.2. Finally, we also refer to Chiappori et al. (2015), Vanhems and Van Keilegom (2017) among others for more details about these sets of conditions.

1.3.2 Estimation of the model

Chiappori et al. (2015) extensively studied the nonparametric transformation model (1.3.1). They worked under the normalization conditions (J1) for theoretical reasons and considered in particular $\alpha_1 = 0$, $a_1 = 0$, $\alpha_2 = 1$ and $a_2 = 1$ without loss of generality. Note that the values 0 and 1 can be replaced by any other values as long as they are in the interior of the support $\mathcal{Y}$ of Y . They developed sufficient conditions so that this nonparametric transformation model is identifiable and constructed nonparametric estimators of the transformation, the regression function and the distribution function of ε and analyzed their asymptotic properties under (J1). In this section, we will focus on the problem of nonparametric estimation of the transformation Λ .

The main idea of the nonparametric estimators of the transformation Λ developed in Chiappori et al. (2015) is to consider the conditional distribution function of Y given X , denoted by $\varphi_C(y, x)$, and to divide its derivative with respect to y , denoted by $\varphi_{C,y}(y, x)$, by its derivative with respect to x_{ρ} , denoted by $\varphi_{C,\rho}(y, x)$, where ρ is some index contained in $\{1, \dots, d\}$ and x_{ρ} is the component of the vector $x = (x_1, \dots, x_d)^t$ corresponding to the index ρ . This particular way of constructing a nonparametric estimator of the transformation was already considered in Horowitz (1996) but in the case of a parametric regression function. More precisely, define $\varphi_C(y, x) = F_{Y|X}(y|x)$. Using the independence between ε and X , we have

$$\begin{aligned} \varphi_C(y, x) &= P(Y \leq y | X = x) \\ &= P(m(X) + \varepsilon \leq \Lambda(y) | X = x) \\ &= F_{\varepsilon}(\Lambda(y) - m(x)) . \end{aligned}$$

Next, the derivative of $\varphi_C(y, x)$ with respect to y is given by

$$\varphi_{C,y}(y, x) = \Lambda'(y) f_{\varepsilon}(\Lambda(y) - m(x)) ,$$

where $\Lambda'(y) = \frac{\partial}{\partial y} \Lambda(y)$. Similarly, the derivative of $\varphi_C(y, x)$ with respect to x_ρ , for some $\rho \in \{1, \dots, d\}$ is given by

$$\varphi_{C,\rho}(y, x) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) f_\varepsilon(\Lambda(y) - m(x)) .$$

Hence, dividing $\varphi_{C,y}(y, x)$ by $\varphi_{C,\rho}(y, x)$, they obtained

$$\Lambda'(y) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) \times \frac{\varphi_{C,y}(y, x)}{\varphi_{C,\rho}(y, x)} .$$

Using this last expression, Chiappori et al. (2015) established in their Theorem 1 that the transformation Λ can be globally identified, under the normalization conditions (J1) with $\alpha_1 = 0$, $a_1 = 0$, $\alpha_2 = 1$ and $a_2 = 1$ as mentioned above, by

$$\Lambda(y) = \lambda_{C,\rho}(y, x) = \frac{S_{C,\rho}(y, x)}{S_{C,\rho}(1, x)} \quad \text{where} \quad S_{C,\rho}(y, x) = \int_0^y \frac{\varphi_{C,y}(w, x)}{\varphi_{C,\rho}(w, x)} dw .$$

Note that this result holds for any $\rho \in \{1, \dots, d\}$ and for any $x \in \chi$ and also the expression $\frac{S_{C,\rho}(y, x)}{S_{C,\rho}(1, x)}$ does not depend on ρ nor x . This result is at the basis of the definition of the two estimators that Chiappori et al. (2015) proposed for $\Lambda(y)$. Indeed, it suffices to estimate the quantity $\lambda_{C,\rho}(y, x)$ to get an estimator of the transformation Λ . Hereafter, it will be assumed that $\rho = 1$ without loss of generality. First, Chiappori et al. (2015) used kernel estimators to estimate the quantity $\varphi_C(y, x)$:

$$\hat{\varphi}_C(y, x) = \frac{\frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_y}(y - Y_i) \mathbf{K}_{h_x}(X_i - x)}{\frac{1}{n} \sum_{i=1}^n \mathbf{K}_{h_x}(X_i - x)} ,$$

where $h_y > 0$ and $h_x > 0$ are two univariate bandwidths, K is a univariate kernel, $\mathcal{K}(y) = \int_{-\infty}^y K(w) dw$ and $\mathbf{K}(x)$ is a multivariate product kernel of the form $\mathbf{K}(x) = \prod_{i=1}^d K(x_i)$ with $x = (x_1, \dots, x_d)^t$. Second, they deduced the following estimator for $S_{C,1}(y, x)$:

$$\hat{S}_{C,1}(y, x) = \int_0^y \frac{\hat{\varphi}_{C,y}(w, x)}{\hat{\varphi}_{C,1}(w, x)} dw ,$$

where

$$\hat{\varphi}_{C,y}(y, x) = \frac{\partial}{\partial y} \hat{\varphi}_C(y, x) \quad \text{and} \quad \hat{\varphi}_{C,1}(y, x) = \frac{\partial}{\partial x_1} \hat{\varphi}_C(y, x) .$$

Finally, they proposed the two following nonparametric estimators for $\Lambda(y)$ respectively based on the least squares loss and the least absolute deviation loss:

$$\hat{\Lambda}_{C,LS}(y) = \int_{\chi} v(x) \hat{\lambda}_{C,1}(y, x) dx , \quad (1.3.2)$$

and

$$\hat{\Lambda}_{C,LAD}(y) = \arg \min_{q_m \in \mathbb{R}} \int_{\mathcal{X}} v(x) \left| \hat{\lambda}_{C,1}(y, x) - q_m \right| dx, \quad (1.3.3)$$

where v is a weighting function and $\hat{\lambda}_{C,1}(y, x) = \frac{\hat{S}_{C,1}(y, x)}{\hat{S}_{C,1}(1, x)}$. Chiappori et al. (2015) also established under (J1) the limiting distributions of these two estimators, see their Theorem 2.

However, the response Y is often skewed with very long tails in transformation models, which implies that the kernel estimators based on Y and proposed by Chiappori et al. (2015) can work very badly in practice. This also naturally impacts the quality of their estimators of the transformation. In Chapter 4, we will define two new nonparametric estimators of the transformation Λ , and we will show that they outperform the ones constructed by Chiappori et al. (2015).

To conclude this section, we would like to mention the works of Breiman and Friedman (1985), Horowitz (2001) and Jacho-Chavez et al. (2010) among others. They developed fully nonparametric estimators but in contexts that are slightly different from model (1.3.1). Breiman and Friedman (1985) constructed an algorithm for estimating the different components of the model (1.3.1) when the regression function m is supposed to be additive. Moreover, Horowitz (2001) proposed a nonparametric estimation of a generalized additive model with an unknown link function and Jacho-Chavez et al. (2010) discussed the identification and the estimation of the unknown functions H , M , G and F , where $r(x, z) = H\{M(x, z)\}$ and $M(x, z) = G(x) + F(z)$.

1.4 Goodness-of-fit tests in regression models

1.4.1 Introduction

We will now focus on the problem of testing in a regression model. We have already mentioned in Section 1.2.4 that several aspects of the model can be tested. In this thesis, we consider in particular the problem of testing the parametric form of the regression function m , i.e., we like to test the hypothesis

$$H_0 : m \in \mathcal{M} \quad \text{against} \quad H_1 : m \notin \mathcal{M}, \quad (1.4.1)$$

where $\mathcal{M} = \{m_\beta : \beta \in \mathcal{B}\}$ is some parametric class of regression functions and $\mathcal{B} \subset \mathbb{R}^q$. In the literature, we can find several ways for defining such test but in the case of regression models without transformation of the response. One main goal in this thesis is to extend tests for (1.4.1) in a context of a semiparametric transformation model of the form (1.2.1). It will be done using two different approaches, the first using estimations of the error distribution function, see Chapter 2 and the second using estimations of the integrated regression function, see Chapter 3.

In Section 1.4.2, we will explain in detail how tests for the parametric form of the regression function m can be constructed using the error distribution function in a context of a nonparametric regression model without transformation of the response. The same will be done in Section 1.4.3 with the integrated regression function. In addition to these two particular approaches, we will also present in Section 1.4.4 other possible ways to construct a test for the problem (1.4.1). We would also like to mention the work of González-Manteiga and Crujeiras (2013), who gave a recent overview on goodness-of-fit tests for regression models.

1.4.2 Using the error distribution function

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 et al. (2008a). Consider the hypotheses H_0 and H_1 introduced in (1.4.1). Their test was developed for the heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$, where X is a 1-dimensional covariate, Y is a univariate response variable, the error term ε has zero mean and is independent of X and the regression function $m(x) = E(Y|X = x)$ and the conditional variance function $\sigma^2(x) = V(Y|X = x)$ are unknown. Let (X_i, Y_i) be independent replications of (X, Y) for $i = 1, \dots, n$. The basic idea of their test is to measure a certain distance between the empirical distribution function of the nonparametric residuals (in a general context) and the empirical distribution function of the parametric residuals estimated under H_0 . The theoretical justification of this testing procedure is given in their Theorem 2.1, where they proved that H_0 is true if and only if

$$\varepsilon = \frac{Y - m(X)}{\sigma(X)} \quad \text{and} \quad \varepsilon_0 = \frac{Y - m_{\tilde{\beta}_0}(X)}{\sigma(X)},$$

have the same distribution, where $\tilde{\beta}_0 = \arg \min_{\beta \in \mathcal{B}} E[(m(X) - m_\beta(X))^2]$. First, they constructed the nonparametric residuals $\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}$ and defined the following empirical distribution function:

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

where $\hat{m}(x)$ and $\hat{\sigma}^2(x)$ are suitable estimators of $m(x)$ and $\sigma^2(x)$, constructed with the Nadaraya (1964) and Watson (1964) approach for instance, see Van Keilegom et al. (2008a). Second, under H_0 , they estimated the parameter β by the least squares method, i.e.,

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^n (Y_i - m_\beta(X_i))^2.$$

Hence, they obtained $m_{\hat{\beta}}(x)$, an estimator of the regression function m under H_0 . Next, Van Keilegom et al. (2008a) decided to follow Härdle and Mammen (1993) and to smooth $m_{\hat{\beta}}(x)$ using the same bandwidth and the same kernel as for the nonparametric estimator $\hat{m}(x)$. They argued that this new smoothed parametric estimator $\hat{m}_0(x)$ has the advantage of having the same bias as its nonparametric counterpart, which implies considerable power improvements for small samples in comparison with the nonsmoothed parametric estimator $m_{\hat{\beta}}(x)$. Then, they defined

$$\hat{F}_{\hat{\varepsilon}_0}(y) = n^{-1} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_{i0} \leq y\}} ,$$

where $\hat{\varepsilon}_{i0} = \frac{Y_i - \hat{m}_0(X_i)}{\hat{\sigma}(X_i)}$ are the residuals estimated under H_0 . Finally, to compare the empirical distribution functions $\hat{F}_{\hat{\varepsilon}}(y)$ and $\hat{F}_{\hat{\varepsilon}_0}(y)$, they used a Kolmogorov-Smirnov test statistic and a Cramér-von Mises test statistic:

$$T_{KS} = n^{1/2} \sup_{y \in \mathbb{R}} \left| \hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_{\hat{\varepsilon}_0}(y) \right| ,$$

and

$$T_{CM} = n \int \left(\hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_{\hat{\varepsilon}_0}(y) \right)^2 d\hat{F}_{\hat{\varepsilon}}(y) .$$

We also refer to their Theorem 3.1 where they established an asymptotic representation for $\hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_{\hat{\varepsilon}_0}(y)$ under H_0 from which they deduced the limiting distribution of the process $n^{1/2}(\hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_{\hat{\varepsilon}_0}(y))$ in their Corollary 3.1 and the limiting distributions of the test statistics T_{KS} and T_{CM} under H_0 in their Corollary 3.2. In their Theorem 3.2, they also proved similar results under the local alternative $H_{1n} : m(x) = m_{\beta_0}(x) + n^{-1/2}r(x)$ for all x , for some function r and where β_0 is the true value of β under H_0 . In Chapter 2, we will see how their ideas and methodology can be carried over to a semiparametric transformation model.

Besides the test for the parametric form of the regression function, several other aspects of a regression model have already been tested in the literature for the same location-scale model and using estimations of the error distribution. We would like to mention the works of Dette et al. (2007), Neumeyer and Dette (2007), Pardo-Fernández et al. (2007) and Einmahl and Van Keilegom (2008) who constructed respectively tests for the parametric form of the variance function, the symmetry of the residual distribution, the equality of regression curves and the validity of the location-scale model. 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.

Note that all these tests are based on Kolmogorov-Smirnov and Cramér-von Mises test statistics that are built using the same empirical distribution

function $\widehat{F}_{\widehat{\varepsilon}}(y)$ of the nonparametric residuals $\widehat{\varepsilon}_i$ as in Van Keilegom et al. (2008a).

First, Dette et al. (2007) constructed a test for the parametric form of the variance function, i.e., $H_0 : \sigma^2 \in \{\sigma_{\beta}^2 : \beta \in \mathcal{B}\}$, and defined test statistics based on the difference between $\widehat{F}_{\widehat{\varepsilon}}(y)$ and the empirical distribution function of $\widehat{\varepsilon}_{i0} = \{Y_i - \widehat{m}(X_i)\}/\widehat{\sigma}_0(X_i)$ where $\widehat{\sigma}_0(x)$ is a smoothed parametric estimator of $\sigma_{\widehat{\beta}}(x)$ and $\widehat{\beta}$ is some suitable estimator of β_0 which is the true value of β under H_0 . Dette et al. (2007) considered in particular a least squares estimator for β_0 .

Second, Neumeyer and Dette (2007) constructed a test for the symmetry of the residual distribution, i.e., $H_0 : F_{\varepsilon}(y) = 1 - F_{\varepsilon}(-y)$ for all $y \in \mathbb{R}$, and defined test statistics based on the difference between $\widehat{F}_{\widehat{\varepsilon}}(y)$ and the empirical distribution function of $\widehat{\varepsilon}_{i0} = -\widehat{\varepsilon}_i$ since $1 - F_{\varepsilon}(-y) = F_{-\varepsilon}(y)$.

Third, Pardo-Fernández et al. (2007) constructed a test for the equality of regression curves. The idea of their testing procedure was to compare, in each population, the empirical distribution functions of the residuals with the same distribution function estimated under the null hypothesis.

Next, Einmahl and Van Keilegom (2008) built a test for the validity of the location-scale model and constructed test statistics based on the difference between $\widehat{F}_{X,\varepsilon}(x, y)$ and $\widehat{F}_X(x)\widehat{F}_{\varepsilon}(y)$ where $\widehat{F}_{X,\varepsilon}$ and $\widehat{F}_X$ are the empirical distribution functions of (X, ε) and X respectively.

Finally, we also refer to the work of Neumeyer and Van Keilegom (2010) and Heuchenne and Van Keilegom (2010) who constructed respectively a test for the additivity of the regression function and a test for a specific parametric form of the error distribution for the same location-scale model $Y = m(X) + \sigma(X)\varepsilon$ but now with a d -dimensional covariate X . Hence, in comparison with the previous mentioned papers, the regression and variance functions are estimated in that context using multivariate local polynomial fits instead of Nadaraya (1964) and Watson (1964) fits.

Neumeyer and Van Keilegom (2010) constructed a test for $H_0 : m(x_1, \dots, x_d) = m_1(x_1) + \dots + m_d(x_d) + c$ for all $(x_1, \dots, x_d) \in \chi$. They defined test statistics based on the difference between $\widehat{F}_{\widehat{\varepsilon}}(y)$ and the empirical distribution function of $\widehat{\varepsilon}_{i0} = \{Y_i - \widehat{m}_0(X_i)\}/\widehat{\sigma}(X_i)$ where $\widehat{m}_0(x_1, \dots, x_d) = \widehat{m}_1(x_1) + \dots + \widehat{m}_d(x_d) + \overline{Y}_n$, $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ and $\widehat{m}_j(x_j)$ are suitable estimators of $m_j(x_j)$ for $j = 1, \dots, d$, see Neumeyer and Van Keilegom (2010) for more details.

Moreover, Heuchenne and Van Keilegom (2010) considered the null hypothesis $H_0 : F_{\varepsilon} \in \{F_{\varepsilon\beta} : \beta \in \mathcal{B}\}$, and defined test statistics based on the difference between $\widehat{F}_{\widehat{\varepsilon}}(y)$ and $F_{\varepsilon\widehat{\beta}}(y)$ where $\widehat{\beta}$ is some suitable estimator of β_0 , like for example a maximum likelihood estimator, see Heuchenne and Van Keilegom (2010).

In conclusion, all these papers can also be carried over to a semiparametric transformation model, which can be subject to future research.

1.4.3 Using the integrated regression function

In this section, we present in detail a second way to construct tests for the parametric form of the regression function in a context of a nonparametric regression model without transformation of the response. Consider again the hypotheses H_0 and H_1 introduced in (1.4.1). The basic idea is now to compare the integrated regression function estimated in a nonparametric way to the integrated regression function estimated under H_0 . The idea of testing the form of the regression function using the integrated regression function has been studied among others by Bierens (1982), Stute (1997) and Escanciano (2006a). We will focus here on these three particular papers who considered the classical nonparametric regression model $Y = m(X) + \varepsilon$, where X is a d -dimensional covariate with compact support $\chi \subset \mathbb{R}^d$, Y is a univariate response variable, the error term ε has zero mean and is independent of X and the regression function $m(x) = E(Y|X = x)$ is unknown. Let (X_i, Y_i) be independent replications of (X, Y) for $i = 1, \dots, n$.

We explain now the general idea behind the tests constructed by Bierens (1982), Stute (1997) and Escanciano (2006a) and we specify afterwards the differences between these three approaches. Consider the following integrated regression function:

$$M(x, \gamma) = \int w(t, x, \gamma) m(t) dF_X(t) = E[w(X, x, \gamma)Y] ,$$

where w is some weighting function that depends on some parameter $\gamma \in \mathbb{R}^{d_\gamma}$. We assume that γ has compact support $\chi_\gamma \subset \mathbb{R}^{d_\gamma}$. The empirical analog of $M(x, \gamma)$ is given by

$$\widehat{M}(x, \gamma) = n^{-1} \sum_{j=1}^n w(X_j, x, \gamma) Y_j .$$

Next, under H_0 , $m \equiv m_{\beta_0}$ where β_0 is the true value of β under H_0 . Then, the integrated regression function becomes

$$M_{\beta_0}(x, \gamma) = \int w(t, x, \gamma) m_{\beta_0}(t) dF_X(t) ,$$

and its empirical analog is given by

$$\widehat{M}_{\beta_0}(x, \gamma) = n^{-1} \sum_{j=1}^n w(X_j, x, \gamma) m_{\beta_0}(X_j) .$$

Hence, the tests of Bierens (1982), Stute (1997) and Escanciano (2006a) will be constructed on the basis of the following residual process:

$$\begin{aligned} R_n(x, \gamma, \beta_0) &= \sqrt{n} \left(\widehat{M}(x, \gamma) - \widehat{M}_{\beta_0}(x, \gamma) \right) \\ &= n^{-1/2} \sum_{j=1}^n w(X_j, x, \gamma) (Y_j - m_{\beta_0}(X_j)) . \end{aligned}$$

Finally, it suffices to replace the unknown parameter β_0 by a suitable estimator, for example by a least squares estimator, i.e.,

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} n^{-1} \sum_{j=1}^n (Y_j - m_{\beta}(X_j))^2 .$$

This gives the following residual process:

$$\begin{aligned} R_n(x, \gamma, \hat{\beta}) &= \sqrt{n} \left(\widehat{M}(x, \gamma) - \widehat{M}_{\hat{\beta}}(x, \gamma) \right) \\ &= n^{-1/2} \sum_{j=1}^n w(X_j, x, \gamma) \left(Y_j - m_{\hat{\beta}}(X_j) \right) . \end{aligned}$$

The main difference between the approaches of Bierens (1982), Stute (1997) and Escanciano (2006a) is the weighting function w that each author used to construct his residual process. The first consistent integrated test was proposed by Bierens (1982) with $w(X, x, \gamma) = \exp(ix^t X)$, where i is the imaginary unit, which gives the residual process

$$R_{n,B}(x, \hat{\beta}) = n^{-1/2} \sum_{j=1}^n \exp(ix^t X_j) \left(Y_j - m_{\hat{\beta}}(X_j) \right) .$$

He also defined the following Cramér-von Mises test statistic:

$$\int R_{n,B}^2(x, \hat{\beta}) \Phi(x) dx ,$$

where Φ is a positive integrating function, for example a d -variate normal density. The limiting distributions of this test statistic under both hypotheses are given in Section 6 in Bierens (1982). Second, the test of Stute (1997) was based on $w(X, x, \gamma) = 1_{\{X \leq x\}}$, which is a component by component indicator and thus on the following residual process:

$$R_{n,S}(x, \hat{\beta}) = n^{-1/2} \sum_{j=1}^n 1_{\{X_j \leq x\}} \left(Y_j - m_{\hat{\beta}}(X_j) \right) .$$

The limiting distribution of this process is given in Corollary 1.3 in Stute (1997). The Kolmogorov-Smirnov and Cramér-von Mises test statistics in Stute (1997) take respectively the following forms:

$$\sup_{x \in \mathcal{X}} \left| R_{n,S}(x, \hat{\beta}) \right| ,$$

and

$$\int_{\mathcal{X}} R_{n,S}^2(x, \hat{\beta}) d\widehat{F}_X(x) .$$

The main advantage of the indicator weighting function considered in Stute (1997) over the exponential weight considered in Bierens (1982) is that it avoids the choice of an arbitrary integrating function Φ in the definition of the Cramér-von Mises test statistic and conversely, the exponential weight has the advantage of being less sensitive to the dimension d than the indicator weight. With the aim of keeping the good properties and avoiding the deficiencies of these two weighting functions, Escanciano (2006a) constructed another test with $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$. Then, Escanciano (2006a) defined the following residual process:

$$R_{n,E}(x, \gamma, \hat{\beta}) = n^{-1/2} \sum_{j=1}^n 1_{\{\gamma^t X_j \leq x\}} \left(Y_j - m_{\hat{\beta}}(X_j) \right) ,$$

and the following Kolmogorov-Smirnov and Cramér-von Mises test statistics:

$$\sup_{(x, \gamma) \in [-\Delta, \Delta] \times \mathbb{S}_d} \left| R_{n,E}(x, \gamma, \hat{\beta}) \right| ,$$

and

$$\int_{[-\Delta, \Delta] \times \mathbb{S}_d} R_{n,E}^2(x, \gamma, \hat{\beta}) d\hat{F}_{n,\gamma}(x) d\gamma ,$$

where $\hat{F}_{n,\gamma}(x)$ is the empirical distribution function of the projected regressor $\{\gamma^t X_i\}$, $d\gamma$ is the uniform density on $\mathbb{S}_d$ which is the unit ball in $\mathbb{R}^d$ and $\Delta = d \max_{1 \leq i \leq d} \sup_{t \in \chi} |t_i|$ and t_i is the i -th component of the vector $t \in \chi$. This particular choice of the integrating measure in the Cramér-von Mises test statistic is mainly motivated by two facts. First, once γ is fixed, x lives in the projected regressor variable's space, which justifies the use of $\hat{F}_{n,\gamma}(x)$ and second, all the directions are in principle equally important which justifies the fact that γ lives in the space $\mathbb{S}_d$. Finally, the limiting distributions of the residual process $R_{n,E}$ and of the test statistics are given in Theorem 2 and Corollary 1 in Escanciano (2006a).

We would also like to mention the work of Stute and Zhu (2002) who used a similar approach as Escanciano (2006a). Stute and Zhu (2002) also considered the weighting function $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$ for model checks of generalized linear models in a iid framework except that they estimated the parameter γ . The approach of Stute and Zhu (2002) is consequently different from that of Escanciano (2006a) since Escanciano (2006a) considered simultaneously all the directions γ in $\mathbb{S}_d$.

More generally, the class of tests of this section is based on the equivalence

$$E(\varepsilon|X) = 0 \text{ a.s.} \iff E(\varepsilon w(X, x, \gamma)) = 0 \quad \forall (x, \gamma) \in \Pi , \quad (1.4.2)$$

where Π is a properly chosen space and $w(\cdot, x, \gamma)$ is a parametric family such that the equivalence (1.4.2) holds. Bierens and Ploberger (1997), Stinchcombe

and White (1998) and Escanciano (2006b) among others propose some primitive conditions on the family of weighting functions $w(\cdot, x, \gamma)$ so that the equivalence (1.4.2) is satisfied, including $w(X, x, \gamma) = \exp(ix^t X)$, $w(X, x, \gamma) = 1_{\{X \leq x\}}$ and $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$, the weighting functions used by Bierens (1982), Stute (1997) and Escanciano (2006a) respectively. Other possibilities are for example $w(X, x, \gamma) = \exp(x^t X)$, $w(X, x, \gamma) = (1 + \exp(c - x^t X))^{-1}$ for some constant c , $w(X, x, \gamma) = \sin(x^t X)$ and $w(X, x, \gamma) = \sin(x^t X) + \cos(x^t X)$. In the context of nonparametric regression without transformation of the response, this class of tests, which is called “the integrated approach”, avoids the use of smoothing methods which is an important advantage.

In Chapter 3, our goal will be to extend this class of tests to the context of semiparametric transformation models of the form (1.2.1). Besides the particular form of the weighting function w used by each author to construct his test, we also observed that the integrated function used to define the Cramér-von Mises test statistic is not always the same. Hence, in Chapter 3, we will construct a residual process based on a general weighting function w and we will define a Cramér-von Mises test statistic based on an arbitrary integrating function Ψ_n . The main advantage of putting a general weighting function w and an arbitrary integrating function Ψ_n in the definition of the Cramér-von Mises test statistic is that we can use in particular the three main approaches of Bierens (1982), Stute (1997) and Escanciano (2006a), and also several other weighting functions, but in a context of a semiparametric transformation model.

Finally, note that many papers in the literature use this integrated approach in other contexts, in time series for example, see among others Bierens (1984) who used the weighting function $w(X, x, \gamma) = \exp(ix^t X)$ to propose a test for the null hypothesis that the expectation of the error conditional on the exogenous variables, all lagged exogenous variables and all lagged dependent variables equals 0 with probability 1 and Koul and Stute (1999) who used the weighting function $w(X, x, \gamma) = 1_{\{X \leq x\}}$ to study a class of tests useful for testing the goodness-of-fit of an autoregressive model. Next, using the weighting function $w(X, x, \gamma) = \exp(ix^t X)$, Bierens (1990) showed that any conditional moment test of functional form of nonlinear regression models can be converted into a chi-square test that is consistent against all deviations from the null hypothesis. Moreover, using the weighting function $w(X, x, \gamma) = 1_{\{X \leq x\}}$, Whang (2000) introduced specification tests of parametric mean-regression models and tested the null hypothesis that the parametric regression function is correctly specified.

1.4.4 Other ways to construct a test

In the literature, besides the approaches using the error distribution function and the integrated regression function presented in Sections 1.4.2 and 1.4.3 respectively, we can also find several other ways to construct a test for the problem (1.4.1). In this section, we will briefly present three additional ap-

proaches, the first is based on estimations of the regression function, the second is based on estimations of the characteristic function and the last is the empirical likelihood ratio tests. However, all these tests are again constructed in the case of a nonparametric regression without transformation of the response and could consequently be extended to the context of semiparametric transformation models of the form (1.2.1). This will not be treated in this thesis and can be subject to future research.

1.4.4.1 Using the regression function

First, consider a nonparametric regression model of the form $Y = m(X) + \varepsilon$, where X is a d -dimensional covariate with compact support $\chi \subset \mathbb{R}^d$, Y is a univariate response variable, the error term ε has zero mean and is independent of X and the regression function $m(x) = E(Y|X = x)$ is unknown. In that context, Härdle and Mammen (1993) constructed a test for the parametric form of the regression function m where the basic idea is to measure the distance between the function m estimated in a nonparametric way to the function m estimated under H_0 . In particular, Härdle and Mammen (1993) considered a L_2 -distance between the nonparametric and the parametric fits and defined the following test statistic:

$$nh^{d/2} \int_{\chi} (\hat{m}(x) - \hat{m}_0(x))^2 \pi(x) dx ,$$

where π is a weighting function, $\hat{m}(x)$ is the Nadaraya (1964) and Watson (1964) estimator of the function m , $\hat{m}_0(x)$ is a smoothed parametric estimator of $m_{\hat{\beta}}(x)$ using the same bandwidth and the same kernel as for the nonparametric estimator $\hat{m}(x)$, and $\hat{\beta}$ is the weighted least squares estimator of β_0 which is the true value of β under H_0 . As mentioned in Section 1.4.2, this smoothed parametric estimator $\hat{m}_0(x)$ has the advantage of having the same bias as its nonparametric counterpart, which implies considerable power improvements for small samples in comparison with the nonsmoothed parametric estimator $m_{\hat{\beta}}(x)$. Note also that Härdle and Mammen (1993) established the limiting distribution of this test statistic in their Proposition 1. We would also like to mention the work of Alcalá et al. (1999) who constructed a similar test as Härdle and Mammen (1993) except that they worked in dimension $d = 1$ and estimated the regression function m and smoothed the parametric estimator $m_{\hat{\beta}}(x)$ by a local polynomial method instead of the Nadaraya (1964) and Watson (1964) method. It is important to notice that the tests constructed by Härdle and Mammen (1993) and Alcalá et al. (1999) can detect alternatives at the rate $n^{-1/2}h^{-d/4}$ which is a disadvantage with respect to the tests based on the error distribution function and the integrated regression function introduced in Sections 1.4.2 and 1.4.3 that can detect alternatives at the rate $n^{-1/2}$.

Instead of estimating the regression function in a nonparametric way and in a parametric way separately and defining the test statistic based on these estimators like Härdle and Mammen (1993) did, other authors based their test for the parametric form on the regression function on estimating the quantity

$$\arg \min_{\beta \in \mathcal{B}} \int_{\mathcal{X}} (m(x) - m_{\beta}(x))^2 dF_X(x) ,$$

where F_X is the distribution function of X . This idea was achieved either by estimating this quantity directly, see for example Dette and Munk (1998) and Dette et al. (2000), or by estimating related quantities, like the difference of the variances under the null and under the alternative, see for example Dette (1999).

1.4.4.2 Using the characteristic function

Consider now the heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$ already introduced in Section 1.4.2, where X is a 1-dimensional covariate, Y is a univariate response variable, the error term ε has zero mean and is independent of X and the regression function $m(x) = E(Y|X = x)$ and the conditional variance function $\sigma^2(x) = V(Y|X = x)$ are unknown. Let (X_i, Y_i) be independent replications of (X, Y) for $i = 1, \dots, n$. We also mentioned in Section 1.4.2 that H_0 in (1.4.1) is true if and only if the distribution functions F_{ε} and F_{ε_0} are the same. The equality of the error distributions can also be interpreted in terms of the characteristic functions. Indeed, if F_{ε} equals F_{ε_0} , the same happens with the corresponding characteristic functions. Hence, in that context, Hušková and Meintanis (2009) constructed a test for (1.4.1) where the basic idea is to compare the error characteristic function $\phi_{\varepsilon}(y) = E[e^{iy\varepsilon}]$ estimated in a nonparametric way to the error characteristic function estimated under H_0 . More precisely, Hušková and Meintanis (2009) estimated the characteristic function of ε in a general context by

$$\hat{\phi}_{\varepsilon}(y) = n^{-1} \sum_{j=1}^n e^{iy\hat{\varepsilon}_j} ,$$

and the characteristic function of ε under H_0 by

$$\hat{\phi}_{\varepsilon_0}(y) = n^{-1} \sum_{j=1}^n e^{iy\hat{\varepsilon}_{j0}} ,$$

where i is the imaginary unit and the definitions of $\hat{\varepsilon}_j$ and $\hat{\varepsilon}_{j0}$ are the same as in Van Keilegom et al. (2008a), see Section 1.4.2, except that Hušková and Meintanis (2009) constructed the residuals under H_0 on the basis of a non smoothed parametric estimator of the regression function. Finally, they defined

the following test statistic based on a L_2 -distance between these two empirical characteristic functions:

$$n \int_{-\infty}^{+\infty} \left| \widehat{\phi}_{\widehat{\varepsilon}}(y) - \widehat{\phi}_{\widehat{\varepsilon}_0}(y) \right|^2 w(y) dy ,$$

where w denotes a weighting function. Hušková and Meintanis (2009) also established the limiting distribution of this test statistic under H_0 and under the local alternative $H_{1n} : m(x) = m_{\beta_0}(x) + n^{-1/2}r(x)$ for all x and some function r in their Theorems 1 and 2 respectively.

In a similar way as for the error distribution function, several other aspects of the location-scale model have been tested in the literature but using the characteristic function. We would like to mention the works of Hušková and Meintanis (2010) and Pardo-Fernández et al. (2015a) who constructed respectively tests for a specific parametric form of the error distribution and for the equality of regression curves. Hušková and Meintanis (2010) constructed a test for $H_0 : F_{\varepsilon} \in \{F_{\varepsilon\beta} : \beta \in \mathcal{B}\}$, which is equivalent to $H_0 : \phi_{\varepsilon} \in \{\phi_{\varepsilon\beta} : \beta \in \mathcal{B}\}$ and defined test statistics based on the difference between $\widehat{\phi}_{\widehat{\varepsilon}}(y)$ and $\phi_{\varepsilon\widehat{\beta}}$ where $\widehat{\beta}$ is some suitable estimator of β_0 . Moreover, the basic idea of the testing procedure in Pardo-Fernández et al. (2015a) is to compare the empirical estimators of the characteristic functions of the regression residuals in each population. Finally, we also refer to the work of Pardo-Fernández et al. (2015b) who constructed test for the equality of conditional variance functions using both estimations of the error distribution function and estimations of the error characteristic function.

1.4.4.3 Empirical likelihood ratio tests

We now present a third approach based on the empirical likelihood ratio to construct tests in nonparametric regression models.

Consider first a nonparametric regression model of the form $Y = m(X) + \varepsilon$, where X is 1-dimensional, Y is a univariate response variable, the error term ε has zero mean and is independent of X and the regression function $m(x) = E(Y|X = x)$ is unknown. Let (X_i, Y_i) be iid replications of (X, Y) and suppose we want to test (1.4.1). In parametric likelihood methods, we assume that the joint distribution of all available data has a known form, apart from one or more unknown parameters. The most popular parametric likelihood methods begin by assuming a normal distribution for the data. For example, if the error term ε in a regression model follows a Gaussian distribution $N(0, \sigma^2)$, it is possible to build a generalized likelihood ratio test for (1.4.1) where the basic idea of the test statistic is to compare the maximum log-likelihood of an observed outcome under the null hypothesis to the maximum log-likelihood of an observed outcome evaluated over the whole parameter space, i.e.,

$$\Lambda_n = l(\widehat{m}, \widehat{\sigma}) - l(m_{\widehat{\beta}}, \widehat{\sigma}_0) ,$$

where

$$l(m, \sigma) = -\frac{n}{2} \log(\sqrt{2\pi}) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \{Y_i - m(X_i)\}^2$$

is the Gaussian log-likelihood, $\hat{m}$ is the Nadaraya (1964) and Watson (1964) estimator of m , $\hat{\beta}$ is the least squares estimator of β_0 which is the true value of β under H_0 , $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \{Y_i - \hat{m}(X_i)\}^2$ is the maximum likelihood estimator of σ^2 and $\hat{\sigma}_0^2 = n^{-1} \sum_{i=1}^n \{Y_i - m_{\hat{\beta}}(X_i)\}^2$ is the maximum likelihood estimator of σ^2 under H_0 . This testing procedure has been studied by Fan et al. (2001) and Fan and Jiang (2007) among others. This generalized likelihood ratio test is an extension of the classical likelihood ratio test where the likelihood function is evaluated in a nonparametric estimator of the regression function under the alternative, or more generally in an appropriate estimator of the function of interest. Finally, an important property of the generalized likelihood ratio tests we would like to mention is that the asymptotic distribution of $-2\Lambda_n$ is a Chi-square distribution that does not depend on nuisance functions under useful assumptions, which is called Wilk's phenomenon. We refer to Fan et al. (2001) and Fan and Jiang (2007) for more details.

However, a problem with parametric likelihood inference is that we might not know which parametric family to use. Indeed there is no reason to suppose that a new encountered data set will belong to the same Gaussian distribution or even to any other studied parametric families. Hence, many papers in the literature have considered nonparametric likelihood inference since they do not require us to specify a family of distributions for the data. In particular, when the empirical distribution of the data plays an important role, we adopt the term "empirical likelihood inference". The empirical likelihood is a technique useful to build a nonparametric likelihood for parameters of interest in a nonparametric or a semiparametric setting but keeping the nice properties of the parametric likelihood, like the Wilk's theorem or the Bartlett correction for example. The Bartlett correction was proved in that context by Chen and Cui (2006) and corresponds to a simple mean adjustment of the likelihood ratio which implies an improvement by one order of magnitude of the approximation of the distribution of this likelihood ratio by the limiting chi-square distribution. We refer to the seminal book of Owen (1998) for more details about the empirical likelihood theory.

In the literature, several papers have used the ideas of empirical likelihood introduced in Owen (1998) to construct likelihood ratio tests. Consider a random vector (X, Y) , where X is a d -variate vector, Y is one-dimensional and $m(x) = E(Y|X = x)$ for $x \in \mathbb{R}^d$. Assume that the data (X_i, Y_i) , $i = 1, \dots, n$, are independent replications of (X, Y) . Van Keilegom et al. (2008b) constructed a test for

$$H_0 : m \in \{\ell(\cdot, \beta, g) : \beta \in \mathcal{B}, g \in \mathcal{G}\}, \quad (1.4.3)$$

where ℓ is some known function, $\mathcal{B} \subset \mathbb{R}^q$ is some compact set and $\mathcal{G}$ is some set

of real valued functions. Testing the parametric form of the regression function is a special case of this general null hypothesis when we consider $m(\cdot) = \ell(\cdot, \beta)$. Other special cases of this general null hypothesis are testing for a generalized linear model, a partial linear model, the selection of explanatory variables, the parametric form of the variance function, the comparison of regression curves, homoscedasticity, conditional independence and autoregressive models among others. Hence, Van Keilegom et al. (2008b) presented a very general procedure for testing different regression models in a unified way. In particular, consider $m(\cdot) = \ell(\cdot, \beta)$ and remind the test for the parametric form of the regression function introduced in Section 1.4.3 based on the following empirical residual process:

$$n^{-1/2} \sum_{j=1}^n W(X_j, x, \gamma) \left[Y_j - \ell(X_j, \hat{\beta}) \right] ,$$

where W is a weighting function, γ is some parameter of dimension d_γ and $\hat{\beta}$ is an appropriate estimator of β , for example a least squares estimator. In that case, note also that the true value β_0 of β under H_0 satisfies

$$E[W(X, x, \gamma)(Y - \ell(X, \beta_0))] = 0 .$$

The main idea of Van Keilegom et al. (2008b) was to combine ideas of empirical regression processes and empirical likelihood to construct their test, i.e., first they extended the above process to a general context including a large number of common regression models and second they used the theory of empirical likelihood for the construction of their test statistic. For the general null hypothesis defined in (1.4.3), the above empirical residual process is extended by

$$R_n(v) = n^{-1/2} \sum_{j=1}^n W(X_j, v, \hat{\beta}, \hat{g}) \left[Y_j - \ell(X_j, \hat{\beta}, \hat{g}) \right] ,$$

where W is a certain appropriate weighting function, $\hat{\beta}$ and $\hat{g}$ are suitable estimators of β and g depending on the particular form of H_0 and v belongs to a compact set V . Next, using Definition 3.2 in Owen (1998) define the nonparametric likelihood of any $d + 1$ -dimensional distribution function F by

$$L(F) = \prod_{i=1}^n F(\{X_i, Y_i\}) ,$$

where $F(\{X_i, Y_i\})$ is the probability of getting exactly the value (X_i, Y_i) when drawing one observation from F . Thus, one consequence of that definition is that $L(F) = 0$ if F is a continuous distribution. Moreover, Owen (1998) proved in his Theorem 3.1 that the nonparametric likelihood of F is maximized for

$F = F_n$ where F_n is the empirical distribution function of F . This gives

$$\begin{aligned}
 \sup\{L(F)\} &= \prod_{i=1}^n F_n(\{X_i, Y_i\}) \\
 &= \prod_{i=1}^n \left[n^{-1} \sum_{j=1}^n 1_{\{X_j=X_i, Y_j=Y_i\}} \right] \\
 &= \prod_{i=1}^n n^{-1} \\
 &= n^{-n}.
 \end{aligned}$$

Using this last result, Van Keilegom et al. (2008b) introduced the following empirical likelihood ratio based on the empirical process $R_n(v)$:

$$\begin{aligned}
 EL(v, \hat{\beta}, \hat{g}) &= \frac{\sup\left\{L(F) : E\left[W(X, v, \hat{\beta}, \hat{g})(Y - \ell(X, \hat{\beta}, \hat{g}))\right] = 0\right\}}{\sup\{L(F)\}} \\
 &= \frac{1}{n^{-n}} \sup\left\{\prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \right. \\
 &\quad \left. \sum_{i=1}^n p_i W(X_i, v, \hat{\beta}, \hat{g}) [Y_i - \ell(X_i, \hat{\beta}, \hat{g})] = 0\right\}.
 \end{aligned}$$

The second equality was obtained since

$$E\left[W(X, v, \hat{\beta}, \hat{g})(Y - \ell(X, \hat{\beta}, \hat{g}))\right] = \sum_{i=1}^n p_i W(X_i, v, \hat{\beta}, \hat{g}) [Y_i - \ell(X_i, \hat{\beta}, \hat{g})],$$

where $p_i \geq 0$ is the probability of getting the value (X_i, Y_i) for $i = 1, \dots, n$. Hence,

$$\begin{aligned}
 EL(v, \hat{\beta}, \hat{g}) &= \sup\left\{n^n \prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \right. \\
 &\quad \left. \sum_{i=1}^n p_i W(X_i, v, \hat{\beta}, \hat{g}) [Y_i - \ell(X_i, \hat{\beta}, \hat{g})] = 0\right\}.
 \end{aligned}$$

This empirical likelihood ratio takes the following particular form when we are interested to test the parametric form of the regression function:

$$\sup\left\{n^n \prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i w(X_i, x, \gamma) [Y_i - m_{\hat{\beta}}(X_i)] = 0\right\},$$

where w is some weighting function that depends on some parameter $\gamma \in \mathbb{R}^{d_\gamma}$, like in Section 1.4.3. Finally, Van Keilegom et al. (2008b) also proposed the following test statistics:

$$\sup_{v \in V} \left[-2 \log EL(v, \hat{\beta}, \hat{g}) \right] ,$$

and

$$-2 \int_V \log EL(v, \hat{\beta}, \hat{g}) d\mu_n(v) ,$$

for some measure μ_n such that $\sup_{v \in V} |\mu_n(v) - \mu(v)| = o_P(1)$ and where the measure μ does not depend on n . Under some conditions, the limiting distributions of these two test statistics are established under H_0 and H_1 , see Theorem 2.1 in Van Keilegom et al. (2008b). We have already mentioned that this test was constructed in a very general context including a large number of common regression models. In particular, Van Keilegom et al. (2008b) verified that the conditions of their Theorem 2.1 are satisfied for testing the parametric form of the regression function, a generalized linear model, the selection of explanatory variables and a partial linear model. Other special cases that could also be treated by the procedure developed in that paper are testing for the parametric form of the variance function, the comparison of regression curves, homoscedasticity, conditional independence and autoregressive models among others.

Finally, note that Hall and Yatchew (2005) presented a different unified approach as the one of Van Keilegom et al. (2008b) for testing a wide class of qualitative hypotheses in regression models. Moreover, we also mention Hjort et al. (2009) who extended the work of Owen (1998) in the presence of both finite and infinite dimensional nuisance parameters and also when the data dimension is high. They investigated the problem of finding conditions under which the properties of the empirical likelihood continue to hold.

1.5 Contents of the thesis

In Chapter 2, based on Colling and Van Keilegom (2016), we will consider a semiparametric transformation model of the form (1.2.1). The goal of this chapter 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 measures 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 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 et al. (2008a). We refer to Section 1.4.2 for more details. The estimated distributions are based on the profile likelihood estimator of θ defined in (1.2.9)

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. Next, a simulation study is carried out to illustrate the performance of our testing procedures as well as the performance of the estimators of the error density function and the error distribution function defined in (1.2.12) and (1.2.13) and that are useful to construct the test statistics. Finally, we apply our test to data on the scattering of sunlight in the atmosphere, where the response Y is the scattering angle at which the polarization of sunlight vanishes and the covariate X is the cube root of a measure of particulate concentration in the atmosphere.

In Chapter 3, based on Colling and Van Keilegom (2017b), we will again consider the semiparametric transformation model defined in (1.2.1), the profile likelihood estimator for the parameter θ defined in (1.2.9) and a local polynomial estimator for m . Our goal is to develop a new test for the parametric form of the regression function m , which has power against local alternatives that converge to the null model at parametric rate, and to compare its performance to that of the test proposed in Chapter 2. The idea of the new test is to compare the integrated regression function estimated in a semiparametric way to the integrated regression function estimated under the null hypothesis. The idea of testing the form of the regression function using the integrated regression function has been studied among others by Bierens (1982), Stute (1997) and Escanciano (2006a). We refer to Section 1.4.3 for more details. We consider two different test statistics, a Kolmogorov-Smirnov and a Cramér-von Mises type statistic, and establish the limiting distributions of these two test statistics under the null hypothesis and under a local alternative. We use the same bootstrap procedure as in Chapter 2 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 procedure, to compare this new test to the previous one and to see under which model conditions which test behaves the best. We also apply both methods on a real data set, where the response Y is ultrasonic response and the covariate X is metal distance.

In Chapter 4, based on Colling and Van Keilegom (2017a), we will consider the nonparametric transformation model defined in (1.3.1). Our goal is to develop two new nonparametric estimators of the transformation Λ , the first one based on the least squares loss and the second one based on the least absolute deviation loss, and to compare their performance with that of the estimators developed by Chiappori et al. (2015). Our proposed estimators are based on an estimator of the conditional distribution of U given X , where U is an appropriate transformation of Y that is uniformly distributed. The main motivation for working with U instead of Y is that, in transformation models, the response Y is often skewed with very long tails, and so kernel smoothing based on Y does not work well. Hence, we expect to obtain better estimators if we pre-transform Y before applying kernel smoothing. We establish the

asymptotic normality of the two proposed estimators. We also carry out a simulation study to illustrate the performance of our estimators, to compare these new estimators with the previous ones of Chiappori et al. (2015) defined in (1.3.2) and (1.3.3) and to see under which model conditions which estimators behave the best.

Chapter 2

Goodness-of-fit tests in semiparametric transformation models using the error distribution function

This chapter is based on the paper

Colling, B. and Van Keilegom, I. (2016). Goodness-of-fit tests in semiparametric transformation models. *Test*, 25(2):291–308.

2.1 Introduction

In this chapter, we consider the semiparametric transformation model defined in (1.2.1), as well as the different assumptions and notations made on it in Section 1.2.1, and we want to construct a test for the parametric form of the regression function m , i.e., we like to test the hypothesis

$$H_0 : m \in \mathcal{M}, \quad (2.1.1)$$

where $\mathcal{M} = \{m_\beta : \beta \in \mathcal{B}\}$ is some parametric class of regression functions and $\mathcal{B} \subset \mathbb{R}^q$. The basic idea of the test introduced in this chapter is to measure a certain distance between the distribution function of the error term ε estimated in a semiparametric way, i.e., in a general context, and the distribution function of ε estimated under the null hypothesis. As a theoretical justification of this testing procedure, we will show that the two error 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 et al. (2008a). Their test was developed for a nonparametric

location-scale model without transforming the response variable. In Section 1.4.2, we explained in detail how Van Keilegom et al. (2008a) constructed their test. In the present chapter, we will see how their ideas and methodology can be carried over to a transformation model.

The chapter is organized as follows. In Section 2.2 we explain in detail the testing procedure and in particular we introduce the definition of the test statistics. Section 2.3 contains the main asymptotic results and in particular the limiting distributions of the proposed test statistics under the null hypothesis and under a local alternative. We also explain how the critical values of these test statistics can be obtained using a bootstrap procedure. In Section 2.4, a simulation study is carried out to illustrate the performance of the estimators of the error density function and of the error distribution function that are needed to construct the test statistics. We also carry out another simulation study to evaluate the performance of our tests. Section 2.5 is devoted to the application of our testing procedures to data on the scattering of sunlight in the atmosphere. Finally, Sections 2.7 and 2.8 contain respectively the technical assumptions and the proofs of the main results.

2.2 The proposed test

2.2.1 Estimation of the model

First, we have explained in detail in Section 1.2.3 how to estimate the parameter θ , and here we will use the profile likelihood estimator introduced by Linton et al. (2008), i.e.,

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

where $\hat{m}(x, \theta)$ and $\hat{f}_{\varepsilon(\theta)}(y)$ are respectively nonparametric estimators of the unknown regression function $m(x, \theta)$ and of the error density function $f_{\varepsilon(\theta)}(y)$. More precisely, here we will estimate the unknown regression function by a local polynomial estimator of degree p (like in Neumeyer and Van Keilegom (2010)), i.e., let $h = (h_1, \dots, h_d)^t$ be a d -dimensional bandwidth vector and let $K_1(u)$ be 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. Then, for an arbitrary point $x = (x_1, \dots, x_d)^t$ in the support χ of X , $\hat{m}(x, \theta) = \hat{b}_0(\theta)$ where $\hat{b}_0(\theta)$ is the first component of the vector $\hat{b}(\theta)$, 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 $0 \leq l \leq p$ factors of the form $X_{ij} - x_j$ for $j = 1, \dots, d$. Moreover, for reasons

of simplicity of notation, we will write $\hat{\theta} \equiv \hat{\theta}_{PL}$ and we will use the notation $\hat{m}(x) = \hat{m}(x, \hat{\theta})$ when there is no ambiguity. Moreover, $\hat{f}_{\varepsilon(\theta)}(y)$ is the following kernel estimator of the error density function:

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

where $\hat{\varepsilon}_i(\theta) = \Lambda_{\theta}(Y_i) - \hat{m}(X_i, \theta)$, k_2 is a kernel and g is a bandwidth.

Note that, we can also use the profile likelihood estimator $\hat{\theta}$ to build the estimated residuals $\hat{\varepsilon}_i = \hat{\varepsilon}_i(\hat{\theta}) = \Lambda_{\hat{\theta}}(Y_i) - \hat{m}(X_i)$ and to define the following estimator of $f_{\varepsilon}(y)$:

$$\hat{f}_{\hat{\varepsilon}}(y) = \frac{1}{ng} \sum_{i=1}^n k_2 \left(\frac{y - \hat{\varepsilon}_i}{g} \right), \quad (2.2.1)$$

We explained in Section 1.2.4 why the error density function f_{ε} plays an important role in regression models. Hence, a suitable estimator of this function is required. In Section 2.4.2, the finite sample performance of $\hat{f}_{\hat{\varepsilon}}(y)$ is investigated through simulations.

It is important to remark that we assume a completely unspecified regression function $m(\cdot)$ which is slightly different from what Linton et al. (2008) assume, since they assume an additive or multiplicative structure on $m(\cdot)$. Moreover, we have estimated this regression function by a local polynomial estimator whereas Linton et al. (2008) used a higher order kernel estimator. This has however no impact on how the profile likelihood estimator of θ is constructed.

Finally, we will prove in Proposition 2.8.2 that the asymptotic representation for $\hat{\theta} - \theta_0$ and the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ obtained by Linton et al. (2008) and introduced in Section 1.2.3 stay valid when $m(\cdot)$ is completely unspecified and is estimated by a local polynomial estimator under appropriate regularity conditions.

2.2.2 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 2.2.1. *Let m be a continuous function. Then, H_0 is valid if and only if the random variables*

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

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

The proof is given in Section 2.8.4. Clearly, when H_0 is true, then $\tilde{\beta}_0 = \beta_0$ where β_0 is the true value of β under H_0 . Remind that $F_{\varepsilon}(y) = F_{\varepsilon(\theta_0)}(y) =$

$P(\Lambda_{\theta_0}(Y) - m(X) \leq y)$, and define $F_{\varepsilon_0}(y) = P(\Lambda_{\theta_0}(Y) - m_{\tilde{\beta}_0}(X) \leq y)$. Next, we explain how to estimate $F_{\varepsilon}(\cdot)$ and $F_{\varepsilon_0}(\cdot)$ in order to construct the test statistics. First, define

$$\hat{F}_{\varepsilon}(y) = n^{-1} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_i \leq y\}} , \quad (2.2.2)$$

where $\hat{\varepsilon}_i = \Lambda_{\hat{\theta}}(Y_i) - \hat{m}(X_i)$ are the semiparametric residuals. We explained in Section 1.2.4 why the error distribution function F_{ε} plays an important role in regression models. Hence, a suitable estimator of this function is required. In Section 2.4.3, the finite sample performance of $\hat{F}_{\varepsilon}(y)$ is investigated through simulations.

Second, estimate $m_{\tilde{\beta}_0}(x)$ by the least squares method for nonlinear regression, i.e., estimate $m_{\tilde{\beta}_0}(x)$ by $m_{\hat{\beta}}(x)$, where $\hat{\beta}$ is a minimizer over $\beta \in \mathcal{B}$ of the expression

$$S_n(\beta) = n^{-1} \sum_{i=1}^n (\Lambda_{\hat{\theta}}(Y_i) - m_{\beta}(X_i))^2 . \quad (2.2.3)$$

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

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

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 $j = 1, \dots, d$. 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 :

$$\hat{F}_{\varepsilon_0}(y) = n^{-1} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_{i0} \leq y\}} , \quad (2.2.4)$$

where $\hat{\varepsilon}_{i0} = \Lambda_{\hat{\theta}}(Y_i) - \hat{m}_{\hat{\beta}}(X_i)$ 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

$$T_{KS} = n^{1/2} \sup_{y \in \mathbb{R}} \left| \hat{F}_{\varepsilon}(y) - \hat{F}_{\varepsilon_0}(y) \right| ,$$

and

$$T_{CM} = n \int \left(\hat{F}_{\varepsilon}(y) - \hat{F}_{\varepsilon_0}(y) \right)^2 d\hat{F}_{\varepsilon}(y) .$$

Next, to study the power of the test statistics, 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.

2.3 Asymptotic results

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

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

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

where

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

is a $(q \times 1)$ -vector and $\beta = (\beta_1, \dots, \beta_q)^t$.

2.3.1 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 $n^{1/2}(\widehat{F}_\varepsilon(\cdot) - \widehat{F}_{\varepsilon_0}(\cdot))$. The identification conditions (I1)-(I2) and the regularity conditions (A1)-(A9) under which the results of this section are valid, can be found in Sections 1.2.2 and 2.7 respectively.

Theorem 2.3.1. *Assume (I1), (I2), (A1)-(A9) and suppose that H_0 holds.*

(i) *Then,*

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

where

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

where $\sup_{y \in \mathbb{R}} |R_n(y)| = o_P(n^{-1/2})$ and $g(X, Y)$ is defined in (1.2.10).

- (ii) Moreover, the process $n^{1/2}(\hat{F}_\varepsilon(y) - \hat{F}_{\hat{\varepsilon}_0}(y))$ ($-\infty < y < +\infty$) converges weakly to $f_\varepsilon(y)\mathbb{W}_\infty$, where $\mathbb{W}_\infty$ is a zero mean normal random variable with variance

$$V(\mathbb{W}_\infty) = E[L^2(X, Y, \theta_0, \beta_0)] .$$

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 negligible 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 $V(\mathbb{W}_\infty)$ 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 2.3.2. *Assume (I1), (I2) and (A1)-(A9). Then, under H_0 ,*

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

2.3.2 Under H_{1n}

First, we define

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

and

$$\tilde{S}_{0n}(\beta) = \sigma^2 + E[(m_\beta(X) - m(X))^2] , \quad (2.3.1)$$

and let $\tilde{\beta}_{0n}$ be a minimizer over $\beta \in \mathcal{B}$ of $\tilde{S}_{0n}(\beta)$, which depends on n under H_{1n} . Similarly to Section 2.3.1, but now under H_{1n} , the following theorem states an asymptotic representation for $\hat{F}_\varepsilon(y) - \hat{F}_{\hat{\varepsilon}_0}(y)$ and gives the limiting distribution of the process $n^{1/2}(\hat{F}_\varepsilon(\cdot) - \hat{F}_{\hat{\varepsilon}_0}(\cdot))$.

Theorem 2.3.3. *Assume (I1), (I2) and (A1)-(A10) and suppose that H_{1n} holds.*

- (i) Then,

$$\hat{F}_\varepsilon(y) - \hat{F}_{\hat{\varepsilon}_0}(y) = f_\varepsilon(y) n^{-1} \sum_{i=1}^n L(X_i, Y_i, \theta_0, \tilde{\beta}_{0n}) + n^{-1/2} f_\varepsilon(y) b_1 + R_n(y) ,$$

where $\sup_{y \in \mathbb{R}} |R_n(y)| = o_P(n^{-1/2})$, $L(X, Y, \theta, \beta)$ is defined in Theorem 2.3.1 and

$$\begin{aligned} b_1 &= - \int \left(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \right)^t dF_X(x) \Omega^{-1} \int r(x) \frac{\partial m_{\beta_0}(x)}{\partial \beta} dF_X(x) \\ &\quad + \int r(x) dF_X(x). \end{aligned}$$

(ii) Moreover, the process $n^{1/2}(\hat{F}_{\varepsilon}(y) - \hat{F}_{\varepsilon_0}(y))$ ($-\infty < y < +\infty$) converges weakly to $f_{\varepsilon}(y)(\mathbb{W}_{\infty} + b_1)$, where $\mathbb{W}_{\infty}$ is the same normal random variable as in Theorem 2.3.1(ii).

Note that the bias term $f_{\varepsilon}(y)b_1$ 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 2.3.4. *Assume (I1), (I2) and (A1)-(A10). Then, under H_{1n} ,*

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

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, see for example the seminal paper by Härdle and Mammen (1993), or any other paper based on their approach. Note however that there are situations in which the random variable $\mathbb{W}_{\infty}$ defined in Theorem 2.3.1 is non degenerate, and the bias term b_1 in Theorem 2.3.3 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 $b_1 = 0$, whereas in general $L(X, Y, \theta_0, \beta_0)$ 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 et al. (2008a) and Pardo-Fernández et al. (2007), among others. Also, remind that our test is consistent in the sense of Theorem 2.2.1.

In order to apply the results of Corollaries 2.3.2 and 2.3.4 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.

2.3.3 Bootstrap

The critical values of the test statistics T_{KS} and T_{CM} will be approximated in practice 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:

$$\check{\varepsilon}_i = \widehat{\varepsilon}_i - \frac{1}{n} \sum_{k=1}^n \widehat{\varepsilon}_k \quad i = 1, \dots, n$$

Let $\check{F}_\varepsilon$ be the empirical distribution of these standardized residuals. Let $\zeta_1^*, \dots, \zeta_n^*$ be bootstrap samples of the errors drawn with replacement from this distribution. Moreover, let $\xi_1, \dots, \xi_n$ be independent normally distributed random variables and independent from the original sample $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. We define the bootstrap errors by $\varepsilon_i^* = \zeta_i^* + b_n \xi_i$, $i = 1, \dots, n$, where b_n is some small bandwidth. We choose here $b_n = 0.1$. Finally, we can see that ε_i^* has a smooth distribution function given by

$$\widetilde{F}_\varepsilon(y) = \frac{1}{n} \sum_{j=1}^n \Phi\left(\frac{y - \check{\varepsilon}_j}{b_n}\right)$$

where Φ is the standard normal distribution function. 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 2.3.1 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 $Y_{ib}^* = \Lambda_{\widehat{\theta}}^{-1}(m_{\widehat{\beta}}(X_{ib}^*) + \varepsilon_{ib}^*)$, $i = 1, \dots, n$, 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} and the test statistics introduced in Chapter 3. We refer to Neumeyer (2009) for the consistency of this bootstrap procedure in the case where one is interested in the distribution of the estimator of the error distribution in a nonparametric location-scale model without transformation of the response.

2.4 Simulations

2.4.1 Estimation of θ , h , g and β in practice

In this section, we explain how we will estimate the parameters θ , h , g and β in practice. For each value of θ , we obtain $h^*(\theta)$ the cross validation bandwidth estimator:

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

where

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

and $k_1(x) = k_2(x) = \frac{3}{4}(1 - x^2) 1_{\{|x| \leq 1\}}$ are the Epanechnikov kernel if we work in dimension $d = 1$. More generally, for $d > 1$, we use the product of d Epanechnikov kernels for the estimator of the regression function. Moreover, we estimate g by $\hat{g}(\theta) = (40\sqrt{\pi})^{1/5} n^{-1/5} \hat{\sigma}_{\hat{\varepsilon}(\theta, h^*(\theta))}$, where $\hat{\sigma}_{\hat{\varepsilon}(\theta, h^*(\theta))}$ is the classical estimator of the standard deviation of the error term $\hat{\varepsilon}(\theta, h^*(\theta)) = \Lambda_\theta(Y) - \hat{m}(X, \theta, h^*(\theta))$, where $\hat{m}(x, \theta, h)$ denotes $\hat{m}(x, \theta)$ constructed with a bandwidth h . Note that we have estimated g by the classical normal reference rule for kernel density estimation. Then, the optimal value of θ is given by

$$\hat{\theta} = \arg \max_{\theta} l_\theta(h^*(\theta), \hat{g}(\theta)) ,$$

where

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

where $\hat{f}_{\varepsilon(\theta), g}(y)$ denotes the estimator $\hat{f}_{\varepsilon(\theta)}(y)$ constructed with a bandwidth g . The estimator of θ is obtained iteratively with the function *optimize* in R over the interval $[\theta_0 - 2, \theta_0 + 2]$. Finally, to estimate β , we minimize the following expression over the interval $[-20, 20]$:

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

Note that the kernels k_1 and k_2 and the bandwidths h and g do not satisfy some of the requirements in assumptions (A1) and (A2). However, we believe that these requirements are sufficient but not strictly necessary, and they help to simplify the technical arguments in the proofs in Sections 2.8 and 3.8.

Also, note that a bandwidth obtained by cross-validation may not be optimal for testing purposes. We will therefore also consider the case where the bandwidth h is fixed by the user in order to compare the results obtained with both methods in Section 2.4.4.

2.4.2 Performance of $\hat{f}_\varepsilon(y)$

In this section, the finite sample performance of the estimator $\hat{f}_\varepsilon(y)$ defined in (2.2.1) is investigated. This is achieved through simulations described hereunder. Consider the following transformation model:

$$\Lambda_{\theta_0}(Y) = b_0 + b_1 X^2 + b_2 \sin(\pi X) + \sigma_e e, \quad (2.4.1)$$

where Λ_θ is the Box and Cox (1964) transformation

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

X is uniformly distributed on the interval $[-1, 1]$ and $\varepsilon = \sigma_e e$ is independent of X . We choose $\theta_0 = 0, 0.5$ or 1 , and for each θ_0 , three different models are considered:

$$\begin{aligned} \text{Model 1:} & \quad b_0 = 6.5, \quad b_1 = 5, \quad \sigma_e = 1.5; \\ \text{Model 2:} & \quad b_0 = 4.5, \quad b_1 = 3.5, \quad \sigma_e = 1; \\ \text{Model 3:} & \quad b_0 = 2.5, \quad b_1 = 2.5, \quad \sigma_e = 0.5. \end{aligned}$$

For each model, $b_2 = b_0 - 3\sigma_e$. We study two simulation settings that correspond to two different distributions of e in (2.4.1): first, a standard normal distribution and, second, a mixture of the normal distributions $N(-1.5, 0.25)$ and $N(1.5, 0.25)$ with equal weights.

In order to avoid negative values of $\Lambda_{\theta_0}(Y_i)$, $i = 1, \dots, n$, these distributions are truncated on $[-3, 3]$ (namely, the corresponding densities are put to zero outside the interval $[-3, 3]$ but their integrals on this support are equal to one).

One hundred samples of sizes $n = 100$ and $n = 200$ are generated. For the estimation of θ_0 and $f_\varepsilon(y)$, we proceed as described in Section 2.4.1, i.e., we obtain

$$\hat{\theta} = \arg \max_{\theta} l_{\theta}(h^*(\theta), \hat{g}(\theta)),$$

where $l_{\theta}(h, g)$, $h^*(\theta)$ and $\hat{g}(\theta)$ are defined in Section 2.4.1. Then, the estimator of $f_\varepsilon(y)$ is given by

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

where $\hat{\varepsilon}(\theta, h^*(\theta))$ is also defined in Section 2.4.1.

Tables 2.1, 2.2, 2.3 and 2.4 show the mean squared error (MSE) of the estimator $\hat{f}_\varepsilon(y)$ of the standardized (pseudo-estimated) error

$$\tilde{\varepsilon} = \frac{\Lambda_{\hat{\theta}}(Y) - \hat{m}_{\hat{\theta}}(X)}{\sigma_e},$$

with known σ_e , for $y = -1$, $y = 0$ and $y = 1$ (respectively $y = -1.5$, $y = -1$, $y = 0$, $y = 1$ and $y = 1.5$) and for the unimodal (respectively bimodal) normal error distribution.

Dividing the residuals by σ_e only aims at comparing models 1, 2 and 3 (modes are the same). In practice, if we would rather construct a standardized version of $\hat{f}_\varepsilon(y)$, a global estimator of σ_e should be introduced in the procedure.

Tables 2.5 and 2.6 show the integrated mean squared error (IMSE) of the estimator $\hat{f}_\varepsilon(\cdot)$ for both error distributions, where the integration is done over the interval $[-3, 3]$.

As expected, in both cases, the estimation is better for the normal density than for the mixture of two normals, and the estimation improves when n increases, and in most cases, when σ_e decreases. In particular, this can be observed from Tables 2.5 and 2.6.

The limiting case $\theta_0 = 0$ (the logarithmic transformation) seems to be more easily captured, especially when the error is normally distributed.

In general, we observe from Tables 2.1, 2.2, 2.3, 2.4 that the estimation is poorer near local maxima and minima of the density, which is not uncommon for kernel smoothing methods. This also suggests that the choice of the smoothing parameters is important and should be the object of further investigation.

Model	θ_0		$\widehat{f}_\varepsilon(y)$		
			$y = -1$	$y = 0$	$y = 1$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_0 = 1.5$	$\theta_0 = 0$	Bias	-.0365	.0484	-.0394
		Var	.0010	.0204	.0005
		MSE	.0023	.0227	.0020
	$\theta_0 = 0.5$	Bias	-.0574	.1477	-.0875
		Var	.0080	.1796	.0071
		MSE	.0113	.2014	.0146
	$\theta_0 = 1$	Bias	-.0832	.2077	-.0874
		Var	.0045	.1841	.0046
		MSE	.0114	.2272	.0123
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_0 = 1$	$\theta_0 = 0$	Bias	-.0352	.0451	-.0382
		Var	.0009	.0179	.0004
		MSE	.0022	.0199	.0019
	$\theta_0 = 0.5$	Bias	-.0587	.1041	-.0641
		Var	.0023	.0708	.0024
		MSE	.0058	.0817	.0065
	$\theta_0 = 1$	Bias	-.0685	.1344	-.0750
		Var	.0029	.1115	.0033
		MSE	.0076	.1295	.0089
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_0 = 0.5$	$\theta_0 = 0$	Bias	-.0303	.0271	-.0363
		Var	.0005	.0114	.0002
		MSE	.0014	.0122	.0015
	$\theta_0 = 0.5$	Bias	-.04641	.0626	-.0515
		Var	.0013	.0402	.0010
		MSE	.0035	.0441	.0037
	$\theta_0 = 1$	Bias	-.0538	.0855	-.0602
		Var	.0021	.0683	.0020
		MSE	.0050	.0756	.0056

Table 2.1: $MSE(\widehat{f}_\varepsilon(y))$ for different models, values of y and $n = 100$, when $f_\varepsilon(\cdot)$ is a standard normal density.

Model	θ_0		$\widehat{f_\varepsilon}(y)$		
			$y = -1$	$y = 0$	$y = 1$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_0 = 1.5$	$\theta_0 = 0$	Bias	-.0385	.0755	-.0288
		Var	.0009	.0130	.0006
		MSE	.0024	.0187	.0014
	$\theta_0 = 0.5$	Bias	-.0522	.1138	-.0513
		Var	.0028	.0565	.0031
		MSE	.0056	.0694	.0058
	$\theta_0 = 1$	Bias	-.0608	.1340	-.0595
		Var	.0034	.0967	.0039
		MSE	.0071	.1147	.0075
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_0 = 1$	$\theta_0 = 0$	Bias	-.0363	.0735	-.0264
		Var	.0009	.0110	.0006
		MSE	.0022	.0164	.0012
	$\theta_0 = 0.5$	Bias	-.0469	.0994	-.0430
		Var	.0022	.0389	.0025
		MSE	.0044	.0488	.0043
	$\theta_0 = 1$	Bias	-.0579	.1267	-.0544
		Var	.0031	.0708	.0037
		MSE	.0065	.0868	.0067
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_0 = 0.5$	$\theta_0 = 0$	Bias	-.0341	.0697	-.0240
		Var	.0006	.0076	.0004
		MSE	.0018	.0124	.0009
	$\theta_0 = 0.5$	Bias	-.0402	.0842	-.0333
		Var	.0012	.0195	.0012
		MSE	.0028	.0260	.0023
	$\theta_0 = 1$	Bias	-.0444	.0912	-.0396
		Var	.0018	.0308	.0020
		MSE	.0037	.0391	.0035

Table 2.2: $MSE(\widehat{f_\varepsilon}(y))$ for different models, values of y and $n = 200$, when $f_\varepsilon(\cdot)$ is a standard normal density.

Model	θ_0		$\widehat{f}_\varepsilon(y)$				
			$y = -1.5$	$y = -1$	$y = 0$	$y = 1$	$y = 1.5$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_0 = 1.5$	$\theta_0 = 0$	Bias	-.2014	-.0415	.1615	-.0108	-.1789
		Var	.0002	.0005	.0005	.0014	.0004
		MSE	.0407	.0022	.0266	.0015	.0324
	$\theta_0 = 0.5$	Bias	-.2062	-.0211	.1831	.0147	-.1894
		Var	.0006	.0021	.0024	.0038	.0015
		MSE	.0431	.0025	.0359	.0040	.0373
	$\theta_0 = 1$	Bias	-.2127	-.0183	.1883	.0161	-.1983
		Var	.0014	.0029	.0037	.0047	.0026
		MSE	.0466	.0032	.0391	.0050	.0419
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_0 = 1$	$\theta_0 = 0$	Bias	-.2026	-.0421	.1636	-.0096	-.1799
		Var	.0002	.0005	.0006	.0014	.0003
		MSE	.0412	.0023	.0274	.0014	.0327
	$\theta_0 = 0.5$	Bias	-.2044	-.0274	.1779	.0091	-.1849
		Var	.0004	.0016	.0018	.0031	.0010
		MSE	.0422	.0023	.0334	.0032	.0351
	$\theta_0 = 1$	Bias	-.2091	-.0224	.1847	.0138	-.1922
		Var	.0008	.0022	.0027	.0038	.0018
		MSE	.0445	.0027	.0368	.0040	.0387
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_0 = 0.5$	$\theta_0 = 0$	Bias	-.2066	-.0443	.1717	-.0017	-.1794
		Var	.0001	.0005	.0006	.0014	.0003
		MSE	.0428	.0024	.0301	.0014	.0325
	$\theta_0 = 0.5$	Bias	-.2093	-.0390	.1779	.0047	-.1848
		Var	.0005	.0013	.0015	.0027	.0009
		MSE	.0443	.0028	.0332	.0027	.0350
	$\theta_0 = 1$	Bias	-.2117	-.0335	.1845	.0107	-.1896
		Var	.0007	.0018	.0024	.0035	.0014
		MSE	.0456	.0029	.0365	.0036	.0374

Table 2.3: $MSE(\widehat{f}_\varepsilon(y))$ for different models, values of y and $n = 100$, when $f_\varepsilon(\cdot)$ is a mixture of two normal densities ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights.

Model	θ_0		$\widehat{f}_\varepsilon(y)$				
			$y = -1.5$	$y = -1$	$y = 0$	$y = 1$	$y = 1.5$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_0 = 1.5$	$\theta_0 = 0$	Bias	-.1272	.0467	.1093	-.0818	-.1914
		Var	.0001	.0010	.0001	.0009	.0007
		MSE	.0163	.0032	.0121	.0076	.0373
	$\theta_0 = 0.5$	Bias	-.1363	.0614	.1163	-.0624	-.1853
		Var	.0004	.0029	.0005	.0032	.0014
		MSE	.0190	.0067	.0140	.0071	.0357
	$\theta_0 = 1$	Bias	-.01439	.0661	.1197	-.0540	-.1862
		Var	.0009	.0039	.0007	.0046	.0016
		MSE	.0216	.0083	.0150	.0076	.0363
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_0 = 1$	$\theta_0 = 0$	Bias	-.1268	.0480	.1101	-.0802	-.1903
		Var	.0001	.0008	.0001	.0010	.0006
		MSE	.0162	.0031	.0122	.0074	.0368
	$\theta_0 = 0.5$	Bias	-.1322	.0571	.1142	-.0685	-.1862
		Var	.0002	.0022	.0003	.0023	.0011
		MSE	.0177	.0054	.0134	.0070	.0358
	$\theta_0 = 1$	Bias	-.1377	.0608	.1166	-.0621	-.1863
		Var	.0005	.0031	.0005	.0034	.0014
		MSE	.0195	.0068	.0141	.0073	.0361
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_0 = 0.5$	$\theta_0 = 0$	Bias	-.1268	.0476	.1112	-.0797	-.1891
		Var	.00003	.0006	.0001	.0006	.0004
		MSE	.0161	.0029	.0125	.0070	.0362
	$\theta_0 = 0.5$	Bias	-.1307	.0519	.1133	-.0732	-.1877
		Var	.0001	.0015	.0003	.0016	.0008
		MSE	.0172	.0042	.0131	.0069	.0360
	$\theta_0 = 1$	Bias	-.1324	.0546	.1148	-.0700	-.1869
		Var	.0002	.0019	.0003	.0019	.0010
		MSE	.0177	.0048	.0135	.0068	.0359

Table 2.4: $MSE(\widehat{f}_\varepsilon(y))$ for different models, values of y and $n = 200$, when $f_\varepsilon(\cdot)$ is a mixture of two normal densities ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights.

Model	θ_0	$n = 100$	$n = 200$
$b_0 = 6.5$	$\theta_0 = 0$	.0036	.0030
$b_1 = 5$	$\theta_0 = 0.5$	.0189	.0081
$\sigma_0 = 1.5$	$\theta_0 = 1$	.0197	.0114
$b_0 = 4.5$	$\theta_0 = 0$	.0032	.0026
$b_1 = 3.5$	$\theta_0 = 0.5$	.0097	.0061
$\sigma_0 = 1$	$\theta_0 = 1$	.0134	.0095
$b_0 = 2.5$	$\theta_0 = 0$	.0022	.0021
$b_1 = 2.5$	$\theta_0 = 0.5$	.0059	.0039
$\sigma_0 = 0.5$	$\theta_0 = 1$	.0086	.0053

Table 2.5: $IMSE(\hat{f}_{\varepsilon})$ for different models and sample sizes, when $f_{\varepsilon}(\cdot)$ is a standard normal density.

Model	θ_0	$n = 100$	$n = 200$
$b_0 = 6.5$	$\theta_0 = 0$	.0134	.0099
$b_1 = 5$	$\theta_0 = 0.5$	.0184	.0119
$\sigma_0 = 1.5$	$\theta_0 = 1$	.0208	.0134
$b_0 = 4.5$	$\theta_0 = 0$	.0137	.0099
$b_1 = 3.5$	$\theta_0 = 0.5$	.0168	.0110
$\sigma_0 = 1$	$\theta_0 = 1$	.0190	.0121
$b_0 = 2.5$	$\theta_0 = 0$	.0145	.0097
$b_1 = 2.5$	$\theta_0 = 0.5$	.0164	.0105
$\sigma_0 = 0.5$	$\theta_0 = 1$	.0182	.0108

Table 2.6: $IMSE(\hat{f}_{\varepsilon})$ for different models and sample sizes, when $f_{\varepsilon}(\cdot)$ is a mixture of two normal densities ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights.

2.4.3 Performance of $\hat{F}_{\varepsilon}(y)$

In this section, the finite sample performance of the estimator $\hat{F}_{\varepsilon}(y)$ defined in (2.2.2) is investigated through simulations.

We consider exactly the same simulated model as in Section 2.4.2, i.e., we consider model (2.4.1) where Λ_{θ} is the Box and Cox (1964) transformation, X is uniformly distributed on the interval $[-1, 1]$ and $\varepsilon = \sigma_e e$ where e represents first a standard normal distribution and, second, a mixture of the normal distributions $N(-1.5, 0.25)$ and $N(1.5, 0.25)$ with equal weights. We again choose $\theta_0 = 0, 0.5$ or 1 , and for each θ_0 , models 1 to 3 are considered.

One hundred samples of sizes $n = 100$ and $n = 200$ are generated. For the estimation of θ_0 and $F_{\varepsilon}(t)$, we proceed as described in Section 2.4.1, i.e., we

obtain

$$\hat{\theta} = \arg \max_{\theta} l_{\theta}(h^*(\theta), \hat{g}(\theta)) ,$$

where $l_{\theta}(h, g)$, $h^*(\theta)$ and $\hat{g}(\theta)$ are defined in Section 2.4.1. Then, the estimator of $F_{\varepsilon}(y)$ is given by

$$\hat{F}_{\varepsilon}(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{\hat{\varepsilon}_i(\hat{\theta}, h^*(\hat{\theta})) \leq y\}} ,$$

where $\hat{\varepsilon}(\theta, h^*(\theta))$ is also defined in Section 2.4.1.

Figures 2.1 and 2.2 show realizations of the estimator $\hat{F}_{\varepsilon}(t)$, the above empirical estimator but based on standardized residuals

$$\tilde{e}_i = \frac{\hat{\varepsilon}_i(\hat{\theta}, h^*(\hat{\theta}))}{\sigma_e}$$

instead of $\hat{\varepsilon}_i(\hat{\theta}, h^*(\hat{\theta}))$, $i = 1, \dots, n$ when $\theta_0 = 0$, $\theta_0 = 0.5$ and when the error distribution is normal and it is a mixture of two normals.

Tables 2.7, 2.8, 2.9 and 2.10 show the bias, the variance (Var) and the mean squared error (MSE) of $\hat{F}_{\varepsilon}(y)$ for these two error distributions and for a number of values of y . The reason for dividing the residuals by σ_e is the same as in Section 2.4.2. Moreover, Tables 2.11 and 2.12 show the integrated mean squared error (IMSE) of $\hat{F}_{\varepsilon}(y)$ for both assumed error distributions.

As expected, we can observe, in particular from Tables 2.11 and 2.12 that the estimation improves for sample sizes going from $n = 100$ to $n = 200$ and is better for the normal error density than for the mixture. These tables also suggest that a larger σ_e globally leads to worse results.

In these simulated examples, the best results are obtained for the logarithmic transformation. This is intuitively clear, because the shape of the logarithmic function is very different from a power function. Indeed, the range of $\log(y)$ equals $(-\infty, +\infty)$ and $\log(y)$ is very steep close to $y = 0$, while $(y^{\theta} - 1)/\theta$ takes values from $-1/\theta$ to $+\infty$ for a given $\theta > 0$ and is less steep close to $y = 0$. Therefore, if the logarithm is the true transformation, then it should be relatively easy to detect. This is in line with the findings in Linton et al. (2008). Indeed, 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 $\theta_0 = 0$. This suggests that the estimations of $\hat{\varepsilon}_i$ and consequently the estimations of $\hat{F}_{\varepsilon}(y)$ are less accurate when θ_0 increases from 0 to 1 since these estimations and values depend on $\hat{\theta}$.

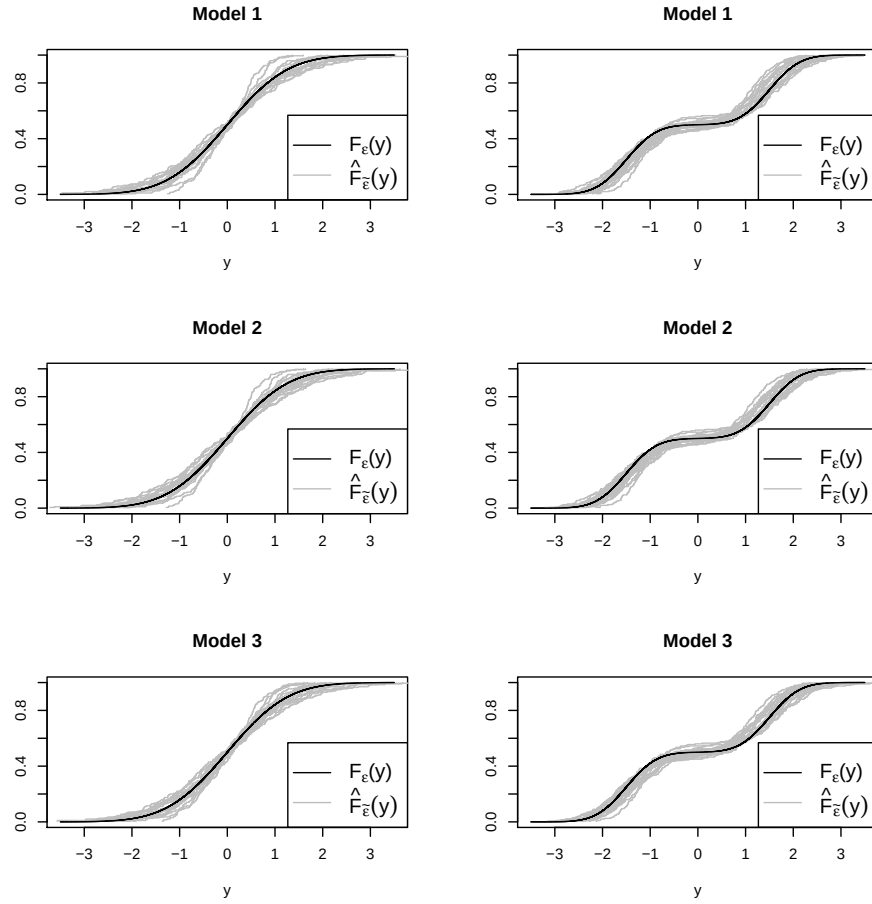


Figure 2.1: Realizations of $\hat{F}_\epsilon(y)$ for $n = 200$ and $\theta_0 = 0$, when the distribution of e is either a standard normal (on the left) or a mixture of two normal distributions ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights (on the right). The first row corresponds to model 1, the second row to model 2 and the third row to model 3.

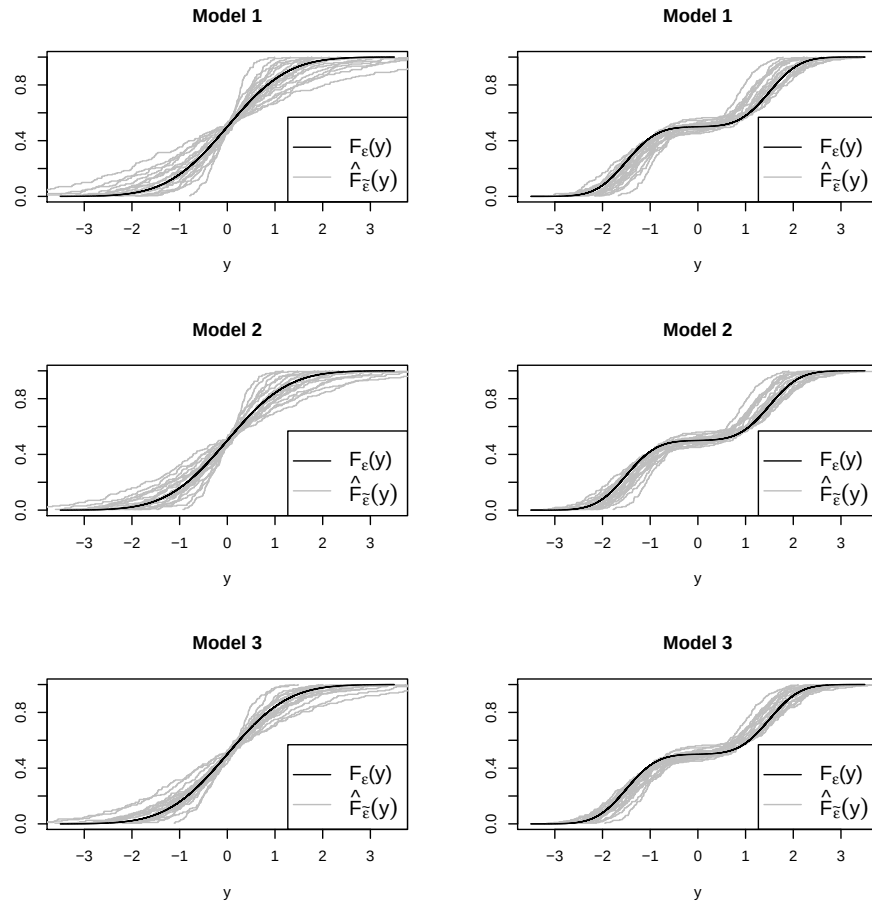


Figure 2.2: Realizations of $\hat{F}_\epsilon(y)$ for $n = 200$ and $\theta_0 = 0.5$, when the distribution of e is either a standard normal (on the left) or a mixture of two normal distributions ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights (on the right). The first row corresponds to model 1, the second row to model 2 and the third row to model 3.

Model	θ_0		$\widehat{F}_{\widehat{\varepsilon}}(y)$		
			$y = -1$	$y = 0$	$y = 1$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_e = 1.5$	$\theta_0 = 0$	Bias	-2.13	-0.74	2.33
		Var	54.86	10.74	60.90
		MSE	59.38	11.29	66.31
	$\theta_0 = 0.5$	Bias	-2.20	-0.88	2.57
		Var	128.71	10.73	126.17
		MSE	133.53	11.51	132.75
	$\theta_0 = 1$	Bias	-2.38	-0.87	2.90
		Var	152.11	10.98	146.01
		MSE	157.75	11.74	154.39
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_e = 1$	$\theta_0 = 0$	Bias	-2.25	-0.74	2.34
		Var	49.29	11.27	52.96
		MSE	54.33	11.81	58.41
	$\theta_0 = 0.5$	Bias	-1.86	-0.75	2.07
		Var	112.15	11.60	107.15
		MSE	115.59	12.17	111.42
	$\theta_0 = 1$	Bias	-1.54	-0.76	2.08
		Var	139.96	11.42	135.58
		MSE	142.35	11.99	139.89
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_e = 0.5$	$\theta_0 = 0$	Bias	-1.48	-0.52	1.29
		Var	40.46	10.98	41.96
		MSE	42.64	11.25	43.62
	$\theta_0 = 0.5$	Bias	-1.55	-0.46	1.14
		Var	78.60	11.44	92.06
		MSE	80.99	11.65	93.35
	$\theta_0 = 1$	Bias	-1.17	-0.58	1.25
		Var	100.62	11.70	103.77
		MSE	101.98	12.04	105.33

Table 2.7: $\text{Bias}(\widehat{F}_{\widehat{\varepsilon}}(y))$ ($\times 10^2$), $\text{Var}(\widehat{F}_{\widehat{\varepsilon}}(y))$ ($\times 10^4$) and $MSE(\widehat{F}_{\widehat{\varepsilon}}(y))$ ($\times 10^4$) for different models, values of y and $n = 100$, when e has a standard normal density

Model	θ_0		$\widehat{F}_{\widehat{e}}(y)$		
			$y = -1$	$y = 0$	$y = 1$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_e = 1.5$	$\theta_0 = 0$	Bias	-0.63	-0.68	0.66
		Var	35.09	4.18	34.56
		MSE	35.49	4.63	34.99
	$\theta_0 = 0.5$	Bias	-0.40	-0.61	0.48
		Var	75.15	4.22	77.49
		MSE	75.31	4.59	77.73
	$\theta_0 = 1$	Bias	-0.11	-0.59	0.31
		Var	101.49	4.65	96.59
		MSE	101.50	5.00	96.68
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_e = 1$	$\theta_0 = 0$	Bias	-0.78	-0.64	0.62
		Var	27.26	4.31	30.33
		MSE	27.87	4.71	30.72
	$\theta_0 = 0.5$	Bias	-0.47	-0.68	0.54
		Var	62.40	4.14	61.93
		MSE	62.62	4.59	62.21
	$\theta_0 = 1$	Bias	-0.64	-0.64	0.65
		Var	88.93	4.29	85.20
		MSE	89.34	4.70	85.62
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_e = 0.5$	$\theta_0 = 0$	Bias	-0.96	-0.69	0.71
		Var	21.11	3.85	22.93
		MSE	22.02	4.32	23.43
	$\theta_0 = 0.5$	Bias	-0.91	-0.72	0.65
		Var	45.07	3.86	46.43
		MSE	45.89	4.37	46.85
	$\theta_0 = 1$	Bias	-0.78	-0.76	0.55
		Var	56.34	3.81	56.16
		MSE	56.95	4.38	56.46

Table 2.8: $\text{Bias}(\widehat{F}_{\widehat{e}}(y))$ ($\times 10^2$), $\text{Var}(\widehat{F}_{\widehat{e}}(y))$ ($\times 10^4$) and $\text{MSE}(\widehat{F}_{\widehat{e}}(y))$ ($\times 10^4$) for different models, values of y and $n = 200$, when e has a standard normal density

Model	θ_0		$\hat{F}_{\hat{e}}(y)$				
			$y = -1.5$	$y = -1$	$y = 0$	$y = 1$	$y = 1.5$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_e = 1.5$	$\theta_0 = 0$	Bias	-4.83	-7.14	-0.09	5.45	3.84
		Var	57.44	27.40	20.95	25.03	49.03
		MSE	80.70	78.34	20.96	54.70	63.77
	$\theta_0 = 0.5$	Bias	-8.33	-11.09	-0.14	9.43	7.85
		Var	106.59	83.41	20.55	87.30	101.71
		MSE	175.98	206.34	20.57	176.18	163.33
	$\theta_0 = 1$	Bias	-9.03	-12.37	-0.09	10.37	8.30
		Var	128.39	103.06	20.10	115.22	138.46
		MSE	209.93	256.01	20.11	222.70	207.35
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_e = 1$	$\theta_0 = 0$	Bias	-4.92	-7.45	-0.16	5.78	4.00
		Var	52.07	33.71	20.04	30.88	46.08
		MSE	76.28	89.17	20.06	64.25	62.08
	$\theta_0 = 0.5$	Bias	-7.29	-9.96	-0.18	8.38	6.85
		Var	90.41	67.51	20.84	68.86	81.26
		MSE	143.55	166.66	20.87	139.04	128.18
	$\theta_0 = 1$	Bias	-8.32	-11.09	-0.18	9.56	7.73
		Var	109.51	85.19	20.67	91.99	102.16
		MSE	178.74	208.12	20.71	183.33	161.91
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_e = 0.5$	$\theta_0 = 0$	Bias	-5.73	-8.30	-0.22	7.00	5.39
		Var	54.40	34.58	21.43	36.19	49.72
		MSE	87.23	103.43	21.47	85.15	78.77
	$\theta_0 = 0.5$	Bias	-6.81	-9.63	-0.30	8.14	6.29
		Var	82.54	59.52	21.04	53.56	75.14
		MSE	128.92	152.21	21.13	119.78	114.70
	$\theta_0 = 1$	Bias	-7.71	-10.80	-0.25	9.35	7.31
		Var	99.68	73.63	21.32	72.65	96.30
		MSE	159.13	190.21	21.38	160.02	149.73

Table 2.9: Bias($\hat{F}_{\hat{e}}(y)$) ($\times 10^2$), Var($\hat{F}_{\hat{e}}(y)$) ($\times 10^4$) and MSE($\hat{F}_{\hat{e}}(y)$) ($\times 10^4$) for different models, values of y and $n = 100$, when the distribution of e is a mixture of two normal densities ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights

Model	θ_0		$\widehat{F}_{\varepsilon}(y)$				
			$y = -1.5$	$y = -1$	$y = 0$	$y = 1$	$y = 1.5$
$b_0 = 6.5$ $b_1 = 5$ $\sigma_e = 1.5$	$\theta_0 = 0$	Bias	-2.47	-3.93	0.01	3.22	1.82
		Var	37.46	12.96	10.34	11.20	28.77
		MSE	43.53	28.42	10.34	21.58	32.06
	$\theta_0 = 0.5$	Bias	-4.69	-5.88	-0.02	4.72	4.43
		Var	81.25	33.96	10.38	32.59	73.38
		MSE	103.20	68.51	10.38	54.89	93.01
	$\theta_0 = 1$	Bias	-5.57	-6.80	-0.04	5.45	5.45
		Var	105.76	51.05	10.32	47.79	97.34
		MSE	136.73	97.26	10.32	77.46	127.04
$b_0 = 4.5$ $b_1 = 3.5$ $\sigma_e = 1$	$\theta_0 = 0$	Bias	-2.65	-4.20	0.04	3.42	1.95
		Var	34.76	13.85	10.40	15.05	28.90
		MSE	41.76	31.46	10.40	26.76	32.70
	$\theta_0 = 0.5$	Bias	-4.18	-5.19	0.02	4.25	3.76
		Var	66.39	24.66	10.32	24.88	55.59
		MSE	83.86	51.62	10.32	42.92	69.69
	$\theta_0 = 1$	Bias	-4.73	-6.06	0.00	4.91	4.40
		Var	86.47	36.52	10.39	37.13	77.11
		MSE	108.80	73.27	10.39	61.21	96.47
$b_0 = 2.5$ $b_1 = 2.5$ $\sigma_e = 0.5$	$\theta_0 = 0$	Bias	-2.73	-4.42	-0.07	3.72	2.52
		Var	26.57	10.74	10.01	10.44	22.37
		MSE	34.00	30.30	10.02	24.26	28.70
	$\theta_0 = 0.5$	Bias	-3.60	-4.91	-0.07	4.17	3.26
		Var	48.62	17.48	10.07	16.90	38.65
		MSE	61.54	41.61	10.08	34.27	49.28
	$\theta_0 = 1$	Bias	-4.10	-5.28	-0.08	4.45	3.79
		Var	57.91	21.34	10.01	21.46	47.42
		MSE	74.68	49.24	10.02	41.28	61.78

Table 2.10: $\text{Bias}(\widehat{F}_{\varepsilon}(y)) (\times 10^2)$, $\text{Var}(\widehat{F}_{\varepsilon}(y)) (\times 10^4)$ and $\text{MSE}(\widehat{F}_{\varepsilon}(y)) (\times 10^4)$ for different models, values of y and $n = 200$, when the distribution of e is a mixture of two normal distributions ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights

Model	θ_0	$n = 100$	$n = 200$
$b_0 = 6.5$	$\theta_0 = 0$	28.20	15.78
$b_1 = 5$	$\theta_0 = 0.5$	82.73	36.80
$\sigma_e = 1.5$	$\theta_0 = 1$	95.59	56.26
$b_0 = 4.5$	$\theta_0 = 0$	25.13	13.04
$b_1 = 3.5$	$\theta_0 = 0.5$	59.91	30.20
$\sigma_e = 1$	$\theta_0 = 1$	80.95	49.67
$b_0 = 2.5$	$\theta_0 = 0$	20.54	10.25
$b_1 = 2.5$	$\theta_0 = 0.5$	43.77	20.59
$\sigma_e = 0.5$	$\theta_0 = 1$	56.73	26.03

Table 2.11: $IMSE(\hat{F}_{\tilde{e}})$ ($\times 10^4$) for different models and sample sizes, when e has a standard normal density

Model	θ_0	$n = 100$	$n = 200$
$b_0 = 6.5$	$\theta_0 = 0$	35.86	17.80
$b_1 = 5$	$\theta_0 = 0.5$	80.74	38.92
$\sigma_e = 1.5$	$\theta_0 = 1$	105.91	51.24
$b_0 = 4.5$	$\theta_0 = 0$	37.23	17.19
$b_1 = 3.5$	$\theta_0 = 0.5$	66.28	30.42
$\sigma_e = 1$	$\theta_0 = 1$	83.10	40.66
$b_0 = 2.5$	$\theta_0 = 0$	43.50	14.87
$b_1 = 2.5$	$\theta_0 = 0.5$	65.54	23.19
$\sigma_e = 0.5$	$\theta_0 = 1$	81.86	27.40

Table 2.12: $IMSE(\hat{F}_{\tilde{e}})$ ($\times 10^4$) for different models and sample sizes, when the distribution of e is a mixture of two normal distributions ($N(-1.5, 0.25)$, $N(1.5, 0.25)$) with equal weights

2.4.4 Performance of the test in dimension 1

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 again the Box and Cox (1964) transformation.

Moreover, $X_1, \dots, X_n$ are independent, of dimension $d = 1$ and uniformly distributed on $[0, 1]$, and $\varepsilon_1, \dots, \varepsilon_n$ 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 $n = 200$.

For the estimation of θ , h , g and β , we proceed as explained in Section 2.4.1. As mentioned in that section, we consider here the case where the bandwidth h is obtained by cross-validation and also where it is fixed by the user in order to compare the results obtained with both methods. We choose here $h = 0.1$, $h = 0.15$ and $h = 0.2$.

Finally, the critical values of the test statistics T_{KS} and T_{CM} will be approximated by means of the bootstrap procedure described in Section 2.3.3 with $B = 250$.

Table 2.13 shows the percentage of rejection under the null hypothesis and under the different deviations $c(x)$ we have introduced above when the bandwidth h is obtained by cross-validation. Table 2.14 shows respectively the percentage of rejection under the null hypothesis and under different deviations $c(x)$ when the bandwidth h is fixed by the user. These percentages of rejection are obtained with the test statistics T_{KS} and T_{CM} for 500 samples. The nominal level is 5%.

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 slightly 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 et al. (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 $\theta_0 = 0$. Finally, 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 .

Moreover, under the null hypothesis and when the bandwidth h is fixed, the best results are in majority obtained for $h = 0.1$, both for T_{KS} and T_{CM} . Under the alternative, the power decreases when h increases and especially for $\theta_0 = 0.5$ and $\theta_0 = 1$, which suggests that among the three tested values of the bandwidth, $h = 0.1$ is the most adapted value for this test. Finally, we notice that sometimes the power is better when the bandwidth h is estimated by cross-validation, for example when $c(x) = 5x^2$, and sometimes the power is better when h is fixed, for example when $c(x) = 7.5x^2$ and $h = 0.1$.

Finally, the only alternative estimator of θ_0 that has been studied rigorously in the literature, is the mean squared distance from independence estimator introduced by Linton et al. (2008) and defined in (1.2.11). We have redone the simulations leading to Table 2.13, but by using the latter estimator instead of the profile likelihood estimator. The results are given in Table 2.15 for the quadratic alternative, and they show that the use of the mean squared distance from independence estimator of θ gives lower power than if we use the profile

likelihood estimator of θ , whereas the behavior under H_0 is comparable.

$c(x)$	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	5.2	7.8	5.8	8.0	6.6	7.4
$5x^2$	100.0	100.0	57.4	60.8	13.4	14.4
$7.5x^2$	99.6	99.6	96.4	96.4	76.0	77.2
$10x^2$	99.0	99.0	96.6	96.0	87.2	86.4
$0.5 \exp(x)$	30.8	38.6	17.2	19.0	10.0	12.6
$\exp(x)$	86.8	92.0	36.4	39.8	20.2	22.6
$2 \exp(x)$	99.8	99.8	68.0	69.2	45.8	46.6
$3 \exp(x)$	98.6	98.4	72.8	72.0	54.8	56.2
$0.25 \sin(2\pi x)$	17.0	19.6	11.2	13.0	11.8	14.2
$0.5 \sin(2\pi x)$	29.2	32.4	19.8	22.0	18.8	20.8
$\sin(2\pi x)$	49.6	50.6	25.4	27.4	23.8	26.0
$1.5 \sin(2\pi x)$	68.6	66.6	26.8	27.8	19.2	19.4

Table 2.13: Percentage of rejection under the null and the alternative hypothesis (nominal level 5%) when the bandwidth h is obtained by cross-validation and for samples of size $n = 200$.

$c(x)$	h	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
		T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	0.1	2.2	4.2	4.4	6.2	6.8	7.0
0	0.15	3.8	7.0	7.2	9.0	7.0	8.2
0	0.2	3.6	6.4	6.6	9.0	7.0	7.4
$5x^2$	0.1	100.0	100.0	57.8	59.8	9.6	9.4
$5x^2$	0.15	100.0	100.0	53.2	53.8	12.6	11.4
$5x^2$	0.2	100.0	100.0	41.6	43.6	8.8	8.8
$7.5x^2$	0.1	100.0	100.0	98.0	98.2	79.2	79.8
$7.5x^2$	0.15	100.0	100.0	96.2	96.4	69.6	73.0
$7.5x^2$	0.2	100.0	100.0	85.8	87.8	50.4	53.2
$10x^2$	0.1	100.0	100.0	99.2	99.2	91.6	91.4
$10x^2$	0.15	100.0	100.0	99.4	99.4	91.6	91.8
$10x^2$	0.2	100.0	100.0	96.6	97.0	75.8	78.0

Table 2.14: Percentage of rejection under the null and the alternative hypothesis (nominal level 5%) for fixed bandwidth h and for samples of size $n = 200$.

$c(x)$	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	5.0	8.2	4.0	3.4	6.6	7.0
$5x^2$	100.0	100.0	21.0	20.6	0.4	0.2
$7.5x^2$	100.0	100.0	81.4	81.2	21.6	21.8
$10x^2$	100.0	100.0	83.6	81.4	51.2	49.2

Table 2.15: Percentage of rejection obtained with the mean squared distance from independence estimator of θ , under the null and the alternative hypothesis (nominal level 5%) when the bandwidth h is obtained by cross-validation and for samples of size $n = 200$.

2.4.5 Performance of the test in dimension 2

In this section, we present the results of a simulation study that shows the performance of our test in higher dimensions. We consider $d = 2$, $q = 2$ and the simulated model is $\Lambda_\theta(Y_i) = 3 + \beta_1 X_{1i} + \beta_2 X_{2i} + c(X_{1i}, X_{2i}) + \varepsilon_i$, where Λ_θ is again the Box and Cox (1964) transformation. Moreover, $X_{11}, \dots, X_{1n}$ and $X_{21}, \dots, X_{2n}$ are independent and uniformly distributed on the unit square, and $\varepsilon_1, \dots, \varepsilon_n$ are independent standard normal random variables truncated on $[-3, 3]$. We consider the following null hypothesis:

$$H_0 : m(x) = 3 + \beta_1 x_1 + \beta_2 x_2 \quad \text{for all } x.$$

We perform simulations for the same three different values of the parameter θ as in Sections 2.4.2 to 2.4.4: $\theta_0 = 0$, $\theta_0 = 0.5$ and $\theta_0 = 1$. The true value of the parameter $\beta = (\beta_1, \beta_2)$ is $(\beta_{10}, \beta_{20}) = (3, 5)$.

The term $c(x_1, x_2)$ represents different deviations from the null hypothesis and we consider here $c(x_1, x_2) = 2.5x_1x_2$, $c(x_1, x_2) = 5x_1x_2$, $c(x_1, x_2) = 7.5x_1x_2$, $c(x_1, x_2) = 10x_1x_2$, $c(x_1, x_2) = 2\frac{x_1+1}{x_2+1}$, $c(x_1, x_2) = 3\frac{x_1+1}{x_2+1}$, $c(x_1, x_2) = 4\frac{x_1+1}{x_2+1}$, $c(x_1, x_2) = 5\frac{x_1+1}{x_2+1}$, for samples of size $n = 200$.

Finally, for the estimation of θ , h , g and β , we proceed as explained in Section 2.4.1 and the critical values of the test statistics T_{KS} and T_{CM} will be approximated by means of the bootstrap procedure described in Section 2.3.3 with $B = 250$. Note that we estimate h by cross validation.

Table 2.16 shows the percentage of rejection under the null hypothesis and under the different deviations $c(x_1, x_2)$ 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%.

Table 2.16 shows that the nominal level is in general too low, especially for $\theta_0 = 0$ and $\theta_0 = 0.5$. This bad behavior can be explained by the poor quality of the estimation of the function $m(\cdot)$. Since we are estimating $m(\cdot)$ in two dimensions in a completely nonparametric way, the method is clearly suffering from curse-of-dimensionality problems, which are very common in

nonparametric regression. This finding is also valid under several different deviations $c(x_1, x_2)$. Indeed, we can see that the power is especially low when the deviation is close to the null hypothesis or when $\theta_0 = 1$. So, a sample size of $n = 200$ is not sufficient for the method to work well.

For that reason, we perform an additional simulation under H_0 and under the same different deviations $c(x_1, x_2)$ as before to check whether these problems disappear for a larger sample size. Table 2.17 shows the percentage of rejection under the null hypothesis and under the different deviations $c(x_1, x_2)$ for sample size $n = 500$. We see in this table that the nominal level is now much better approximated and that the different powers have increased with respect to the case of samples of size $n = 200$.

$c(x_1, x_2)$	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	0.8	0.4	2.0	2.0	5.2	5.4
$2.5x_1x_2$	20.2	7.4	5.4	4.6	5.0	5.6
$5x_1x_2$	52.6	39.8	11.4	10.8	6.2	6.6
$7.5x_1x_2$	66.4	67.0	24.2	26.2	11.6	11.4
$10x_1x_2$	67.6	74.2	41.0	41.6	17.8	18.6
$2\frac{x_1+1}{x_2+1}$	94.8	96.0	9.0	10.0	6.8	7.0
$3\frac{x_1+1}{x_2+1}$	96.6	97.6	34.6	38.4	11.4	13.2
$4\frac{x_1+1}{x_2+1}$	97.0	97.8	50.6	53.8	21.2	23.0
$5\frac{x_1+1}{x_2+1}$	94.6	95.0	65.8	67.8	33.4	35.4

Table 2.16: Percentage of rejection under the null and the alternative hypothesis (nominal level 5%) for samples of size $n = 200$.

$c(x_1, x_2)$	$\theta_0 = 0$		$\theta_0 = 0.5$		$\theta_0 = 1$	
	T_{KS}	T_{CM}	T_{KS}	T_{CM}	T_{KS}	T_{CM}
0	2.4	2.0	6.2	6.8	6.4	6.4
$2.5x_1x_2$	71.4	67.0	23.0	23.0	12.0	11.0
$5x_1x_2$	97.4	98.2	47.4	46.6	19.4	18.8
$7.5x_1x_2$	97.2	97.4	68.4	67.8	33.0	32.4
$10x_1x_2$	97.8	98.6	87.2	86.8	49.0	49.2
$2\frac{x_1+1}{x_2+1}$	98.0	98.2	37.4	41.8	11.4	12.4
$3\frac{x_1+1}{x_2+1}$	100.0	100.0	77.6	79.0	35.6	35.2
$4\frac{x_1+1}{x_2+1}$	100.0	100.0	83.2	83.4	56.6	57.2
$5\frac{x_1+1}{x_2+1}$	100.0	100.0	82.6	83.2	63.0	63.4

Table 2.17: Percentage of rejection under the null hypothesis (nominal level 5%) for samples of size $n = 500$.

2.5 Application on sunlight data

2.5.1 Previous analyses of the data

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 et al. (2008a). A test for linearity of the underlying regression function was realized in Hart (1997), while different tests for l th degree polynomial regression ($l = 1, 2, 3, 4$) were realized in Zhang (2003) and in Van Keilegom et al. (2008a), both with their own testing procedure.

2.5.2 Results

Here, a Box-Cox transformation of the response variable is considered and we check the goodness-of-fit of the l th degree polynomial regression ($l = 1, 2, 3, 4$) by using the Kolmogorov-Smirnov and the Cramér-von Mises test statistics defined in this chapter. The distributions and p-values of these two test stat-

istics are approximated by bootstrap on the basis of 1000 replicates. The results are given in Table 2.18.

First, note that the profile likelihood estimator of θ is equal to $\hat{\theta} = 1.9428$. This implies that we transform the response variable Y by taking approximately its square. Table 2.18 indicates that there is no evidence against a polynomial fit of order $l = 2, 3, 4$, similarly as in Van Keilegom et al. (2008a). Moreover, there is also no evidence against a linear fit, which is different from the conclusions in Hart (1997), Zhang (2003) and Van Keilegom et al. (2008a). This can be explained by the transformation realized on the response variable Y , which has an important impact on the regression function.

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

Table 2.18: p-values of the l th degree polynomial fit for the sunlight data.

2.6 Conclusions and recommendations

First, under H_0 and when $d = 1$, the different estimations of the nominal level are globally good and we note that the results for T_{KS} are slightly better. When $d = 2$, the estimations of the nominal level are in general too low, especially for $\theta_0 = 0$ and $\theta_0 = 0.5$, and is due to the poor quality of the estimation of the function $m(\cdot)$. Then, the method clearly suffers from curse-of-dimensionality problems.

Next, under the alternative and when $d = 1$ and $d = 2$, the power is largest for $\theta_0 = 0$, followed by $\theta_0 = 0.5$ and then $\theta_0 = 1$ and the percentages of rejection are slightly larger for T_{CM} than for T_{KS} , which is in line with what happens under H_0 .

When we increase the sample size from $n = 200$ to $n = 500$, the nominal level is much better approximated and the power increases for $\theta_0 = 0$, $\theta_0 = 0.5$ and $\theta_0 = 1$, which are expected conclusions.

Moreover, under the null hypothesis and under the alternative, the results are sometimes better when the bandwidth h is estimated by cross-validation and are sometimes better when h is fixed by the user, which suggests that both approaches give comparable results. Then, in Chapter 3, we will only consider the case when the bandwidth h is estimated by cross-validation.

Finally, under the alternative, the use of the mean squared distance from independence estimator of θ gives lower power than if we use the profile likelihood estimator of θ , whereas the behavior under H_0 is comparable.

2.7 Technical assumptions

For $0 < \alpha < v/2$, where v is defined as in condition (A2) (see below), let $C_1^{d+\alpha}(\chi)$, be the set of d -times differentiable functions $f : \chi \rightarrow \mathbb{R}$ 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 $j = (j_1, \dots, j_d)$, $j \cdot = \sum_{i=1}^d j_i$, $D^j = \partial^{j \cdot} / \partial x_1^{j_1} \dots \partial x_d^{j_d}$ and $\|\cdot\|$ 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 $k = 1, \dots, q_2 - 1$ 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 $l = 0, \dots, d-1$ and k_2 is twice continuously differentiable.
- (A2) h_l (for $l = 1, \dots, d$) satisfies $h_l/h \rightarrow c_l$ for some $0 < c_l < \infty$ and the bandwidths h and g satisfy $nh^{2p+2} \rightarrow 0$ for some $p \geq 3$, $nh^{3d+v} \rightarrow \infty$ for some $v > 0$, $ng^6(\log g^{-1})^{-2} \rightarrow \infty$ and $ng^{2q_2} \rightarrow 0$ when $n \rightarrow \infty$, 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) $\inf_{x \in \chi} f_X(x) > 0$.
- (A4) The distribution function $F_{\varepsilon(\theta)|X}(y|x)$ is three times continuously differentiable with respect to y and θ , and

$$\sup_{\theta, y, x} \left| \frac{\partial^{i+j}}{\partial y^i \partial \theta_1^{j_1} \dots \partial \theta_k^{j_k}} F_{\varepsilon(\theta)|X}(y|x) \right| < \infty$$

for all i and j such that $0 \leq i + j \leq 2$ where $j = \sum_{l=1}^k j_l$.

- (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:

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

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

- (ii) $\sup_{x \in \chi} E[(\dot{\Lambda}_{\theta_0}(Y))_l^2 | X = x] < \infty$ for $l = 1, \dots, k$.

- (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 neighborhood 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 Ξ is of full rank.

- (A9) (i) The error term $\varepsilon = \Lambda_{\theta_0}(Y) - m(X)$ has finite fourth moment and is independent of X .

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

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

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

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

Note that conditions (A6) and (A10) come from Van Keilegom et al. (2008a), conditions (A4), (A5)(i) and (A8) come from Linton et al. (2008), condition (A3)(ii) come from Neumeyer and Van Keilegom (2010) and condition (A9) come from Heuchenne et al. (2015). Finally, note that conditions (A1) and (A2), which are assumptions on the different kernels and bandwidths and condition (A7) come partially from Linton et al. (2008) and partially from Neumeyer and Van Keilegom (2010).

2.8 Proofs

2.8.1 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 (1964) and Watson (1964) 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

$$\sum_{i=1}^n W_0^n\left(\frac{x - X_i}{h}, x\right) = 1 \quad \text{and} \quad W_0^n(u, x) = \frac{1}{n \prod_{j=1}^d h_j} \frac{1}{f_X(x)} K_1^*(u) (1 + o_P(1)) \quad (2.8.1)$$

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 $K_1^*(u_1, \dots, u_d) = \prod_{j=1}^d k_1^*(u_j)$ where k_1^* is a univariate kernel that satisfies $\int u^{p'} k_1^*(u) du = 1_{\{p'=0\}}$ ($0 \leq p' \leq p$). 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 et al. (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 et al. (2008) to our context.

Lemma 2.8.1. *Assume (I1), (I2) and (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 $\sup_{y \in \mathbb{R}} |\widehat{r}(y)| = o_P(n^{-1/2})$.

Proof. First, we have by a Taylor expansion:

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

Next, using the equivalent kernels introduced above, we have

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

where

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

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

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

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

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

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

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

By a Taylor expansion, the first expectation is equal to

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

The second expectation is equal to

$$\int \prod_{k=1}^d \frac{1}{h_k} k_1^* \left(\frac{t - X_{jk}}{h_k} \right) dt = 1 .$$

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. $\square$

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

Proposition 2.8.2. *Assume (I1), (I2) and (A1)-(A9). Then,*

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

where $g(X, Y)$ is defined in (1.2.10). Moreover,

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

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

2.8.2 Results regarding the estimation of β

The following three lemmas state the consistency of $\hat{\beta}$, the consistency of $\tilde{\beta}_{0n}$, defined at the beginning of Section 2.3.2, and an asymptotic representation for $\hat{\beta} - \beta_0$, 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 2.8.3. *Assume (I1), (I2) and (A1)-(A10). Then, under H_{1n} ,*

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

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 et al. (2008a). 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

$$\sup_{\beta \in \mathcal{B}} |S_n(\beta) - S_0(\beta)| \xrightarrow{P} 0 .$$

The expression $S_n(\beta)$ is equal to

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

Hence, we have:

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

From Proposition 2.8.2, we know that $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$. By a Taylor expansion, $\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i) = (\dot{\Lambda}_{\xi}(Y_i))^t (\hat{\theta} - \theta_0)$ for some ξ between $\hat{\theta}$ 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 $\beta \in \mathcal{B}$. Next, using Theorem 2 in Jennrich (1969), we have uniformly in $\beta \in \mathcal{B}$:

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

and

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

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 $\sigma^2 + E[(m_{\beta_0}(X) - m_{\beta}(X))^2] = S_0(\beta)$. $\square$

Lemma 2.8.4. *Assume (I1), (I2) and (A1)-(A10). Then, under H_{1n} ,*

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

Proof. Similarly as in the proof of Lemma 2.8.3, we will use Theorem 5.7 in Van der Vaart (1998), this time with $M_n(\beta) = -\tilde{S}_{0n}(\beta)$ 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 2.8.3. Moreover,

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

and this converges to zero, because

$$\begin{aligned} & E[r(X)(m(X) + m_{\beta_0}(X) - 2m_\beta(X))] \\ &\leq \{E[r^2(X)]\}^{1/2} \{E[(m(X) + m_{\beta_0}(X) - 2m_\beta(X))^2]\}^{1/2} \\ &< \infty , \end{aligned}$$

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. $\square$

Lemma 2.8.5. *Assume (I1), (I2) and (A1)-(A10). Then, under H_{1n} ,*

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

where $g(X, Y)$ is defined in (1.2.10).

Proof. The proof of this Lemma is somewhat similar to the proof of Lemma 4.2 in Van Keilegom et al. (2008a), but we have to take care of the estimation of the transformation 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 $\hat{\beta} - \tilde{\beta}_{0n}$ and second we will search an asymptotic expression for $\tilde{\beta}_{0n} - \beta_0$. By a Taylor expansion, we obtain:

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

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

$$\hat{\beta} - \tilde{\beta}_{0n} = - \left(\frac{\partial^2 S_n(\beta_{1n})}{\partial \beta \partial \beta^t} \right)^{-1} \frac{\partial S_n(\tilde{\beta}_{0n})}{\partial \beta} . \quad (2.8.2)$$

First note that

$$\frac{\partial S_n(\tilde{\beta}_{0n})}{\partial \beta} = -2n^{-1} \sum_{i=1}^n (\Lambda_{\tilde{\theta}}(Y_i) - m_{\tilde{\beta}_{0n}}(X_i)) \frac{\partial m_{\tilde{\beta}_{0n}}(X_i)}{\partial \beta} . \quad (2.8.3)$$

Next, using that $\hat{\beta} - \tilde{\beta}_{0n} = o_P(1)$ by Lemmas 2.8.3 and 2.8.4 and thus $\beta_{1n} = \tilde{\beta}_{0n} + o_P(1) = \beta_0 + o_P(1)$, we have by condition (A6)(ii) that:

$$\begin{aligned} \frac{\partial^2 S_n(\beta_{1n})}{\partial \beta \partial \beta^t} &= 2n^{-1} \sum_{i=1}^n \frac{\partial m_{\beta_{1n}}(X_i)}{\partial \beta} \left(\frac{\partial m_{\beta_{1n}}(X_i)}{\partial \beta} \right)^t \\ &\quad - 2n^{-1} \sum_{i=1}^n (\Lambda_{\hat{\theta}}(Y_i) - m_{\beta_{1n}}(X_i)) \frac{\partial^2 m_{\beta_{1n}}(X_i)}{\partial \beta \partial \beta^t} \\ &= 2n^{-1} \sum_{i=1}^n \frac{\partial m_{\beta_0}(X_i)}{\partial \beta} \left(\frac{\partial m_{\beta_0}(X_i)}{\partial \beta} \right)^t \\ &\quad - 2n^{-1} \sum_{i=1}^n (\Lambda_{\hat{\theta}}(Y_i) - m_{\beta_0}(X_i)) \frac{\partial^2 m_{\beta_0}(X_i)}{\partial \beta \partial \beta^t} + o_P(1) . \end{aligned}$$

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

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

Note that to obtain the last equality, we used Proposition 2.8.2, and conditions (A2), (A5)(i) and (A6)(ii). Hence,

$$\frac{\partial^2 S_n(\beta_{1n})}{\partial \beta \partial \beta^t} = 2\Omega + o_P(1) . \quad (2.8.4)$$

Inserting (2.8.3) and (2.8.4) in (2.8.2), we obtain:

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

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

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

Hence, using Proposition 2.8.2,

$$\begin{aligned}\hat{\beta} - \tilde{\beta}_{0n} &= -\Omega^{-1} E \left[\frac{\partial m_{\tilde{\beta}_{0n}}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] n^{-1} \sum_{i=1}^n g(X_i, Y_i) \\ &\quad + n^{-1} \sum_{i=1}^n \eta_{\tilde{\beta}_{0n}}(X_i, Y_i) + o_P(n^{-1/2}) .\end{aligned}\tag{2.8.5}$$

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

$$\tilde{\beta}_{0n} - \beta_0 = \Omega^{-1} n^{-1/2} \int r(x) \frac{\partial m_{\tilde{\beta}_{0n}}(x)}{\partial \beta} dF_X(x) + o_P(n^{-1/2}) .\tag{2.8.6}$$

Finally, we get the result by combining (2.8.5) and (2.8.6). $\square$

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

We state now some results concerning $\hat{m}(x)$ that will be useful to establish the limiting distributions of the test statistics. The three following propositions are generalizations of Propositions 6.1-6.3 in Heuchenne et al. (2015) when the dimension of the variable X is equal to $d \geq 1$:

Proposition 2.8.6. *Assume (I1), (I2) and (A1)-(A9). Then*

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

Proof. First, we have

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

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

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

and this is $O_P(n^{-1/2})$ because $\sup_{\theta \in \Theta, x \in \chi} |E(\dot{\Lambda}_{\theta}(Y)|X = x)|$ is finite by condition (A9)(iii) and $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ by Proposition 2.8.2. Concerning the second term on the right hand side, i.e., $\hat{m}(x, \theta_0) - m(x)$, we see that the transformation is fixed, and hence we can use Theorem 6 in Masry (1996) with $j = 0$, which gives

$$\sup_{x \in \chi} |\hat{m}(x, \theta_0) - m(x)| = O((nh^d)^{-1/2}(\log n)^{1/2}) + O(h^{p+1}) .$$

This concludes the proof. $\square$

Proposition 2.8.7. *Assume (I1), (I2) and (A1)-(A9). Then for all vectors $j = (j_1, \dots, j_d)$ such that $j = \sum_{i=1}^d j_i \leq d$,*

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

where D^j is defined in the beginning of Section 2.7.

Proposition 2.8.8. *Assume (I1), (I2) and (A1)-(A9). Then,*

$$\begin{aligned} &\sup_{x, x' \in \chi} \frac{|D^d \hat{m}(x) - D^d m(x) - D^d \hat{m}(x') + D^d m(x')|}{||x - x'||^\alpha} \\ &= O_P((nh^{3d+2\alpha})^{-1/2}(\log n)^{1/2}) + O(h^{p+1-d-\alpha}) . \end{aligned}$$

The proofs of Propositions 2.8.7 and 2.8.8 are similar to the one of Proposition 2.8.6 and are therefore omitted. For Proposition 2.8.8, 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 2.8.9. *Assume (I1), (I2) and (A1)-(A9). Then,*

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

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 2.8.6, we have shown that

$$\begin{aligned} \widehat{m}(x) - m(x) &= \sum_{i=1}^n W_0^n \left(\frac{x - X_i}{h}, x \right) (\dot{\Lambda}_{\xi}(Y_i))^t (\widehat{\theta} - \theta_0) \\ &\quad + \sum_{i=1}^n W_0^n \left(\frac{x - X_i}{h}, x \right) (\Lambda_{\theta_0}(Y_i) - m(x)) , \end{aligned}$$

where ξ is some value between $\widehat{\theta}$ and θ_0 . Using (2.8.1), we have that

$$\begin{aligned} & \int (\widehat{m}(x) - m(x)) dF_X(x) \\ &= \sum_{i=1}^n \int \frac{1}{nh} \frac{1}{f_X(x)} k_1^* \left(\frac{x - X_i}{h} \right) (\Lambda_{\theta_0}(Y_i) - m(x)) dF_X(x) (1 + o_P(1)) \\ &\quad + \sum_{i=1}^n \int \frac{1}{nh} \frac{1}{f_X(x)} k_1^* \left(\frac{x - X_i}{h} \right) (\dot{\Lambda}_{\xi}(Y_i))^t (\widehat{\theta} - \theta_0) dF_X(x) (1 + o_P(1)) . \end{aligned} \tag{2.8.7}$$

The first term on the right hand side of (2.8.7) is equal to

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

By a Taylor expansion of order $p + 1$, with a certain ξ_i between X_i and x , this

last term is equal to:

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

Next, we remind that $k_{1h}^*(u) = k_1^*(u/h)/h$ where k_1^* is a univariate kernel that satisfies $\int u^{p'} k_1^*(u) du = 1_{\{p'=0\}}$ for $0 \leq p' \leq p$. This implies that for $i = 1, \dots, n$:

$$\int k_{1h}^*(x - X_i) dx = \int \frac{1}{h} k_1^*\left(\frac{x - X_i}{h}\right) dx = \int k_1^*(u) du = 1 .$$

Similarly, for $i = 1, \dots, n$, we have

$$\sum_{l=1}^p \frac{1}{l!} m^{(l)}(X_i) \int k_{1h}^*(x - X_i) (x - X_i)^l dx = 0 ,$$

because $\int u^{p'} k_1^*(u) du = 0$ for $1 \leq p' \leq p$. Finally, as $\sup_x |m^{(p+1)}(x)| < \infty$ by condition (A7) and $(x - X_i)^{p+1} = O(h^{p+1})$ which is $o_P(n^{-1/2})$ by condition (A2), the first term on the right hand side of (2.8.7) 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 2.8.2, the second term on the right hand side of (2.8.7) is equal to

$$\begin{aligned} & n^{-1} \sum_{i=1}^n (\dot{\Lambda}_{\theta_0}(Y_i))^t \left(-n^{-1} \sum_{j=1}^n g(X_j, Y_j) \right) + o_P(n^{-1/2}) \\ & = -n^{-1} \sum_{j=1}^n E((\dot{\Lambda}_{\theta_0}(Y))^t) g(X_j, Y_j) + o_P(n^{-1/2}) , \end{aligned}$$

because $-n^{-1} \sum_{j=1}^n g(X_j, Y_j) = O_P(n^{-1/2})$ by Proposition 2.8.2 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. $\square$

2.8.4 Proofs of the main results

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

Proof of Theorem 2.2.1. If H_0 is true, then

$$F_\varepsilon(y) = P(\Lambda_{\theta_0}(Y) - m(X) \leq y) = P(\Lambda_{\theta_0}(Y) - m_{\beta_0}(X) \leq y) = F_{\varepsilon_0}(y) .$$

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

$$E(m(X) - m_{\beta_0}(X)) = 0 . \quad (2.8.8)$$

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

$$V(m(X) - m_{\beta_0}(X)) = 0 \quad (2.8.9)$$

Finally, (2.8.8) and (2.8.9) 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 2.3.1.

Lemma 2.8.10. *Assume (I1), (I2) and (A1)-(A9). Then,*

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

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

The proof of this lemma can be found in Proposition 6.4 in Heuchenne et al. (2015). We also need the following technical result, the proof of which is based on different results in Van der Vaart and Wellner (1996):

Lemma 2.8.11. *Assume (I1), (I2) and (A1)-(A9). Then,*

$$n^{-1} \sum_{i=1}^n \left\{ 1_{\{\Lambda_{\hat{\theta}}(Y_i) - \hat{m}(X_i) \leq t\}} - 1_{\{\Lambda_{\theta_0}(Y_i) - m(X_i) \leq t\}} \right. \\ \left. - P(\Lambda_{\hat{\theta}}(Y) - \hat{m}(X) \leq t) + P(\Lambda_{\theta_0}(Y) - m(X) \leq t) \right\} = o_P(n^{-1/2})$$

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

Proof. The proof is a combination of the proof of Lemma 1 in Heuchenne et al. (2015) 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 et al. (2015) 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 $\Lambda_{\hat{\theta}}(Y) - \hat{m}(X) = \Lambda_{\hat{\theta}}(Y) - m(X) - h_n(X)$ where $h_n(X) = \hat{m}(X) - m(X)$. We define:

$$\mathcal{F}_1 = \left\{ (x, y) \longrightarrow 1_{\{\Lambda_{\theta}(y) \leq t + m(x) + h(x)\}} , \Lambda_{\theta} : \mathbb{R} \rightarrow \mathbb{R} \right. \\ \left. \text{strictly increasing, } \theta \in \Theta \subset \mathbb{R}^k, t \in \mathbb{R}, h \in C_1^{d+\alpha}(\chi) \right\}.$$

By Propositions 2.8.6, 2.8.7 and 2.8.8 we have that $P(h_n \in C_1^{d+\alpha}(\chi)) \rightarrow 1$ as $n \rightarrow \infty$, 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

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

where $N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P))$ is the $\tilde{\varepsilon}$ -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

$$N_{[]}(\tilde{\varepsilon}^2, C_1^{d+\alpha}(\chi), L_2(P)) \leq \exp(C\tilde{\varepsilon}^{-2d/(d+\alpha)}),$$

where C is a certain positive constant. Using the same arguments as in Lemma 1 in Heuchenne et al. (2015), it follows that the $\tilde{\varepsilon}$ -bracketing number of the class $\mathcal{F}_1$ is at most $O(\tilde{\varepsilon}^{-2(k+1)} \exp(C\tilde{\varepsilon}^{-2d/(d+\alpha)}))$ 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

$$\mathcal{F} = \left\{ (x, y) \longrightarrow 1_{\{\Lambda_{\theta}(y) \leq t + m(x) + h(x)\}} - 1_{\{\Lambda_{\theta_0}(y) \leq t + m(x)\}} \right. \\ \left. - P(\Lambda_{\theta}(Y) \leq t + m(X) + h(X)) + P(\Lambda_{\theta_0}(Y) \leq t + m(X)) , \right. \\ \left. \Lambda_{\theta} : \mathbb{R} \rightarrow \mathbb{R} \text{ strictly increasing, } \theta \in \Theta \subset \mathbb{R}^k, t \in \mathbb{R}, h \in C_1^{d+\alpha}(\chi) \right\}$$

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

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

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. $\square$

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

Proof of Theorem 2.3.1. The proof will be divided in two parts: first we will develop an asymptotic representation for $\widehat{F}_{\widehat{\varepsilon}}(y) - F_{\varepsilon}(y)$ and second we will obtain an asymptotic representation for $F_{\varepsilon}(y) - \widehat{F}_{\widehat{\varepsilon}_0}(y)$. Using the definition of the semiparametric transformation model, we have that:

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

Next, using Lemma 2.8.11 and the last expression above, we have that

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

where $\sup_{y \in \mathbb{R}} |R_{n1}(y)| = o_P(n^{-1/2})$. By a Taylor expansion with a certain y_x between y and $y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})$

$$\begin{aligned} &\widehat{F}_{\widehat{\varepsilon}}(y) - F_{\varepsilon}(y) \\ &= n^{-1} \sum_{i=1}^n 1_{\{\varepsilon_i \leq y\}} - F_{\varepsilon}(y) + f_{\varepsilon}(y) \int (\widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})) dF_X(x) \\ &\quad + \frac{1}{2} \int f'_{\varepsilon}(y_x) (\widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta}))^2 dF_X(x) \\ &\quad + \int \left[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right. \\ &\quad \left. - F_{\varepsilon}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right] dF_X(x) + R_{n1}(y), \end{aligned}$$

By conditions (A2), (A4), (A5) and Propositions 2.8.2 and 2.8.6, the third term on the right hand side is $o_P(n^{-1/2})$ and we obtain the following asymptotic

representation for $\widehat{F}_\varepsilon(y) - F_\varepsilon(y)$:

$$\begin{aligned} & \widehat{F}_\varepsilon(y) - F_\varepsilon(y) \\ &= n^{-1} \sum_{i=1}^n 1_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y) + f_\varepsilon(y) \int (\widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})) dF_X(x) \\ & \quad + \int \left[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right. \\ & \quad \left. - F_\varepsilon(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right] dF_X(x) + R_{n1}(y). \end{aligned} \quad (2.8.10)$$

Next, we need a similar representation for $\widehat{F}_{\varepsilon_0}(y) - F_\varepsilon(y)$. We can easily prove the equivalence of Propositions 2.8.6 to 2.8.10 when the function $\widehat{m}(x)$ is replaced by $\widehat{m}_{\widehat{\beta}}(x)$ (using condition (A6)). Then, Lemma 2.8.11 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:

$$\begin{aligned} & \widehat{F}_{\varepsilon_0}(y) - F_\varepsilon(y) \\ &= n^{-1} \sum_{i=1}^n 1_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y) + f_\varepsilon(y) \int (\widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})) dF_X(x) \\ & \quad + \int \left[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right. \\ & \quad \left. - F_\varepsilon(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right] dF_X(x) + R_{n2}(y), \end{aligned} \quad (2.8.11)$$

where $\sup_{y \in \mathbb{R}} |R_{n2}(y)| = o_P(n^{-1/2})$. Next, by a Taylor expansion with some value ξ between $\widehat{\theta}$ and θ_0 , we obtain:

$$\begin{aligned} & \int \left[F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) - F_\varepsilon(y + \widehat{m}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right. \\ & \quad \left. - F_{\varepsilon(\widehat{\theta})}(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) + F_\varepsilon(y + \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta}) - m(x, \widehat{\theta})|x) \right] dF_X(x) \\ &= \int \left[\left(\dot{F}_{\varepsilon(\theta)}(z|x) \Big|_{z=y+\widehat{m}(x, \widehat{\theta})-m(x, \widehat{\theta})}^{\theta=\xi} \right)^t \right. \\ & \quad \left. - \left(\dot{F}_{\varepsilon(\theta)}(z|x) \Big|_{z=y+\widehat{m}_{\widehat{\beta}}(x, \widehat{\theta})-m(x, \widehat{\theta})}^{\theta=\xi} \right)^t \right] (\widehat{\theta} - \theta_0) dF_X(x). \end{aligned}$$

By another Taylor expansion, this last expression is equal to

$$\int \left[\left(\frac{\partial}{\partial z} \dot{F}_{\varepsilon(\theta)}(z|x) \Big|_{z=y-m(x, \widehat{\theta})+\eta}^{\theta=\xi} \right)^t (\widehat{\theta} - \theta_0) (\widehat{m}(x, \widehat{\theta}) - \widehat{m}_{\widehat{\beta}}(x, \widehat{\theta})) \right] dF_X(x) \quad (2.8.12)$$

with η some value between $\widehat{m}(x, \widehat{\theta})$ and $\widehat{m}_{\widehat{\beta}}(x, \widehat{\theta})$. Consequently, using assumption (A4), Lemma 2.8.3 and Propositions 2.8.2 and 2.8.6, we conclude that expression (2.8.12) is $o_P(n^{-1/2})$. This shows that

$$\widehat{F}_{\widehat{\varepsilon}}(y) - \widehat{F}_{\widehat{\varepsilon}_0}(y) = f_{\varepsilon}(y) \int [\widehat{m}(x) - \widehat{m}_{\widehat{\beta}}(x)] dF_X(x) + R_n(y),$$

where $\sup_{y \in \mathbb{R}} |R_n(y)| = o_P(n^{-1/2})$.

Next, in a very similar way as in the proof of Proposition 2.8.9, 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 2.8.3, we obtain:

$$\begin{aligned} f_{\varepsilon}(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)] dF_X(x) \\ = f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n \left(\frac{\partial m_{\beta_0}(X_i)}{\partial \beta} \right)^t (\widehat{\beta} - \beta_0) + o_P(n^{-1/2}). \end{aligned} \quad (2.8.13)$$

Now, using Lemma 2.8.5 and Proposition 2.8.9, we have that

$$\begin{aligned} \widehat{F}_{\widehat{\varepsilon}}(y) - \widehat{F}_{\widehat{\varepsilon}_0}(y) \\ = f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n \left(\Lambda_{\theta_0}(Y_i) - m(X_i) - E((\dot{\Lambda}_{\theta_0}(Y))^t) g(X_i, Y_i) \right. \\ \left. - \int \left(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \right)^t dF_X(x) \eta_{\beta_0}(X_i, Y_i) \right. \\ \left. + \int \left(\frac{\partial m_{\beta_0}(x)}{\partial \beta} \right)^t dF_X(x) \Omega^{-1} E \left[\frac{\partial m_{\beta_0}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] g(X_i, Y_i) \right) \\ + R_n(y), \end{aligned}$$

where $\sup_{y \in \mathbb{R}} |R_n(y)| = o_P(n^{-1/2})$.

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 $n^{1/2}(\widehat{F}_{\widehat{\varepsilon}}(\cdot) - \widehat{F}_{\widehat{\varepsilon}_0}(\cdot))$ ($-\infty < y < +\infty$) converges to $f_{\varepsilon}(\cdot) \mathbb{W}_{\infty}$ where $\mathbb{W}_{\infty}$ is a zero-mean normal random variable. That $E(\mathbb{W}_{\infty}) = 0$ 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 $g(X, Y) = \Xi^{-1} \xi(\theta_0, X, Y)$ and $\xi(\theta_0, X, Y)$ is the derivative of the likelihood. This concludes the proof. $\square$

Proof of Corollary 2.3.2. The proof follows along the same lines as the proof of Corollary 3.2 in Van Keilegom et al. (2008a). We refer to their proof

for more details. $\square$

Proof of Theorem 2.3.3. In a very similar way as in the proof of Theorem 2.3.1(i), but now under H_{1n} , we can show that:

$$\begin{aligned}\widehat{F}_{\widehat{\varepsilon}}(y) - \widehat{F}_{\widehat{\varepsilon}_0}(y) &= f_{\varepsilon}(y) \int [\widehat{m}(x) - m(x)] dF_X(x) \\ &\quad - f_{\varepsilon}(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)] dF_X(x) + R_n(y),\end{aligned}$$

and

$$f_{\varepsilon}(y) \int [\widehat{m}_{\widehat{\beta}}(x) - m(x)] dF_X(x) = f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n (m_{\widehat{\beta}}(X_i) - m(X_i)) + o_P(n^{-1/2}). \quad (2.8.14)$$

Next, under H_{1n} : $m(x) = m_{\beta_0}(x) + n^{-1/2}r(x)$, expression (2.8.14) becomes:

$$f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n (m_{\widehat{\beta}}(X_i) - m_{\beta_0}(X_i)) - n^{-1/2} f_{\varepsilon}(y) \int r(x) dF_X(x) + o_P(n^{-1/2})$$

Using Lemma 2.8.3 and a Taylor expansion, this last expression is equal to

$$f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n \left(\frac{\partial m_{\beta_0}(X_i)}{\partial \beta} \right)^t (\widehat{\beta} - \beta_0) - n^{-1/2} f_{\varepsilon}(y) \int r(x) dF_X(x) + o_P(n^{-1/2}),$$

and by Lemmas 2.8.4 and 2.8.5, it is also equal to

$$\begin{aligned}& f_{\varepsilon}(y) \int \left(\frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial \beta} \right)^t dF_X(x) \\ & \times \left(-\Omega^{-1} n^{-1} \sum_{i=1}^n E \left[\frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] g(X_i, Y_i) \right. \\ & \quad \left. + n^{-1} \sum_{i=1}^n \eta_{\widetilde{\beta}_{0n}}(X_i, Y_i) + \Omega^{-1} n^{-1/2} \int r(x) \frac{\partial m_{\widetilde{\beta}_{0n}}(x)}{\partial \beta} dF_X(x) \right) \\ & - n^{-1/2} f_{\varepsilon}(y) \int r(x) dF_X(x) + o_P(n^{-1/2}).\end{aligned} \quad (2.8.15)$$

Using Proposition 2.8.9 and (2.8.15), this gives

$$\widehat{F}_{\widehat{\varepsilon}}(y) - \widehat{F}_{\widehat{\varepsilon}_0}(y) = f_{\varepsilon}(y) n^{-1} \sum_{i=1}^n L(X_i, Y_i, \theta_0, \widetilde{\beta}_{0n}) + n^{-1/2} f_{\varepsilon}(y) b_1 + R_n(y), \quad (2.8.16)$$

where $\sup_{y \in \mathbb{R}} |R_n(y)| = o_P(n^{-1/2})$.

We will now prove (ii). It follows from (2.8.16) that if we can prove that the limiting distribution under H_{1n} of the expression

$$n^{-1/2} \sum_{i=1}^n L(X_i, Y_i, \theta_0, \tilde{\beta}_{0n}) \quad (2.8.17)$$

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

$$n^{-1/2} \sum_{i=1}^n L(X_i, Y_i, \theta_0, \beta_0) , \quad (2.8.18)$$

then we can conclude that the process $n^{1/2}(\hat{F}_{\varepsilon}(\cdot) - \hat{F}_{\varepsilon_0}(\cdot))$ converges weakly to $f_{\varepsilon}(\cdot)(\mathbb{W}_{\infty} + b_1)$, where $\mathbb{W}_{\infty}$ is the same normal distribution as in Theorem 2.3.1(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 (2.8.17) and (2.8.18) are asymptotically the same. We proved in Theorem (2.3.1)(ii) that the mean of (2.8.18) under H_0 is equal to 0. Moreover, as $\tilde{\beta}_{0n}$ is by definition a minimizer over β of $E[(m(X) - m_{\beta}(X))^2]$, we have that $E[(m(X) - m_{\tilde{\beta}_{0n}}(X)) \frac{\partial m_{\tilde{\beta}_{0n}}(X)}{\partial \beta}] = 0$. Consequently, under H_{1n} :

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

Hence, under H_{1n} , the mean of (2.8.17) is also equal to 0. Next, under H_{1n} , the variance of $\eta_{\tilde{\beta}_{0n}}(X, Y)$ is equal to:

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

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 2.8.4, the last

expression on the right hand side is equal to:

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

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

Proof of Corollary 2.3.4. The proof follows from Theorem 2.3.3 and using similar techniques as in the proof of Corollary 2.3.2. $\square$

Chapter 3

Goodness-of-fit tests in semiparametric transformation models using the integrated regression function

This chapter is based on the paper

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

3.1 Introduction

Consider again the semiparametric transformation model defined in Section 1.2.1. The main objective of this chapter is to develop a second test for the parametric form of the regression function, see the null hypothesis (2.1.1) introduced in Section 2.1.

The basic idea of the new test is to compare the integrated regression function estimated in a semiparametric way to the integrated regression function estimated under H_0 . The idea of testing the form of the regression function using the integrated regression function has been studied among others by Bierens (1982), Stute (1997) and Escanciano (2006a) as introduced in Section 1.4.3. These three articles worked in a context of a nonparametric regression model without transformation of the response variable. In this chapter, our goals will be to extend this class of tests to the context of semiparametric transformation models of the form (1.2.1) and to compare its performance to that of the test based on the error distribution function and constructed in Chapter 2.

The main difference between the three approaches of Bierens (1982), Stute (1997) and Escanciano (2006a) is the weighting function that each author uses

to construct his residual process. In Section 1.4.3, we also observed that the integrated function used to define the Cramér-von Mises test statistic is not always the same. Hence, in this chapter, we will construct a residual process based on a general weighting function w and we will define a Cramér-von Mises test statistic based on an arbitrary integrating function Ψ_n . The main advantage of putting a general weighting function w and an arbitrary integrating function Ψ_n in the definition of the Cramér-von Mises test statistic is that we can use in particular the three main approaches of Bierens (1982), Stute (1997) and Escanciano (2006a), and also several other weighting functions, but in a context of a semiparametric transformation model as we will see in the next section.

The chapter is organized as follows. In Section 3.2, we explain in detail the construction of our test statistics and how we can apply the methods of Bierens (1982), Stute (1997) and Escanciano (2006a) in particular. In Section 3.3, we present the main results of the asymptotic theory and in particular the limiting distributions of the proposed test statistics under the null hypothesis and under a local alternative. Section 3.4 contains a simulation study that shows the performance of the proposed test and compares this new test to the previous test developed in Chapter 2. In Section 3.5, we apply our method to a real data set and Section 3.7 contains the technical assumptions. Finally, Section 3.8 contains the proofs of the main results.

3.2 The proposed test

First, note that we will proceed exactly in the same way as in Chapter 2 to estimate the semiparametric transformation model (1.2.1), i.e., we will estimate the transformation parameter θ by the profile likelihood estimator introduced in (1.2.9) and we will estimate the unknown regression function $m(x, \theta)$ by a local polynomial estimator of degree p , see Section 2.2.1 for a precise definition.

We will introduce two new test statistics where the basic idea is to compare the integrated regression function estimated in a semiparametric way to the integrated regression function estimated under H_0 . We consider the following integrated regression function:

$$M(x, \gamma, \theta) = \int w(t, x, \gamma) m(t, \theta) dF_X(t) = E[w(X, x, \gamma) \Lambda_\theta(Y)] ,$$

where w is some weighting function that depends on some parameter $\gamma \in \mathbb{R}^{d_\gamma}$ and that satisfies the equivalence (1.4.2). We assume that γ has compact support $\chi_\gamma \subset \mathbb{R}^{d_\gamma}$. The empirical analog of $M(x, \gamma, \theta)$ is given by

$$\widehat{M}(x, \gamma, \theta) = n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) \Lambda_\theta(Y_i) .$$

Next, under H_0 , $m \equiv m_{\beta_0}$ where β_0 is the true value of β under H_0 . Then, the integrated regression function becomes

$$M_{\beta_0}(x, \gamma, \theta_0) = \int w(t, x, \gamma) m_{\beta_0}(t) dF_X(t) ,$$

and its empirical analog is given by

$$\widehat{M}_{\beta_0}(x, \gamma, \theta_0) = n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) m_{\beta_0}(X_i) .$$

Hence, our test will be constructed on the basis of the following residual process:

$$\begin{aligned} R_n(x, \gamma, \theta_0, \beta_0) &= \sqrt{n}(\widehat{M}(x, \gamma, \theta_0) - \widehat{M}_{\beta_0}(x, \gamma, \theta_0)) \\ &= n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma)(\Lambda_{\theta_0}(Y_i) - m_{\beta_0}(X_i)) . \end{aligned}$$

Finally, as the parameters θ and β are unknown, we will estimate θ by the profile likelihood estimator defined in (1.2.9) and we will estimate β by a least squares estimator, i.e., we consider $\widehat{\beta}$ which is a minimizer over $\beta \in \mathcal{B}$ of the expression

$$S_n(\beta) = n^{-1} \sum_{i=1}^n (\Lambda_{\widehat{\theta}}(Y_i) - m_{\beta}(X_i))^2 . \quad (3.2.1)$$

This gives the following residual process:

$$\begin{aligned} R_n(x, \gamma, \widehat{\theta}, \widehat{\beta}) &= \sqrt{n}(\widehat{M}(x, \gamma, \widehat{\theta}) - \widehat{M}_{\widehat{\beta}}(x, \gamma, \widehat{\theta})) \\ &= n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma)(\Lambda_{\widehat{\theta}}(Y_i) - m_{\widehat{\beta}}(X_i, \widehat{\theta})) . \end{aligned}$$

For an easier readability, we will use the notations $m_{\widehat{\beta}}(X_i) = m_{\widehat{\beta}}(X_i, \widehat{\theta})$, $R_n(x, \gamma) = R_n(x, \gamma, \theta_0, \beta_0)$ and $R_n^1(x, \gamma) = R_n(x, \gamma, \widehat{\theta}, \widehat{\beta})$ when there is no ambiguity. It is important to remark that we will follow the idea of Escanciano (2006a) and we will not estimate the parameter γ unlike Stute and Zhu (2002) for example. We consider a process that depends both on x and γ . The test statistics that we will use are Kolmogorov-Smirnov and Cramér-von Mises type statistics defined by

$$D_n = \sup_{(x, \gamma) \in \Pi} |R_n^1(x, \gamma)| \quad \text{and} \quad W_n^2 = \int_{\Pi} [R_n^1(x, \gamma)]^2 d\Psi_n(x, \gamma) ,$$

where Π is a properly chosen compact space and $\Psi_n(x, \gamma)$ is a certain estimator of an arbitrary integrating function $\Psi(x, \gamma)$ that is absolutely continuous and that satisfies regularity condition (B2) given in Section 3.7. The main advantage of putting a general weighting function w and an arbitrary integrating

function Ψ_n in the definition of the Cramér-von Mises test statistic is that we can use the three main approaches in the literature based on the integrated regression function but in a context of a semiparametric transformation model:

1. Bierens (1982): take $\Pi = \chi$, $w(X, x, \gamma) = \exp(ix^t X)$ where i is the imaginary unit, $\Psi_n(x, \gamma) = \Psi(x, \gamma)$ and $d\Psi(x, \gamma) = \Phi(x)dx$. Here, we will choose $\Phi(x)$ in such a way that the imaginary part of the Cramér-von Mises test statistic is equal to 0. For example, if $\Phi(x)$ is the standard d -variate normal density function, i.e., $\Phi(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}x^t x) = (2\pi)^{-d/2} \exp(-\frac{1}{2} \sum_{l=1}^d x_l^2)$, we have

$$\begin{aligned} W_{\exp_i}^2 &= \int_{\mathbb{R}^d} \left(n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) e^{iu^t X_j} \right)^2 \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} \sum_{l=1}^d u_l^2\right) du \\ &= \frac{n^{-1}}{(2\pi)^{d/2}} \sum_{j=1}^n \sum_{k=1}^n e_j(\hat{\theta}, \hat{\beta}) e_k(\hat{\theta}, \hat{\beta}) \\ &\quad \times \int_{\mathbb{R}^d} e^{iu^t(X_j + X_k)} \exp\left(-\frac{1}{2} \sum_{l=1}^d u_l^2\right) du, \end{aligned}$$

where $u = (u_1, \dots, u_d)^t$ and $e_j(\hat{\theta}, \hat{\beta}) = \Lambda_{\hat{\theta}}(Y_j) - m_{\hat{\beta}}(X_j)$. Next,

$$\begin{aligned} &\int_{\mathbb{R}^d} e^{iu^t(X_j + X_k)} \exp\left(-\frac{1}{2} \sum_{l=1}^d u_l^2\right) du \\ &= \int_{\mathbb{R}^d} \cos(u^t(X_j + X_k)) \exp\left(-\frac{1}{2} \sum_{l=1}^d u_l^2\right) du \\ &\quad + i \int_{\mathbb{R}^d} \sin(u^t(X_j + X_k)) \exp\left(-\frac{1}{2} \sum_{l=1}^d u_l^2\right) du. \quad (3.2.2) \end{aligned}$$

The second term on the right hand side is equal to zero because the function to integrate is odd and symmetric around 0. Hence, expression (3.2.2) is equal to

$$\begin{aligned} &\int_{-\infty}^{+\infty} \exp(-u_d^2/2) \dots \int_{-\infty}^{+\infty} \exp(-u_1^2/2) \cos\left\{u_1(X_{j1} + X_{k1}) \right. \\ &\quad \left. + \dots + u_d(X_{jd} + X_{kd})\right\} du_1 \dots du_d. \quad (3.2.3) \end{aligned}$$

Next, we can prove by some trivial calculations that

$$\int_{-\infty}^{+\infty} \cos(au + b) e^{-u^2/2} du = \sqrt{2\pi} \cos(b) e^{-a^2/2},$$

for any constants a and b . Hence, applying this last result d times, expression (3.2.3) becomes

$$(2\pi)^{d/2} \exp \left(-\frac{1}{2} \sum_{l=1}^d (X_{jl} + X_{kl})^2 \right).$$

Consequently, in this particular case, the Cramér-von Mises test statistic takes the following form:

$$W_{\text{exp}_i}^2 = n^{-1} \sum_{j=1}^n \sum_{k=1}^n e_j(\hat{\theta}, \hat{\beta}) e_k(\hat{\theta}, \hat{\beta}) \exp \left(-\frac{1}{2} \sum_{l=1}^d (X_{jl} + X_{kl})^2 \right).$$

We will use this particular expression of $W_{\text{exp}_i}^2$ in our simulation study when the complex exponential weight will be used.

2. Stute (1997): take $\Pi = \chi$, $w(X, x, \gamma) = 1_{\{X \leq x\}}$ where $1_{\{X \leq x\}}$ is a component by component indicator and $\Psi_n(x, \gamma) = \hat{F}_X(x)$ where $\hat{F}_X(x)$ is the empirical distribution function of the data $\{X_i\}_{i=1, \dots, n}$. Moreover, $\Psi(x, \gamma) = F_X(x)$ is the true distribution function of X . Hence, the two test statistics take the following form:

$$D_1 = \max_{1 \leq k \leq n} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) 1_{\{X_j \leq X_k\}} \right|,$$

and

$$W_1^2 = n^{-2} \sum_{k=1}^n \left(\sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) 1_{\{X_j \leq X_k\}} \right)^2.$$

3. Escanciano (2006a): take $d_\gamma = d$, $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$ and $d\Psi_n(x, \gamma) = d\hat{F}_{n, \gamma}(x) d\gamma$ where $\hat{F}_{n, \gamma}(x)$ is the empirical distribution function of the projected regressor $\{\gamma^t X_i\}_{i=1, \dots, n}$ and $d\gamma$ is the uniform density on $\mathbb{S}_d$ which is the unit ball of dimension d and ensures that all directions are equally important. Then, $\Pi = [-\Delta, \Delta] \times \mathbb{S}_d$ where $\Delta = d \max_{1 \leq i \leq d} \sup_{t \in \chi} |t_i|$ and t_i is the i -th component of the vector $t \in \chi$. Indeed, if $t = (t_1, \dots, t_d)^t \in \chi$ and $\gamma = (\gamma_1, \dots, \gamma_d)^t \in \mathbb{S}_d$, we have

$$|\gamma^t t| = |\gamma_1 t_1 + \dots + \gamma_d t_d| \leq \sum_{i=1}^d |\gamma_i t_i| \leq \sum_{i=1}^d |t_i| \leq d \max_{1 \leq i \leq d} |t_i|,$$

where we have use the fact that $|\gamma_i| \leq 1$, for $i = 1, \dots, d$, by definition of $\mathbb{S}_d$. Moreover, $\Psi(x, \gamma) = F_\gamma(x)$ is the true cumulative distribution function of $\gamma^t X$. In that case, the two test statistics take the following form:

$$D_\gamma = \sup_{1 \leq k \leq n, \gamma \in \mathbb{S}_d} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) 1_{\{\gamma^t X_j \leq \gamma^t X_k\}} \right|,$$

and

$$\begin{aligned}
W_\gamma^2 &= \int_{\Pi} \left(n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) 1_{\{\gamma^t X_j \leq u\}} \right)^2 d\hat{F}_{n,\gamma}(u) d\gamma \\
&= n^{-1} \sum_{j=1}^n \sum_{k=1}^n e_j(\hat{\theta}, \hat{\beta}) e_k(\hat{\theta}, \hat{\beta}) \int_{\Pi} 1_{\{\gamma^t X_j \leq u\}} 1_{\{\gamma^t X_k \leq u\}} d\hat{F}_{n,\gamma}(u) d\gamma \\
&= n^{-2} \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n e_j(\hat{\theta}, \hat{\beta}) e_k(\hat{\theta}, \hat{\beta}) \int_{\mathbb{S}_d} 1_{\{\gamma^t X_j \leq \gamma^t X_l\}} 1_{\{\gamma^t X_k \leq \gamma^t X_l\}} d\gamma .
\end{aligned}$$

In practice, to compute these test statistics, we will consider a random sample $\gamma_1, \dots, \gamma_{n_\gamma}$ from $\mathbb{S}_d$. Hence, we can approximate both test statistics respectively by

$$\tilde{D}_\gamma = \sup_{1 \leq k \leq n, 1 \leq m \leq n_\gamma} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) 1_{\{\gamma_m^t X_j \leq \gamma_m^t X_k\}} \right| ,$$

and

$$\tilde{W}_\gamma^2 = n^{-2} n_\gamma^{-1} \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \sum_{m=1}^{n_\gamma} e_j(\hat{\theta}, \hat{\beta}) e_k(\hat{\theta}, \hat{\beta}) 1_{\{\gamma_m^t X_j \leq \gamma_m^t X_l\}} 1_{\{\gamma_m^t X_k \leq \gamma_m^t X_l\}} .$$

Note that, for the Cramér-von Mises test statistic, it is possible to find an analytic expression for the following integration:

$$A_{jkl} = \int_{\mathbb{S}_d} 1_{\{\gamma^t X_j \leq \gamma^t X_l\}} 1_{\{\gamma^t X_k \leq \gamma^t X_l\}} d\gamma .$$

We refer to Escanciano (2006a), who showed that A_{jkl} is proportional to the volume of a spherical wedge and is equal to

$$A_{jkl} = A_{jkl}^{(0)} \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)} ,$$

where

$$A_{jkl}^{(0)} = \left| \pi - \arccos \left(\frac{(X_j - X_l)^t (X_k - X_l)}{|X_j - X_l| |X_k - X_l|} \right) \right| .$$

Note that $A_{jkl}^{(0)}$ is the complementary angle between the vectors $X_j - X_l$ and $X_k - X_l$. Hence, Escanciano (2006a) provided closed-form formulas for the Cramér-von Mises test statistic. However, we have decided here to evaluate this Cramér-von Mises test statistic through a random sample $\gamma_1, \dots, \gamma_{n_\gamma}$ to ensure that the Kolmogorov-Smirnov and Cramér-von Mises test statistics are evaluated in the same way. Indeed, it is not possible to find a closed-form formula for the Kolmogorov-Smirnov test statistic.

In the context of nonparametric regression without transformation of the response, the advantage of taking the indicator weight over the exponential weight is that it avoids the choice of an arbitrary function Ψ and the advantage of taking the exponential weight over the indicator weight of Stute (1997) is that the test is less sensitive to the dimension d . Moreover, note that the tests proposed by Stute (1997) and Escanciano (2006a) are equivalent when $d = 1$. However, in the context of goodness-of-fit in nonparametric regression without transformation, the method of Escanciano (2006a) is known to avoid the curse of dimensionality.

Note that our test can also be applied with several other weighting functions we can find in the literature, see Stinchcombe and White (1998) and Escanciano (2007) among others. In this chapter, we will also consider the three additional following approaches in our simulation study:

- $w(X, x, \gamma) = \exp(x^t X)$, $\Pi = \chi$, $\Psi_n(x, \gamma) = \hat{F}_X(x)$ and $\Psi(x, \gamma) = F_X(x)$ which gives the following test statistics:

$$D_{\text{exp}} = \max_{1 \leq k \leq n} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) \exp(X_k^t X_j) \right|,$$

and

$$W_{\text{exp}}^2 = n^{-2} \sum_{k=1}^n \left(\sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) \exp(X_k^t X_j) \right)^2.$$

- $w(X, x, \gamma) = (1 + \exp(-x^t X))^{-1}$, $\Pi = \chi$, $\Psi_n(x, \gamma) = \hat{F}_X(x)$ and $\Psi(x, \gamma) = F_X(x)$ which gives the following test statistics:

$$D_{1/\text{exp}} = \max_{1 \leq k \leq n} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) (1 + \exp(-X_k^t X_j))^{-1} \right|,$$

and

$$W_{1/\text{exp}}^2 = n^{-2} \sum_{k=1}^n \left(\sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) (1 + \exp(-X_k^t X_j))^{-1} \right)^2.$$

- $w(X, x, \gamma) = \sin(x^t X)$, $\Pi = \chi$, $\Psi_n(x, \gamma) = \hat{F}_X(x)$ and $\Psi(x, \gamma) = F_X(x)$ which gives the following test statistics:

$$D_{\text{sin}} = \max_{1 \leq k \leq n} \left| n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) \sin(X_k^t X_j) \right|,$$

and

$$W_{\text{sin}}^2 = n^{-2} \sum_{k=1}^n \left(\sum_{j=1}^n e_j(\hat{\theta}, \hat{\beta}) \sin(X_k^t X_j) \right)^2.$$

3.3 Asymptotic results

Recall the following notations already introduced at the beginning of Section 2.3:

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

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

where

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

is a $(q \times 1)$ -vector and $\beta = (\beta_1, \dots, \beta_q)^t$. Moreover, we also define $h(x, \beta) = \frac{\partial m_\beta(x)}{\partial \beta}$ and

$$H(x, \gamma, \beta) = E[w(X, x, \gamma)h(X, \beta)] .$$

3.3.1 Under H_0

To start, we introduce three theorems. The first one establishes the limiting process of R_n , the second one states that the process R_n^1 can be expressed in terms of the process R_n and a sum of iid terms up to a negligible term, and the last one states that we can “replace” $\Psi_n(x, \gamma)$ by $\Psi(x, \gamma)$ in the definition of W_n^2 plus a negligible term. Combining these three results, we will next easily obtain the limiting process of R_n^1 and the limiting distributions of the test statistics. The identification conditions (I1)-(I2) and the regularity conditions (A1)-(A8) and (B1)-(B2) under which the results of this section are valid, can be found in Sections 1.2.2, 2.7 and 3.7 respectively. The proofs of these results are given in Section 3.8.

Theorem 3.3.1. *Assume (I1), (I2), (A1)-(A8), (B1) and (B2). Then, under H_0 , the process R_n converges weakly to R_∞ , where R_∞ is a centered Gaussian process with covariance function given by*

$$C(x_1, \gamma_1, x_2, \gamma_2) = \sigma^2 E[w(X, x_1, \gamma_1)w(X, x_2, \gamma_2)] ,$$

for $(x_1, \gamma_1), (x_2, \gamma_2) \in \Pi$.

Theorem 3.3.2. *Assume (I1), (I2), (A1)-(A8), (B1) and (B2). Then, under H_0 ,*

$$R_n^1(x, \gamma) = R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n G(x, \gamma, X_i, Y_i, \theta_0, \beta_0) + o_P(1) ,$$

where

$$\begin{aligned} G(x, \gamma, X, Y, \theta, \beta) &= H^t(x, \gamma, \beta) \eta_\beta(X, Y) + E[w(X, x, \gamma)(\dot{\Lambda}_\theta(Y))^t] g(X, Y) \\ &\quad - H^t(x, \gamma, \beta) \Omega^{-1} E \left[\frac{\partial m_\beta(X)}{\partial \beta} (\dot{\Lambda}_\theta(Y))^t \right] g(X, Y) . \end{aligned} \quad (3.3.1)$$

Theorem 3.3.3. *Assume (I1), (I2), (A1)-(A8), (B1) and (B2). Then, under H_0 ,*

$$W_n^2 = \int_{\Pi} [R_n^1(x, \gamma)]^2 d\Psi_n(x, \gamma) = \int_{\Pi} [R_n^1(x, \gamma)]^2 d\Psi(x, \gamma) + o_P(1) .$$

As a consequence of the previous theorems, we obtain successively the limiting process of R_n^1 and the limiting distributions of the Kolmogorov-Smirnov and Cramér-von Mises test statistics under the null hypothesis in the two following corollaries.

Corollary 3.3.4. *Assume (I1), (I2), (A1)-(A8), (B1) and (B2). Then, under H_0 , the process R_n^1 converges weakly to R_∞^1 , where R_∞^1 is a centered Gaussian process with covariance function given by*

$$\begin{aligned} C_1(x_1, \gamma_1, x_2, \gamma_2) &= C(x_1, \gamma_1, x_2, \gamma_2) - E[G(x_2, \gamma_2, X, Y, \theta_0, \beta_0) w(X, x_1, \gamma_1) \varepsilon] \\ &\quad - E[G(x_1, \gamma_1, X, Y, \theta_0, \beta_0) w(X, x_2, \gamma_2) \varepsilon] \\ &\quad + E[G(x_1, \gamma_1, X, Y, \theta_0, \beta_0) G(x_2, \gamma_2, X, Y, \theta_0, \beta_0)] , \end{aligned}$$

for $(x_1, \gamma_1), (x_2, \gamma_2) \in \Pi$.

Corollary 3.3.5. *Assume (I1), (I2), (A1)-(A8), (B1) and (B2). Then, under H_0 ,*

$$D_n \xrightarrow{d} \sup_{(x, \gamma) \in \Pi} |R_\infty^1(x, \gamma)| \quad \text{and} \quad W_n^2 \xrightarrow{d} \int_{\Pi} [R_\infty^1(x, \gamma)]^2 d\Psi(x, \gamma) .$$

3.3.2 Under H_{1n}

We consider the following local alternative hypothesis in order to study the power of the test statistics:

$$H_{1n} : m(x) = m_{\beta_0}(x) + n^{-1/2} r(x) \text{ for all } x \quad (3.3.2)$$

for some fixed function $r \neq 0$. First, we obtain the analog of Theorem 3.3.2 under H_{1n} .

Theorem 3.3.6. *Assume (I1), (I2), (A1)-(A8), (A10), (B1) and (B2). Then, under H_{1n} ,*

$$R_n^1(x, \gamma) = R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n G(x, \gamma, X_i, Y_i, \theta_0, \tilde{\beta}_{0n}) + b_2(x, \gamma) + o_P(1) ,$$

where $\tilde{\beta}_{0n}$ is a minimizer over $\beta \in \mathcal{B}$ of $E[(m_\beta(X) - m(X))^2]$ and

$$b_2(x, \gamma) = -H^t(x, \gamma, \beta_0)\Omega^{-1} \int r(u) \frac{\partial m_{\beta_0}(u)}{\partial \beta} dF_X(u) + E[w(X, x, \gamma)r(X)] .$$

We remark the presence of an additional bias term b that depends on the deviation function r in comparison with Theorem 3.3.2. Note that this bias term is exactly the same as in the case where the transformation of the response would be known (see formula (3.3) in Stute (1997) for the case where $w(t, x, \gamma) = 1_{\{t \leq x\}}$). In other words, the estimation of the transformation parameter θ has no impact on the asymptotic bias under H_{1n} . This is because θ is estimated based on a nonparametric estimator of $m(\cdot)$ (see formula (1.2.9)).

Finally, the following corollaries give respectively the limiting process of R_n^1 and the limiting distributions of the two test statistics under the local alternative.

Corollary 3.3.7. *Assume (I1), (I2), (A1)-(A8), (A10), (B1) and (B2). Then, under H_{1n} , the process R_n^1 converges weakly to $R_\infty^1 + b_2$, where R_∞^1 is the same centered Gaussian process as in Corollary 3.3.4.*

Corollary 3.3.8. *Assume (I1), (I2), (A1)-(A8), (A10), (B1) and (B2). Then, under H_{1n} ,*

$$D_n \xrightarrow{d} \sup_{(x, \gamma) \in \Pi} |R_\infty^1(x, \gamma) + b_2(x, \gamma)| ,$$

and

$$W_n^2 \xrightarrow{d} \int_{\Pi} [R_\infty^1(x, \gamma) + b_2(x, \gamma)]^2 d\Psi(x, \gamma) .$$

Since the bias term $b_2(x, \gamma)$ is the same as in the case without transformation of the response, we can directly use the results that have been obtained in Stute (1997) and Escanciano (2006a), who studied the bias term when the response is not transformed and when $w(t, x, \gamma) = 1_{\{t \leq x\}}$ and $w(t, x, \gamma) = 1_{\{\gamma^t t \leq x\}}$, respectively. They showed that the tests have power against all alternatives r that are such that the projection of r onto the orthogonal complement of the functions $\int \frac{\partial m_{\beta_0}}{\partial \beta_l}(x) dF_X(x)$, $l = 1, \dots, q$ stays away from $\text{span}\{\frac{\partial m_{\beta_0}}{\partial \beta_1}, \dots, \frac{\partial m_{\beta_0}}{\partial \beta_q}\}$.

In practice, we need to estimate the limiting distributions of D_n and W_n^2 obtained in Corollaries 3.3.5 and 3.3.8. However, this implies the estimation of f_ε , $\dot{f}_\varepsilon$, f'_ε , $\dot{m}$ and f_X and consequently the introduction of new bandwidths

which is possible but not easy. Therefore, we prefer to use a bootstrap procedure in order to approximate the limiting distributions of D_n and W_n^2 under H_0 in practice. This bootstrap procedure is the same as in Chapter 2 and is described in Section 2.3.3.

3.4 Simulations

3.4.1 Comparison of both tests in dimension 1

In this section, we perform some simulations in order to evaluate the performance of our test statistics for small samples and also to compare the results given by these tests to those given by the tests based on the error distribution function developed in Chapter 2.

As the different Cramér-von Mises test statistics give in most cases similar or better results than the corresponding Kolmogorov-Smirnov test statistics, we decide in this section to compare only the results given by the Cramér-von Mises test statistics, i.e., T_{CM} , W_1^2 , $W_{\exp_i}^2$, $\widehat{W}_\gamma^2$, $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$.

For the estimation of θ , β , h and g in practice, we proceed exactly as explained in Section 2.4.1. Moreover, the critical values of the different test statistics are obtained with the residual bootstrap procedure explained in Section 2.3.3. We will use $B = 250$ in particular.

First, we perform simulations in dimension $d = 1$. The simulated model is $\Lambda_\theta(Y_i) = \beta_1 + \beta_2 X_i + c(X_i) + \varepsilon_i$, where $\Lambda_\theta(Y)$ is the Yeo and Johnson (2000) transformation:

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

Note that the Yeo and Johnson (2000) transformation is an extension of the Box and Cox (1964) transformation that allows the response variable Y to be negative. We will consider three different values of the parameter transformation 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 $\beta_0 = (\beta_{10}, \beta_{20}) = (3, 5)$. Moreover, $X_1, \dots, X_n$ are independent uniform random variables on $[0, 1]$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent standard normal random variables truncated on $[-3, 3]$. We will also consider the cases where $\varepsilon_1, \dots, \varepsilon_n$ are independent normal random variables with zero mean and standard deviation 0.5 truncated on $[-3, 3]$ and where $\varepsilon_1, \dots, \varepsilon_n$ are independent student-t random variables with degrees of freedom equal to 10. We consider the following null hypothesis:

$$H_0 : m(x) = \beta_1 + \beta_2 x \quad \text{for all } x \text{ and for some } (\beta_1, \beta_2) \in \mathbb{R}^2.$$

We consider different deviation functions $c(x)$ from the null hypothesis: $c(x) = 2x^2$, $c(x) = 3x^2$, $c(x) = 4x^2$, $c(x) = 5x^2$, $c(x) = 2\exp(x)$, $c(x) = 3\exp(x)$, $c(x) = 4\exp(x)$, $c(x) = 5\exp(x)$, $c(x) = 0.25\sin(2\pi x)$, $c(x) = 0.5\sin(2\pi x)$, $c(x) = 0.75\sin(2\pi x)$, $c(x) = \sin(2\pi x)$.

Tables 3.1 to 3.3 show respectively the percentages of rejection obtained with the test statistics T_{CM} , W_1^2 , $W_{\exp_i}^2$, $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$ for 500 samples of size $n = 200$ under the null hypothesis and under the different deviations $c(x)$ we have introduced above. The nominal level is 10%. We also remind that the test statistics W_1^2 and $\widetilde{W}_\gamma^2$ are equivalent when $d = 1$, hence Tables 3.1 to 3.3 only include the results obtained with W_1^2 and not with $\widetilde{W}_\gamma^2$.

First, under H_0 , we observe that the different estimations of the nominal level are globally good for all test statistics, choices of θ_0 and distributions of ε except if we use the test statistic W_1^2 when θ_0 is increasing. In that case, the estimation of the nominal level is increasingly small.

Next, under the alternative, the power is largest when $\varepsilon \sim N(0, 0.5^2)$, followed by $\varepsilon \sim N(0, 1)$ and finally by $\varepsilon \sim t_{10}$. This seems logical because we increase consequently the variance of ε when we change from one situation to another. In a similar way, we observe generally that the power is largest for $\theta_0 = 0$, followed by $\theta_0 = 0.5$ and finally by $\theta_0 = 1$. This conclusion is the same as in Sections 2.4.4 and 2.4.5 and is logical with respect to the results obtained by Linton et al. (2008). Indeed, 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 $\theta_0 = 0$. This suggests that the estimations of β_0 and the values of the test statistics are less accurate when θ_0 increases from 0 to 1 since these estimations and values depend on $\widehat{\theta}$. Hence, it seems logical that the power is largest for $\theta_0 = 0$, followed by $\theta_0 = 0.5$ and finally by $\theta_0 = 1$.

Moreover, again under H_1 , when the deviation from the null hypothesis is monotone, for example $c(x) = cx^2$ and $c(x) = c\exp(x)$, the highest power is generally obtained with the test statistics $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$. This last conclusion is valid for all tested values of θ_0 . On the other hand, if the deviation from the null hypothesis is non monotone, for example $c(x) = c\sin(2\pi x)$, the highest power is obtained with the test statistic $W_{\exp_i}^2$ when the deviation is close to the null hypothesis ($c = 0.25$ and sometimes $c = 0.5$) and is obtained either with the test statistic T_{CM} or with the test statistic W_1^2 when the deviation is less close to the null hypothesis ($c = 0.75$ and $c = 1$). This depends on the value of θ_0 . If $\theta_0 = 0$, we will prefer W_1^2 and when θ_0 increases, we will prefer T_{CM} which is in line with what happens under H_0 . Finally, note that the test statistics $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$ give very small power when the deviation from H_0 is non monotone.

$c(x)$	$\theta_0 = 0$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	10.2	10.2	13.4	12.2	11.4	12.4
$2x^2$	27.2	46.6	37.6	50.0	50.6	51.2
$3x^2$	48.8	75.4	68.0	78.8	79.0	78.6
$4x^2$	66.2	89.2	86.4	91.6	91.0	91.0
$5x^2$	78.2	94.4	91.8	94.4	94.4	94.6
$2 \exp(x)$	23.6	33.2	26.4	38.2	39.2	38.8
$3 \exp(x)$	34.2	55.0	42.0	60.8	60.2	60.0
$4 \exp(x)$	49.0	69.6	60.0	71.6	73.2	72.2
$5 \exp(x)$	62.8	78.2	72.2	81.0	81.2	80.8
$0.25 \sin(2\pi x)$	21.4	21.0	29.4	12.4	13.8	13.4
$0.5 \sin(2\pi x)$	56.2	59.0	56.4	13.4	16.8	15.6
$0.75 \sin(2\pi x)$	86.0	92.4	70.0	16.0	20.4	20.6
$\sin(2\pi x)$	98.4	99.8	76.8	17.8	26.0	27.6

$c(x)$	$\theta_0 = 0.5$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	10.4	7.2	12.4	12.8	11.4	12.4
$2x^2$	27.2	41.2	34.6	50.4	48.8	49.2
$3x^2$	49.2	72.0	62.6	78.8	78.2	78.8
$4x^2$	66.0	88.2	82.8	91.4	90.8	91.0
$5x^2$	77.4	95.2	91.2	97.0	96.4	96.4
$2 \exp(x)$	22.8	22.8	19.8	32.2	33.0	31.6
$3 \exp(x)$	34.6	36.2	27.4	52.8	53.6	53.0
$4 \exp(x)$	47.0	52.4	36.4	64.8	67.6	66.0
$5 \exp(x)$	61.2	63.4	46.6	72.6	74.4	73.2
$0.25 \sin(2\pi x)$	24.8	13.8	28.2	12.0	13.2	12.4
$0.5 \sin(2\pi x)$	54.8	45.0	56.6	12.8	14.8	13.6
$0.75 \sin(2\pi x)$	84.4	82.8	69.8	12.8	16.4	17.8
$\sin(2\pi x)$	98.0	94.0	76.2	13.8	21.2	23.4

$c(x)$	$\theta_0 = 1$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	9.8	5.0	10.8	10.6	9.6	10.4
$2x^2$	29.2	38.6	32.0	49.2	47.8	48.2
$3x^2$	47.6	67.8	57.8	77.4	78.8	78.4
$4x^2$	65.8	86.8	82.0	90.4	89.8	89.8
$5x^2$	76.8	93.2	87.8	96.4	96.2	95.8
$2 \exp(x)$	22.0	17.8	16.4	29.4	30.0	29.4
$3 \exp(x)$	34.4	30.8	20.8	47.0	47.6	45.6
$4 \exp(x)$	47.6	44.6	30.0	61.8	62.4	61.6
$5 \exp(x)$	61.6	56.0	36.4	67.6	69.2	68.8
$0.25 \sin(2\pi x)$	22.8	10.6	27.8	10.4	10.6	11.2
$0.5 \sin(2\pi x)$	55.4	34.0	55.4	10.6	12.8	11.8
$0.75 \sin(2\pi x)$	84.0	69.2	68.8	11.2	14.0	14.8
$\sin(2\pi x)$	98.2	87.8	75.6	11.6	15.8	18.0

Table 3.1: Percentage of rejection for $\theta_0 = 0, 0.5, 1$ and for $\varepsilon \sim N(0, 1)$.

$c(x)$	$\theta_0 = 0$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	11.0	10.8	11.8	12.4	12.0	12.6
$2x^2$	42.4	63.6	52.2	66.4	66.8	66.6
$3x^2$	65.0	85.4	79.4	87.8	87.4	87.0
$4x^2$	83.4	95.6	91.2	96.2	95.4	95.4
$5x^2$	92.4	98.2	96.2	98.2	98.2	98.2
$2 \exp(x)$	32.4	46.6	35.2	51.2	51.2	51.0
$3 \exp(x)$	47.6	63.2	52.0	67.6	68.2	67.8
$4 \exp(x)$	64.8	76.4	65.2	78.6	79.0	78.8
$5 \exp(x)$	74.4	83.2	73.6	85.0	85.4	85.0
$0.25 \sin(2\pi x)$	36.0	32.8	42.2	14.0	14.2	13.8
$0.5 \sin(2\pi x)$	81.8	88.8	65.6	16.2	24.4	22.8
$0.75 \sin(2\pi x)$	98.2	99.8	78.2	21.0	29.6	29.4
$\sin(2\pi x)$	99.8	99.8	83.4	24.4	34.6	36.6

$c(x)$	$\theta_0 = 0.5$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	12.2	6.4	10.4	11.2	10.2	10.4
$2x^2$	43.0	55.4	44.8	64.0	64.4	64.2
$3x^2$	63.2	80.0	72.6	86.0	86.0	86.0
$4x^2$	82.4	92.8	87.2	94.8	94.2	94.4
$5x^2$	91.8	96.8	93.2	97.4	97.2	97.2
$2 \exp(x)$	32.8	30.8	23.2	40.4	40.4	40.8
$3 \exp(x)$	47.4	48.2	31.6	57.8	58.8	58.2
$4 \exp(x)$	62.2	57.8	39.0	66.2	68.0	67.0
$5 \exp(x)$	73.6	67.8	46.0	74.4	74.8	74.8
$0.25 \sin(2\pi x)$	35.4	20.4	38.0	13.0	14.2	14.2
$0.5 \sin(2\pi x)$	83.4	75.6	62.4	14.8	18.0	18.0
$0.75 \sin(2\pi x)$	98.6	96.2	76.4	17.0	23.4	23.4
$\sin(2\pi x)$	100.0	99.4	80.2	19.0	27.0	29.4

$c(x)$	$\theta_0 = 1$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	12.2	4.4	8.6	8.8	8.2	9.0
$2x^2$	41.2	49.2	39.0	62.2	61.4	60.8
$3x^2$	63.8	80.8	68.2	85.6	86.2	86.2
$4x^2$	82.0	92.4	84.8	93.8	93.2	93.8
$5x^2$	91.6	96.8	92.8	97.2	96.6	96.6
$2 \exp(x)$	31.2	27.4	17.6	37.6	38.6	37.8
$3 \exp(x)$	47.2	38.2	24.0	51.6	53.0	52.2
$4 \exp(x)$	62.8	51.0	35.0	60.2	61.4	60.6
$5 \exp(x)$	73.4	61.4	36.4	69.2	69.8	69.0
$0.25 \sin(2\pi x)$	34.4	12.0	38.8	9.8	11.4	11.6
$0.5 \sin(2\pi x)$	82.0	57.4	61.2	12.0	15.6	16.0
$0.75 \sin(2\pi x)$	98.6	90.6	75.0	14.4	20.4	20.6
$\sin(2\pi x)$	100.0	96.8	78.6	15.2	22.0	23.6

Table 3.2: Percentage of rejection for $\theta_0 = 0, 0.5, 1$ and for $\varepsilon \sim N(0, 0.5^2)$.

$c(x)$	$\theta_0 = 0$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	11.0	10.6	11.4	12.6	12.2	12.8
$2x^2$	24.4	34.4	24.4	38.0	37.4	37.0
$3x^2$	36.2	54.8	44.8	61.2	59.0	58.8
$4x^2$	54.8	75.0	65.4	79.4	78.4	78.6
$5x^2$	71.0	87.4	80.4	90.2	89.6	90.0
$2 \exp(x)$	21.4	30.6	22.2	34.0	35.0	34.8
$3 \exp(x)$	31.2	46.4	34.0	50.4	51.4	50.4
$4 \exp(x)$	44.6	60.0	49.4	63.0	64.4	63.4
$5 \exp(x)$	54.4	66.8	55.0	68.4	68.8	68.6
$0.25 \sin(2\pi x)$	18.6	17.8	25.2	13.2	12.6	12.6
$0.5 \sin(2\pi x)$	46.6	47.2	52.0	13.4	15.0	15.0
$0.75 \sin(2\pi x)$	81.6	85.6	67.4	16.2	18.4	18.8
$\sin(2\pi x)$	96.2	97.4	75.0	17.8	23.4	24.4
$c(x)$	$\theta_0 = 0.5$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	11.2	7.4	10.4	13.0	12.2	12.8
$2x^2$	28.8	38.2	29.2	46.8	45.8	45.2
$3x^2$	42.4	60.8	48.8	69.2	68.6	68.8
$4x^2$	59.8	78.4	70.2	84.8	84.8	84.8
$5x^2$	74.2	88.8	81.4	91.2	91.4	91.4
$2 \exp(x)$	23.0	18.6	16.0	28.6	30.4	29.6
$3 \exp(x)$	31.4	30.2	22.4	41.8	42.6	42.0
$4 \exp(x)$	43.6	40.6	27.8	51.8	52.0	50.8
$5 \exp(x)$	54.6	49.4	31.8	58.4	59.2	58.4
$0.25 \sin(2\pi x)$	18.6	12.8	24.6	12.6	12.2	12.8
$0.5 \sin(2\pi x)$	47.6	32.4	51.2	13.0	15.2	15.6
$0.75 \sin(2\pi x)$	81.0	64.2	64.8	15.2	17.4	15.4
$\sin(2\pi x)$	95.8	83.4	71.4	13.4	19.0	20.4
$c(x)$	$\theta_0 = 1$					
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	11.0	5.6	9.0	11.6	10.6	11.2
$2x^2$	26.0	34.6	30.0	43.4	42.8	43.6
$3x^2$	40.8	55.4	47.6	64.8	64.6	65.0
$4x^2$	55.0	74.8	65.4	81.8	81.6	82.0
$5x^2$	74.6	86.6	80.2	90.6	89.8	90.2
$2 \exp(x)$	22.4	13.4	14.4	27.2	27.4	26.8
$3 \exp(x)$	32.0	23.0	17.8	37.6	38.6	37.8
$4 \exp(x)$	43.0	32.0	23.0	45.2	46.6	45.8
$5 \exp(x)$	55.0	41.8	27.0	54.2	55.6	54.2
$0.25 \sin(2\pi x)$	19.0	9.4	25.2	12.6	10.8	11.4
$0.5 \sin(2\pi x)$	47.2	24.2	51.6	11.8	13.0	12.6
$0.75 \sin(2\pi x)$	80.8	52.8	66.0	11.0	14.2	13.8
$\sin(2\pi x)$	95.4	68.6	70.4	11.6	16.6	17.2

Table 3.3: Percentage of rejection for $\theta_0 = 0, 0.5, 1$ and for $\varepsilon \sim t_{10}$.

3.4.2 Comparison of both tests in dimension 2

Finally, we perform simulations in dimension $d = 2$. The simulated model is $\Lambda_\theta(Y_i) = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + c(X_{1i}, X_{2i}) + \varepsilon_i$, where $\Lambda_\theta(Y)$ is again the Yeo and Johnson (2000) transformation. We will consider the same three different values of the parameter transformation θ : $\theta_0 = 0$, $\theta_0 = 0.5$ and $\theta_0 = 1$. The true value of the parameter β is $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30}) = (3, 2, 1)$. Moreover, $X_{11}, \dots, X_{1n}$ and $X_{21}, \dots, X_{2n}$ are independent and uniformly distributed on the unit square and $\varepsilon_1, \dots, \varepsilon_n$ are independent normal random variables with zero mean and standard deviation 0.5 truncated on $[-3, 3]$. We consider the following null hypothesis:

$$H_0 : m(x) = \beta_1 + \beta_2 x_1 + \beta_3 x_2 \quad \text{for all } x \text{ and for some } (\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3.$$

We consider different deviations $c(x_1, x_2)$ from the null hypothesis: $c(x_1, x_2) = 2x_1x_2$, $c(x_1, x_2) = 3x_1x_2$, $c(x_1, x_2) = 4x_1x_2$, $c(x_1, x_2) = 5x_1x_2$, $c(x_1, x_2) = 0.5x_1 \sin(2\pi x_2)$, $c(x_1, x_2) = x_1 \sin(2\pi x_2)$, $c(x_1, x_2) = 1.5x_1 \sin(2\pi x_2)$, $c(x_1, x_2) = 2x_1 \sin(2\pi x_2)$, for 300 samples of size $n = 300$ and with $n_\gamma = 200$. For the estimation of θ , β , h and g and the bootstrap procedure, we again proceed as described in Sections 2.4.1 and 2.3.3. Table 3.4 shows the percentage of rejection obtained with the test statistics T_{CM} , W_1^2 , $W_{\exp_i}^2$, $\widetilde{W}_\gamma^2$, $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$ under the null hypothesis and under the different deviations $c(x_1, x_2)$ we have introduced above. Note that the results given by W_1^2 and $\widetilde{W}_\gamma^2$ will be here different since $d \neq 1$.

Table 3.4 shows that the estimations of the nominal level are generally too low, especially for $\theta_0 = 0$ when we consider the different new test statistics developed in this chapter. This problem was already encountered in Section 2.4.5 and is due to the poor nonparametric 2-dimensional estimation of the function $m(\cdot)$. This suggests that the method suffers from curse-of-dimensionality problems, implying that samples of size $n = 300$ are not always large enough.

Next, under the alternative, the highest power is obtained with the test statistic $W_{\exp_i}^2$ when the deviation from the null hypothesis is monotone, for example $c(x_1, x_2) = cx_1x_2$, and is obtained with the test statistic T_{CM} when the deviation from the null hypothesis is non monotone, for example $c(x_1, x_2) = cx_1 \sin(2\pi x_2)$. We can also observe that the test statistics W_1^2 and $\widetilde{W}_\gamma^2$ give very poor power, even if it is a little bit better when the deviation is non monotone. This suggests that it is not a good idea to use indicator weights when the dimension d of X is increasing. Finally, the test statistics $W_{\exp}^2$, $W_{1/\exp}^2$ and $W_{\sin}^2$ give a higher power than the one obtained with the indicator weight, but smaller than the one obtained with T_{CM} and $W_{\exp_i}^2$ with a monotone deviation.

$c(x_1, x_2)$	$\theta_0 = 0$						
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$\tilde{W}_\gamma^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	10.0	2.3	3.7	2.0	2.3	2.3	2.3
$2x_1x_2$	9.7	2.0	16.7	1.3	9.3	8.3	9.3
$3x_1x_2$	17.3	2.3	33.3	2.0	13.7	13.7	14.0
$4x_1x_2$	28.3	3.0	55.3	1.7	17.7	18.0	18.3
$5x_1x_2$	31.0	4.0	69.7	3.0	26.3	26.3	26.3
$0.5x_1 \sin(2\pi x_2)$	33.0	5.0	8.7	6.3	7.3	7.3	7.7
$x_1 \sin(2\pi x_2)$	68.0	14.3	17.3	19.7	26.0	25.3	26.0
$1.5x_1 \sin(2\pi x_2)$	88.3	16.0	16.7	24.0	38.3	37.0	39.0
$2x_1 \sin(2\pi x_2)$	90.0	17.3	17.7	27.7	44.0	41.3	45.3
$c(x_1, x_2)$	$\theta_0 = 0.5$						
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$\tilde{W}_\gamma^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	7.0	7.7	10.7	9.0	9.0	9.3	9.0
$2x_1x_2$	6.3	6.7	21.7	6.3	15.3	15.7	16.0
$3x_1x_2$	13.0	4.7	38.3	5.0	16.7	16.7	16.7
$4x_1x_2$	18.7	8.7	59.0	6.7	21.3	21.0	21.0
$5x_1x_2$	23.0	9.7	73.7	7.0	20.7	22.0	20.3
$0.5x_1 \sin(2\pi x_2)$	19.0	14.3	19.3	16.7	18.0	18.7	18.3
$x_1 \sin(2\pi x_2)$	51.7	23.3	26.3	28.0	37.3	38.0	37.3
$1.5x_1 \sin(2\pi x_2)$	82.3	22.0	26.3	26.7	46.3	47.3	47.3
$2x_1 \sin(2\pi x_2)$	91.0	25.0	23.3	32.3	52.3	50.7	53.7
$c(x_1, x_2)$	$\theta_0 = 1$						
	T_{CM}	W_1^2	$W_{\exp_i}^2$	$\tilde{W}_\gamma^2$	$W_{\exp}^2$	$W_{1/\exp}^2$	$W_{\sin}^2$
0	8.7	8.7	8.7	10.3	7.3	7.0	7.7
$2x_1x_2$	5.3	5.0	20.0	4.0	15.0	15.0	15.3
$3x_1x_2$	10.3	4.3	30.7	4.0	18.3	19.7	19.0
$4x_1x_2$	19.3	9.7	51.0	8.3	22.3	23.3	23.0
$5x_1x_2$	22.0	10.3	67.7	9.0	23.0	24.0	24.0
$0.5x_1 \sin(2\pi x_2)$	30.0	12.0	14.3	15.0	16.0	16.0	16.7
$x_1 \sin(2\pi x_2)$	65.7	19.3	24.0	22.3	32.7	34.3	33.3
$1.5x_1 \sin(2\pi x_2)$	86.3	20.7	22.3	24.3	41.7	41.0	42.7
$2x_1 \sin(2\pi x_2)$	92.3	19.3	18.7	25.3	42.7	42.3	44.3

Table 3.4: Percentage of rejection $\theta_0 = 0, 0.5, 1$ and for $\varepsilon \sim N(0, 0.5^2)$.

3.5 Application on ultrasonic data

3.5.1 Previous analyses of the data

We apply our testing procedure to a ultrasonic calibration data set composed of 214 observations. The data can be found on the website <http://www.itl.nist.gov/div898/handbook/pmd/section6/pmd631.htm> and comes from the NIST/SEMATECH e-Handbook of Statistical Methods. The response Y is ultrasonic response and the covariate X is metal distance.

This data set has already been analyzed in the e-Handbook and in Neumeyer et al. (2016). A simple plot of the data shows that the ultrasonic response decreases exponentially when the metal distance increases. Consequently, for this data set, the scientists tried to fit the following model: $Y = \frac{\exp(\beta_1 X)}{\beta_2 + \beta_3 X} + \varepsilon$. In the e-Handbook, three different fits are proposed: a fit based on the initial data, a fit using a transformation of the response and a weighted fit in a sense that we give less weight to less precise measurements. It has been shown that these three fits are quite close but the transformed fit and the weighted fit have the advantage of giving a regression model that satisfies the assumptions. Indeed, the assumption of homogeneous variance seems to be violated when a fit using the initial data is performed. Three transformed fits, i.e., using respectively square root, logarithmic and reciprocal transformations of the response, were analyzed in order to stabilize the variance of the residuals. In conclusion, the square root transformation was the most appropriate transformation based on the constancy of the variation of the residuals.

Using previous analyses, we can consequently find that the data satisfy the model $\sqrt{Y_i} = m(X_i) + \varepsilon_i$, $i = 1, \dots, n$. The goal of Neumeyer et al. (2016) was to verify such validity with their own procedure and without using the square root transformation. They estimated the transformation and they concluded that this data set satisfies the assumption of a homoscedastic transformation model using a Box-Cox transformation.

3.5.2 Results

In the e-Handbook, we can find that $\widehat{\sqrt{Y}} = \frac{\exp(-0.015X)}{0.0807+0.0639X}$ is the fitted transformed model. In this section, we consider a Box and Cox (1964) transformation of the response variable. The estimated transformation parameter is equal to $\hat{\theta} = 0.43$. Note that this transformation is very similar to the square root transformation of the e-Handbook. We will check the following natural goodness-of-fit:

$$H_0 : m(x) = \frac{\exp(\beta_1 x)}{\beta_2 + \beta_3 x} \text{ for all } x \quad (\text{test 1}) .$$

We will also check the following exponential, inverse linear, linear and quadratic goodness-of-fits:

- $H_0 : m(x) = \beta_1 + \beta_2 \exp(\beta_3 x)$ for all x (test 2) ,
- $H_0 : m(x) = \frac{\beta_1}{\beta_2 + \beta_3 x}$ for all x (test 3) ,
- $H_0 : m(x) = \beta_1 + \beta_2 x$ for all x (test 4) ,
- $H_0 : m(x) = \beta_1 + \beta_2 x + \beta_3 x^2$ for all x (test 5) .

We use the Cramér-von Mises test statistic defined in Section 2.2.2 and the ones in chapter. The distributions and p-values of these test statistics are approximated by the bootstrap on the basis of 1000 replicates. The results are given in Table 3.5.

	test 1	test 2	test 3	test 4	test 5
T_{CM}	0.375	0.525	0.090	0	0
W_1^2	0.271	0.093	0.017	0.001	0.004
$W_{\exp_i}^2$	0.130	0.103	0.033	0	0.004
$W_{\exp}^2$	0.024	0.835	0.291	0.002	0.011
$W_{1/\exp}^2$	0.176	0.085	0.028	0	0.004
$W_{\sin}^2$	0.379	0.067	0.036	0	0.004

Table 3.5: p-values of the different goodness-of-fit tests for the ultrasonic calibration data.

Table 3.5 indicates that there is no evidence against the fitted transformed model introduced in the e-Handbook when $\alpha = 0.01$ whereas only the test statistic $W_{\exp}^2$ rejects this model when $\alpha = 0.05$. Note that, in this case, our estimations of β_i , $i = 1, 2, 3$ are $\hat{\beta}_1 = -0.0569$, $\hat{\beta}_2 = 0.0568$ and $\hat{\beta}_3 = 0.0367$, the small differences with respect to the e-Handbook are due to the use of a Box-Cox transformation with $\hat{\theta} = 0.43$ instead of the square root transformation. Next, we observe that there is also no evidence against the exponential model introduced in test 2 when $\alpha = 0.05$. Finally, for most test statistics, the model introduced in test 3 is rejected as well as the linear and the quadratic fits of tests 4 and 5.

3.6 Conclusions and recommendations

First, under H_0 and when $d = 1$, the different estimations of the nominal level are globally good for all test statistics, choices of θ_0 and distributions of ε except if we use the weighting function $1_{\{X \leq x\}}$ when θ_0 is increasing.

Moreover, under H_0 and when $d = 2$, the estimations of the nominal level are generally too low, especially for $\theta_0 = 0$. This problem is due to the poor nonparametric 2-dimensional estimation of the function $m(\cdot)$ and suggests that

the method suffers from curse-of-dimensionality problems, which was already encountered in Chapter 2.

Next, under the alternative and when $d = 1$, the power is largest when $\varepsilon \sim N(0, 0.5^2)$, followed by $\varepsilon \sim N(0, 1)$ and finally by $\varepsilon \sim t_{10}$ and is largest for $\theta_0 = 0$, followed by $\theta_0 = 0.5$ and finally by $\theta_0 = 1$, which was already encountered in Chapter 2.

Moreover, when the deviation from the null hypothesis is monotone, the highest power is generally obtained with the test statistics based on the weights $\exp(x^t X)$, $(1 + \exp(-x^t X))^{-1}$ and $\sin(x^t X)$ but these test statistics give very small power when the deviation from H_0 is non monotone. If the deviation from the null hypothesis is non monotone and close to the null hypothesis, the highest power is obtained with the test statistic based on the weight $\exp(ix^t X)$ and when the deviation is far away from the null hypothesis, the highest power is obtained with the weight $1_{\{X \leq x\}}$ for $\theta_0 = 0$ and with the test constructed in Chapter 2 for $\theta_0 = 0.5$ and $\theta_0 = 1$.

Finally, under the alternative and when $d = 2$, the highest power is obtained with the weighting function $\exp(ix^t X)$ when the deviation from the null hypothesis is monotone and is obtained with the test constructed in Chapter 2 when the deviation from the null hypothesis is non monotone.

Moreover, the test statistics based on the weights $1_{\{X \leq x\}}$ and $1_{\{\gamma^t X \leq x\}}$ give very poor power, which suggests that it is not a good idea to use indicator weights when the dimension d of X is increasing, and the test statistics based on the weights $\exp(x^t X)$, $(1 + \exp(-x^t X))^{-1}$ and $\sin(x^t X)$ give intermediate power.

3.7 Technical assumptions

In this chapter, we have estimated the semiparametric transformation model exactly in the same way as in Chapter 2. Hence, the main results of the asymptotic theory require several regularity conditions already introduced in Section 2.7.

Indeed, we have to assume Assumptions (A1) to (A8) and (A10) in this chapter since these conditions are mainly necessary to prove the consistency of $\hat{\theta}$ and $\hat{\beta}$, to establish an asymptotic representation of $\hat{\theta} - \theta_0$ and $\hat{\beta} - \beta_0$ and we need the same results concerning $\hat{\theta}$ and $\hat{\beta}$ to establish an iid representation of $R_n^1(x, \gamma)$.

The differences between Chapter 2 and Chapter 3 regarding the assumptions are due to the particular definitions of the test statistics when we want to establish their limiting distributions, i.e., in Chapter 2, we needed Assumption (A9) to establish the asymptotic representation of $\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon_0}(y)$. This assumption was no longer necessary here and conversely, we need in this chapter new assumptions, (B1) and (B2) introduced below, to establish an iid representation of $R_n^1(x, \gamma)$.

The new conditions (B1) and (B2), respectively on the weighting function w and the integrating functions Ψ_n and Ψ , are the following:

(B1) (i) We assume that

$$\int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} < \infty ,$$

where

$$\mathcal{W} = \{u \rightarrow w(u, x, \gamma), (x, \gamma) \in \Pi\} ,$$

and $N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))$ is the $\tilde{\varepsilon}$ -bracketing number of the class $\mathcal{W}$, P is the probability measure corresponding to the distribution of X , and $L_2(P)$ is the L_2 -norm.

- (ii) The weighting function w satisfies $\sup_{(x, \gamma) \in \Pi, t \in \chi} |w(t, x, \gamma)| < \infty$.
- (iii) The weighting function w satisfies $\sup_{t \in \chi} T_V[w(t, \cdot, \cdot)] < \infty$, where $T_V[w(t, \cdot, \cdot)]$ is the total variation of $w(t, \cdot, \cdot)$ defined on the compact set Π .

(B2) The integrating functions $\Psi_n(x, \gamma)$ and $\Psi(x, \gamma)$ satisfy the two following regularity conditions: there exists a random function $\tilde{\Psi}$ such that

$$\sup_{(x, \gamma) \in \Pi} \left| \Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma) \right| = o_P(n^{-1/2}) ,$$

and such that

$$\sup_{(x, \gamma) \in \Pi} \left| \tilde{\Phi}(x, \gamma) - \Phi(x, \gamma) \right| = o_P(1) ,$$

where

$$\Phi(x, \gamma) = \frac{\partial^{d+d_\gamma}}{\partial x_1 \dots \partial x_d \partial \gamma_1 \dots \partial \gamma_{d_\gamma}} \Psi(x, \gamma) ,$$

$$\tilde{\Phi}(x, \gamma) = \frac{\partial^{d+d_\gamma}}{\partial x_1 \dots \partial x_d \partial \gamma_1 \dots \partial \gamma_{d_\gamma}} \tilde{\Psi}(x, \gamma)$$

if $d_\gamma \neq 0$ and

$$\Phi(x, \gamma) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} \Psi(x, \gamma) ,$$

$$\tilde{\Phi}(x, \gamma) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} \tilde{\Psi}(x, \gamma) ,$$

if $d_\gamma = 0$.

Finally, the two following propositions establish that Conditions (B1)(i) and (B2) are satisfied for the different weighting functions w and the corresponding integrating functions Ψ_n and Ψ that we have used in the simulations and in the application in Sections 3.4 and 3.5.

Proposition 3.7.1. Assume that $\sup_{x \in \chi_r} f_\gamma(x) < \infty$, where f_γ is the density function of $\gamma^t X$, χ_r is the support of $\gamma^t X$ and is a compact subset of $\mathbb{R}$. Then, Condition (B1)(i) is satisfied for the weighting functions $w(X, x, \gamma) = 1_{\{X \leq x\}}$, $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$, $w(X, x, \gamma) = \exp(ix^t X)$, $w(X, x, \gamma) = \sin(x^t X)$, $w(X, x, \gamma) = \exp(x^t X)$ and $w(X, x, \gamma) = (1 + \exp(-x^t X))^{-1}$.

Proposition 3.7.2. Condition (B2) is satisfied for the three following cases:

- (a) $\Psi_n(x, \gamma) = \Psi(x, \gamma)$, when the weighting function used in Bierens (1982) is considered.
- (b) $\Psi_n(x, \gamma) = \hat{F}_X(x)$ and $\Psi(x, \gamma) = F_X(x)$, when the weighting function used in Stute (1997) is considered and also used when $w(X, x, \gamma) = \sin(x^t X)$, $w(X, x, \gamma) = \exp(x^t X)$ and $w(X, x, \gamma) = (1 + \exp(-x^t X))^{-1}$.
- (c) $d\Psi_n(x, \gamma) = d\hat{F}_{n, \gamma}(x)d\gamma$ and $\Psi(x, \gamma) = F_\gamma(x)$, when the weighting function used in Escanciano (2006a) is considered.

The proofs of these two propositions are given in Section 3.8.1.

3.8 Proofs

3.8.1 Proofs regarding Conditions (B1) and (B2)

Proof of Proposition 3.7.1. Consider first $w(X, x, \gamma) = \sin(x^t X)$, $w(X, x, \gamma) = \exp(x^t X)$ and $w(X, x, \gamma) = (1 + \exp(-x^t X))^{-1}$. These three weighting functions are infinitely differentiable and all derivatives are uniformly bounded on the compact set χ . Hence, applying Corollary 2.7.2 in Van der Vaart and Wellner (1996), we get that $\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P)) \leq C\tilde{\varepsilon}^{-d/\kappa}$, where C is some positive constant, κ is the smoothness of the class which can be taken arbitrarily large, $N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))$ is the $\tilde{\varepsilon}$ -bracketing number of the class $\mathcal{W}$, P is the probability measure corresponding to the distribution of X , and $L_2(P)$ is the L_2 -norm. This gives

$$\begin{aligned}
 \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} &= \int_0^{2T} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} \\
 &\leq \int_0^{2T} \sqrt{C\tilde{\varepsilon}^{-d/\kappa}} d\tilde{\varepsilon} \\
 &= \frac{\sqrt{C}(2T)^{\frac{-d}{2\kappa}+1}}{\frac{-d}{2\kappa}+1} \\
 &< \infty,
 \end{aligned}$$

provided $\kappa > d/2$, and where T is a uniform upper bound for the above three w -functions. We obtained the first equality since only one $\tilde{\varepsilon}$ -bracket suffices to

cover $\mathcal{W}$ if $\tilde{\varepsilon} > 2T$. Next, if $w(X, x, \gamma) = \exp(ix^t X)$, the proof is exactly the same, taking into account that $\exp(ix^t X) = \cos(x^t X) + i \sin(x^t X)$.

Second, consider $w(X, x, \gamma) = 1_{\{X \leq x\}}$. This function is not differentiable, so we can not apply the same proof as above. However, $1_{\{X_i \leq x_i\}}$ is an increasing function in x_i for each $i = 1, \dots, d$ and we remind that $1_{\{X \leq x\}} = \prod_{i=1}^d 1_{\{X_i \leq x_i\}}$. Consequently, for $i = 1, \dots, d$, we can apply Theorem 2.7.5 in Van der Vaart and Wellner (1996) to obtain that $N_{[]}(\tilde{\varepsilon}, \mathcal{W}_i, L_2(P)) \leq \exp(K_i \tilde{\varepsilon}^{-1})$ where K_i is some positive constant, $\mathcal{W}_i = \{t \rightarrow 1_{\{t \leq x_i\}} : x_i \in \chi_i\}$ and χ_i is a compact subset of $\mathbb{R}$. Moreover,

$$N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P)) \leq \prod_{i=1}^d N_{[]}(\tilde{\varepsilon}, \mathcal{W}_i, L_2(P)) = \exp(C_1 \tilde{\varepsilon}^{-1}),$$

where $C_1 = \sum_{i=1}^d K_i$. Note that one bracket suffices to cover $\mathcal{W}$ if $\tilde{\varepsilon} \geq 1$. We conclude that

$$\begin{aligned} \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} &\leq \int_0^1 \sqrt{C_1 \tilde{\varepsilon}^{-1}} d\tilde{\varepsilon} \\ &= 2\sqrt{C_1} \\ &< \infty, \end{aligned}$$

Third, consider $w(X, x, \gamma) = 1_{\{\gamma^t X \leq x\}}$. The proof is somewhat similar to the proof of Lemma 1 in Akritas and Van Keilegom (2001). Let

$$\mathcal{W} = \{t \rightarrow 1_{\{\gamma^t t \leq x\}} : (x, \gamma) \in \Pi\},$$

and also define the classes

$$\mathcal{W}_1 = \{t \rightarrow \gamma^t t : \gamma \in \chi_\gamma\},$$

and

$$\mathcal{W}_2 = \{t \rightarrow 1_{\{t \leq x\}} : x \in \chi\}.$$

Then, there exists $S = O(\tilde{\varepsilon}^{-2d})$ $\tilde{\varepsilon}^2$ -brackets for $\mathcal{W}_1$ with respect to the L_∞ -norm, say $d_1^L \leq d_1^U, \dots, d_S^L \leq d_S^U$. For a given $\gamma \in \chi_\gamma$ and $j = 1, \dots, S$, let $d_j^L(X) \leq \gamma^t X \leq d_j^U(X)$, and define

$$F_j^U(x) = P(d_j^U(X) \leq x) \quad \text{and} \quad F_j^L(x) = P(d_j^L(X) \leq x).$$

Moreover, for $j = 1, \dots, S$, let $x_{j1}^L, \dots, x_{jS_L}^L$, $S_L = O(\tilde{\varepsilon}^{-2})$, be a partition of the real line in segments having F_j^L -probability bounded by a fraction of $\tilde{\varepsilon}^2$. Similarly, for $j = 1, \dots, S$, let $x_{j1}^U, \dots, x_{jS_U}^U$, $S_U = O(\tilde{\varepsilon}^{-2})$, be a partition of the real line in segments having F_j^U -probability bounded by a fraction of $\tilde{\varepsilon}^2$. Next, for a given $x \in \chi$ and for each $j = 1, \dots, S$, we define the following bracket: $x_{js_1}^L \leq x \leq x_{js_2}^U$, where $x_{js_1}^L$ is the largest of the x_{js}^L , $s = 1, \dots, S_L$, with the

property of being less than or equal to x , and $x_{js_2}^U$ is the smallest of the x_{js}^U , $s = 1, \dots, S_U$, with the property of being greater than or equal to x . Hence, we claim that the bracket for $1_{\{\gamma^t X \leq x\}}$ is given by

$$1_{\{d_j^U(X) \leq x_{js_1}^L\}} \leq 1_{\{\gamma^t X \leq x\}} \leq 1_{\{d_j^L(X) \leq x_{js_2}^U\}} .$$

Indeed, using the definitions of F_j^L and F_j^U , we get

$$\begin{aligned} & \left\| 1_{\{d_j^L(X) \leq x_{js_2}^U\}} - 1_{\{d_j^U(X) \leq x_{js_1}^L\}} \right\|_2^2 \\ &= E \left[\left(1_{\{d_j^L(X) \leq x_{js_2}^U\}} - 1_{\{d_j^U(X) \leq x_{js_1}^L\}} \right)^2 \right] \\ &= P(d_j^L(X) \leq x_{js_2}^U) - P(d_j^U(X) \leq x_{js_1}^L) \\ &= F_j^U(x_{js_2}^U) - F_j^L(x_{js_1}^L) \\ &= \left[F_j^U(x_{js_2}^U) - F_j^U(x) \right] + \left[F_j^U(x) - F_j^L(x) \right] + \left[F_j^L(x) - F_j^L(x_{js_1}^L) \right] . \end{aligned}$$

Next, by construction of $x_{js_1}^L$ and $x_{js_2}^U$, we know that $F_j^U(x_{js_2}^U) - F_j^U(x)$ and $F_j^L(x) - F_j^L(x_{js_1}^L)$ are $O(\tilde{\varepsilon}^2)$. Consequently, for some positive constant K , we have

$$\begin{aligned} & \left\| 1_{\{d_j^L(X) \leq x_{js_2}^U\}} - 1_{\{d_j^U(X) \leq x_{js_1}^L\}} \right\|_2^2 \\ &\leq F_j^U(x) - F_j^L(x) + K\tilde{\varepsilon}^2 \\ &= P(\gamma^t X + d_j^L(X) - \gamma^t X \leq x) - P(\gamma^t X + d_j^U(X) - \gamma^t X \leq x) + K\tilde{\varepsilon}^2 \\ &\leq P\left(\gamma^t X \leq x + \sup_{(x,\gamma) \in \Pi} |d_j^L(x) - \gamma^t x|\right) \\ &\quad - P\left(\gamma^t X \leq x - \sup_{(x,\gamma) \in \Pi} |d_j^U(x) - \gamma^t x|\right) + K\tilde{\varepsilon}^2 . \end{aligned} \tag{3.8.1}$$

Finally, using a Taylor expansion for some η between $x + \sup_{(x,\gamma) \in \Pi} |d_j^L(x) - \gamma^t x|$ and $x - \sup_{(x,\gamma) \in \Pi} |d_j^U(x) - \gamma^t x|$, expression (3.8.1) is equal to

$$f_\gamma(\eta) \left[\sup_{(x,\gamma) \in \Pi} |d_j^L(x) - \gamma^t x| + \sup_{(x,\gamma) \in \Pi} |d_j^U(x) - \gamma^t x| \right] + K\tilde{\varepsilon}^2 ,$$

where f_γ is the density function of $\gamma^t X$. Since $d_j^L(X) \leq \gamma^t X \leq d_j^U(X)$, we know that

$$\sup_{(x,\gamma) \in \Pi} |d_j^L(x) - \gamma^t x| \leq \sup_{x \in \mathcal{X}} |d_j^U(x) - d_j^L(x)| ,$$

and

$$\sup_{(x,\gamma) \in \Pi} |d_j^U(x) - \gamma^t x| \leq \sup_{x \in \mathcal{X}} |d_j^U(x) - d_j^L(x)| ,$$

which implies that

$$\left\| 1_{\{d_j^L(X) \leq x_{js_2}^U\}} - 1_{\{d_j^U(X) \leq x_{js_1}^L\}} \right\|_2^2 \leq 2 \sup_{x \in \chi_r} f_\gamma(x) \sup_{x \in \chi} |d_j^U(x) - d_j^L(x)| + K\tilde{\varepsilon}^2 .$$

Using the assumption of this Proposition, we know that $\sup_{x \in \chi_r} f_\gamma(x) < \infty$ and by construction of the brackets, we also have $\sup_{x \in \chi} |d_j^U(x) - d_j^L(x)| \leq \tilde{\varepsilon}^2$. Hence, we need at most $O(\tilde{\varepsilon}^{-2d-4})$ brackets to cover the class $\mathcal{W}$, i.e., $N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P)) \leq C_2 \tilde{\varepsilon}^{-2d-4}$ for some constant C_2 . This implies

$$\begin{aligned} \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} &= \int_0^1 \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))} d\tilde{\varepsilon} \\ &\leq \int_0^1 \sqrt{\log(C_2 \tilde{\varepsilon}^{-2d-4})} d\tilde{\varepsilon} \\ &= \int_0^1 \sqrt{\log(C_2) - (2d+4) \log(\tilde{\varepsilon})} d\tilde{\varepsilon} \\ &\leq \int_0^1 \sqrt{\log(C_2)} d\tilde{\varepsilon} \\ &\quad + \int_0^1 \sqrt{-(2d+4) \log(\tilde{\varepsilon})} d\tilde{\varepsilon} \\ &< \infty , \end{aligned}$$

since $\int_0^1 \sqrt{-\log(\tilde{\varepsilon})} d\tilde{\varepsilon} = \frac{\sqrt{\pi}}{2} < \infty$. We obtained the first equality since only one bracket suffices to cover $\mathcal{W}$ when $\tilde{\varepsilon} > 1$. This concludes the proof. $\square$

Proof of Proposition 3.7.2. First, the result is trivial in case (a), since we can take $\tilde{\Psi} = \Psi_n = \Psi$. Second, consider case (b). In that case, we define

$$\tilde{\Psi}(x) = \tilde{F}_X(x) = b_n^{-d} \int \hat{F}_X(t) L\left(\frac{x-t}{b_n}\right) dt ,$$

where $t = (t_1, \dots, t_d)$, $dt = (dt_1, \dots, dt_d)$, $\frac{x-t}{b_n} = (\frac{x_1-t_1}{b_n}, \dots, \frac{x_d-t_d}{b_n})$, L is a kernel of order $\tau > \frac{d}{2}$ and b_n is a bandwidth such that $nb_n^d(\log n)^{-1} \rightarrow \infty$ and $nb_n^{2\tau} \rightarrow 0$. By definition of the kernel L , with $u_i = \frac{x_i-t_i}{b_n}$ for $i = 1, \dots, d$, we know that

$$\begin{aligned} b_n^{-d} \int L\left(\frac{x-t}{b_n}\right) dt &= b_n^{-d} \int \dots \int L\left(\frac{x_1-t_1}{b_n}, \dots, \frac{x_d-t_d}{b_n}\right) dt_1 \dots dt_d \\ &= b_n^{-d} \int \dots \int L(u_1, \dots, u_d) (b_n du_1) \dots (b_n du_d) \\ &= 1 . \end{aligned}$$

Consequently $\widehat{F}_X(x) = b_n^{-d} \int \widehat{F}_X(t) L\left(\frac{x-t}{b_n}\right) dt$ and

$$\begin{aligned} \widetilde{F}_X(x) - F_X(x) &= b_n^{-d} \int \widehat{F}_X(t) L\left(\frac{x-t}{b_n}\right) dt - F_X(x) \\ &= b_n^{-d} \int (\widehat{F}_X(t) - F_X(t) - \widehat{F}_X(x)) L\left(\frac{x-t}{b_n}\right) dt - F_X(x) \\ &\quad + \widehat{F}_X(x) + b_n^{-d} \int F_X(t) L\left(\frac{x-t}{b_n}\right) dt . \end{aligned}$$

Next, using a Taylor expansion, we have

$$F_X(t) = F_X(x) + \sum_{j=1}^d \frac{\partial F_X(x)}{\partial x_j} (x_j - t_j) + \dots + O(b_n^\tau) ,$$

and also

$$b_n^{-d} \int F_X(t) L\left(\frac{x-t}{b_n}\right) dt = b_n^{-d} \int F_X(x) L\left(\frac{x-t}{b_n}\right) dt + O(b_n^\tau) ,$$

since F_X is τ times continuously differentiable by condition (A3)(ii) (if we take $\tau \leq 2d+1$), and since L is a kernel of order τ . We get

$$\begin{aligned} \widetilde{F}_X(x) - F_X(x) &= \widehat{F}_X(x) - F_X(x) + b_n^{-d} \int \left\{ \widehat{F}_X(t) - F_X(t) \right. \\ &\quad \left. - \widehat{F}_X(x) + F_X(x) \right\} L\left(\frac{x-t}{b_n}\right) dt + O(b_n^\tau) . \end{aligned}$$

Using Theorem 2.14 in Stute (1982), we know that

$$b_n^{-d} \int (\widehat{F}_X(t) - F_X(t) - \widehat{F}_X(x) + F_X(x)) L\left(\frac{x-t}{b_n}\right) dt = O_P(n^{-1/2} b_n^{1/2}) ,$$

which is also $o_P(n^{-1/2})$ uniformly in $x \in \chi$. Note also that $O(b_n^\tau)$ is $o(n^{-1/2})$ since $nb_n^{2\tau} \rightarrow 0$. This implies that

$$\widetilde{F}_X(x) - F_X(x) = \widehat{F}_X(x) - F_X(x) + o_P(n^{-1/2}) ,$$

uniformly in $x \in \chi$, i.e., $\sup_{x \in \chi} |\widehat{F}_X(x) - \widetilde{F}_X(x)| = o_P(n^{-1/2})$. We will now check the second condition defined in (B2) in case (b). The function $\widetilde{\Phi}(x, \gamma) = \widetilde{f}_X(x)$ is here given by

$$\widetilde{f}_X(x) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} \widetilde{F}_X(x) = (nb_n)^{-d} \sum_{i=1}^n L\left(\frac{X_i - x}{b_n}\right) ,$$

and $\Phi(x, \gamma) = f_X(x)$. Using the asymptotic properties of the classical kernel estimator of a density function, we obtain

$$\sup_{x \in \mathcal{X}} |\tilde{f}_X(x) - f_X(x)| = O_P((nb_n^d)^{-1/2}(\log n)^{1/2}) + O(b_n^{2\tau}).$$

The last expression is $o_P(1)$ since $nb_n^d(\log n)^{-1} \rightarrow \infty$. This concludes the proof for case (b). Finally, the proof of case (c) follows the same way as the proof of case (b). $\square$

3.8.2 Proofs of the main results under H_0

Proof of Theorem 3.3.1. First, we will prove that the class

$$\mathcal{F}_1 = \{(u, v) \rightarrow w(u, x, \gamma)(\Lambda_{\theta_0}(v) - m_{\beta_0}(u)), (x, \gamma) \in \Pi\}, \quad (3.8.2)$$

is Donsker. Using assumption (B1)(i), we know that there exists a finite number of $\tilde{\varepsilon}$ -brackets, say M , to cover $\mathcal{W}$. Let $w_1^L \leq w_1^U, \dots, w_M^L \leq w_M^U$ be the functions defining the M $\tilde{\varepsilon}$ -brackets to cover $\mathcal{W}$.

Next, the functions $y_1^L \leq y_1^U, \dots, y_M^L \leq y_M^U$, where

$$y_j^L(u, v) = \min \left[w_j^L(u) \{ \Lambda_{\theta_0}(v) - m_{\beta_0}(u) \}, w_j^U(u) \{ \Lambda_{\theta_0}(v) - m_{\beta_0}(u) \} \right],$$

and

$$y_j^U(u, v) = \max \left[w_j^L(u) \{ \Lambda_{\theta_0}(v) - m_{\beta_0}(u) \}, w_j^U(u) \{ \Lambda_{\theta_0}(v) - m_{\beta_0}(u) \} \right],$$

for $j = 1, \dots, M$, define the M $\tilde{\varepsilon}$ -brackets to cover $\mathcal{F}_1$. Indeed, for $j = 1, \dots, M$, we have

$$\begin{aligned} \left\| y_j^U(X, Y) - y_j^L(X, Y) \right\|_2^2 &= E \left[(w_j^U(X) - w_j^L(X))^2 (\Lambda_{\theta_0}(Y) - m_{\beta_0}(X))^2 \right] \\ &= E \left[(w_j^U(X) - w_j^L(X))^2 \right] E(\varepsilon^2), \end{aligned}$$

where the last equality is obtained by independence between X and ε . The last expression is $O(\tilde{\varepsilon}^2)$ since $E(\varepsilon^2) = \sigma^2 < \infty$ and $E[(w_j^U(X) - w_j^L(X))^2] \leq \tilde{\varepsilon}^2$ by definition of the brackets $w_j^L \leq w_j^U$ for $j = 1, \dots, M$.

Consequently, $N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P)) = N_{[]}(\tilde{\varepsilon}, \mathcal{W}, L_2(P))$. Using assumption (B1)(i), we obtain that $\int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P))} d\tilde{\varepsilon} < \infty$. Next, applying Theorem 2.5.6 in Van der Vaart and Wellner (1996), we get that the class $\mathcal{F}_1$ is Donsker. Hence, the process R_n converges weakly to a limiting Gaussian process.

Finally, by independence between X and ε , we have

$$E[w(X, x, \gamma)(\Lambda_{\theta_0}(Y) - m(X))] = E[w(X, x, \gamma)]E(\varepsilon) = 0,$$

as $\sup_{(x,\gamma) \in \Pi, t \in \chi} |w(t, x, \gamma)| < \infty$ by assumption (B1)(ii). Finally,

$$\begin{aligned} & \text{Cov}[w(X, x_1, \gamma_1)(\Lambda_{\theta_0}(Y) - m(X)), w(X, x_2, \gamma_2)(\Lambda_{\theta_0}(Y) - m(X))] \\ &= E[w(X, x_1, \gamma_1)(\Lambda_{\theta_0}(Y) - m(X))w(X, x_2, \gamma_2)(\Lambda_{\theta_0}(Y) - m(X))] \\ &= E[w(X, x_1, \gamma_1)w(X, x_2, \gamma_2)\varepsilon^2] . \end{aligned}$$

By independence between X and ε and since $E(\varepsilon^2) = \sigma^2$, this last expression is equal to $\sigma^2 E[w(X, x_1, \gamma_1)w(X, x_2, \gamma_2)]$. $\square$

Proof of Theorem 3.3.2. First, note that

$$R_n^1(x, \gamma) = I + R_n(x, \gamma) - II , \quad (3.8.3)$$

where

$$I = n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma)(\Lambda_{\hat{\theta}}(Y_i) - \Lambda_{\theta_0}(Y_i)) , \quad (3.8.4)$$

and

$$II = n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma)(m_{\hat{\beta}}(X_i) - m(X_i)) . \quad (3.8.5)$$

Using successively a Taylor expansion with some ξ between $\hat{\theta}$ and θ_0 and Proposition 2.8.2, the expression I is equal to

$$\begin{aligned} I &= n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)(\dot{\Lambda}_{\xi}(Y_i))^t n^{1/2}(\hat{\theta} - \theta_0) \\ &= n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y_i))^t \left(-n^{1/2} \sum_{j=1}^n g(X_j, Y_j) + o_P(1) \right) \\ &= -n^{-1/2} \sum_{j=1}^n [g(X_j, Y_j)]^t \\ &\quad \times \left(n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y_i)) - E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))] \right) \\ &\quad - n^{-1/2} \sum_{j=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(X_j, Y_j) + o_P(1) \\ &= -n^{-1/2} \sum_{j=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(X_j, Y_j) + o_P(1) . \end{aligned} \quad (3.8.6)$$

We have obtained the last two equalities because

$$n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y_i))^t = O_P(1) ,$$

and

$$n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y_i)) - E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))] = O_P(n^{-1/2}) ,$$

uniformly in $(x, \gamma) \in \Pi$, and because $n^{-1/2} \sum_{j=1}^n g(X_j, Y_j) = O_P(1)$ by Proposition 2.8.2. Indeed, for $l = 1, \dots, k$, we define the classes

$$\mathcal{G}_l = \{(u, v) \rightarrow w(u, x, \gamma)(\dot{\Lambda}_{\theta_0}(v))_l, (x, \gamma) \in \Pi\} ,$$

where $(\dot{\Lambda}_{\theta_0}(v))_l$ is the l -th component of the vector $\dot{\Lambda}_{\theta_0}(v)$. Define

$$z_{lj}^L(u, v) = \min \left[w_j^L(u) \{ \dot{\Lambda}_{\theta_0}(v) \}_l, w_j^U(u) \{ \dot{\Lambda}_{\theta_0}(v) \}_l \right] ,$$

and

$$z_{lj}^U(u, v) = \max \left[w_j^L(u) \{ \dot{\Lambda}_{\theta_0}(v) \}_l, w_j^U(u) \{ \dot{\Lambda}_{\theta_0}(v) \}_l \right] ,$$

for $j = 1, \dots, M$, where $w_1^L \leq w_1^U, \dots, w_M^L \leq w_M^U$ are the M functions defining the $\tilde{\varepsilon}$ -brackets to cover $\mathcal{W}$ that have been introduced in the proof of Theorem 3.3.1. Consequently, for $l = 1, \dots, k$ and $j = 1, \dots, M$, we have

$$\begin{aligned} & \left\| z_{lj}^U(X, Y) - z_{lj}^L(X, Y) \right\|_2^2 \\ &= E \left[\left(w_j^U(X) - w_j^L(X) \right)^2 \left(\dot{\Lambda}_{\theta_0}(Y) \right)_l^2 \right] \\ &= E \left[\left(w_j^U(X) - w_j^L(X) \right)^2 E \left[\left(\dot{\Lambda}_{\theta_0}(Y) \right)_l^2 \middle| X \right] \right] \\ &\leq \sup_{x \in \mathcal{X}} E \left[\left(\dot{\Lambda}_{\theta_0}(Y) \right)_l^2 \middle| X = x \right] E \left[\left(w_j^U(X) - w_j^L(X) \right)^2 \right] . \end{aligned}$$

The last expression is $O(\tilde{\varepsilon}^2)$ since $\sup_{x \in \mathcal{X}} E[(\dot{\Lambda}_{\theta_0}(Y))_l^2 | X = x] < \infty$ by condition (A5)(ii) and $E(w_j^U(X) - w_j^L(X))^2 \leq \tilde{\varepsilon}^2$ by definition of the brackets $w_j^L \leq w_j^U$ for $j = 1, \dots, M$. Hence, in a similar way as in the proof of Theorem 3.3.1, we conclude that $\mathcal{G}_l$, for $l = 1, \dots, k$, is Donsker using Theorem 2.5.6 in Van der Vaart and Wellner (1996), and then (3.8.6) follows.

Next, as we are under H_0 , $m(x) = m_{\beta_0}(x)$ and using a Taylor expansion

for some value ζ between $\hat{\beta}$ and β_0 , expression II is equal to

$$\begin{aligned}
II &= n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma) (\hat{\beta} - \beta_0)^t \frac{\partial}{\partial \beta} m_\zeta(X_i) \\
&= n^{1/2} (\hat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) h(X_i, \zeta) \\
&= n^{1/2} (\hat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) (h(X_i, \zeta) - h(X_i, \beta_0)) \\
&\quad + n^{1/2} (\hat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n \left[w(X_i, x, \gamma) h(X_i, \beta_0) - H(x, \gamma, \beta_0) \right] \\
&\quad + n^{1/2} (\hat{\beta} - \beta_0)^t H(x, \gamma, \beta_0) .
\end{aligned} \tag{3.8.7}$$

Note that $n^{1/2}(\hat{\beta} - \beta_0) = O_P(1)$ by Lemma 2.8.5. Moreover, as $h(X, \beta)$ is a differentiable function in β with a uniformly bounded derivative (see condition (A6)(ii)), and as $\hat{\beta} - \beta_0 = o_P(1)$ by Lemma 2.8.3, we have $\sup_{x \in \mathcal{X}} |h(x, \zeta) - h(x, \beta_0)| = o_P(1)$. Consequently,

$$\begin{aligned}
&\left| n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) (h(X_i, \zeta) - h(X_i, \beta_0)) \right| \\
&\leq n^{-1} \sum_{i=1}^n \left| w(X_i, x, \gamma) \right| \left| h(X_i, \zeta) - h(X_i, \beta_0) \right| \\
&= o_P(1) ,
\end{aligned}$$

since $\sup_{(x, \gamma) \in \Pi, t \in \mathcal{X}} |w(t, x, \gamma)| < \infty$ by condition (B1)(ii). Hence, the first term on the right hand side of (3.8.7) is $o_P(1)$. Next, for $l = 1, \dots, q$, we define the classes

$$\mathcal{H}_l = \{u \rightarrow w(u, x, \gamma) h_l(u, \beta_0), (x, \gamma) \in \Pi\} ,$$

where $h_l(u, \beta_0)$ is the l -th component of the vector $h(u, \beta_0)$, i.e., $h_l(u, \beta_0) = \frac{\partial m_{\beta_0}(X)}{\partial \beta_l}$, and

$$v_{lj}^L(u) = \min \left[w_j^L(u) h_l(u, \beta_0), w_j^U(u) h_l(u, \beta_0) \right] ,$$

and

$$v_{lj}^U(u) = \max \left[w_j^L(u) h_l(u, \beta_0), w_j^U(u) h_l(u, \beta_0) \right] ,$$

for $j = 1, \dots, M$. Consequently, for $l = 1, \dots, q$ and $j = 1, \dots, M$, we have

$$\begin{aligned} \left\| v_{lj}^U(X) - v_{lj}^L(X) \right\|_1 &= E \left[\left\| w_j^U(X) - w_j^L(X) \right\| h_l(X, \beta_0) \right] \\ &\leq \left(E[h_l^2(X, \beta_0)] \right)^{1/2} \left(E[(w_j^U(X) - w_j^L(X))^2] \right)^{1/2} \\ &\leq \sup_{x \in \mathcal{X}} |h_l(x, \beta_0)| \left(E[(w_j^U(X) - w_j^L(X))^2] \right)^{1/2}. \end{aligned}$$

This last expression is finite by condition (A6)(ii) and since $E[(w_j^U(X) - w_j^L(X))^2] \leq \varepsilon^2$. Hence, for $l = 1, \dots, q$, $N_{[]}(\varepsilon, \mathcal{H}_l, L_1(P)) < \infty$, which implies that $\mathcal{H}_l$ is Glivenko-Cantelli by Theorem 2.4.1 in Van der Vaart and Wellner (1996), i.e.,

$$\sup_{f \in \mathcal{H}_l} \left| n^{-1} \sum_{i=1}^n f(X_i) - E(f(X)) \right| = o_P(1).$$

Then, the second term on the right hand side of (3.8.7) is $o_P(1)$. In conclusion, combining (3.8.3), (3.8.6) and (3.8.7), we get

$$\begin{aligned} R_n^1(x, \gamma) &= R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(X_i, Y_i) \\ &\quad - n^{1/2}(\hat{\beta} - \beta_0)^t H(x, \gamma, \beta_0) + o_P(1), \end{aligned}$$

uniformly in $(x, \gamma) \in \Pi$. Finally, we conclude the proof using Lemma 2.8.5 (with $r \equiv 0$):

$$\begin{aligned} R_n^1(x, \gamma) &= R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(X_i, Y_i) \\ &\quad - n^{-1/2} \sum_{i=1}^n H^t(x, \gamma, \beta_0) \left\{ \eta_{\beta_0}(X_i, Y_i) \right. \\ &\quad \left. - \Omega^{-1} E \left[\frac{\partial m_{\beta_0}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] g(X_i, Y_i) \right\} + o_P(1). \end{aligned}$$

The last equality was obtained using the fact that $H(x, \gamma, \beta_0)$ is bounded uniformly in $(x, \gamma) \in \Pi$. Indeed, for each $l = 1, \dots, q$, if we denote by $H_l(x, \gamma, \beta)$ the l -th component of the vector $H(x, \gamma, \beta)$, i.e., $H_l(x, \gamma, \beta) = E[w(X, x, \gamma) \frac{\partial m_{\beta}(X)}{\partial \beta_l}]$, we have that

$$\sup_{(x, \gamma) \in \Pi} |H_l(x, \gamma, \beta_0)| \leq \sup_{(x, \gamma) \in \Pi, t \in \mathcal{X}} |w(t, x, \gamma)| \sup_{t \in \mathcal{X}} \left| \frac{\partial m_{\beta_0}(t)}{\partial \beta_l} \right|.$$

Using conditions (B1)(ii) and (A6)(ii), we conclude that the last expression is finite. $\square$

Proof of Theorem 3.3.3. First, note that

$$\left| \int_{\Pi} [R_n^1(x, \gamma)]^2 d(\Psi_n(x, \gamma) - \Psi(x, \gamma)) \right| \leq A + B ,$$

where

$$A = \left| \int_{\Pi} [R_n^1(x, \gamma)]^2 d(\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)) \right| ,$$

and

$$B = \left| \int_{\Pi} [R_n^1(x, \gamma)]^2 d(\tilde{\Psi}(x, \gamma) - \Psi(x, \gamma)) \right| .$$

We will first prove that the term A is $o_P(1)$ uniformly in $(x, \gamma) \in \Pi$. Using integration by parts, the term A is equal to

$$\left| \left[(\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)) R_n^1(x, \gamma) \right]_{\Pi} - 2 \int_{\Pi} (\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)) R_n^1(x, \gamma) dR_n^1(x, \gamma) \right| .$$

Note that the first term on the right hand side is $o_P(1)$, since $\sup_{(x, \gamma) \in \Pi} |R_n^1(x, \gamma)| = O_P(1)$ by Corollary 3.3.4 and since $\sup_{(x, \gamma) \in \Pi} |\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)| = o_P(n^{-1/2}) = o_P(1)$ by condition (B2). Consequently,

$$\begin{aligned} A &= 2 \left| \int_{\Pi} (\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)) R_n^1(x, \gamma) dR_n^1(x, \gamma) \right| \\ &\leq 2 \sup_{(x, \gamma) \in \Pi} |\Psi_n(x, \gamma) - \tilde{\Psi}(x, \gamma)| \sup_{(x, \gamma) \in \Pi} |R_n^1(x, \gamma)| \\ &\quad \times \sup_{t \in \chi} T_V[w(t, \cdot, \cdot)] n^{-1/2} \sum_{i=1}^n |\Lambda_{\hat{\theta}}(Y_i) - m_{\hat{\beta}}(X_i)| . \end{aligned} \quad (3.8.8)$$

Next, the first factor on the right hand side of (3.8.8) is $o_P(n^{-1/2})$ using condition (B2), the second factor is $O_P(1)$ as explained just above and $\sup_{t \in \chi} T_V[w(t, \cdot, \cdot)] < \infty$ using condition (B1)(iii). Moreover, it follows easily from the proofs of Theorem 3.3.2 and Corollary 3.3.4 that

$$n^{-1/2} \sum_{i=1}^n |\Lambda_{\hat{\theta}}(Y_i) - m_{\hat{\beta}}(X_i)| = O_P(n^{1/2}) ,$$

which implies that A is $o_P(1)$ uniformly in $(x, \gamma) \in \Pi$. Finally, we will prove that the term B is $o_P(1)$ uniformly in $(x, \gamma) \in \Pi$. As $\tilde{\Psi}(x, \gamma) - \Psi(x, \gamma)$ is a differentiable function with respect to (x, γ) , we have that

$$\begin{aligned} B &= \left| \int_{\Pi} [R_n^1(x, \gamma)]^2 (\tilde{\Phi}(x, \gamma) - \Phi(x, \gamma)) dx d\gamma \right| \\ &\leq K \sup_{(x, \gamma) \in \Pi} |R_n^1(x, \gamma)|^2 \sup_{(x, \gamma) \in \Pi} |\tilde{\Phi}(x, \gamma) - \Phi(x, \gamma)| , \end{aligned} \quad (3.8.9)$$

for some $K < \infty$. We conclude the proof using the facts that the first factor on the right hand side of (3.8.9) is $O_P(1)$ as explained just above and that the second factor is $o_P(1)$ using condition (B2). $\square$

Proof of Corollary 3.3.4. First, we prove that $R_n^1(x, \gamma)$ converges to a limiting Gaussian process. In the proof of Theorem 3.3.1 we have shown that the class $\mathcal{F}_1$ (corresponding to the process $R_n(x, \gamma)$) is Donsker. Hence, using Theorem 3.3.2, it suffices to show that the class

$$\mathcal{F}_2 = \left\{ (u, v) \rightarrow G(x, \gamma, u, v, \theta_0, \beta_0), (x, \gamma) \in \Pi \right\}$$

is Donsker. Recall that

$$\begin{aligned} G(x, \gamma, u, v, \theta_0, \beta_0) &= H^t(x, \gamma, \beta_0) \eta_{\beta_0}(u, v) + E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(u, v) \\ &\quad - H^t(x, \gamma, \beta_0) \Omega^{-1} E \left[\frac{\partial m_{\beta_0}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] g(u, v). \end{aligned}$$

Hence, each term of $G(x, \gamma, u, v, \theta_0, \beta_0)$ can be decomposed in a factor that depends on (x, γ) but not on (u, v) , and another factor that depends on (u, v) but not on (x, γ) . Hence, it can be easily seen using similar arguments as before that the class $\mathcal{F}_2$ is Donsker. In fact, we only need to prove that $\sup_{(x, \gamma) \in \Pi} |H(x, \gamma, \beta_0)| < \infty$ and that $\sup_{(x, \gamma) \in \Pi} |E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))_l]| < \infty$ for $l = 1, \dots, k$. The former property has been shown at the end of the proof of Theorem 3.3.2, whereas for the latter note that for $l = 1, \dots, k$, we have

$$\begin{aligned} &\sup_{(x, \gamma) \in \Pi} \left| E \left[w(X, x, \gamma) (\dot{\Lambda}_{\theta_0}(Y))_l \right] \right| \\ &= \sup_{(x, \gamma) \in \Pi} \left| E \left[w(X, x, \gamma) E \left[(\dot{\Lambda}_{\theta_0}(Y))_l \middle| X \right] \right] \right| \\ &\leq \sup_{(x, \gamma) \in \Pi, t \in \chi} \left| w(t, x, \gamma) \right| \sup_{x \in \chi} E \left[\left| (\dot{\Lambda}_{\theta_0}(Y))_l \right| \middle| X = x \right]. \end{aligned}$$

This last expression is finite by conditions (B1)(ii) and (A5)(ii). It now follows that $R_n^1(x, \gamma)$ converges to a Gaussian process R_∞^1 .

We can easily see that R_∞^1 has zero mean, because R_n converges to a centered Gaussian process by Theorem 3.3.1, $E(\eta_{\beta_0}(X, Y)) = 0$ and $E(g(X, Y)) = 0$ since $g(X, Y) = \Xi^{-1} \xi(\theta_0, X, Y)$ and $\xi(\theta_0, X, Y)$ is the derivative of the like-

lihood. Moreover,

$$\begin{aligned}
& \text{Cov}[w(X, x_1, \gamma_1)(\Lambda_{\theta_0}(Y) - m(X)) - G(x_1, \gamma_1, X, Y, \theta_0, \beta_0), \\
& \quad w(X, x_2, \gamma_2)(\Lambda_{\theta_0}(Y) - m(X)) - G(x_2, \gamma_2, X, Y, \theta_0, \beta_0)] \\
&= E[\{w(X, x_1, \gamma_1)(\Lambda_{\theta_0}(Y) - m(X)) - G(x_1, \gamma_1, X, Y, \theta_0, \beta_0)\} \\
& \quad \times \{w(X, x_2, \gamma_2)(\Lambda_{\theta_0}(Y) - m(X)) - G(x_2, \gamma_2, X, Y, \theta_0, \beta_0)\}] \\
&= E[w(X, x_1, \gamma_1)w(X, x_2, \gamma_2)\varepsilon^2] - E[G(x_2, \gamma_2, X, Y, \theta_0, \beta_0)w(X, x_1, \gamma_1)\varepsilon] \\
& \quad - E[G(x_1, \gamma_1, X, Y, \theta_0, \beta_0)w(X, x_2, \gamma_2)\varepsilon] \\
& \quad + E[G(x_1, \gamma_1, X, Y, \theta_0, \beta_0)G(x_2, \gamma_2, X, Y, \theta_0, \beta_0)].
\end{aligned}$$

The first term on the right hand side of the last expression is $C(x_1, \gamma_1, x_2, \gamma_2)$. This concludes the proof. $\square$

Proof of Corollary 3.3.5. First, to obtain the limiting distribution of D_n , it suffices to apply Corollary 3.3.4 and the continuous mapping theorem. Next, for W_n^2 , we use Theorem 3.3.3, which states that we can replace $d\Psi_n(x, \gamma)$ by $d\Psi(x, \gamma)$ up to a negligible term. Hence, we obtain the limiting distribution of W_n^2 applying Corollary 3.3.4 and the continuous mapping theorem. $\square$

3.8.3 Proofs of the main results under H_{1n}

Proof of Theorem 3.3.6. First, we remind that

$$R_n^1(x, \gamma) = I + R_n(x, \gamma) - II, \quad (3.8.10)$$

where I and II are given in (3.8.4) and (3.8.5) respectively. Exactly in the same way as in the proof of Theorem 3.3.2, expression I is equal to

$$I = -n^{-1/2} \sum_{i=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t]g(X_i, Y_i) + o_P(1). \quad (3.8.11)$$

Next, as we are under H_{1n} , $m(x) = m_{\beta_0}(x) + n^{-1/2}r(x)$ and using a Taylor

expansion for some value ζ between $\widehat{\beta}$ and β_0 , expression II is equal to

$$\begin{aligned}
II &= n^{-1/2} \sum_{i=1}^n w(X_i, x, \gamma) (\widehat{\beta} - \beta_0)^t \frac{\partial}{\partial \beta} m_\zeta(X_i) - n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) r(X_i) \\
&= n^{1/2} (\widehat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) h(X_i, \zeta) - n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) r(X_i) \\
&= n^{1/2} (\widehat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) (h(X_i, \zeta) - h(X_i, \widetilde{\beta}_{0n})) \\
&\quad + n^{1/2} (\widehat{\beta} - \beta_0)^t n^{-1} \sum_{i=1}^n \left[w(X_i, x, \gamma) h(X_i, \widetilde{\beta}_{0n}) - H(x, \gamma, \widetilde{\beta}_{0n}) \right] \\
&\quad + n^{1/2} (\widehat{\beta} - \beta_0)^t H(x, \gamma, \widetilde{\beta}_{0n}) - n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) r(X_i) . \quad (3.8.12)
\end{aligned}$$

Note that $n^{1/2}(\widehat{\beta} - \beta_0) = O_P(1)$. Indeed, using the asymptotic representation of $\widehat{\beta} - \beta_0$ proved in Lemma 2.8.5, expression

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

is $O_P(1)$ using the central limit theorem, and

$$\Omega^{-1} \int r(u) \frac{\partial m_{\widetilde{\beta}_{0n}}(u)}{\partial \beta} dF_X(u) = \Omega^{-1} E \left[r(X) \frac{\partial m_{\widetilde{\beta}_{0n}}(X)}{\partial \beta} \right]$$

is also $O_P(1)$ using conditions (A6)(ii) and (A10). Moreover, as the function h is a differentiable function in β with a uniformly bounded derivative (see condition (A6)(ii)) and $\widetilde{\beta}_{0n} - \beta_0 = o_P(1)$ using Lemma 2.8.4, we can prove in exactly the same way as in the proof of Theorem 3.3.2 that expressions

$$n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) (h(X_i, \zeta) - h(X_i, \widetilde{\beta}_{0n})) ,$$

and

$$n^{-1} \sum_{i=1}^n \left[w(X_i, x, \gamma) h(X_i, \widetilde{\beta}_{0n}) - H(x, \gamma, \widetilde{\beta}_{0n}) \right] ,$$

are $o_P(1)$ uniformly in $(x, \gamma) \in \Pi$, which implies that the first and the second terms on the right hand side of (3.8.12) are also $o_P(1)$ uniformly in $(x, \gamma) \in \Pi$. Then, combining (3.8.10), (3.8.11) and (3.8.12), we get

$$\begin{aligned}
R_n^1(x, \gamma) &= R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n E[w(X, x, \gamma) (\dot{\Lambda}_{\theta_0}(Y))^t] g(X_i, Y_i) \\
&\quad - n^{1/2} (\widehat{\beta} - \beta_0)^t H(x, \gamma, \widetilde{\beta}_{0n}) + n^{-1} \sum_{i=1}^n w(X_i, x, \gamma) r(X_i) + o_P(1) .
\end{aligned}$$

Finally, we define the class

$$\mathcal{F}_3 = \{u \rightarrow w(u, x, \gamma)r(u), (x, \gamma) \in \Pi\} ,$$

and we let

$$r_j^L(u) = \min \left[w_j^L(u)r(u), w_j^U(u)r(u) \right] ,$$

and

$$r_j^U(u) = \max \left[w_j^L(u)r(u), w_j^U(u)r(u) \right] ,$$

for $j = 1, \dots, M$, where $w_1^L \leq w_1^U, \dots, w_M^L \leq w_M^U$ are the M functions defining the $\tilde{\varepsilon}$ -brackets to cover $\mathcal{W}$ that have been introduced in the proof of Theorem 3.3.1. Consequently, for $j = 1, \dots, M$, we have

$$\begin{aligned} \left\| r_j^U(X) - r_j^L(X) \right\|_1 &= E \left[\left| w_j^U(X) - w_j^L(X) \right| r(X) \right] \\ &\leq \left(E[r^2(X)] \right)^{1/2} \left(E \left[(w_j^U(X) - w_j^L(X))^2 \right] \right)^{1/2} . \end{aligned}$$

This last expression is finite by condition (A10) and since $E[(w_j^U(X) - w_j^L(X))^2] \leq \tilde{\varepsilon}^2$. Hence, $N_{[]}(\tilde{\varepsilon}, \mathcal{F}_3, L_1(P)) < \infty$, which implies that $\mathcal{F}_3$ is Glivenko-Cantelli by Theorem 2.4.1 in Van der Vaart and Wellner (1996), i.e.,

$$\sup_{f \in \mathcal{F}_3} \left| n^{-1} \sum_{i=1}^n f(X_i) - E(f(X)) \right| = o_P(1) .$$

This means that $n^{-1} \sum_{i=1}^n w(X_i, x, \gamma)r(X_i) = E[w(X, x, \gamma)r(X)] + o_P(1)$ uniformly in $(x, \gamma) \in \Pi$. We conclude the proof using Lemma 2.8.5:

$$\begin{aligned} R_n^1(x, \gamma) &= R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n E[w(X, x, \gamma)(\dot{\Lambda}_{\theta_0}(Y))^t] g(X_i, Y_i) \\ &\quad - n^{-1/2} \sum_{i=1}^n H^t(x, \gamma, \tilde{\beta}_{0n}) \left\{ \eta_{\tilde{\beta}_{0n}}(X_i, Y_i) \right. \\ &\quad \left. - \Omega^{-1} E \left[\frac{\partial m_{\tilde{\beta}_{0n}}(X)}{\partial \beta} (\dot{\Lambda}_{\theta_0}(Y))^t \right] g(X_i, Y_i) \right\} \\ &\quad - H^t(x, \gamma, \tilde{\beta}_{0n}) \Omega^{-1} \int r(u) \frac{\partial m_{\tilde{\beta}_{0n}}(u)}{\partial \beta} dF_X(u) \\ &\quad + E[w(X, x, \gamma)r(X)] + o_P(1). \end{aligned}$$

The last equality was obtained using the fact that $H(x, \gamma, \tilde{\beta}_{0n})$ is bounded uniformly in $(x, \gamma) \in \Pi$, the proof is exactly the same as in Theorem 3.3.2. $\square$

Proof of Corollary 3.3.7. To prove this result, we will show that expressions

$$R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n G(x, \gamma, X_i, Y_i, \theta_0, \beta_0) ,$$

under H_0 and

$$R_n(x, \gamma) - n^{-1/2} \sum_{i=1}^n G(x, \gamma, X_i, Y_i, \theta_0, \tilde{\beta}_{0n}) ,$$

under H_{1n} have the same limiting Gaussian process R_∞^1 . Since the bias term $b_2(x, \gamma)$ was already obtained in Theorem 3.3.6, this will conclude the proof.

Recall the definition of $G(x, \gamma, X, Y, \theta_0, \beta_0)$ given in (3.3.1). First note that Theorem 2.3.3(ii) shows that $n^{-1/2} \sum_{i=1}^n \eta_{\beta_0}(X_i, Y_i)$ under H_0 and $n^{-1/2} \sum_{i=1}^n \eta_{\tilde{\beta}_{0n}}(X_i, Y_i)$ under H_{1n} have the same limiting distribution.

Moreover, combining assumption (A6)(ii) and the fact that $\tilde{\beta}_{0n} - \beta_0 = o_P(1)$ by Lemma 2.8.4, we have that

$$\sup_{t \in \mathcal{X}} \left| \frac{\partial m_{\tilde{\beta}_{0n}}(t)}{\partial \beta} - \frac{\partial m_{\beta_0}(t)}{\partial \beta} \right| = o(1) .$$

Next, for $l = 1, \dots, q$, we denote by $H_l(x, \gamma, \beta)$ the l -th component of the vector $H(x, \gamma, \beta)$, i.e., $H_l(x, \gamma, \beta) = E[w(X, x, \gamma) \frac{\partial m_\beta(X)}{\partial \beta_l}]$. For $l = 1, \dots, q$, we have

$$\begin{aligned} & \sup_{(x, \gamma) \in \Pi} \left| H_l(x, \gamma, \tilde{\beta}_{0n}) - H_l(x, \gamma, \beta_0) \right| \\ &= \sup_{(x, \gamma) \in \Pi} \left| E \left[w(X, x, \gamma) \left(\frac{\partial m_{\tilde{\beta}_{0n}}(X)}{\partial \beta_l} - \frac{\partial m_{\beta_0}(X)}{\partial \beta_l} \right) \right] \right| \\ &\leq \sup_{(x, \gamma) \in \Pi, t \in \mathcal{X}} \left| w(t, x, \gamma) \right| \sup_{t \in \mathcal{X}} \left| \frac{\partial m_{\tilde{\beta}_{0n}}(t)}{\partial \beta_l} - \frac{\partial m_{\beta_0}(t)}{\partial \beta_l} \right| . \end{aligned}$$

The last expression is $o_P(1)$ using condition (B1)(ii). Hence,

$$\sup_{(x, \gamma) \in \Pi} |H_l(x, \gamma, \tilde{\beta}_{0n}) - H_l(x, \gamma, \beta_0)| = o(1) ,$$

for $l = 1, \dots, q$. This shows the above statement, and hence finishes the proof. $\square$

Proof of Corollary 3.3.8. First, note that we can prove by very similar arguments the same result as in Theorem 3.3.3 but under H_{1n} . The proof of this Corollary is similar to the proof of Corollary 3.3.5. It suffices to apply Corollary 3.3.7 and the continuous mapping theorem. $\square$

Chapter 4

Estimation of fully nonparametric transformation models

This chapter is based on the paper

Colling, B. and Van Keilegom, I. (2017). Estimation of fully nonparametric transformation models. (*submitted*).

4.1 Introduction and motivation

In this chapter, we consider the nonparametric transformation model defined in (1.3.1), as well as the different assumptions and notations made on it in Section 1.3.1, and we will focus on the problem of nonparametric estimation of the transformation Λ . We will define two new nonparametric estimators of the transformation Λ , and we will show that it often outperforms the ones constructed by Chiappori et al. (2015).

As explained in detail in Section 1.3.2, the main idea of the nonparametric estimators of the transformation Λ developed in Chiappori et al. (2015) is to consider the conditional distribution function of Y given X and to divide its derivative with respect to y by its derivative with respect to x_1 where x_1 is the first component of the vector x . As we have observed in Section 1.3.2, to estimate the transformation in a nonparametric way, Chiappori et al. (2015) introduced kernel estimators and in particular smoothing on Y . However, the distribution of Y can be very asymmetric, depending on which transformation Λ is used. If ε is for instance normal and Λ is the logarithmic transform, then Y is log-normal and hence smoothing for large values of Y will not work well at all. In that case, the kernel estimators based on Y can work very badly in practice. This naturally impacts the quality of the estimator of the conditional distribution function of Y given X and consequently also those of its derivatives

and of the nonparametric estimators of the transformation as we can see in the simulations in Section 4.6.

In this chapter, we will propose a way to overcome this problem. The main idea will be to rewrite the transformation $\Lambda(Y)$ as $\Gamma(T(Y))$ where Γ is an increasing function and T is a certain function such that we can control the distribution of $T(Y)$. This will be presented more precisely in Section 4.2. In Section 4.3, we explain how we can identify the function $\Gamma(\cdot)$. Sections 4.4 and 4.5 contain respectively the definitions of the new nonparametric estimators of the transformation Λ and the theorems that establish their asymptotic normality. In Section 4.6, we realize a simulation study in order to compare the performance of our estimators to the ones developed in Chiappori et al. (2015). Finally, Section 4.8 contains the assumptions needed to establish the limiting theorems and Section 4.9 contains the different proofs.

4.2 Main idea of the estimation procedure

In this chapter, we will work under the conditions (J1) and (J2), introduced in Section 1.3.1, for theoretical reasons and we will consider $\alpha_1 = 0$, $a_1 = 0$, $\alpha_2 = 1$ and $a_2 = 1$ without loss of generality. Note that the values 0 and 1 can be replaced by any other values as long as they are in the interior of the support $\mathcal{Y}$ of Y , where $\mathcal{Y}$ is a connected subset of $\mathbb{R}$.

As explained in the introduction, kernel smoothing based on Y can work very badly in practice. We will now explain how this can be avoided. The main idea is to rewrite the transformation $\Lambda(Y)$ as $\Gamma(T(Y))$, i.e., $\Lambda = \Gamma \circ T$ where Γ is an increasing function and

$$T(Y) = \frac{F_Y(Y) - F_Y(0)}{F_Y(1) - F_Y(0)} ,$$

where $F_Y(\cdot)$ is the distribution function of Y . Note that we can always write $\Lambda(Y)$ in this way, since $T(Y)$ is invertible. Moreover, we have to work with $T(Y)$ instead of the more simple transformation $F_Y(Y)$, since we want to guarantee that $T(0) = 0$ and $T(1) = 1$, and hence $\Gamma(0) = \Lambda(T^{-1}(0)) = 0$ and $\Gamma(1) = \Lambda(T^{-1}(1)) = 1$. Consequently, the set of normalization conditions (J1) does not change and we will now consider the following nonparametric transformation model:

$$\Gamma(U) = m(X) + \varepsilon , \tag{4.2.1}$$

where $U = T(Y)$ and $\Gamma(U) = \Lambda(Y)$. We will denote by $\mathcal{U} \subset \mathbb{R}$ the compact support of U . As $F_Y(Y)$ is uniformly distributed on the interval $[0, 1]$, it is clear that U is uniformly distributed on the interval $\left[\frac{-F_Y(0)}{F_Y(1) - F_Y(0)}, \frac{1 - F_Y(0)}{F_Y(1) - F_Y(0)} \right]$. The advantage of working with U instead of Y is that U has a uniform distribution, and hence kernel smoothing based on U should work much better than when it is based on Y , which can have very long tails.

4.3 Identification of $\Gamma(\cdot)$

In this section, we will explain how we can identify the transformation $\Gamma(\cdot)$ defined in (4.2.1). The way we will do it is similar as in Chiappori et al. (2015) except that we will now work with $U = T(Y)$ instead of Y .

Let $x = (x_1, \dots, x_d)^t$ and define $\varphi(u, x) = F_{U|X}(u, x)$ the conditional distribution function of U given X . Using the independence between X and ε , we have

$$\varphi(u, x) = P(U \leq u | X = x) = P(m(X) + \varepsilon \leq \Gamma(u) | X = x) = F_\varepsilon(\Gamma(u) - m(x)) .$$

Next, Conditions (C2) and (C3) in Section 4.8 imply that the derivatives of $\varphi(u, x)$ with respect to u and x_ρ for some $\rho \in \{1, \dots, d\}$ exist. The derivative of $\varphi(u, x)$ with respect to u is denoted by $\varphi_u(u, x)$ and is given by

$$\varphi_u(u, x) = \frac{\partial}{\partial u} \varphi(u, x) = \Gamma'(u) f_\varepsilon(\Gamma(u) - m(x)) , \quad (4.3.1)$$

where $\Gamma'(u) = \frac{\partial}{\partial u} \Gamma(u)$ exists by Condition (C2). Similarly, the derivative of $\varphi(u, x)$ with respect to x_ρ is denoted by $\varphi_\rho(u, x)$ and is given by

$$\varphi_\rho(u, x) = \frac{\partial}{\partial x_\rho} \varphi(u, x) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) f_\varepsilon(\Gamma(u) - m(x)) . \quad (4.3.2)$$

We have used the fact that the function $m(x)$ is differentiable with respect to x_ρ by condition (C3), see Section 4.8. Moreover, we choose ρ such that the set $\mathcal{A}_\rho = \{x \in \chi : \varphi_\rho(u, x) \neq 0 \ \forall u \in \mathcal{U}_0\}$ is non empty, where $\mathcal{U}_0$ is a compact subset in the interior of $\mathcal{U}$. This last condition is imposed in (C4), see Section 4.8. Hence, dividing (4.3.1) by (4.3.2), we obtain

$$\Gamma'(u) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) \times \frac{\varphi_u(u, x)}{\varphi_\rho(u, x)} . \quad (4.3.3)$$

This last expression will be very important to prove the following Theorem which states how we can identify the transformation $\Gamma(u)$.

Theorem 4.3.1. *Assume (C1)-(C4). Then, for any $\rho \in \{1, \dots, d\}$ such that $\mathcal{A}_\rho$ is non empty, Γ is identified under (J1) and (J2) as*

$$\Gamma(u) = \lambda_\rho(u, x) = \frac{S_\rho(u, x)}{S_\rho(1, x)} ,$$

where $u \in \mathcal{U}_0$, $s_\rho(u, x) = \frac{\varphi_u(u, x)}{\varphi_\rho(u, x)}$ and $S_\rho(u, x) = \int_0^u s_\rho(w, x) dw$. Moreover, the expression $\lambda_\rho(u, x)$ does not depend on ρ nor x .

The proof of this theorem is the same as the proof of Theorem 1 in Chiappori et al. (2015). This theorem implies that we can arbitrarily choose the

component of the vector $x = (x_1, \dots, x_d)^t$ with respect to which we will derive to construct the nonparametric estimators of the transformation in Section 4.4. In practice, we recommend to choose ρ and x such that $\frac{\partial}{\partial x_\rho} m(x)$ is as far as possible from 0, by visualizing this function on a graph for instance. In the following, to construct the estimators of $\Gamma(u)$, we will consider the derivative with respect to x_1 without loss of generality.

Note that we established Theorem 4.3.1 under the particular identification conditions $\Lambda(0) = 0$ and $\Lambda(1) = 1$, which requires that 0 and 1 are in the interior of the support $\mathcal{Y}$ of Y . However, if we consider for instance the square root transform $\Lambda(Y) = \sqrt{Y}$ with $\mathcal{Y} = [0, +\infty[$, it is clearly not the case since the point 0 is at the boundary of $\mathcal{Y}$. To overcome this problem, as we mentioned at the beginning of Section 4.2, the points 0 and 1 can be replaced by any other values as long as they are in the interior of the support $\mathcal{Y}$ of Y . In general, if $\Lambda(\alpha_1) = a_1$ and $\Lambda(\alpha_2) = a_2$ for some $\alpha_1 < \alpha_2$ and $a_1 < a_2$ such that α_1 and α_2 are in the interior of $\mathcal{Y}$, it suffices to define Γ such that $\Lambda = \Gamma \circ \tilde{T}$ where

$$\tilde{U} = \tilde{T}(Y) = (\alpha_2 - \alpha_1) \frac{F_Y(Y) - F_Y(\alpha_1)}{F_Y(\alpha_2) - F_Y(\alpha_1)} + \alpha_1 .$$

This ensures that $T(\alpha_1) = \alpha_1$ and $T(\alpha_2) = \alpha_2$ and Γ satisfies consequently exactly the same identification conditions as Λ . Indeed, $\Gamma(\alpha_1) = \Lambda(T^{-1}(\alpha_1)) = a_1$ and $\Gamma(\alpha_2) = \Lambda(T^{-1}(\alpha_2)) = a_2$. In this situation, following exactly the same reasoning as in the proof of Theorem 4.3.1, Γ is identified under (J1) and (J2) as

$$\Gamma(u) = a_1 + (a_2 - a_1) \frac{\tilde{S}_1(u, x)}{\tilde{S}_1(\alpha_2, x)} ,$$

where $\tilde{S}_1(u, x) = \int_{\alpha_1}^u \frac{\tilde{\varphi}_{\tilde{u}(w, x)}}{\tilde{\varphi}_1(w, x)} dw$ and $\tilde{\varphi}_{\tilde{u}}$ and $\tilde{\varphi}_1$ are the derivatives of the conditional distribution function of $\tilde{U}$ given X with respect to $\tilde{u}$ and x_1 respectively.

4.4 Definition of the new estimators

In this section, we will construct our new nonparametric estimators of the transformation $\Gamma(u)$ that will also be estimators of $\Lambda(y)$ provided we can replace F_Y by its empirical distribution function $\hat{F}_Y$. The idea is to construct an estimator of the function $\lambda_1(u, x)$ introduced in Theorem 4.3.1.

First, for $i = 1, \dots, n$, let $X_i = (X_{i1}, \dots, X_{id})$ and assume that we have randomly drawn an *iid* sample $(X_1, Y_1), \dots, (X_n, Y_n)$ from the nonparametric transformation model (1.3.1). Moreover, we define $U_i = T(Y_i) = \frac{F_Y(Y_i) - F_Y(0)}{F_Y(1) - F_Y(0)}$.

We can rewrite the conditional distribution function of U given X defined in Section 4.3 in the following way:

$$\varphi(u, x) = \frac{p(u, x)}{f_X(x)} , \quad (4.4.1)$$

where

$$p(u, x) = \int_{-\infty}^u f_{U,X}(w, x) dw \quad , \quad f_X(x) = \int_{\mathcal{U}} f_{U,X}(w, x) dw \quad , \quad (4.4.2)$$

and $f_{U,X}(u, x)$ is the joint density function of U and X . Let K be a univariate kernel in $\mathbb{R}$ that satisfies Condition (C5) in Section 4.8, $\mathcal{K}(u) = \int_{-\infty}^u K(w) dw$ and $\mathbf{K}(x)$ be a multivariate product kernel of the form $\mathbf{K}(x) = \prod_{i=1}^d K(x_i)$ with $x = (x_1, \dots, x_d)^t$. We also introduce $h_u > 0$ and $h_x > 0$ two univariate bandwidths that satisfy Condition (C6) in Section 4.8. Consequently, a natural kernel estimator of $\varphi(u, x)$ is given by

$$\hat{\varphi}(u, x) = \frac{\hat{p}(u, x)}{\hat{f}_X(x)} \quad ,$$

where

$$\hat{p}(u, x) = \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(u - \hat{U}_i) \mathbf{K}_{h_x}(X_i - x) \quad \text{and} \quad \hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{K}_{h_x}(X_i - x) \quad , \quad (4.4.3)$$

with $\mathcal{K}_{h_u} = \mathcal{K}(u/h_u)$ and $\mathbf{K}_{h_x}(x) = \mathbf{K}(x/h_x)/h_x^d$. Moreover, $\hat{U}_i$ is defined by

$$\hat{U}_i = \hat{T}(Y_i) = \frac{\hat{F}_Y(Y_i) - \hat{F}_Y(0)}{\hat{F}_Y(1) - \hat{F}_Y(0)} \quad ,$$

where $\hat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq y\}}$ is the empirical distribution of $Y_1, \dots, Y_n$. It is important to notice that we have to work with $\hat{U}_1, \dots, \hat{U}_n$ instead of $U_1, \dots, U_n$ since the U_i 's depend on the distribution function of Y which is an unknown function. This will make an important difference when we will develop the asymptotic properties of our estimators in comparison with Chiappori et al. (2015) who constructed their estimators on the basis of $Y_1, \dots, Y_n$.

As mentioned in Chiappori et al. (2015), we have to use $\mathcal{K}_{h_u}(u - \hat{U}_i)$ instead of $1_{\{\hat{U}_i \leq u\}}$ because we need $\hat{\varphi}(u, x)$ to be differentiable with respect to u . Also, note that we could work with a separate bandwidth for each component X_i of the vector X but wanted to keep the presentation simple.

Next, we consider

$$\varphi_u(u, x) = \frac{\partial}{\partial u} \varphi(u, x) = \frac{p_u(u, x)}{f_X(x)} \quad , \quad (4.4.4)$$

and

$$\varphi_1(u, x) = \frac{\partial}{\partial x_1} \varphi(u, x) = \frac{p_1(u, x)}{f_X(x)} - \frac{p(u, x) f_{X,1}(x)}{f_X^2(x)} \quad , \quad (4.4.5)$$

where $p_u(u, x) = \frac{\partial}{\partial u} p(u, x)$, $p_1(u, x) = \frac{\partial}{\partial x_1} p(u, x)$ and $f_{X,1}(x) = \frac{\partial}{\partial x_1} f_X(x)$. We estimate these quantities respectively by

$$\hat{\varphi}_u(u, x) = \frac{\hat{p}_u(u, x)}{\hat{f}_X(x)} \quad \text{and} \quad \hat{\varphi}_1(u, x) = \frac{\hat{p}_1(u, x)}{\hat{f}_X(x)} - \frac{\hat{p}(u, x) \hat{f}_{X,1}(x)}{\hat{f}_X^2(x)} \quad ,$$

where $\hat{p}_u(u, x) = \frac{\partial}{\partial u} \hat{p}(u, x)$, $\hat{p}_1(u, x) = \frac{\partial}{\partial x_1} \hat{p}(u, x)$ and $\hat{f}_{X,1}(x) = \frac{\partial}{\partial x_1} \hat{f}_X(x)$ and $\hat{p}(u, x)$ and $\hat{f}_X(x)$ are defined in (4.4.3). Finally, we define the following estimator of $\lambda_1(u, x)$:

$$\hat{\lambda}_1(u, x) = \frac{\hat{S}_1(u, x)}{\hat{S}_1(1, x)} \quad \text{where} \quad \hat{S}_1(u, x) = \int_0^u \frac{\hat{\varphi}_u(w, x)}{\hat{\varphi}_1(w, x)} dw . \quad (4.4.6)$$

It is important to remark that the estimator $\hat{\lambda}_1(u, x)$ depends on x despite the fact that it is not the case for the true function $\lambda_1(u, x)$ by Theorem 4.3.1. Consequently, we will have to integrate the function $\hat{\lambda}_1(u, x)$ over χ . We also consider a weighting function $v(x)$ that satisfies Condition (C9) in Section 4.8. Moreover, Theorem 4.3.1 implies that $\Gamma(u) = \lambda_1(u, x)$ and then

$$\Gamma(u) = \arg \min_{q_m \in \mathbb{R}} \int_{\chi} v(x) L(\lambda_1(u, x) - q_m) dx ,$$

where $L(\cdot)$ is a given loss function. In this paper, we will consider $L(z) = z^2$ and $L(z) = |z|$ which gives more precisely the two following nonparametric estimators for the transformation $\Gamma(u)$:

$$\hat{\Gamma}_{LS}(u) = \int_{\chi} v(x) \hat{\lambda}_1(u, x) dx , \quad (4.4.7)$$

and

$$\hat{\Gamma}_{LAD}(u) = \arg \min_{q_m \in \mathbb{R}} \int_{\chi} v(x) \left| \hat{\lambda}_1(u, x) - q_m \right| dx . \quad (4.4.8)$$

For example, if we want to estimate $\Lambda(2)$, we just have to calculate $\hat{\Gamma}_{LS}(\hat{T}(2))$ and $\hat{\Gamma}_{LAD}(\hat{T}(2))$ where $\hat{T}(y) = \frac{\hat{F}_Y(y) - \hat{F}_Y(0)}{\hat{F}_Y(1) - \hat{F}_Y(0)}$.

In summary, the construction of our estimators is similar as in Chiappori et al. (2015) except that we proposed a way to overcome the facts that the distribution of Y can be very asymmetric and can have very long tails, which can lead to bad estimations of $\Lambda(y)$ due to the poor quality of the kernel smoothing. In Section 4.6, we will compare the performance of our new estimators to those of Chiappori et al. (2015) and we will see that the transformation of $U = T(Y)$ is recommended, especially for large values of Y . However, the disadvantage of transforming Y and considering kernel smoothing on $\hat{U}$ is the asymptotic theory that will be considerably more difficult in comparison with Chiappori et al. (2015). Indeed, we have a lot of technical results to prove before establishing the limiting distributions of the estimators $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$, see Section 4.9.1.

4.5 Asymptotic results

In this section, we will present the theorems that establish the limiting distributions of the estimators $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$ introduced in (4.4.7) and (4.4.8)

respectively. First, we need to introduce the following notations:

$$D_{p,0}(u, x) = \frac{\varphi_u(u, x)f_{X,1}(x)}{\varphi_1^2(u, x)f_X^2(x)} \quad , \quad D_{p,u}(u, x) = \frac{1}{f_X(x)\varphi_1(u, x)} \quad ,$$

$$D_{p,1}(u, x) = \frac{-\varphi_u(u, x)}{f_X(x)\varphi_1^2(u, x)} \quad , \quad D_{f,0}(u, x) = \frac{-\varphi_u(u, x)\varphi(u, x)f_{X,1}(x)}{\varphi_1^2(u, x)f_X^2(x)} \quad ,$$

and

$$D_{f,1}(u, x) = \frac{\varphi_u(u, x)\varphi(u, x)}{\varphi_1^2(u, x)f_X(x)} \quad .$$

For any functions $\tilde{p}(u, x)$ and $\tilde{f}(x)$, let $\tilde{p}_u(u, x)$ and $\tilde{p}_1(u, x)$ be respectively the derivatives of $\tilde{p}(u, x)$ with respect to u and x_1 , where x_1 is the first component of the vector $x \in \mathbb{R}^d$, and let $\tilde{f}_1(x)$ be the derivative of the function $\tilde{f}(x)$ with respect to x_1 . We also define the following functionals

$$\begin{aligned} & \nabla_p S_1(u, x)[\tilde{p}] \\ &= \int_0^u \left\{ D_{p,0}(w, x)\tilde{p}(w, x) + D_{p,u}(w, x)\tilde{p}_u(w, x) + D_{p,1}(w, x)\tilde{p}_1(w, x) \right\} dw \quad , \end{aligned} \quad (4.5.1)$$

and

$$\nabla_f S_1(u, x)[\tilde{f}] = \int_0^u \left\{ D_{f,0}(u, x)\tilde{f}(x) + D_{f,1}(u, x)\tilde{f}_1(x) \right\} du \quad . \quad (4.5.2)$$

The following two theorems establish respectively the limiting distributions of $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$. However, it is important to remark that the derivative of $\varphi(u, x)$ with respect to a component of x is zero for all $x \in \chi$ if u is at the lower boundary of the support $\mathcal{U}$ of U . Hence, Assumptions (C4) and (C8), given in Section 4.8, cannot be fulfilled for all $u \in \mathcal{U}$. Consequently, we have to restrict Assumptions (C4) and (C8) and also the following two theorems to any $u \in \mathcal{U}_0$, where $\mathcal{U}_0$ is a compact subset in the interior of $\mathcal{U}$. In particular, if 0 is the lower boundary of $\mathcal{Y}$, it can be replaced by any other value in the interior of $\mathcal{Y}$, see the remark at the end of Section 4.3. Note that this will also avoid boundary problems when we will do kernel smoothing on U .

Theorem 4.5.1. *Assume (C1)-(C9). Then, under (J1) and (J2), the process $\sqrt{n}(\hat{\Gamma}_{LS}(u) - \Gamma(u))$ converges weakly to $\mathcal{N}(u)$, where $u \in \mathcal{U}_0$ and $\mathcal{N}$ is a centered Gaussian process with covariance function given by $\Delta(u_1, u_2) = E[\delta_i^v(u_1)\delta_i^v(u_2)]$, where $\delta_i^v(u) = \delta_i^{v_1}(1, u) - \delta_i^{v_2}(u, 1)$, $v_1(u_0, x) = \frac{v(x)}{S_1(u_0, x)}$, $v_2(u_0, x) = \frac{v(x)S_1(u_0, x)}{S_1^2(1, x)}$, $\delta_i^{\tilde{v}}(u_0, u)$ is defined in (4.9.33) for $u_0 \in \mathcal{U}_0$ and $\tilde{v}$ equals either v_1 or v_2 .*

Theorem 4.5.2. *Assume (C1)-(C9). Then, under (J1) and (J2), the process $\sqrt{n}(\hat{\Gamma}_{LAD}(u) - \Gamma(u))$ converges weakly to $\mathcal{N}(u)$, where $u \in \mathcal{U}_0$ and $\mathcal{N}(u)$ is the same centered Gaussian process as in Theorem 4.5.1.*

The proofs of these two theorems are given in Section 4.9.2. It is important to notice that the function $\delta_i^{\tilde{v}}(u_0, u)$ is not the same as the one obtained in Chiappori et al. (2015). The first three terms of the expression $\delta_i^{\tilde{v}}(u_0, u)$ are the same as in Chiappori et al. (2015) and the last two terms in $\delta_i^{\tilde{v}}(u_0, u)$ are new contributions to the final asymptotic representation that are due to the estimation of U by $\hat{U}$. We refer to Proposition 4.9.3 that shows in detail where these last two terms come from. Chiappori et al. (2015) did not need this proposition since they worked directly with Y .

4.6 Simulations

In this section, we perform simulations in order to compare the performance of the new nonparametric estimators $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$ of the transformation with the estimators $\hat{\Lambda}_{C,LS}(y)$ and $\hat{\Lambda}_{C,LAD}(y)$ defined in (1.3.2) and (1.3.3) and given in Chiappori et al. (2015).

We consider $d = 1$ and the simulated model is $\Lambda(Y_i) = 6X_i - 3 + \varepsilon_i$, where $X_1, \dots, X_n$ are independent uniform random variables on $[0, 1]$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent standard normal random variables truncated on $[-3, 3]$. Here, we will consider the four following different transformations:

$$\Lambda_1(Y) = \begin{cases} \frac{\log(Y+1)}{\log 2} & \text{if } Y \geq 0 \\ \frac{1-(1-Y)^2}{2 \log 2} & \text{if } Y < 0, \end{cases}, \quad \Lambda_2(Y) = \begin{cases} \frac{\sqrt{Y+1}-1}{\sqrt{2}-1} & \text{if } Y \geq 0 \\ \frac{1-(1-Y)^{3/2}}{3(\sqrt{2}-1)} & \text{if } Y < 0, \end{cases},$$

$$\Lambda_3(Y) = Y \quad \text{and} \quad \Lambda_4(Y) = \begin{cases} \frac{(Y+1)^{3/2}-1}{2\sqrt{2}-1} & \text{if } Y \geq 0 \\ \frac{3(1-\sqrt{1-Y})}{2\sqrt{2}-1} & \text{if } Y < 0, \end{cases}.$$

Hence, we can easily verify that Y takes values on the intervals $[-2.05, 63]$, $[-3.15, 11.15]$, $[-6, 6]$ and $[-20.69, 4.23]$ when we use the above simulated model with the transformations $\Lambda_j(Y)$, for $j = 1, \dots, 4$, respectively.

Note also that these four transformations are particular cases of the Yeo and Johnson (2000) transformation up to a multiplicative constant. Indeed, the Yeo and Johnson (2000) is given by

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

and $\Lambda_1(Y) = \frac{\Lambda_{\theta=0}(Y)}{\log 2}$, $\Lambda_2(Y) = \frac{\Lambda_{\theta=0.5}(Y)}{2(\sqrt{2}-1)}$, $\Lambda_3(Y) = \Lambda_{\theta=1}(Y)$ and $\Lambda_4(Y) = \frac{3\Lambda_{\theta=1.5}(Y)}{2(2\sqrt{2}-1)}$. These different multiplicative constants were introduced in order to ensure that $\Lambda_j(1) = 1$ for $j = 1, \dots, 4$. Moreover, the Yeo and Johnson

transformation is such that $\Lambda_\theta(0) = 0$ for all θ . Hence, $\Lambda_j(0) = 0$ for $j = 1, \dots, 4$ and the normalization conditions (J1) are satisfied.

We will now explain how we estimate the expressions $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$ in practice. The reasoning will be the same for $\hat{\Lambda}_{C,LS}(y)$ and $\hat{\Lambda}_{C,LAD}(y)$. Note that $K(x) = \frac{3}{4}(1-x^2)1_{\{|x| \leq 1\}}$ is the Epanechnikov kernel and we select the bandwidths h_x and h_u by a classical bandwidth selection rule for kernel density estimation. For simplicity, we choose here the normal reference rule, i.e., $\hat{h}_x = (40\sqrt{\pi})^{1/5}n^{-1/5}\hat{\sigma}_x$ and $\hat{h}_u = (40\sqrt{\pi})^{1/5}n^{-1/5}\hat{\sigma}_u$ where $\hat{\sigma}_x$ and $\hat{\sigma}_u$ are the classical estimator of the standard deviations of X and U respectively. First, we generate a grid $x_1^*, \dots, x_{N_x}^*$ of N_x equidistant points between $\min_{1 \leq j \leq n} X_j$ and $\max_{1 \leq j \leq n} X_j$. Hence, we can compute

$$\hat{S}_1(u, x_\ell^*) = \int_0^u \frac{\hat{\varphi}_u(w, x_\ell^*)}{\hat{\varphi}_1(w, x_\ell^*)} dw \quad \text{and} \quad \hat{\lambda}_1(u, x_\ell^*) = \frac{\hat{S}_1(u, x_\ell^*)}{\hat{S}_1(1, x_\ell^*)},$$

for each $\ell = 1, \dots, N_x$. However, in practice, the expression $\hat{\varphi}_1(w, x_\ell^*)$ could be very close to 0 for some $\ell = 1, \dots, N_x$, and consequently, the integral $\int_0^u \frac{\hat{\varphi}_u(w, x_\ell^*)}{\hat{\varphi}_1(w, x_\ell^*)} dw$ could diverge. To overcome this problem, we will remove from the original sample $x_1^*, \dots, x_{N_x}^*$ the values x_ℓ^* such that $\hat{S}_1(u, x_\ell^*)$ diverges. We will also remove the x^* -values that are within 0.01 of the values x_ℓ^* such that $\hat{S}_1(u, x_\ell^*)$ diverges, even if the corresponding integrals do not diverge. The purpose is to avoid too small values of $\hat{\varphi}_1(w, x_\ell^*)$ and thus too large values of $\hat{\lambda}_1(\cdot)$ that would deteriorate the final estimation of the transformation. This is allowed since $\hat{\lambda}_1(u, x)$ estimates $\Gamma(u)$ for all $x \in \chi$. We assume finally that n_x values remain, denoted by $\tilde{x}_1, \dots, \tilde{x}_{n_x}$, from our original sample $x_1^*, \dots, x_{N_x}^*$. In the following, we will use $N_x = 100$. Finally, we consider the weighting function $v(x) = 1$, for all x such that the above integral does not diverge, and $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$ are consequently approximated by

$$\tilde{\Gamma}_{LS}(u) = \frac{1}{n_x} \sum_{i=1}^{n_x} \hat{\lambda}_1(u, \tilde{x}_i),$$

and

$$\tilde{\Gamma}_{LAD}(u) = \text{median}(\hat{\lambda}_1(u, \tilde{x}_1), \dots, \hat{\lambda}_1(u, \tilde{x}_{n_x})).$$

Then, the transformations $\Lambda_1(y)$, $\Lambda_2(y)$, $\Lambda_3(y)$ and $\Lambda_4(y)$ will be estimated by $\tilde{\Gamma}_{LS}(\hat{T}(y))$ and $\tilde{\Gamma}_{LAD}(\hat{T}(y))$. Using a similar procedure, we can construct $\tilde{\Lambda}_{C,LS}(y)$ and $\tilde{\Lambda}_{C,LAD}(y)$ the approximations of $\hat{\Lambda}_{C,LS}(y)$ and $\hat{\Lambda}_{C,LAD}(y)$.

Each graph of Figure 4.1 shows the true transformation $\Lambda_1(y)$ in black and each line in grey represents one estimation of $\Lambda_1(y)$ obtained with one sample of size $n = 200$. The top graphs on the left and on the right of Figure 4.1 were respectively obtained with the estimators $\tilde{\Gamma}_{LS}(\hat{T}(y))$ and $\tilde{\Gamma}_{LAD}(\hat{T}(y))$ developed in this paper and the bottom graphs on the left and on the right of Figure 4.1 were respectively obtained with the estimators $\tilde{\Lambda}_{C,LS}(y)$ and

$\tilde{\Lambda}_{C,LAD}(y)$, each of the four cases on the basis of 200 samples of size $n = 200$. Figures 4.2 to 4.4 are constructed exactly in the same way but for $\Lambda_2(y)$, $\Lambda_3(y)$ and $\Lambda_4(y)$ respectively.

First, we observe on these four figures that the estimators constructed with the median, i.e., the LAD-estimators, outperform in all tried scenarios the corresponding estimators constructed with the mean, i.e., the LS-estimators. This can be explained by the fact that the mean version of the estimator is more sensitive to outliers in $\hat{\lambda}_1(u, \tilde{x})$ as we vary $\tilde{x}$. Hence, it is not surprising if the LAD-estimators behave better than the corresponding LS-estimators from the biases and the variances points of view. In the following, we will thus concentrate our conclusions on the LAD versions of the estimators.

We also observed in Figures 4.1 to 4.4 that the estimations of the transformation Λ are not necessarily monotone and especially the ones obtained with the mean versions of the estimators. As future research, it would be possible to make these estimators monotone in a second step. We could for example follow the spirit of the paper of Dette et al. (2006), who proposed a method for a monotone estimation of the regression function in a heteroscedastic regression model.

Next, we also introduce Tables 4.1 to 4.4 that show the estimations of the biases, variances and mean squared errors of the LAD-estimators of the transformations $\Lambda_j(y)$, $j = 1, \dots, 4$ evaluated at some specific values of y .

Note that $\tilde{\Gamma}_{LAD}(\hat{T}(0)) = \tilde{\Gamma}_{LAD}(0) = 0$, $\tilde{\Lambda}_{C,LAD}(0) = 0$, $\tilde{\Gamma}_{LAD}(\hat{T}(1)) = \tilde{\Gamma}_{LAD}(1) = 1$ and $\tilde{\Lambda}_{C,LAD}(1) = 1$ by construction, which justifies the fact that all the grey lines from Figures 4.1 to 4.4 cross the points $(0, 0)$ and $(1, 1)$. At these points, the variances of all estimators are equal to 0 and we can see in Tables 4.1 to 4.4 that these variances increase for both LAD-estimators when y gets away from 0 in negative values and from 1 in positive values.

Moreover, when y is close to 0 and 1, the estimations of $\Lambda_j(y)$, for $j = 1, \dots, 4$, are slightly better in terms of variances and biases when we use $\tilde{\Lambda}_{C,LAD}(y)$, see for example $\Lambda_j(0.5)$. However, when we estimate $\Lambda_1(y)$, $\Lambda_2(y)$ and $\Lambda_4(y)$ and when y gets away from 0 in negative values and from 1 in positive values, the absolute value of the biases of the $\tilde{\Lambda}_{C,LAD}(y)$ estimator become significantly bigger and bigger, in comparison with the biases of $\tilde{\Gamma}_{LAD}(y)$ that remain globally constant. This gives very poor estimations of the transformations $\Lambda_j(y)$ for large values of $|y|$ when we use $\tilde{\Lambda}_{C,LAD}(y)$, see for examples the estimations of $\Lambda_1(10)$, $\Lambda_2(5)$ and $\Lambda_4(-4)$, even if the estimations of $\Lambda_j(y)$, for $j = 1, \dots, 4$, are globally better in terms of variance when we use $\tilde{\Lambda}_{C,LAD}(y)$. As explained in the introduction, it is due to the fact that $\tilde{\Lambda}_{C,LAD}(y)$ is constructed by doing smoothing on $Y_1, \dots, Y_n$ which is not uniformly distributed. As a consequence, the kernel estimators behaves not always well in that cases and especially for large values of y .

In conclusion, as we can see in Figures 4.1 to 4.4 and in Tables 4.1 to 4.4, $\tilde{\Gamma}_{LAD}(\hat{T}(y))$ clearly outperforms $\tilde{\Lambda}_{C,LAD}(y)$ when we consider $\Lambda_1(y)$, $\Lambda_2(y)$

and $\Lambda_4(y)$ and large values of $|y|$ and $\tilde{\Lambda}_{C,LAD}(y)$ only outperforms $\tilde{\Gamma}_{LAD}(\hat{T}(y))$ when we consider the identity transformation $\Lambda_3(y)$. Moreover, we observe that $\tilde{\Lambda}_{C,LAD}(y)$ gives very poor estimations of $\Lambda_1(y)$, $\Lambda_2(y)$ and $\Lambda_4(y)$ when $|y|$ increases, which suggests that $\tilde{\Lambda}_{C,LAD}(y)$ is clearly inappropriate when we really transform the response. Indeed, the only case where $\tilde{\Lambda}_{C,LAD}(y)$ performed well corresponds to a nonparametric regression model without transformation of the response.

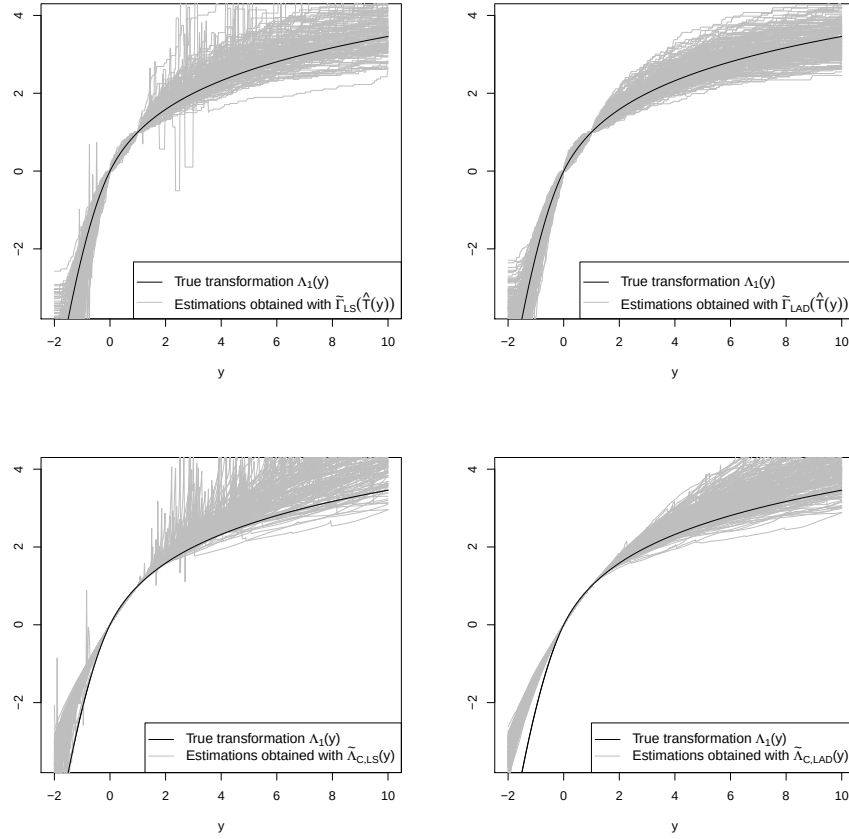


Figure 4.1: Comparison of the different nonparametric estimators of the transformation $\Lambda_1(Y)$.

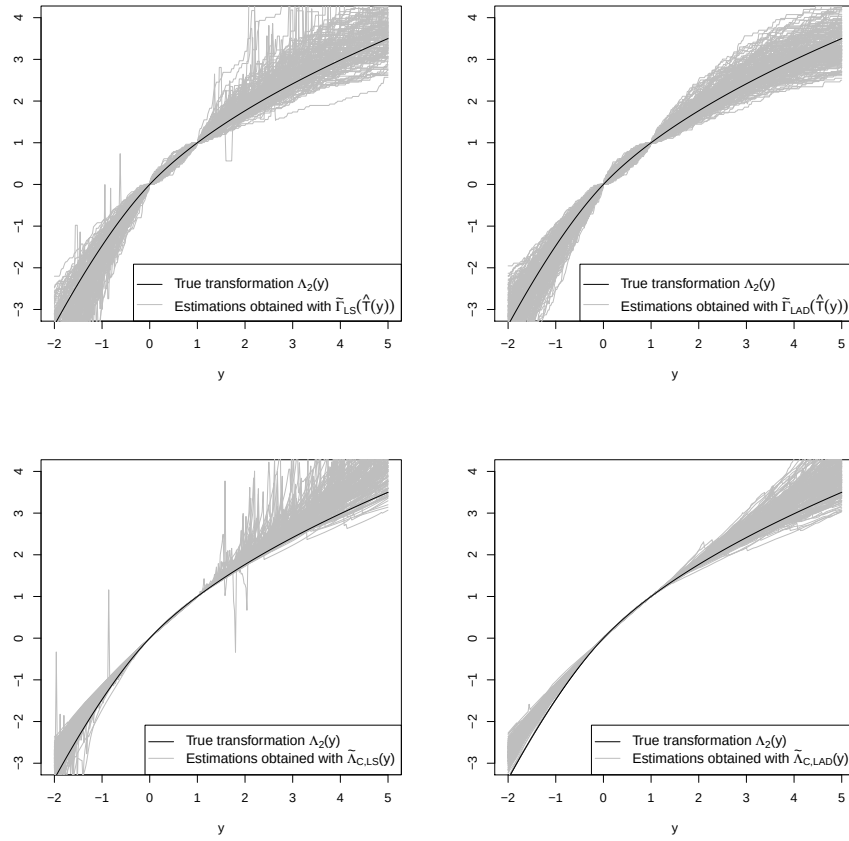


Figure 4.2: Comparison of the different nonparametric estimators of the transformation $\Lambda_2(Y)$.

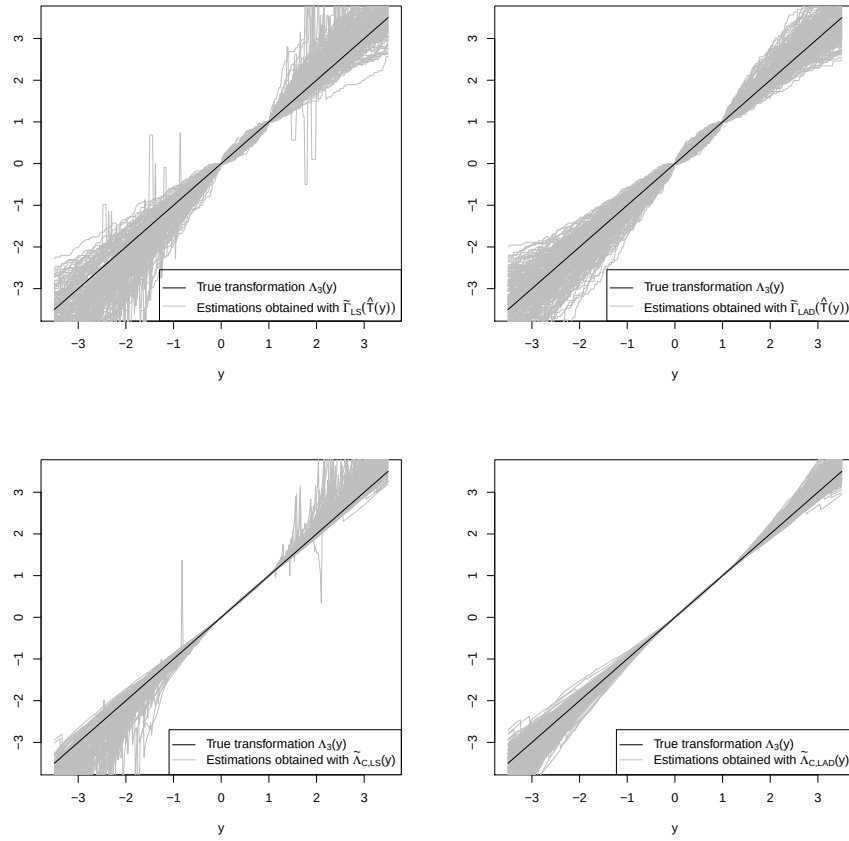


Figure 4.3: Comparison of the different nonparametric estimators of the transformation $\Lambda_3(Y)$.

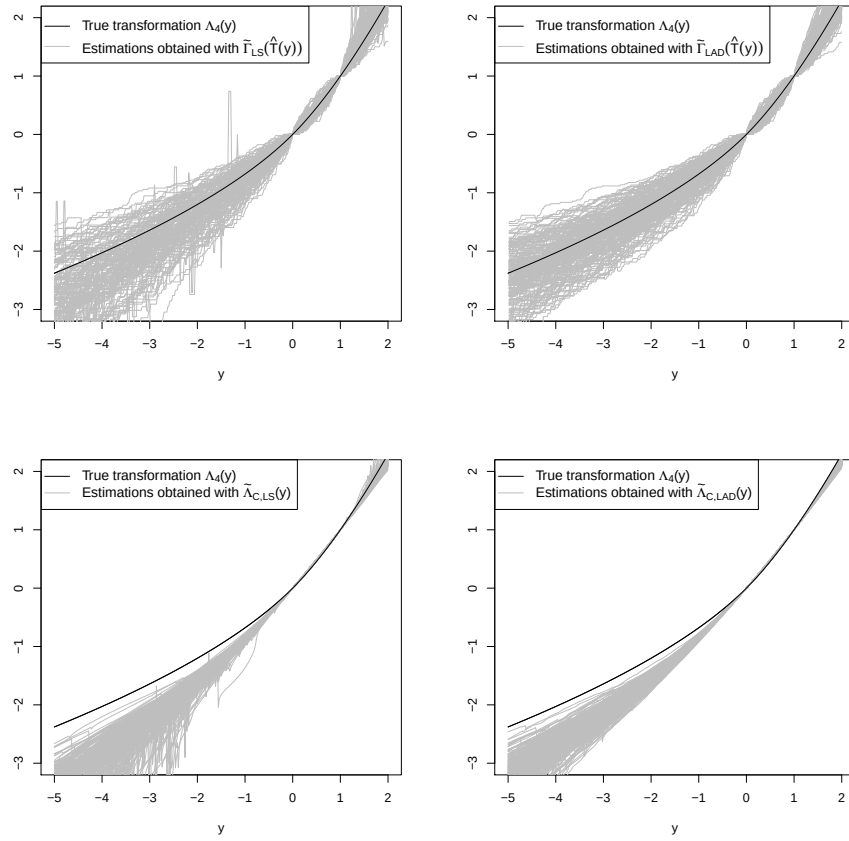


Figure 4.4: Comparison of the different nonparametric estimators of the transformation $\Lambda_4(Y)$.

Some values of y	$\tilde{\Gamma}_{LAD}(\hat{T}(y))$			$\tilde{\Lambda}_{C,LAD}(y)$		
	Bias	Var	MSE	Bias	Var	MSE
-1.5	0.2440	0.4682	0.5277	1.5102	0.0278	2.3087
-1	-0.1338	0.2151	0.2330	0.7600	0.0073	0.5850
0.5	0.0012	0.0076	0.0076	-0.0365	0.0001	0.0014
2	0.0577	0.0364	0.0397	0.0809	0.0033	0.0099
6	0.1121	0.1706	0.1831	0.3777	0.1420	0.2846
8	0.0222	0.2096	0.2101	0.5501	0.3234	0.6260
10	-0.0682	0.2466	0.2513	0.7581	0.5713	1.1460

Table 4.1: Estimations of the biases, variances and mean squared errors of the two LAD-estimators of $\Lambda_1(y)$ evaluated at some specific values of y .

Some values of y	$\tilde{\Gamma}_{LAD}(\hat{T}(y))$			$\tilde{\Lambda}_{C,LAD}(y)$		
	Bias	Var	MSE	Bias	Var	MSE
-2	0.0348	0.4137	0.4149	0.6138	0.0418	0.4185
-1.5	-0.1346	0.2417	0.2598	0.4054	0.0144	0.1788
0.5	-0.0470	0.0079	0.0079	-0.0156	0.0001	0.0003
2	0.0775	0.0479	0.0539	0.0642	0.0042	0.0083
4.5	-0.0012	0.2204	0.2204	0.3747	0.1665	0.3069
5	-0.0897	0.2440	0.2520	0.4769	0.2568	0.4842
6	-0.3565	0.2980	0.4251	0.7784	0.6872	1.2931

Table 4.2: Estimations of the biases, variances and mean squared errors of the two LAD-estimators of $\Lambda_2(y)$ evaluated at some specific values of y .

Some values of y	$\tilde{\Gamma}_{LAD}(\hat{T}(y))$			$\tilde{\Lambda}_{C,LAD}(y)$		
	Bias	Var	MSE	Bias	Var	MSE
-3	-0.0830	0.3663	0.3732	-0.1350	0.1078	0.1260
-2	-0.1262	0.1836	0.1995	-0.0433	0.0298	0.0317
-1	-0.0530	0.0590	0.0618	-0.0063	0.0038	0.0039
0.5	-0.0062	0.0080	0.0080	-0.0016	0.0001	0.0001
1.5	0.0497	0.0324	0.0348	0.0107	0.0007	0.0008
2	0.1039	0.0783	0.0891	0.0360	0.0049	0.0062
3	0.0686	0.1915	0.1962	0.1361	0.0400	0.0585

Table 4.3: Estimations of the biases, variances and mean squared errors of the two LAD-estimators of $\Lambda_3(y)$ evaluated at some specific values of y .

Some values of y	$\tilde{\Gamma}_{LAD}(\hat{T}(y))$			$\tilde{\Lambda}_{C,LAD}(y)$		
	Bias	Var	MSE	Bias	Var	MSE
-5	-0.1344	0.2411	0.2591	-0.8359	0.0917	0.7904
-4	-0.1235	0.1884	0.2037	-0.6624	0.0503	0.4892
-3	-0.0989	0.1357	0.1455	-0.5121	0.0225	0.2847
-2	-0.0640	0.0755	0.0796	-0.3534	0.0062	0.1311
-1	-0.0341	0.0349	0.0361	-0.1689	0.0008	0.0293
0.5	-0.0054	0.0082	0.0082	0.0230	0.00002	0.0006
1.5	0.0595	0.0381	0.0416	-0.0575	0.0002	0.0035

Table 4.4: Estimations of the biases, variances and mean squared errors of the two LAD-estimators of $\Lambda_4(y)$ evaluated at some specific values of y .

4.7 Conclusions and recommendations

First, we observed that the LAD-estimators outperform in all tried scenarios the corresponding LS-estimators, which is due to the fact that the mean versions of the estimator are more sensitive to outliers in $\hat{\lambda}_1(u, \tilde{x})$ as we vary $\tilde{x}$.

Moreover, when y is close to 0 and 1, or more generally when y is close to the points that define the identification conditions, the estimations of the transformation are slightly better in terms of variances and biases when we use the estimators given in Chiappori et al. (2015). However, when y gets away from these points, even if the variances of all the estimators increase, the absolute value of the biases of the estimators given in Chiappori et al. (2015) become significantly bigger and bigger, in comparison with the biases of our new estimators that remain globally constant.

In conclusion, our LAD-estimator clearly outperforms the LAD-estimator given in Chiappori et al. (2015) when the transformation of the response is not the identity and when we consider values of y that are far away from the points that define the identification conditions. Moreover, the LAD-estimator given in Chiappori et al. (2015) is clearly inappropriate when the transformation of the response is not the identity and only outperforms our LAD-estimator when we consider values of y close to these identification points or when we consider a nonparametric regression model without transformation of the response.

An optimal nonparametric estimator of the transformation could be obtained by combining the estimator given in Chiappori et al. (2015) when y is close to the points that define the identification conditions and our new estimator when y is far away from these points, so that we get the best from the two estimators. However, important questions arise such as the determination of the “critical” point(s), depending on the identification conditions and the true transformation, from which we have to switch from one estimator to the other. This could be subject to future research.

Finally, another way to improve the performance of these estimators could

be to fix the points that define the identification conditions according to the interval on which we are interested to estimate the transformation.

4.8 Assumptions

The main results of the asymptotic theory require the following regularity conditions on the distributions of ε given X , the transformation $\Gamma(u)$, the regression function $m(x)$, the kernel K , the bandwidths h_x and h_u , the joint density function of U and X and the weighting function v :

- (C1) The distribution function F_ε of ε is absolutely continuous and has a density f_ε that is continuous on its support. Moreover, X and ε are independent and the support $\mathcal{Y}$ of Y is a connected subset of $\mathbb{R}$.
- (C2) The transformation $\Gamma(u)$ is strictly increasing and continuously differentiable on $\mathcal{U}$.
- (C3) The regression function $m(x)$ is continuously differentiable with respect to x_j on χ for $j = 1, \dots, d$.
- (C4) The set $\mathcal{A}_\rho = \{x \in \chi : \varphi_\rho(u, x) \neq 0 \ \forall u \in \mathcal{U}_0\}$ is nonempty for some $\rho \in \{1, \dots, d\}$ and some compact subset $\mathcal{U}_0$ in the interior of $\mathcal{U}$.
- (C5) The kernel K is symmetric, has support $[-1, 1]$, $K(-1) = K(1) = 0$, $\int K(z) dz = 1$, $\int z^\ell K(z) dz = 0$ for $\ell = 1, \dots, m-1$ and $\int z^m K(z) dz < \infty$. Moreover, K is m -times continuously differentiable and K and K' are of bounded variation.
- (C6) The bandwidths h_x and h_u satisfy $\sqrt{n}h_x^m \rightarrow 0$, $\sqrt{n}h_u^m \rightarrow 0$, $\frac{\sqrt{n}h_x^{d+2}}{\log n} \rightarrow \infty$ and $\frac{\sqrt{n}h_x^2 h_x^d}{\log n} \rightarrow \infty$.
- (C7) The joint density function $f_{Y,X}(y, x)$ of (Y, X) is uniformly bounded, $m+2$ -times continuously differentiable and all these derivatives are uniformly bounded, i.e.,

$$\sup_{(y,x): T(y) \in \mathcal{U}_0, x \in \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial y^{\alpha_0} \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} f_{Y,X}(y, x) \right| < \infty ,$$

for all $\alpha_0, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 0, \dots, d$ and where $|\alpha| = \sum_{j=0}^d \alpha_j$ is such that $|\alpha| \leq m+2$. We also assume that $\inf_{y: T(y) \in \mathcal{U}_0} f_Y(y) > 0$, where f_Y is the density function of Y .

- (C8) We assume that $\inf_{x \in \chi_0} f_X(x) > 0$, $\inf_{(u,x) \in \mathcal{U}_0 \times \chi_0} |\varphi_1(u, x)| > 0$ and $\inf_{x \in \chi_0} |S_1(1, x)| > 0$ for some compact subset $\mathcal{U}_0$ in the interior of $\mathcal{U}$ and where $\chi_0 \subseteq \mathcal{A}_\rho$ is the compact support of the weighting function $v(x)$ defined in (C9).

- (C9) The weighting function $v(x)$ has compact support $\chi_0 \subseteq \mathcal{A}_\rho$ with nonempty interior and satisfies $\int_{\chi_0} v(x) dx = 1$, v is continuous on χ , is m -times continuously differentiable and all these derivatives are uniformly bounded in the interior of χ_0 , i.e.,

$$\sup_{x \in \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} v(x) \right| < \infty ,$$

for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m$.

Note that several of these assumptions are the same as in Chiappori et al. (2015). It is important to remark that $m > d+2$ is a consequence of Assumption (C6), which means that a higher-order kernel is required to ensure that the four conditions of Assumption (C6) are satisfied when the dimension d of X becomes large.

Moreover, Condition (C7) implies that the joint density $f_{U,X}(u, x)$ is uniformly bounded, $m+2$ -times continuously differentiable and all these derivatives are uniformly bounded, i.e.,

$$\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial u^{\alpha_0} \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} f_{U,X}(u, x) \right| < \infty ,$$

since $f_{U,X}(u, x)$ can be rewritten as

$$f_{U,X}(u, x) = \frac{f_{Y,X}(T^{-1}(u), x)}{f_Y(T^{-1}(u))} (F_Y(1) - F_Y(0)) .$$

Hence, the functions $p(u, x)$, $p_u(u, x)$, $p_1(u, x)$, $f_X(x)$ and $f_{X,1}(x)$ are uniformly bounded, at least m -times continuously differentiable and all these derivatives are uniformly bounded by Condition (C7). Indeed, these five functions can be rewritten as functions depending only of $f_{U,X}(u, x)$, see (4.4.2). As a consequence, using (4.4.1), (4.4.4), (4.4.5) and the fact that $\inf_{x \in \chi_0} f_X(x) > 0$ by condition (C8), the functions $\varphi(u, x)$, $\varphi_u(u, x)$ and $\varphi_1(u, x)$ are also uniformly bounded, at least m -times continuously differentiable and all these derivatives are uniformly bounded. Finally, these conclusions are again the same for the functions $D_{p,0}(u, x)$, $D_{p,u}(u, x)$, $D_{p,1}(u, x)$, $D_{f,0}(u, x)$, $D_{f,1}(u, x)$ and $S_1(u, x)$ using the definitions at the beginning of Section 4.5, the definition of $S_1(u, x)$ in Theorem 4.3.1 and the facts that $\inf_{x \in \chi_0} f_X(x) > 0$ and $\inf_{(u,x) \in \mathcal{U}_0 \times \chi_0} |\varphi_1(u, x)| > 0$ by condition (C8). All these consequences of Condition (C7) will be constantly used in the different proofs in Sections 4.9.1 and 4.9.2.

Finally, since $v(x) = 0$ for $x \notin \chi_0$, Condition (C9) implies that $\int_{\chi} v(x) dx = \int_{\chi_0} v(x) dx = 1$ and also $v(x) = 0$ for values of x at the boundary of χ_0 since v is assumed to be continuous on χ . The last two properties of v will be necessary

to prove Theorem 4.5.1. Indeed, to prove Theorem 4.5.1, we will need to apply Propositions 4.9.6 and 4.9.7 given in Section 4.9.1 with $v_p(u_0, x) = v_f(u_0, x) = v(x)/S_1(u_0, x)$ and with $v_p(u_0, x) = v_f(u_0, x) = v(x)S_1(u_0, x)/S_1^2(1, x)$ and these two propositions will require that the functions v_p and v_f are continuous on χ and equal 0 for values of x on the boundary of χ_0 .

4.9 Proofs

4.9.1 Useful technical results

In this section, we introduce some technical propositions that will be useful to prove the main results in the next sections. First, we show the three following results that will be constantly used in the different proofs of the other propositions below.

Proposition 4.9.1. *Let $a, \hat{a}, b \neq 0$ and $\hat{b} \neq 0$ be some parameters or functions. Then,*

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) - \frac{(\hat{b} - b)}{b\hat{b}} \left\{ \hat{a} - a - \frac{a(\hat{b} - b)}{b} \right\}.$$

Proof. First, note that

$$\begin{aligned} \frac{\hat{a}}{\hat{b}} - \frac{a}{b} &= \left(\frac{\hat{a}}{\hat{b}} - \frac{a}{b} \right) + \left(\frac{a}{b} - \frac{a\hat{b}}{b^2} \right) + \left(\frac{\hat{a}}{\hat{b}} - \frac{\hat{a}}{b} + \frac{a\hat{b}}{b^2} - \frac{a}{b} \right) \\ &= \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) + \left(\frac{\hat{a}b^2 - \hat{a}b\hat{b} + a\hat{b}^2 - ab\hat{b}}{b^2\hat{b}} \right). \end{aligned}$$

Next, we also have

$$\begin{aligned} \hat{a}b^2 - \hat{a}b\hat{b} + a\hat{b}^2 - ab\hat{b} &= b(\hat{a}b - a\hat{b}) - \hat{b}(\hat{a}b - a\hat{b}) \\ &= (b - \hat{b})(\hat{a}b - a\hat{b} + ab) \\ &= (b - \hat{b}) \left\{ b(\hat{a} - a) - a(\hat{b} - b) \right\}. \end{aligned}$$

Hence,

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) - \frac{(\hat{b} - b)}{b^2\hat{b}} \left\{ b(\hat{a} - a) - a(\hat{b} - b) \right\}.$$

This concludes the proof. $\square$

Proposition 4.9.2. *Assume (C1)-(C9), (J1) and (J2). Then,*

$$\frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ \mathcal{K}_{h_u}(w - U_i) - 1_{\{U_i \leq w\}} \right\} dw = o_P(n^{-1/2}) ,$$

uniformly in $u, u_0 \in \mathcal{U}_0$ and where ℓ is some m -times continuously differentiable function with respect to each component and all these derivatives are uniformly bounded.

Proof. We define the following notation:

$$T(u_0, u) = \frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ \mathcal{K}_{h_u}(w - U_i) - 1_{\{U_i \leq w\}} \right\} dw .$$

First, note that

$$\mathcal{K}_{h_u}(w - U_i) = \int_{-\infty}^{\frac{w - U_i}{h_u}} K(t) dt = \int_{-\infty}^{+\infty} 1_{\{U_i \leq w - th_u\}} K(t) dt .$$

Moreover, since $\int K(t) dt = 1$ by condition (C5), we have $\ell(u_0, w, X_i) 1_{\{U_i \leq w\}} = \int \ell(u_0, w, X_i) 1_{\{U_i \leq w\}} K(t) dt$ and consequently

$$\begin{aligned} T(u_0, u) &= \frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \int_{-\infty}^{+\infty} \left\{ 1_{\{U_i \leq w - th_u\}} - 1_{\{U_i \leq w\}} \right\} K(t) dt dw \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} \left\{ q(X_i, U_i | u_0, u, th_u) - q(X_i, U_i | u_0, u, 0) \right\} K(t) dt , \end{aligned} \tag{4.9.1}$$

where

$$q(x, z | u_0, u, a) = \int_0^u \ell(u_0, w, x) 1_{\{w \leq z + a\}} dw .$$

We also define the class

$$\mathcal{F} = \left\{ (x, z) \rightarrow q(x, z | u_0, u, a) : u_0, u \in \mathcal{U}_0, a \in [-1, 1] \right\} .$$

Next, $q(x, z | u_0, u, a) = \int_{z+a}^u \ell(u_0, w, x) dw$ if $z + a \leq u$ and $q(x, z | u_0, u, a) = 0$ if $z + a > u$. Then, all the functions that compose $\mathcal{F}$ are m times continuously differentiable which implies that m is the smoothness of the class $\mathcal{F}$. Hence, applying Corollary 2.7.2 in Van der Vaart and Wellner (1996), we get that $\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}, L_2(P)) \leq C \tilde{\varepsilon}^{-(d+1)/m}$, where C is some positive constant, $N_{[]}(\tilde{\varepsilon}, \mathcal{F}, L_2(P))$ is the $\tilde{\varepsilon}$ -bracketing number of the class $\mathcal{F}$, P is the probability

measure corresponding to the distribution of (U, X) , and $L_2(P)$ is the L_2 -norm. Consequently,

$$\begin{aligned}
 \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}, L_2(P))} d\tilde{\varepsilon} &= \int_0^{2R} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}, L_2(P))} d\tilde{\varepsilon} \\
 &\leq \int_0^{2R} \sqrt{C\tilde{\varepsilon}^{-(d+1)/m}} d\tilde{\varepsilon} \\
 &= \frac{\sqrt{C}(2R)^{\frac{-d-1}{2m}+1}}{\frac{-d-1}{2m}+1} \\
 &< \infty,
 \end{aligned}$$

provided $m > (d+1)/2$, which is the case since $m > d+2$ by Condition (C6), and where R is a uniform upper bound for f when $f \in \mathcal{F}$. We obtained the first equality since only one $\tilde{\varepsilon}$ -bracket suffices to cover $\mathcal{F}$ if $\tilde{\varepsilon} > 2R$. Moreover, the envelope function of $\mathcal{F}$ possesses a weak second moment because, if $f \in \mathcal{F}$, f is uniformly bounded by a continuous function and the sets $\mathcal{U}_0$ and χ are compact. Hence, the class $\mathcal{F}$ is Donsker by Theorem 2.5.6 in Van der Vaart and Wellner (1996). Using the mean-value theorem, for some positive constant M , we also have

$$\sup_{\substack{u, u_0 \in \mathcal{U}_0 \\ -1 \leq t \leq 1}} V \left[q(X, U|u_0, u, th_u) - q(X, U|u_0, u, 0) \right] \leq Mh_u,$$

which tends to 0 when $h_u \rightarrow 0$. Consequently, it follows from Corollary 2.3.12 in Van der Vaart and Wellner (1996) that

$$\begin{aligned}
 \sup_{\substack{u, u_0 \in \mathcal{U}_0 \\ -1 \leq t \leq 1}} \left| \frac{1}{n} \sum_{i=1}^n \left\{ q(X_i, U_i|u_0, u, th_u) - q(X_i, U_i|u_0, u, 0) \right. \right. \\
 \left. \left. - E[q(X, U|u_0, u, th_u)] + E[q(X, U|u_0, u, 0)] \right\} \right| = o_P(n^{-1/2}).
 \end{aligned}$$

Combining the last result and expression (4.9.1), we obtain

$$\begin{aligned}
 \sup_{u, u_0 \in \mathcal{U}_0} |T(u_0, u)| &= \sup_{u, u_0 \in \mathcal{U}_0} \left| \int_{-\infty}^{+\infty} \left\{ E[q(X, U|u_0, u, th_u)] \right. \right. \\
 &\quad \left. \left. - E[q(X, U|u_0, u, 0)] \right\} K(t) dt \right| + o_P(n^{-1/2}) \\
 &= \sup_{u, u_0 \in \mathcal{U}_0} \left| \int_{-\infty}^{+\infty} E \left\{ \int_0^u \ell(u_0, w, X) [F_{U|X}(w - th_u, X) \right. \right. \\
 &\quad \left. \left. - F_{U|X}(w, X)] dw \right\} K(t) dt \right| + o_P(n^{-1/2}).
 \end{aligned}$$

Finally, using a Taylor expansion of order m , we have

$$F_{U|X}(w - th_u, X) - F_{U|X}(w, X) = \sum_{j=1}^{m-1} \frac{1}{j!} F_{U|X}^{(j)}(w, X) (-th_u)^j + O_P(h_u^m) .$$

Using assumption (C7) and the fact that $\int z^\ell K(z) dz = 0$ for all $\ell = 1, \dots, m-1$ by Condition (C5), we can conclude that

$$\sup_{u, u_0 \in \mathcal{U}_0} |T(u_0, u)| = O_P(h_u^m) + o_P(n^{-1/2}) ,$$

which is also $o_P(n^{-1/2})$ since $nh_u^{2m} \rightarrow 0$ by Condition (C6). $\square$

Proposition 4.9.3. *Assume (C1)-(C9), (J1) and (J2).*

(i) *Then,*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ \mathcal{K}_{h_u}(w - \widehat{U}_i) - \mathcal{K}_{h_u}(w - U_i) \right\} dw \\ &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{F_Y(1) - F_Y(0)} - w \right] \int_{\mathcal{X}} \ell(u_0, w, x) f_{U,X}(w, x) dx dw \\ & \quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\mathcal{X}} \ell(u_0, w, x) f_{U,X}(w, x) dx dw \\ & \quad + o_P(n^{-1/2}) , \end{aligned}$$

(ii) *and*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ K_{h_u}(w - \widehat{U}_i) - K_{h_u}(w - U_i) \right\} dw \\ &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{F_Y(1) - F_Y(0)} - w \right] \int_{\mathcal{X}} \ell(u_0, w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) dx dw \\ & \quad + \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\mathcal{X}} \frac{\partial}{\partial w} \ell(u_0, w, x) f_{U,X}(w, x) dx dw \\ & \quad + o_P(n^{-1/2}) , \end{aligned}$$

uniformly in $u, u_0 \in \mathcal{U}_0$ and where ℓ is some m -times continuously differentiable function with respect to each component and all these derivatives are uniformly bounded.

Proof. We start by proving (i). First, using a Taylor expansion with some ξ_i between U_i and $\widehat{U}_i$, we have

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ \mathcal{K}_{h_u}(w - \widehat{U}_i) - \mathcal{K}_{h_u}(w - U_i) \right\} dw \\
&= \frac{1}{nh_u} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) (\widehat{U}_i - U_i) dw \\
&\quad + \frac{1}{2nh_u^2} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) K'\left(\frac{w - \xi_i}{h_u}\right) (\widehat{U}_i - U_i)^2 dw \\
&= T_1(u_0, u) + T_2(u_0, u) .
\end{aligned}$$

Using Proposition 4.9.1, we have

$$\begin{aligned}
& \widehat{U}_i - U_i \\
&= \frac{\widehat{F}_Y(Y_i) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - \frac{F_Y(Y_i) - F_Y(0)}{F_Y(1) - F_Y(0)} \\
&= \frac{1}{F_Y(1) - F_Y(0)} \left[\widehat{F}_Y(Y_i) - \widehat{F}_Y(0) - F_Y(Y_i) + F_Y(0) \right] \\
&\quad - \frac{F_Y(Y_i) - F_Y(0)}{\{F_Y(1) - F_Y(0)\}^2} \left[\widehat{F}_Y(1) - \widehat{F}_Y(0) - F_Y(1) + F_Y(0) \right] + o_P(n^{-1/2}) .
\end{aligned}$$

Hence, $\widehat{U}_i - U_i = O_P(n^{-1/2})$ uniformly in i and consequently, the term T_2 is $O_P(n^{-1}h_u^{-1})$ uniformly in $u_0, u \in \mathcal{U}_0$ which is also $o_P(n^{-1/2})$ since $nh_u^2 \rightarrow \infty$ by Condition (C6). Moreover,

$$\begin{aligned}
T_1(u_0, u) &= \frac{1}{nh_u\{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \left[\int_0^u \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) dw \right. \\
&\quad \times \left(\widehat{F}_Y(Y_i) - \widehat{F}_Y(0) - F_Y(Y_i) + F_Y(0) \right) \Big] + o_P(n^{-1/2}) \\
&\quad - \frac{1}{nh_u\{F_Y(1) - F_Y(0)\}^2} \sum_{i=1}^n \left[\{F_Y(Y_i) - F_Y(0)\} \right. \\
&\quad \times \int_0^u \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) dw \Big] \\
&\quad \times \left(\widehat{F}_Y(1) - \widehat{F}_Y(0) - F_Y(1) + F_Y(0) \right) + o_P(n^{-1/2}) \\
&= T_{1,1}(u_0, u) + T_{1,2}(u_0, u) + o_P(n^{-1/2}) . \tag{4.9.2}
\end{aligned}$$

Next, for $i = 1, \dots, n$, we have

$$\begin{aligned} & \frac{F_Y(Y_i) - F_Y(0)}{\{F_Y(1) - F_Y(0)\}^2} \left[\widehat{F}_Y(1) - \widehat{F}_Y(0) - F_Y(1) + F_Y(0) \right] \\ &= \frac{\{F_Y(Y_i) - F_Y(0)\} \{\widehat{F}_Y(1) - \widehat{F}_Y(0)\}}{\{F_Y(1) - F_Y(0)\}^2} - \frac{F_Y(Y_i) - F_Y(0)}{F_Y(1) - F_Y(0)} \\ &= U_i \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right), \end{aligned}$$

which implies that

$$\begin{aligned} T_{1,2}(u_0, u) &= -\frac{1}{nh_u} \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ &\quad \times \sum_{i=1}^n U_i \int_0^u \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) dw \\ &= -\frac{1}{n} \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ &\quad \times \sum_{i=1}^n \ell(u_0, U_i, X_i) U_i (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) + O_P(h_u^m). \end{aligned}$$

The last equality was obtained by a Taylor expansion of order m of the function $\ell(u_0, w, X_i)$ at U_i and the fact that $\int K(t) dt = 1$, $\int t^j K(t) dt = 0$ for $j = 1, \dots, m-1$ by Condition (C5). Note also that $O_P(h_u^m)$ is $o_P(n^{-1/2})$ since $nh_u^{2m} \rightarrow 0$ by Condition (C6). Moreover,

$$\begin{aligned} & T_{1,2}(u_0, u) \\ &= -\left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \left\{ \frac{1}{n} \sum_{i=1}^n \ell(u_0, U_i, X_i) U_i (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) \right. \\ &\quad \left. - E \left[\ell(u_0, U, X) U (1_{\{0 \leq U \leq u\}}) \right] \right\} \\ &\quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) E \left[\ell(u_0, U, X) U (1_{\{U \leq u\}} - 1_{\{U \leq 0\}}) \right] + o_P(n^{-1/2}) \\ &= -\left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) E \left[\ell(u_0, U, X) U (1_{\{U \leq u\}} - 1_{\{U \leq 0\}}) \right] + o_P(n^{-1/2}) \\ &= -\left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\mathcal{X}} \ell(u_0, w, x) f_{U,X}(w, x) dx dw + o_P(n^{-1/2}), \end{aligned} \tag{4.9.3}$$

where the second equality was obtained using the fact that $\frac{\hat{F}_Y(1) - \hat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1$ is $O_P(n^{-1/2})$, which implies that the first term on the right hand side of $T_{1,2}$ is $o_P(n^{-1/2})$, by Glivenko-Cantelli argument. Next,

$$\begin{aligned} T_{1,1}(u_0, u) &= \frac{1}{nh_u\{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) dw \\ &\quad \times \left[\frac{1}{n} \sum_{j=1}^n 1_{\{Y_j \leq Y_i\}} - 1_{\{Y_j \leq 0\}} - F_Y(Y_i) + F_Y(0) \right] \\ &= \frac{1}{n^2 h_u\{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \sum_{j=1}^n \int_0^u q(V_i, V_j, u_0, w, h_u) dw, \end{aligned}$$

where $V_i = (X_i, U_i, Y_i)^t$ and

$$\begin{aligned} &q(V_i, V_j, u_0, w, h_u) \\ &= \ell(u_0, w, X_i) K\left(\frac{w - U_i}{h_u}\right) \left[1_{\{Y_j \leq Y_i\}} - 1_{\{Y_j \leq 0\}} - F_Y(Y_i) + F_Y(0) \right], \end{aligned}$$

for $i, j = 1, \dots, n$. Hence,

$$\begin{aligned} T_{1,1}(u_0, u) &= \frac{1}{n^2 h_u\{F_Y(1) - F_Y(0)\}} \left[\sum_{i=1}^n \sum_{j=1, j \neq i}^n \int_0^u q(V_i, V_j, u_0, w, h_u) dw \right. \\ &\quad \left. + \sum_{i=1}^n \int_0^u q(V_i, V_i, u_0, w, h_u) dw \right]. \end{aligned}$$

The second term on the right hand side above is $O_P(n^{-1}h_u^{-1})$ which is also $o_P(n^{-1/2})$ since $nh_u^2 \rightarrow \infty$ by Condition (C6). Moreover,

$$\begin{aligned} &T_{1,1}(u_0, u) \\ &= \frac{1}{n(n-1)h_u\{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \int_0^u q(V_i, V_j, u_0, w, h_u) dw \\ &\quad - \frac{1}{n^2(n-1)h_u\{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \int_0^u q(V_i, V_j, u_0, w, h_u) dw \\ &\quad + o_P(n^{-1/2}). \end{aligned}$$

The second term on the right hand side is $o_P(n^{-1/2})$ again for the same reason. Finally, we get

$$T_{1,1}(u_0, u) = \frac{1}{h_u\{F_Y(1) - F_Y(0)\}} \tilde{T}_{1,1}(u_0, u) + o_P(n^{-1/2}), \quad (4.9.4)$$

where

$$\tilde{T}_{1,1}(u_0, u) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \int_0^u q(V_i, V_j, u_0, w, h_u) dw .$$

Note that $\tilde{T}_{1,1}$ is a U -statistic of order 2 and we can thus apply equation (6) in Sherman (1994) which gives

$$\begin{aligned} & \tilde{T}_{1,1}(u_0, u) \\ &= \int_0^u \left[\frac{1}{n} \sum_{i=1}^n E \{q(V, V_i, u_0, w, h_u)\} + \frac{1}{n} \sum_{i=1}^n E \{q(V_i, V', u_0, w, h_u)\} \right. \\ & \quad \left. - E \{q(V, V', u_0, w, h_u)\} + \tilde{R}_n(u_0, w, h_u) \right] dw , \end{aligned} \quad (4.9.5)$$

where V, V' are iid, have the same distribution as V_1 and $\tilde{R}_n(u_0, w, h_u)$ is a degenerate U -statistic of order 2. We would like to apply Corollary 4(ii) in Sherman (1994) to decompose $\tilde{R}_n(u_0, w, h_u)$. To apply this Corollary, we have to prove that the class $\mathcal{F}$, defined by

$$\begin{aligned} & \left\{ (v, v') \rightarrow q(v, v', u_0, w, h) - E \{q(V, v', u_0, w, h)\} - E \{q(v, V', u_0, w, h)\} \right. \\ & \quad \left. + E \{q(V, V', u_0, w, h)\} : u_0, w \in \mathcal{U}_0, 0 < h < 1 \right\} , \end{aligned}$$

is Euclidean. We refer to Sherman (1994) for a precise definition of an Euclidean class. First, note that the class

$$\left\{ (v, v') \rightarrow q(v, v', u_0, w, h) : u_0, w \in \mathcal{U}_0, 0 < h < 1 \right\} ,$$

is Euclidean using Lemma 2.14 in Pakes and Pollard (1989) and the facts that the classes

$$\left\{ u' \rightarrow K \left(\frac{u - u'}{h} \right) : u \in \mathcal{U}_0, 0 < h < 1 \right\} ,$$

and

$$\left\{ (y, y', x) \rightarrow \ell(u_0, w, x) [1_{\{y' \leq y\}} - 1_{\{y' \leq 0\}} - F_Y(y) + F_Y(0)] : u_0, w \in \mathcal{U}_0 \right\} ,$$

are Euclidean by Examples 2.10 in Pakes and Pollard (1989), Condition (C5) and the assumptions imposed on the function ℓ . Then, we deduce by Lemma 6 in Sherman (1994) that the class $\mathcal{F}$ is Euclidean. Consequently, Corollary 4(ii) in Sherman (1994) implies that $\sup_{u_0, w, h} |\tilde{R}_n(u_0, w, h)| = O_P(n^{-1})$ and

$$\frac{1}{h_u \{F_Y(1) - F_Y(0)\}} \sup_{u_0, u, h} \int_0^u |\tilde{R}_n(u_0, w, h)| dw = O_P(n^{-1} h_u^{-1}) = o_P(n^{-1/2}) .$$

Hence, using (4.9.4) and (4.9.5), expression $T_{1,1}$ becomes

$$\begin{aligned}
& T_{1,1}(u_0, u) \\
&= \frac{1}{h_u \{F_Y(1) - F_Y(0)\}} \int_0^u \left[\frac{1}{n} \sum_{i=1}^n E \{q(V, V_i, u_0, w, h_u)\} \right. \\
&\quad \left. + \frac{1}{n} \sum_{i=1}^n E \{q(V_i, V', u_0, w, h_u)\} - E \{q(V, V', u_0, w, h_u)\} \right] dw \\
&\quad + o_P(n^{-1/2}) . \tag{4.9.6}
\end{aligned}$$

Moreover,

$$\begin{aligned}
& E \{q(v, V', u_0, w, h_u)\} \\
&= E \left[\ell(u_0, w, x) K \left(\frac{w - u}{h_u} \right) \left\{ 1_{\{Y' \leq y\}} - 1_{\{Y' \leq 0\}} - F_Y(y) + F_Y(0) \right\} \right] \\
&= \ell(u_0, w, x) K \left(\frac{w - u}{h_u} \right) \left[P(Y' \leq y) - P(Y' \leq 0) - F_Y(y) + F_Y(0) \right] \\
&= 0 ,
\end{aligned}$$

and

$$\begin{aligned}
& E \{q(V, V', u_0, w, h_u)\} \\
&= E \left[\ell(u_0, w, X) K \left(\frac{w - U}{h_u} \right) \left\{ 1_{\{Y' \leq Y\}} - 1_{\{Y' \leq 0\}} - F_Y(Y) + F_Y(0) \right\} \right] \\
&= E \left[\ell(u_0, w, X) K \left(\frac{w - U}{h_u} \right) \right. \\
&\quad \left. \times E \left\{ 1_{\{Y' \leq Y\}} - 1_{\{Y' \leq 0\}} - F_Y(Y) + F_Y(0) \middle| X, U \right\} \right] \\
&= 0 ,
\end{aligned}$$

for the same reason. We also have

$$\begin{aligned}
& E \{q(V, v', u_0, w, h_u)\} \\
&= E \left[\ell(u_0, w, X) K \left(\frac{w - U}{h_u} \right) \left\{ 1_{\{y' \leq Y\}} - 1_{\{y' \leq 0\}} - F_Y(Y) + F_Y(0) \right\} \right] \\
&= E \left[\ell(u_0, w, X) K \left(\frac{w - U}{h_u} \right) \left\{ 1_{\{u' \leq U\}} - 1_{\{u' \leq 0\}} - U(F_Y(1) - F_Y(0)) \right\} \right] \\
&= \int_{\mathcal{X}} \int_{\mathcal{U}} \left[\ell(u_0, w, x) K \left(\frac{w - u}{h_u} \right) \right. \\
&\quad \left. \times \left\{ 1_{\{u' \leq u\}} - 1_{\{u' \leq 0\}} - u(F_Y(1) - F_Y(0)) \right\} f_{U,X}(u, x) du dx \right] .
\end{aligned}$$

Applying the substitution $t = \frac{w-u}{h_u}$, the last expression becomes

$$h_u \int_{\mathcal{X}} \int \ell(u_0, w, x) K(t) \left\{ 1_{\{u' \leq w - th_u\}} - 1_{\{u' \leq 0\}} \right. \\ \left. - (w - th_u)(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w - th_u, x) dt dx .$$

Inserting this last expression in (4.9.6), we obtain

$$\begin{aligned} & T_{1,1}(u_0, u) \\ &= \frac{1}{F_Y(1) - F_Y(0)} \frac{1}{n} \sum_{i=1}^n \int_0^u \int_{\mathcal{X}} \int \ell(u_0, w, x) K(t) \left\{ 1_{\{U_i \leq w - th_u\}} - 1_{\{U_i \leq 0\}} \right. \\ & \quad \left. - (w - th_u)(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w - th_u, x) dt dx dw + o_P(n^{-1/2}) \\ &= \frac{1}{F_Y(1) - F_Y(0)} \int_0^u \int_{\mathcal{X}} \int \ell(u_0, w, x) K(t) \left\{ \widehat{F}_U(w - th_u) - \widehat{F}_U(0) \right. \\ & \quad \left. - (w - th_u)(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w - th_u, x) dt dx dw + o_P(n^{-1/2}) . \end{aligned} \tag{4.9.7}$$

Next, recall that $U \sim \text{Un} \left[\frac{-F_Y(0)}{F_Y(1) - F_Y(0)}, \frac{1 - F_Y(0)}{F_Y(1) - F_Y(0)} \right]$ and using Theorem 2.14 in Stute (1982), we have

$$\begin{aligned} \widehat{F}_U(w - th_u) - \widehat{F}_U(w) &= F_U(w - th_u) - F_U(w) + O_P(n^{-1/2} h_u^{1/2}) \\ &= (w - th_u) \{F_Y(1) - F_Y(0)\} + F_Y(0) \\ & \quad - w \{F_Y(1) - F_Y(0)\} - F_Y(0) + O_P(n^{-1/2} h_u^{1/2}) \\ &= -th_u \{F_Y(1) - F_Y(0)\} + o_P(n^{-1/2}) . \end{aligned}$$

Hence,

$$\begin{aligned} & \widehat{F}_U(w - th_u) - \widehat{F}_U(0) - (w - th_u)(F_Y(1) - F_Y(0)) \\ &= \widehat{F}_U(w) - th_u \{F_Y(1) - F_Y(0)\} - \widehat{F}_U(0) - (w - th_u)(F_Y(1) - F_Y(0)) \\ &= \widehat{F}_U(w) - \widehat{F}_U(0) - w(F_Y(1) - F_Y(0)) . \end{aligned}$$

Inserting this last expression in (4.9.7), the term $T_{1,1}$ becomes

$$\begin{aligned}
T_{1,1}(u_0, u) &= \frac{1}{F_Y(1) - F_Y(0)} \int_0^u \int_{\mathcal{X}} \int \ell(u_0, w, x) K(t) \left\{ \widehat{F}_U(w) - \widehat{F}_U(0) \right. \\
&\quad \left. - w(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w - th_u, x) dt dx dw + o_P(n^{-1/2}) \\
&= \frac{1}{F_Y(1) - F_Y(0)} \int_{\mathcal{X}} \int_0^u \ell(u_0, w, x) \left\{ \widehat{F}_U(w) - \widehat{F}_U(0) \right. \\
&\quad \left. - w(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \quad (4.9.8)
\end{aligned}$$

We obtained the last equality by a Taylor expansion and using the properties of the kernel K introduced in (C5). It suffices to insert (4.9.3) and (4.9.8) in (4.9.2) to conclude the proof of Part (i).

The proof of Part (ii) follows exactly the same way as the proof of Part (i). First, using a Taylor expansion with some ξ_i between U_i and $\widehat{U}_i$, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) \left\{ K_{h_u}(w - \widehat{U}_i) - K_{h_u}(w - U_i) \right\} dw \\
&= \frac{1}{nh_u^2} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) K' \left(\frac{w - U_i}{h_u} \right) (\widehat{U}_i - U_i) dw \\
&\quad + \frac{1}{2nh_u^3} \sum_{i=1}^n \int_0^u \ell(u_0, w, X_i) K'' \left(\frac{w - \xi_i}{h_u} \right) (\widehat{U}_i - U_i)^2 dw \\
&= T_3(u_0, u) + T_4(u_0, u) .
\end{aligned}$$

The expression T_4 is $O_P(n^{-1}h_u^{-2})$ which is also $o_P(n^{-1/2})$ since $nh_u^4 \rightarrow \infty$ by condition (C6). Moreover,

$$\begin{aligned}
&T_3(u_0, u) \\
&= \frac{1}{nh_u^2 \{F_Y(1) - F_Y(0)\}} \sum_{i=1}^n \left[\int_0^u \ell(u_0, w, X_i) K' \left(\frac{w - U_i}{h_u} \right) dw \right. \\
&\quad \left. \times \left(\widehat{F}_Y(Y_i) - \widehat{F}_Y(0) - F_Y(Y_i) + F_Y(0) \right) \right] + o_P(n^{-1/2}) \\
&\quad - \frac{1}{nh_u^2 \{F_Y(1) - F_Y(0)\}^2} \\
&\quad \times \sum_{i=1}^n \{F_Y(Y_i) - F_Y(0)\} \int_0^u \ell(u_0, w, X_i) K' \left(\frac{w - U_i}{h_u} \right) dw \\
&\quad \times \left(\widehat{F}_Y(1) - \widehat{F}_Y(0) - F_Y(1) + F_Y(0) \right) + o_P(n^{-1/2}) \\
&= T_{3,1}(u_0, u) + T_{3,2}(u_0, u) + o_P(n^{-1/2}) . \quad (4.9.9)
\end{aligned}$$

Next, exactly as in (4.9.3),

$$\begin{aligned} T_{3,2}(u_0, u) &= -\frac{1}{nh_u^2} \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \sum_{i=1}^n U_i \int_0^u \ell(u_0, w, X_i) K' \left(\frac{w - U_i}{h_u} \right) dw, \end{aligned}$$

and using integration by parts, we have

$$\begin{aligned} & \int_0^u \ell(u_0, w, X_i) K' \left(\frac{w - U_i}{h_u} \right) dw \\ &= \left[h_u \ell(u_0, w, X_i) (1_{\{w \leq u\}} - 1_{\{w \leq 0\}}) K \left(\frac{w - U_i}{h_u} \right) \right]_{\mathcal{U}} \\ & \quad - h_u \int_0^u \frac{\partial}{\partial w} \ell(u_0, w, X_i) K \left(\frac{w - U_i}{h_u} \right) dw \\ &= -h_u \int_0^u \frac{\partial}{\partial w} \ell(u_0, w, X_i) K \left(\frac{w - U_i}{h_u} \right) dw \\ &= -h_u^2 \frac{\partial}{\partial w} \ell(u_0, w, X_i) \Big|_{w=U_i} (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) + O_P(h_u^m). \end{aligned}$$

The second equality was obtained since we assumed that $0 \in \mathcal{U}_0$, $u \in \mathcal{U}_0$ which implies that $1_{\{w \leq u\}} = 1_{\{w \leq 0\}}$ for values of w on the boundary of $\mathcal{U}$ and the last equality was obtained by a Taylor expansion. Hence,

$$\begin{aligned} & T_{3,2}(u_0, u) \\ &= \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ & \quad \times \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial w} \ell(u_0, w, X_i) \Big|_{w=U_i} U_i (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) + o_P(n^{-1/2}) \\ &= \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ & \quad \times E \left[\frac{\partial}{\partial w} \ell(u_0, w, X) \Big|_{w=U} U (1_{\{U \leq u\}} - 1_{\{U \leq 0\}}) \right] + o_P(n^{-1/2}) \\ &= \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ & \quad \times \int_0^u w \int_{\mathcal{X}} \frac{\partial}{\partial w} \ell(u_0, w, x) f_{U,X}(w, x) dx dw + o_P(n^{-1/2}). \quad (4.9.10) \end{aligned}$$

Moreover, doing exactly the same calculations as for $T_{1,1}$, expression $T_{3,1}$ is

equal to

$$\begin{aligned} T_{3,1}(u_0, u) &= \frac{1}{h_u(F_Y(1) - F_Y(0))} \int_0^u \int_{\mathcal{X}} \int \ell(u_0, w, x) K'(t) \left\{ \widehat{F}_U(w) - \widehat{F}_U(0) \right. \\ &\quad \left. - w(F_Y(1) - F_Y(0)) \right\} f_{U,X}(w - th_u, x) dt dx dw + o_P(n^{-1/2}) . \end{aligned}$$

Finally, in a same way as above, it suffices to apply integration by parts:

$$\begin{aligned} T_{3,1}(u_0, u) &= \frac{1}{F_Y(1) - F_Y(0)} \int_{\mathcal{X}} \int_0^u \ell(u_0, w, x) \left\{ \widehat{F}_U(w) - \widehat{F}_U(0) \right. \\ &\quad \left. - w(F_Y(1) - F_Y(0)) \right\} \frac{\partial}{\partial w} f_{U,X}(w, x) dw dx + o_P(n^{-1/2}) . \end{aligned} \quad (4.9.11)$$

It suffices to insert (4.9.10) and (4.9.11) in (4.9.9) to conclude the proof of Part (ii). $\square$

The four following propositions are adaptations of Lemmas 1, 2 and 3 in Chiappori et al. (2015) taking into account that we are now working with $U = T(Y)$ instead of Y . The proofs of these four results will follow the same way as the proofs given in Chiappori et al. (2015). However, we have noticed some mistakes in the proofs of these three Lemmas in the latter paper and we will correct them here.

Proposition 4.9.4. *Assume (C1)-(C9), (J1) and (J2).*

- (i) *Then, the expressions $\|\widehat{p} - p\|_{\infty}$, $\|\widehat{p}_1 - p_1\|_{\infty}$, $\|\widehat{p}_u - p_u\|_{\infty}$, $\|\widehat{f}_X - f_X\|_{\infty}$ and $\|\widehat{f}_{X,1} - f_{X,1}\|_{\infty}$ are $o_P(n^{-1/4})$ uniformly in $\mathcal{U}_0 \times \chi_0$, where $\|\cdot\|_{\infty}$ denotes the supremum norm here and in the following, i.e., $\|\widehat{p} - p\|_{\infty} = \sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} |\widehat{p}(u, x) - p(u, x)|$ for example.*

- (ii) *Moreover,*

$$\|\nabla_p S[\widehat{p} - p]\|_{\infty}^2 = o_P(n^{-1/2}) \quad \text{and} \quad \|\nabla_f S[\widehat{f}_X - f_X]\|_{\infty}^2 = o_P(n^{-1/2}) ,$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, where the functionals $\nabla_p S(u, x)[\cdot]$ and $\nabla_f S(u, x)[\cdot]$ are defined in (4.5.1) and (4.5.2) respectively.

Proof. We start by proving statement (i). First,

$$\begin{aligned}
& |\widehat{p}(u, x) - p(u, x)| \\
&= \left| \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(u - \widehat{U}_i) \mathbf{K}_{h_x}(X_i - x) - p(u, x) \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^n \left\{ \mathcal{K}_{h_u}(u - \widehat{U}_i) - \mathcal{K}_{h_u}(u - U_i) \right\} \mathbf{K}_{h_x}(X_i - x) \right| \\
&\quad + \left| \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(u - U_i) \mathbf{K}_{h_x}(X_i - x) - p(u, x) \right|. \tag{4.9.12}
\end{aligned}$$

Using the mean-value theorem with some ξ_i between U_i and $\widehat{U}_i$, the first term on the right hand side of (4.9.12) is equal to

$$\begin{aligned}
& \left| \frac{1}{nh_u h_x^d} \sum_{i=1}^n K\left(\frac{u - \xi_i}{h_u}\right) \mathbf{K}\left(\frac{X_i - x}{h_x}\right) (\widehat{U}_i - U_i) \right| \\
&\leq \max_{1 \leq i \leq n} |\widehat{U}_i - U_i| \times \left| \frac{1}{nh_u h_x^d} \sum_{i=1}^n K\left(\frac{u - \xi_i}{h_u}\right) \mathbf{K}\left(\frac{X_i - x}{h_x}\right) \right|.
\end{aligned}$$

This last expression is $O_P(n^{-1/2}) \times O_P(1)$. Indeed, using Proposition 4.9.1, we can easily check that $\widehat{U}_i - U_i$ is $O_P(n^{-1/2})$ uniformly in i . We can conclude that the first term on the right hand side of (4.9.12) is $o_P(n^{-1/4})$. Moreover, using Theorem 6 in Hansen (2008) in the same way as in the proof of Lemma 1 in Chiappori et al. (2015), the second term on the right hand side of (4.9.12) is $O_P(\max(h_x, h_u)^m) + O_P(\sqrt{\log n / (nh_x^d)})$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, which is also $o_P(n^{-1/4})$ since $n^{1/2}h_x^m \rightarrow 0$, $n^{1/2}h_u^m \rightarrow 0$ and $n^{1/2}h_x^d / \log n \rightarrow \infty$ by condition (C6). Consequently, $\|\widehat{p} - p\|_\infty = o_P(n^{-1/4})$.

Second,

$$\begin{aligned}
& |\widehat{p}_u(u, x) - p_u(u, x)| \\
&= \left| \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial u} \mathcal{K}_{h_u}(u - \widehat{U}_i) \mathbf{K}_{h_x}(X_i - x) - p_u(u, x) \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^n \left\{ K_{h_u}(u - \widehat{U}_i) - K_{h_u}(u - U_i) \right\} \mathbf{K}_{h_x}(X_i - x) \right| \\
&\quad + \left| \frac{1}{n} \sum_{i=1}^n K_{h_u}(u - U_i) \mathbf{K}_{h_x}(X_i - x) - p_u(u, x) \right|. \tag{4.9.13}
\end{aligned}$$

Using the mean-value theorem with some ξ_i between U_i and $\widehat{U}_i$, the first term

on the right hand side of (4.9.13) is equal to

$$\begin{aligned} & \left| \frac{1}{nh_u^2 h_x^d} \sum_{i=1}^n K' \left(\frac{u - \xi_i}{h_u} \right) \mathbf{K} \left(\frac{X_i - x}{h_x} \right) (\hat{U}_i - U_i) \right| \\ & \leq \max_{1 \leq i \leq n} |\hat{U}_i - U_i| \times \left| \frac{1}{nh_u^2 h_x^d} \sum_{i=1}^n K' \left(\frac{u - \xi_i}{h_u} \right) \mathbf{K} \left(\frac{X_i - x}{h_x} \right) \right|. \end{aligned}$$

This last expression is $O_P(n^{-1/2}) \times O_P(h_u^{-1})$ and then also $o_P(n^{-1/4})$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$ since $nh_u^4 \rightarrow \infty$ by condition (C6). Moreover, using Theorem 6 in Hansen (2008), the second term on the right hand side of (4.9.13) is $O_P(\max(h_x, h_u)^m) + O_P(\sqrt{\log n / (nh_u^2 h_x^d)})$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, which is also $o_P(n^{-1/4})$ since $n^{1/2} h_x^m \rightarrow 0$, $n^{1/2} h_u^m \rightarrow 0$ and $n^{1/2} h_u^2 h_x^d / \log n \rightarrow \infty$ by Condition (C6). Consequently, $\|\hat{p}_u - p_u\|_\infty = o_P(n^{-1/4})$.

Third,

$$\begin{aligned} & |\hat{p}_1(u, x) - p_1(u, x)| \\ & = \left| \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(u - \hat{U}_i) \frac{\partial}{\partial x_1} \mathbf{K}_{h_x}(X_i - x) - p_1(u, x) \right| \\ & \leq \left| \frac{1}{n} \sum_{i=1}^n \left\{ \mathcal{K}_{h_u}(u - \hat{U}_i) - \mathcal{K}_{h_u}(u - U_i) \right\} \mathbf{K}_{h_x,1}(X_i - x) \right| \\ & \quad + \left| \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(u - U_i) \mathbf{K}_{h_x,1}(X_i - x) - p_1(u, x) \right|, \quad (4.9.14) \end{aligned}$$

where $\mathbf{K}_{h_x,1}(X_i - x) = \frac{\partial}{\partial x_1} \mathbf{K}_{h_x}(X_i - x)$. Using Theorem 6 in Hansen (2008), the second term on the right hand side of (4.9.14) is $O_P(\max(h_x, h_u)^m) + O_P(\sqrt{\log n / (nh_x^{d+2})})$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, which is also $o_P(n^{-1/4})$ since $n^{1/2} h_x^m \rightarrow 0$, $n^{1/2} h_u^m \rightarrow 0$ and $n^{1/2} h_x^{d+2} / \log n \rightarrow \infty$ by Condition (C6). Using similar arguments as before, the first term on the right hand side of (4.9.14) is $O_P(n^{-1/2} h_x^{-1})$ which is also $o_P(n^{-1/4})$ since $nh_x^4 \rightarrow \infty$ by Condition (C6).

Finally, the proofs for $\|\hat{f}_X - f_X\|_\infty$ and $\|\hat{f}_{X,1} - f_{X,1}\|_\infty$ are exactly the same as in Chiappori et al. (2015). It suffices to apply Theorem 6 in Hansen (2008), which gives

$$\|\hat{f}_X - f_X\|_\infty = O_P(h_x^m) + O_P \left(\sqrt{\frac{\log n}{nh_x^d}} \right),$$

and

$$\|\hat{f}_{X,1} - f_{X,1}\|_\infty = O_P(h_x^m) + O_P \left(\sqrt{\frac{\log n}{nh_x^{d+2}}} \right).$$

These expressions are also $o_P(n^{-1/4})$ using Condition (C6). This concludes the proof of Part (i).

We will now prove Part (ii). The proof is similar to the proof of Lemma 3 in Chiappori et al. (2015). First, using the definition of $\nabla_p S(u, x)[\cdot]$ in (4.5.1), we have

$$\|\nabla_p S[\hat{p} - p]\|_\infty \leq \|D_{p,0}\|_\infty \|\hat{p} - p\|_\infty + \|D_{p,u}\|_\infty \|\hat{p}_u - p_u\|_\infty + \|D_{p,1}\|_\infty \|\hat{p}_1 - p_1\|_\infty.$$

Next, the expressions $\|D_{p,0}\|_\infty$, $\|D_{p,u}\|_\infty$ and $\|D_{p,1}\|_\infty$ are finite as a consequence of Condition (C7). Then, it suffices to use the facts that $\|\hat{p} - p\|_\infty$, $\|\hat{p}_u - p_u\|_\infty$ and $\|\hat{p}_1 - p_1\|_\infty$ are $o_P(n^{-1/4})$ given Part (i), which implies that $\|\nabla_p S[\hat{p} - p]\|_\infty^2$ is $o_P(n^{-1/2})$. Similarly, using the definition of $\nabla_f S(u, x)[\cdot]$ in (4.5.2), we have

$$\|\nabla_f S[\hat{f}_X - f_X]\|_\infty \leq \|D_{f,0}\|_\infty \|\hat{f}_X - f_X\|_\infty + \|D_{f,1}\|_\infty \|\hat{f}_{X,1} - f_{X,1}\|_\infty.$$

Condition (C7) implies that $\|D_{f,0}\|_\infty$ and $\|D_{f,1}\|_\infty$ are finite. Consequently, we simply apply Part (i) to conclude that $\|\nabla_f S[\hat{f}_X - f_X]\|_\infty^2$ is $o_P(n^{-1/2})$. This concludes the proof of Part (ii). $\square$

Proposition 4.9.5. *Assume (C1)-(C9), (J1) and (J2). Then,*

$$\hat{S}_1(u, x) - S_1(u, x) = \nabla_p S_1(u, x)[\hat{p} - p] + \nabla_f S_1(u, x)[\hat{f}_X - f_X] + o_P(n^{-1/2}),$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, where $\nabla_p S_1(u, x)[\cdot]$ and $\nabla_f S_1(u, x)[\cdot]$ are defined in (4.5.1) and (4.5.2) respectively.

Proof. Throughout this proof, we will omit the dependences of the different functions on u and x for an easier readability, e.g., we will write φ_u instead of $\varphi_u(u, x)$ and $f_{X,1}$ instead of $f_{X,1}(x)$. First, using Proposition 4.9.1, we have

$$\begin{aligned} \frac{\hat{\varphi}_u}{\hat{\varphi}_1} - \frac{\varphi_u}{\varphi_1} &= \frac{1}{\varphi_1} \left(\hat{\varphi}_u - \varphi_u \right) - \frac{\varphi_u}{\varphi_1^2} \left(\hat{\varphi}_1 - \varphi_1 \right) \\ &\quad - \frac{(\hat{\varphi}_1 - \varphi_1)}{\hat{\varphi}_1 \varphi_1} \left\{ (\hat{\varphi}_u - \varphi_u) - \frac{\varphi_u(\hat{\varphi}_1 - \varphi_1)}{\varphi_1} \right\}. \end{aligned} \quad (4.9.15)$$

Using a second time Proposition 4.9.1, we have

$$\hat{\varphi}_u - \varphi_u = \frac{\hat{p}_u}{\hat{f}_X} - \frac{p_u}{f_X} = \frac{1}{f_X} (\hat{p}_u - p_u) - \frac{p_u}{f_X^2} (\hat{f}_X - f_X) + R_1, \quad (4.9.16)$$

where

$$R_1 = -\frac{(\hat{f}_X - f_X)}{\hat{f}_X f_X} \left\{ \hat{p}_u - p_u - \frac{p_u}{f_X} (\hat{f}_X - f_X) \right\}.$$

Moreover, the expression R_1 is $O(\|\hat{f}_X - f_X\|_\infty^2) + O(\|\hat{p}_u - p_u\|_\infty^2)$ because $\frac{-1}{\hat{f}_X f_X}(\hat{f}_X - f_X)(\hat{p}_u - p_u) \leq \frac{1}{2\hat{f}_X f_X}\{(\hat{f}_X - f_X)^2 + (\hat{p}_u - p_u)^2\}$ using Cauchy-Schwarz inequality and $\sup_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} |p_u(u, x)| < \infty$ and $\inf_{x \in \mathcal{X}_0} f_X(x) > 0$ using Conditions (C7) and (C8). Note also that there exists $\zeta > 0$ such that $P(\inf_{x \in \mathcal{X}_0} |\hat{f}_X(x)| > \zeta) \rightarrow 1$ when $n \rightarrow \infty$ since $\|\hat{f}_X - f_X\|_\infty = o_P(1)$ and $\inf_{x \in \mathcal{X}_0} f_X(x) > 0$. Next,

$$\begin{aligned} \hat{\varphi}_1 - \varphi_1 &= \frac{\hat{p}_1}{\hat{f}_X} - \frac{p_1}{f_X} - \left(\frac{\hat{p}\hat{f}_{X,1}}{\hat{f}_X^2} - \frac{pf_{X,1}}{f_X^2} \right) \\ &= \frac{\hat{p}_1}{\hat{f}_X} - \frac{p_1}{f_X} - \left\{ \frac{\hat{p}}{\hat{f}_X} \left(\frac{\hat{f}_{X,1}}{\hat{f}_X} - \frac{f_{X,1}}{f_X} \right) + \frac{f_{X,1}}{f_X} \left(\frac{\hat{p}}{\hat{f}_X} - \frac{p}{f_X} \right) \right\} \\ &= \frac{\hat{p}_1}{\hat{f}_X} - \frac{p_1}{f_X} - \left\{ \frac{p}{f_X} \left(\frac{\hat{f}_{X,1}}{\hat{f}_X} - \frac{f_{X,1}}{f_X} \right) + \frac{f_{X,1}}{f_X} \left(\frac{\hat{p}}{\hat{f}_X} - \frac{p}{f_X} \right) \right. \\ &\quad \left. + \left(\frac{\hat{p}}{\hat{f}_X} - \frac{p}{f_X} \right) \left(\frac{\hat{f}_{X,1}}{\hat{f}_X} - \frac{f_{X,1}}{f_X} \right) \right\}. \end{aligned} \quad (4.9.17)$$

Next, we apply three more times Proposition 4.9.1 and we obtain

$$\begin{aligned} \frac{\hat{p}_1}{\hat{f}_X} - \frac{p_1}{f_X} &= \frac{1}{f_X} (\hat{p}_1 - p_1) - \frac{p_1}{f_X^2} (\hat{f}_X - f_X) \\ &\quad - \frac{(\hat{f}_X - f_X)}{\hat{f}_X f_X} \left\{ \hat{p}_1 - p_1 - \frac{p_1(\hat{f}_X - f_X)}{f_X} \right\}, \end{aligned} \quad (4.9.18)$$

$$\begin{aligned} \frac{p}{f_X} \left(\frac{\hat{f}_{X,1}}{\hat{f}_X} - \frac{f_{X,1}}{f_X} \right) &= \frac{p}{f_X} \left[\frac{1}{f_X} (\hat{f}_{X,1} - f_{X,1}) - \frac{f_{X,1}}{f_X^2} (\hat{f}_X - f_X) \right. \\ &\quad \left. - \frac{(\hat{f}_X - f_X)}{\hat{f}_X f_X} \left\{ \hat{f}_{X,1} - f_{X,1} - \frac{f_{X,1}(\hat{f}_X - f_X)}{f_X} \right\} \right], \end{aligned} \quad (4.9.19)$$

and

$$\begin{aligned} \frac{f_{X,1}}{f_X} \left(\frac{\hat{p}}{\hat{f}_X} - \frac{p}{f_X} \right) &= \frac{f_{X,1}}{f_X} \left[\frac{1}{f_X} (\hat{p} - p) - \frac{p}{f_X^2} (\hat{f}_X - f_X) \right. \\ &\quad \left. - \frac{(\hat{f}_X - f_X)}{\hat{f}_X f_X} \left\{ \hat{p} - p - \frac{p(\hat{f}_X - f_X)}{f_X} \right\} \right]. \end{aligned} \quad (4.9.20)$$

Combining (4.9.17) with (4.9.18), (4.9.19) and (4.9.20), we obtain

$$\begin{aligned}\widehat{\varphi}_1 - \varphi_1 &= -\frac{f_{X,1}}{f_X^2}(\widehat{p} - p) + \frac{1}{f_X}(\widehat{p}_1 - p_1) + \left(\frac{2pf_{X,1}}{f_X^3} - \frac{p_1}{f_X^2}\right)(\widehat{f}_X - f_X) \\ &\quad - \frac{p}{f_X^2}(\widehat{f}_{X,1} - f_{X,1}) + R_2, \end{aligned} \quad (4.9.21)$$

where

$$\begin{aligned}R_2 &= -\frac{(\widehat{f}_X - f_X)}{\widehat{f}_X f_X} \left\{ \widehat{p}_1 - p_1 - \frac{p_1(\widehat{f}_X - f_X)}{f_X} \right\} \\ &\quad - \frac{p(\widehat{f}_X - f_X)}{\widehat{f}_X f_X^2} \left\{ \widehat{f}_{X,1} - f_{X,1} - \frac{f_{X,1}(\widehat{f}_X - f_X)}{f_X} \right\} \\ &\quad + \frac{f_{X,1}(\widehat{f}_X - f_X)}{\widehat{f}_X f_X^2} \left\{ \widehat{p} - p - \frac{p(\widehat{f}_X - f_X)}{f_X} \right\} \\ &\quad - \left(\frac{\widehat{p}}{\widehat{f}_X} - \frac{p}{f_X} \right) \left(\frac{\widehat{f}_{X,1}}{\widehat{f}_X} - \frac{f_{X,1}}{f_X} \right). \end{aligned}$$

Moreover, the last term in R_2 is equal to

$$\begin{aligned}&\left(\frac{\widehat{p}}{\widehat{f}_X} - \frac{p}{f_X} \right) \left(\frac{\widehat{f}_{X,1}}{\widehat{f}_X} - \frac{f_{X,1}}{f_X} \right) \\ &= \left(\frac{\widehat{p}f_X - pf_X}{\widehat{f}_X f_X} \right) \left(\frac{\widehat{f}_{X,1}f_X - f_{X,1}\widehat{f}_X}{\widehat{f}_X f_X} \right) \\ &= \left(\frac{f_X(\widehat{p} - p) - p(\widehat{f}_X - f_X)}{\widehat{f}_X f_X} \right) \left(\frac{f_X(\widehat{f}_{X,1} - f_{X,1}) - f_{X,1}(\widehat{f}_X - f_X)}{\widehat{f}_X f_X} \right) \\ &= \frac{1}{\widehat{f}_X^2}(\widehat{p} - p)(\widehat{f}_{X,1} - f_{X,1}) + \frac{pf_{X,1}}{\widehat{f}_X^2 f_X^2}(\widehat{f}_X - f_X)^2 \\ &\quad - \frac{p}{\widehat{f}_X^2 f_X}(\widehat{f}_X - f_X)(\widehat{f}_{X,1} - f_{X,1}) - \frac{f_{X,1}}{\widehat{f}_X^2 f_X}(\widehat{p} - p)(\widehat{f}_X - f_X). \end{aligned}$$

Using the Cauchy-Schwarz inequality, i.e., $\pm ab \leq (a^2 + b^2)/2$ for all real a and b , the facts that $\inf_{x \in \chi_0} f_X(x) > 0$, there exists $\zeta > 0$ such that $P(\inf_{x \in \chi_0} |\widehat{f}_X(x)| > \zeta) \rightarrow 1$ when $n \rightarrow \infty$, $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} |p_1(u,x)| < \infty$, $\sup_{x \in \chi_0} |f_{X,1}(x)| < \infty$ by Assumptions (C7) and (C8) and $0 \leq p(u,x) \leq 1$ for all $(u,x) \in \mathcal{U}_0 \times \chi_0$ since p is a probability function, we can conclude exactly in the same way as above that

$$R_2 = O(\|\widehat{p} - p\|_\infty^2) + O(\|\widehat{p}_1 - p_1\|_\infty^2) + O(\|\widehat{f}_X - f_X\|_\infty^2) + O(\|\widehat{f}_{X,1} - f_{X,1}\|_\infty^2).$$

Next, combining (4.9.15) with (4.9.16) and (4.9.21), we have

$$\begin{aligned} \frac{\widehat{\varphi}_u}{\widehat{\varphi}_1} - \frac{\varphi_u}{\varphi_1} &= \frac{\varphi_u f_{X,1}}{\varphi_1^2 f_X^2} (\widehat{p} - p) + \frac{1}{\varphi_1 f_X} (\widehat{p}_u - p_u) - \frac{\varphi_u}{\varphi_1^2 f_X} (\widehat{p}_1 - p_1) \\ &\quad + \left\{ -\frac{p_u}{\varphi_1 f_X^2} - \frac{\varphi_u}{\varphi_1^2} \left(\frac{2p f_{X,1}}{f_X^3} - \frac{p_1}{f_X^2} \right) \right\} (\widehat{f}_X - f_X) \\ &\quad + \frac{\varphi_u p}{\varphi_1^2 f_X^2} (\widehat{f}_{X,1} - f_{X,1}) + R, \end{aligned}$$

where

$$R = -\frac{R_1}{\varphi_1} - \frac{\varphi_u R_2}{\varphi_1^2} - \frac{(\widehat{\varphi}_1 - \varphi_1)}{\widehat{\varphi}_1 \varphi_1} \left\{ (\widehat{\varphi}_u - \varphi_u) - \frac{\varphi_u (\widehat{\varphi}_1 - \varphi_1)}{\varphi_1} \right\}.$$

The last term in R is $O(\|\widehat{\varphi}_u - \varphi_u\|_\infty^2) + O(\|\widehat{\varphi}_1 - \varphi_1\|_\infty^2)$ and then also $O(\|\widehat{p}_u - p_u\|_\infty^2) + O(\|\widehat{p} - p\|_\infty^2) + O(\|\widehat{p}_1 - p_1\|_\infty^2) + O(\|\widehat{f}_X - f_X\|_\infty^2) + O(\|\widehat{f}_{X,1} - f_{X,1}\|_\infty^2)$ using exactly the same justifications as before. Hence, as $\sup_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} |\varphi_u(u, x)| < \infty$ and $\inf_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} |\varphi_1(u, x)| > 0$ by Assumptions (C7) and (C8) and $\|\widehat{\varphi}_1 - \varphi_1\|_\infty = o_P(1)$, the expression R is equal to

$$\begin{aligned} R &= O(\|\widehat{p}_u - p_u\|_\infty^2) + O(\|\widehat{p} - p\|_\infty^2) + O(\|\widehat{p}_1 - p_1\|_\infty^2) \\ &\quad + O(\|\widehat{f}_X - f_X\|_\infty^2) + O(\|\widehat{f}_{X,1} - f_{X,1}\|_\infty^2). \end{aligned}$$

We also have $\sup_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} |R(u, x)| = o_P(n^{-1/2})$ using Proposition 4.9.4(i). We will now use the notations introduced at the beginning of Section 4.5. We have directly that $\frac{\varphi_u f_{X,1}}{\varphi_1^2 f_X^2} = D_{p,0}$, $\frac{1}{\varphi_1 f_X} = D_{p,u}$ and $-\frac{\varphi_u}{\varphi_1^2 f_X} = D_{p,1}$. Moreover,

$$\frac{\varphi_u p}{\varphi_1^2 f_X^2} = \frac{\varphi_u \varphi}{\varphi_1^2 f_X} = D_{f,1},$$

and

$$\begin{aligned} &-\frac{p_u}{\varphi_1 f_X^2} - \frac{\varphi_u}{\varphi_1^2} \left(\frac{2p f_{X,1}}{f_X^3} - \frac{p_1}{f_X^2} \right) \\ &= \frac{\varphi_u}{\varphi_1 f_X} \left(-1 - \frac{2p f_{X,1}}{\varphi_1 f_X^2} + \frac{p_1}{\varphi_1 f_X} \right) \\ &= \frac{\varphi_u}{\varphi_1 f_X} \left\{ -1 - \frac{2\varphi f_{X,1}}{\varphi_1 f_X} + \frac{1}{\varphi_1} \left(\varphi_1 + \frac{\varphi f_{X,1}}{f_X} \right) \right\} \\ &= -\frac{\varphi_u \varphi f_{X,1}}{\varphi_1^2 f_X^2} \\ &= D_{f,0}. \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\widehat{\varphi}_u}{\widehat{\varphi}_1} - \frac{\varphi_u}{\varphi_1} &= D_{p,0} (\widehat{p} - p) + D_{p,u} (\widehat{p}_u - p_u) + D_{p,1} (\widehat{p}_1 - p_1) \\ &\quad + D_{f,0} (\widehat{f}_X - f_X) + D_{f,1} (\widehat{f}_{X,1} - f_{X,1}) + R. \end{aligned}$$

Finally, the following concludes the proof:

$$\begin{aligned}
& \widehat{S}_1(u, x) - S_1(u, x) \\
&= \int_0^u \left\{ \frac{\widehat{\varphi}_u(w, x)}{\widehat{\varphi}_1(w, x)} - \frac{\varphi_u(w, x)}{\varphi_1(w, x)} \right\} dw \\
&= \int_0^u \left\{ D_{p,0}(w, x) (\widehat{p}(w, x) - p(w, x)) + D_{p,u}(w, x) (\widehat{p}_u(w, x) - p_u(w, x)) \right. \\
&\quad \left. + D_{p,1}(w, x) (\widehat{p}_1(w, x) - p_1(w, x)) \right\} dw \\
&\quad + \int_0^u \left\{ D_{f,0}(w, x) (\widehat{f}_X(x) - f_X(x)) \right. \\
&\quad \left. + D_{f,1}(w, x) (\widehat{f}_{X,1}(x) - f_{X,1}(x)) \right\} dw + R(u, x) \\
&= \nabla_p S_1(u, x) [\widehat{p} - p] + \nabla_f S_1(u, x) [\widehat{f}_X - f_X] + R(u, x) .
\end{aligned}$$

□

Proposition 4.9.6. Assume (C1)-(C9), (J1) and (J2). Then,

$$\begin{aligned}
& \int_{\mathcal{X}} v_p(u_0, x) \nabla_p S_1(u, x) [\widehat{p} - p] dx \\
&= \frac{1}{n} \sum_{i=1}^n \left[\int_{\max(0, U_i)}^u \left\{ v_p(u_0, X_i) D_{p,0}(w, X_i) \right. \right. \\
&\quad \left. \left. - \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \right|_{x=X_i} \right\} dw \\
&\quad \left. + (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) v_p(u_0, X_i) D_{p,u}(U_i, X_i) \right] \\
&\quad + \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - w \right] \int_{\mathcal{X}} \left(\left\{ v_p(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)] \right\} f_{U,X}(w, x) \right. \\
&\quad \left. + v_p(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw \\
&\quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - 1 \right) \int_0^u w \int_{\mathcal{X}} \left(\left\{ v_p(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. - v_p(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) + \frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)] \right\} \right. \\
&\quad \left. \times f_{U,X}(w, x) dx dw \right) + o_P(n^{-1/2}) ,
\end{aligned}$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$ and where v_p is some continuous function on χ , equals 0 for values of x on the boundary of χ_0 , is m -times continuously differentiable and all these derivatives are uniformly bounded in the interior of χ_0 .

Proof. First, omitting the dependences of the different functions on u and x , we have

$$\begin{aligned} D_{p,0} \times p + D_{p,u} \times p_u + D_{p,1} \times p_1 &= \frac{\varphi_u f_{X,1}}{\varphi_1^2 f_X^2} \times p + \frac{1}{f_X \varphi_1} \times p_u - \frac{\varphi_u}{f_X \varphi_1^2} \times p_1 \\ &= \frac{\varphi_u f_{X,1} p}{\varphi_1^2 f_X^2} + \frac{\varphi_u}{\varphi_1} - \frac{\varphi_u}{\varphi_1^2} \left(\varphi_1 + \frac{p f_{X,1}}{f_X^2} \right) \\ &= 0, \end{aligned}$$

where in the second equality, we have used the fact that $\frac{p_1}{f_X} = \varphi_1 + \frac{p f_{X,1}}{f_X^2}$. Consequently, $\nabla_p S_1(u, x)[p] = 0$ and

$$\nabla_p S_1(u, x)[\hat{p} - p] = \nabla_p S_1(u, x)[\hat{p}] - \nabla_p S_1(u, x)[p] = \nabla_p S_1(u, x)[\hat{p}]$$

We will now develop the expression $\nabla_p S_1(u, x)[\hat{p}]$:

$$\begin{aligned} \nabla_p S_1(u, x)[\hat{p}] &= \int_0^u D_{p,0}(w, x) \hat{p}(w, x) dw + \int_0^u D_{p,u}(w, x) \hat{p}_u(w, x) dw \\ &\quad + \int_0^u D_{p,1}(w, x) \hat{p}_1(w, x) dw \\ &= A_1(u, x) + A_2(u, x) + A_3(u, x). \end{aligned} \tag{4.9.22}$$

First, using the definition of $\hat{p}$ in (4.4.3), we have

$$\begin{aligned} &\int_{\chi} v_p(u_0, x) A_1(u, x) dx \\ &= \int_{\chi} v_p(u_0, x) \int_0^u D_{p,0}(w, x) \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(w - \hat{U}_i) \mathbf{K}_{h_x}(X_i - x) \right\} dw dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \hat{U}_i) \int_{\chi} v_p(u_0, x) D_{p,0}(w, x) \mathbf{K}_{h_x}(X_i - x) dx dw. \end{aligned}$$

Next, using a Taylor expansion of order m and in dimension d , we have

$$\begin{aligned} &v_p(u_0, x) D_{p,0}(w, x) \\ &= \sum_{|\alpha|=0}^{m-1} \left[\frac{1}{|\alpha|!} \frac{\partial^{|\alpha|} \{v_p(u_0, x) D_{p,0}(w, x)\}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \Big|_{x=X_i} \prod_{j=1}^d (X_{ij} - x_j)^{\alpha_j} \right] + O_P(h_x^m), \end{aligned}$$

where $x = (x_1, \dots, x_d)^t$, $\alpha = (\alpha_1, \dots, \alpha_d)^t$ where $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and $|\alpha| = \sum_{j=1}^d \alpha_j$. Consequently, recalling that $\mathbf{K}_{h_x}(X_i - x) =$

$\frac{1}{h_x^d} \prod_{j=1}^d K\left(\frac{X_{ij}-x_j}{h_x}\right)$, we obtain

$$\begin{aligned}
& \int_{\mathcal{X}} v_p(u_0, x) D_{p,0}(w, x) \mathbf{K}_{h_x}(X_i - x) dx \\
&= \sum_{|\alpha|=0}^{m-1} \left[\frac{1}{|\alpha|!} \frac{\partial^{|\alpha|} \{v_p(u_0, x) D_{p,0}(w, x)\}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \right]_{x=X_i} \\
&\quad \times \int \frac{(X_{i1} - x_1)^{\alpha_1}}{h_x} K\left(\frac{X_{i1} - x_1}{h_x}\right) dx_1 \\
&\quad \times \dots \times \int \frac{(X_{id} - x_d)^{\alpha_d}}{h_x} K\left(\frac{X_{id} - x_d}{h_x}\right) dx_d \Big] + O_P(h_x^m) .
\end{aligned} \tag{4.9.23}$$

Condition (C7) implies that $\sup_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} D_{p,0}(u, x) \right| < \infty$ for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m-1$. Moreover, using the fact $\int z^\ell K(z) dz = 0$ for all $\ell = 1, \dots, m-1$ by Condition (C5) and the assumptions imposed on the function v_p , we can conclude that all the terms of the sum in (4.9.23) are equal to 0 except when $\alpha_1 = \dots = \alpha_d = 0$. Consequently,

$$\begin{aligned}
& \int_{\mathcal{X}} v_p(u_0, x) A_1(u, x) dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \left\{ v_p(u_0, X_i) D_{p,0}(w, X_i) + O_P(h_x^m) \right\} dw ,
\end{aligned}$$

uniformly in $(u_0, u) \in \mathcal{U}_0^2$. Moreover, the expression $n^{-1} \sum_{i=1}^n O_P(h_x^m) \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) dw$ is $o_P(n^{-1/2})$ since $\int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) dw \leq \int_{\mathcal{U}} \mathcal{K}_{h_u}(u - \widehat{U}_i) du$, $\mathcal{U}$ is compact, $\mathcal{K}_{h_u}(\cdot)$ is a function taking values between 0 and 1 and $n^{1/2} h_x^m \rightarrow 0$ by Condition (C6). Hence,

$$\begin{aligned}
& \int_{\mathcal{X}} v_p(u_0, x) A_1(u, x) dx \\
&= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) D_{p,0}(w, X_i) dw + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u \left\{ \mathcal{K}_{h_u}(w - \widehat{U}_i) - \mathcal{K}_{h_u}(w - U_i) \right\} D_{p,0}(w, X_i) dw \\
&\quad + \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u \mathcal{K}_{h_u}(w - U_i) D_{p,0}(w, X_i) dw + o_P(n^{-1/2}) \\
&= A_{1,1}(u_0, u) + A_{1,2}(u_0, u) + o_P(n^{-1/2}) .
\end{aligned}$$

Using Condition (C7), the assumptions imposed on v_p and Proposition 4.9.3(i) with $\ell(u_0, w, X_i) = v_p(u_0, X_i)D_{p,0}(w, X_i)$, the term $A_{1,1}(u_0, u)$ is equal to

$$\begin{aligned} & A_{1,1}(u_0, u) \\ &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - w \right] \int_{\mathcal{X}} v_p(u_0, x) D_{p,0}(w, x) f_{U,X}(w, x) dx dw \\ & \quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - 1 \right) \\ & \quad \times \int_0^u w \int_{\mathcal{X}} v_p(u_0, x) D_{p,0}(w, x) f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \end{aligned}$$

Moreover, the term $A_{1,2}(u_0, u)$ is equal to

$$\begin{aligned} A_{1,2}(u_0, u) &= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u D_{p,0}(w, X_i) 1_{\{U_i \leq w\}} dw + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_{\max(0, U_i)}^u D_{p,0}(w, X_i) dw + o_P(n^{-1/2}) , \end{aligned}$$

since $n^{1/2}h_u^m \rightarrow 0$ by Condition (C6). The first equality above was obtained using Proposition 4.9.2 with $\ell(u_0, w, x) = v_p(u_0, x)D_{p,0}(w, x)$. Consequently,

$$\begin{aligned} & \int_{\mathcal{X}} v_p(u_0, x) A_1(u, x) dx \\ &= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_{\max(0, U_i)}^u D_{p,0}(w, X_i) dw \\ & \quad + \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - w \right] \int_{\mathcal{X}} v_p(u_0, x) D_{p,0}(w, x) f_{U,X}(w, x) dx dw \\ & \quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - 1 \right) \\ & \quad \times \int_0^u w \int_{\mathcal{X}} v_p(u_0, x) D_{p,0}(w, x) f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \end{aligned} \tag{4.9.24}$$

We will now work on the expression $A_2(u, x)$. First, we have

$$\begin{aligned} & \int_{\chi} v_p(u_0, x) A_2(u, x) dx \\ &= \int_{\chi} v_p(u_0, x) \int_0^u D_{p,u}(w, x) \frac{\partial}{\partial w} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(w - \widehat{U}_i) \mathbf{K}_{h_x}(X_i - x) \right\} dw dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^u K_{h_u}(w - \widehat{U}_i) \int_{\chi} v_p(u_0, x) D_{p,u}(w, x) \mathbf{K}_{h_x}(X_i - x) dx dw . \end{aligned}$$

We apply now the same Taylor expansion as in the expression (4.9.23) but on the term $v_p(u_0, x) D_{p,u}(w, x)$. Hence, using the same reasoning that followed (4.9.23), the assumptions imposed on v_p and Condition (C7) that ensure that $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} D_{p,u}(u, x) \right| < \infty$ for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m-1$, we obtain

$$\begin{aligned} & \int_{\chi} v_p(u_0, x) A_2(u, x) dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^u K_{h_u}(w - \widehat{U}_i) \left\{ v_p(u_0, X_i) D_{p,u}(w, X_i) + O_P(h_x^m) \right\} dw , \end{aligned}$$

Moreover,

$$\frac{1}{n} \sum_{i=1}^n O_P(h_x^m) \int_0^u K_{h_u}(w - \widehat{U}_i) dw = o_P(n^{-1/2})$$

using the definition and the properties of $K_{h_u}(\cdot)$ and since $n^{1/2} h_x^m \rightarrow 0$ by Condition (C6). Hence,

$$\begin{aligned} & \int_{\chi} v_p(u_0, x) A_2(u, x) dx \\ &= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u K_{h_u}(w - \widehat{U}_i) D_{p,u}(w, X_i) dw + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u \left\{ K_{h_u}(w - \widehat{U}_i) - K_{h_u}(w - U_i) \right\} D_{p,u}(w, X_i) dw \\ &\quad + \frac{1}{n} \sum_{i=1}^n v_p(u_0, X_i) \int_0^u K_{h_u}(w - U_i) D_{p,u}(w, X_i) dw + o_P(n^{-1/2}) \\ &= A_{2,1}(u_0, u) + A_{2,2}(u_0, u) + o_P(n^{-1/2}) . \end{aligned}$$

Using Condition (C7), the assumptions imposed on v_p and Proposition 4.9.3(ii)

with $\ell(u_0, w, X_i) = v_p(u_0, X_i)D_{p,u}(w, X_i)$, the term $A_{2,1}(u_0, u)$ is equal to

$$\begin{aligned} A_{2,1}(u_0, u) &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{F_Y(1) - F_Y(0)} - w \right] \int_{\chi} v_p(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) dx dw \\ &\quad + \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \\ &\quad \times \int_0^u w \int_{\chi} v_p(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \end{aligned}$$

Moreover, the term $A_{2,2}(u_0, u)$ is equal to

$$A_{2,2}(u_0, u) = \frac{1}{n} \sum_{i=1}^n (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) v_p(u_0, X_i) D_{p,u}(U_i, X_i) + O_P(h_u^m) .$$

Note that $O_P(h_u^m)$ is also $o_P(n^{-1/2})$ since $nh_u^{2m} \rightarrow 0$ by Condition (C6). Consequently,

$$\begin{aligned} &\int_{\chi} v_p(u_0, x) A_2(u, x) dx \\ &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{F_Y(1) - F_Y(0)} - w \right] \int_{\chi} v_p(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) dx dw \\ &\quad + \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\chi} v_p(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) f_{U,X}(w, x) dx dw \\ &\quad + \frac{1}{n} \sum_{i=1}^n (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) v_p(u_0, X_i) D_{p,u}(U_i, X_i) + o_P(n^{-1/2}) . \end{aligned} \tag{4.9.25}$$

Finally, we develop the expression $A_3(u, x)$. First, note that

$$\begin{aligned} &\int_{\chi} v_p(u_0, x) A_3(u, x) dx \\ &= \int_{\chi} v_p(u_0, x) \int_0^u D_{p,1}(w, x) \frac{\partial}{\partial x_1} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_u}(w - \widehat{U}_i) \mathbf{K}_{h_x}(X_i - x) \right\} dw dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \int_{\chi} v_p(u_0, x) D_{p,1}(w, x) \mathbf{K}_{h_x,1}(X_i - x) dx dw , \end{aligned}$$

where $\mathbf{K}_{h_x,1}(X_i - x) = \frac{\partial}{\partial x_1} \mathbf{K}_{h_x}(X_i - x)$. Let $\chi_1 \subset \mathbb{R}$ be the compact support of X_1 , $\chi_{-1} \subset \mathbb{R}^{d-1}$ be the compact support of $(X_2, \dots, X_d)^t$ and $x_{-1} =$

$(x_2, \dots, x_d)^t$. Using integration by parts, the last expression is equal to

$$\begin{aligned}
& \int_{\chi} v_p(u_0, x) A_3(u, x) dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_{\chi_{-1}} \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \left[v_p(u_0, x) D_{p,1}(w, x) \mathbf{K}_{h_x}(X_i - x) \right]_{\chi_1} dw dx_{-1} \\
&\quad - \frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \int_{\chi} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \mathbf{K}_{h_x}(X_i - x) dx dw \\
&= -\frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \int_{\chi} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \mathbf{K}_{h_x}(X_i - x) dx dw .
\end{aligned}$$

We obtained the last equality since Condition (C7) implies that $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} |D_{p,1}(u, x)| < \infty$ and v_p is assumed to be continuous on χ and equals 0 for values of x on the boundary of χ_0 . We apply now the same Taylor expansion as in (4.9.23) but on the function $\frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)]$. Consequently, using the same reasoning that followed (4.9.23), the assumptions imposed on v_p and Condition (C7) that ensure that $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \frac{\partial^\tau}{\partial x_1^\tau} D_{p,1}(u, x) \right| < \infty$ for $\tau \in \{0, 1\}$ and for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m-1$, we obtain

$$\begin{aligned}
& \int_{\chi} v_p(u_0, x) A_3(u, x) dx \\
&= -\frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \left\{ \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} + O_P(h_x^m) \right\} dw .
\end{aligned}$$

Using the same reasoning as before, we obtain

$$\begin{aligned}
& \int_{\chi} v_p(u_0, x) A_3(u, x) dx \\
&= -\frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - \widehat{U}_i) \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} dw + o_P(n^{-1/2}) \\
&= -\frac{1}{n} \sum_{i=1}^n \int_0^u \left[\left\{ \mathcal{K}_{h_u}(w - \widehat{U}_i) - \mathcal{K}_{h_u}(w - U_i) \right\} \right. \\
&\quad \times \left. \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} \right] dw \\
&\quad - \frac{1}{n} \sum_{i=1}^n \int_0^u \mathcal{K}_{h_u}(w - U_i) \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} dw + o_P(n^{-1/2}) \\
&= A_{3,1}(u_0, u) + A_{3,2}(u_0, u) + o_P(n^{-1/2}) .
\end{aligned}$$

Using the assumptions imposed on v_p , Condition (C7) and Proposition 4.9.3(i) with $\ell(u_0, w, X_i) = \frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)]|_{x=X_i}$, the term $A_{3,1}(u_0, u)$ is equal to

$$\begin{aligned} & A_{3,1}(u_0, u) \\ &= \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - w \right] \int_{\mathcal{X}} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] f_{U,X}(w, x) dx dw \\ &\quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - 1 \right) \\ &\quad \times \int_0^u w \int_{\mathcal{X}} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \end{aligned}$$

Moreover, using Proposition 4.9.2 and the same reasoning as for $A_{1,2}(u_0, u)$, the term $A_{3,2}(u_0, u)$ is equal to

$$\begin{aligned} & A_{3,2}(u_0, u) \\ &= -\frac{1}{n} \sum_{i=1}^n \int_0^u \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} 1_{\{U_i \leq w\}} dw + o_P(n^{-1/2}) \\ &= -\frac{1}{n} \sum_{i=1}^n \int_{\max(0, U_i)}^u \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} dw + o_P(n^{-1/2}) . \end{aligned}$$

Consequently,

$$\begin{aligned} & \int_{\mathcal{X}} v_p(u_0, x) A_3(u, x) dx \\ &= -\frac{1}{n} \sum_{i=1}^n \int_{\max(0, U_i)}^u \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} dw \\ &\quad + \int_0^u \left[\frac{\widehat{F}_U(w) - \widehat{F}_U(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - w \right] \int_{\mathcal{X}} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] f_{U,X}(w, x) dx dw \\ &\quad - \left(\frac{\widehat{F}_Y(1) - \widehat{F}_Y(0)}{\widehat{F}_Y(1) - \widehat{F}_Y(0)} - 1 \right) \\ &\quad \times \int_0^u w \int_{\mathcal{X}} \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] f_{U,X}(w, x) dx dw + o_P(n^{-1/2}) . \end{aligned} \tag{4.9.26}$$

Combining (4.9.22), (4.9.24), (4.9.25) and (4.9.26), we obtain

$$\begin{aligned}
& \int_{\chi} v_p(u_0, x) \nabla_p S_1(u, x) [\hat{p} - p] dx \\
&= \int_{\chi} v_p(u_0, x) \{A_1(u, x) + A_2(u, x) + A_3(u, x)\} dx \\
&= \frac{1}{n} \sum_{i=1}^n \left[\int_{\max(0, U_i)}^u \left\{ v_p(u_0, X_i) D_{p,0}(w, X_i) \right. \right. \\
&\quad \left. \left. - \frac{\partial}{\partial x_1} \left[v_p(u_0, x) D_{p,1}(w, x) \right] \right|_{x=X_i} \right\} dw \\
&\quad + (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) v_p(u_0, X_i) D_{p,u}(U_i, X_i) \Big] \\
&+ \int_0^u \left[\frac{\hat{F}_U(w) - \hat{F}_U(0)}{F_Y(1) - F_Y(0)} - w \right] \int_{\chi} \left(\left\{ v_p(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)] \right\} f_{U,X}(w, x) \right. \\
&\quad \left. + v_p(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw \\
&- \left(\frac{\hat{F}_Y(1) - \hat{F}_Y(0)}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\chi} \left(\left\{ v_p(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. - v_p(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) + \frac{\partial}{\partial x_1} [v_p(u_0, x) D_{p,1}(w, x)] \right\} \right. \\
&\quad \left. \times f_{U,X}(w, x) dx dw \right) + o_P(n^{-1/2}) ,
\end{aligned}$$

□

Proposition 4.9.7. Assume (C1)-(C9), (J1) and (J2). Then,

$$\begin{aligned}
& \int_{\chi} v_f(u_0, x) \nabla_f S_1(u, x) [\hat{f}_X - f_X] dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^u \left\{ v_f(u_0, X_i) D_{f,0}(w, X_i) \right. \\
&\quad \left. - \frac{\partial}{\partial x_1} [v_f(u_0, x) D_{f,1}(w, x)] \right|_{x=X_i} \Big\} dw + o_P(n^{-1/2}) ,
\end{aligned}$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$ and where v_f is some continuous function on χ , equals 0 for values of x on the boundary of χ_0 , is m -times continuously differentiable and all these derivatives are uniformly bounded in the interior of χ_0 .

Proof. The proof of this proposition follows the same way as the proof of Proposition 4.9.6. First, omitting the dependences of the different functions on u and x , we have

$$D_{f,0} \times f_X + D_{f,1} \times f_{X,1} = -\frac{\varphi_u \varphi f_{X,1}}{\varphi_1^2 f_X^2} \times f_X + \frac{\varphi_u \varphi}{\varphi_1^2 f_X} \times f_{X,1} = 0 .$$

Hence, $\nabla_f S_1(u, x)[f_X] = 0$ and

$$\nabla_f S_1(u, x)[\widehat{f}_X - f_X] = \nabla_f S_1(u, x)[\widehat{f}_X] - \nabla_f S_1(u, x)[f_X] = \nabla_f S_1(u, x)[\widehat{f}_X] .$$

Next,

$$\begin{aligned} \nabla_p S_1(u, x)[\widehat{f}_X] &= \int_0^u D_{f,0}(w, x) \widehat{f}_X(x) dw + \int_0^u D_{f,1}(w, x) \widehat{f}_{X,1}(x) dw \\ &= B_1(u, x) + B_2(u, x) . \end{aligned}$$

We have

$$\int_{\mathcal{X}} v_f(u_0, x) B_1(u, x) dx = \frac{1}{n} \sum_{i=1}^n \int_0^u \int_{\mathcal{X}} v_f(u_0, x) D_{f,0}(w, x) \mathbf{K}_{h_x}(X_i - x) dx dw .$$

We apply now the same Taylor expansion as in the expression (4.9.23) but on the term $v_f(u_0, x) D_{f,0}(w, x)$. Hence, using the same reasoning that followed (4.9.23), the assumptions imposed on v_f and Condition (C7) that ensure that $\sup_{(u,x) \in \mathcal{U}_0 \times \mathcal{X}_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} D_{f,0}(u, x) \right| < \infty$ for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m-1$, we obtain

$$\begin{aligned} &\int_{\mathcal{X}} v_f(u_0, x) B_1(u, x) dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^u \left(v_f(u_0, X_i) D_{f,0}(w, X_i) + O_P(h_x^m) \right) dw \\ &= \frac{1}{n} \sum_{i=1}^n v_f(u_0, X_i) \int_0^u D_{f,0}(w, X_i) dw + o_P(n^{-1/2}) , \quad (4.9.27) \end{aligned}$$

since $n^{1/2} h_x^m \rightarrow 0$ by condition (C6). Second,

$$\int_{\mathcal{X}} v_f(u_0, x) B_2(u, x) dx = \frac{1}{n} \sum_{i=1}^n \int_0^u \int_{\mathcal{X}} v_f(u_0, x) D_{f,1}(w, x) \mathbf{K}_{h_{x,1}}(X_i - x) dx dw ,$$

where $\mathbf{K}_{h_{x,1}}(X_i - x) = \frac{\partial}{\partial x_1} \mathbf{K}_{h_x}(X_i - x)$. Next, the development of the expression $\int_{\mathcal{X}} v_f(u_0, x) B_2(u, x) dx$ is similar to that of the expression $\int_{\mathcal{X}} v_f(u_0, x)$

$A_3(u, x) dx$ in the proof of Proposition 4.9.6, i.e., we apply first an integration by parts and second a similar Taylor expansion as in equation (4.9.23). Hence, we obtain

$$\begin{aligned}
& \int_{\chi} v_f(u_0, x) B_2(u, x) dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^u \left\{ \int_{\chi_{-1}} \left[v_f(u_0, x) D_{f,1}(w, x) \mathbf{K}_{h_x}(X_i - x) \right]_{\chi_1} dx_{-1} \right. \\
&\quad \left. - \int_{\chi} \frac{\partial}{\partial x_1} [v_f(u_0, x) D_{f,1}(w, x)] \mathbf{K}_{h_x}(X_i - x) dx \right\} dw \\
&= -\frac{1}{n} \sum_{i=1}^n \int_0^u \frac{\partial}{\partial x_1} \left[v_f(u_0, x) D_{f,1}(w, x) \right] \Big|_{x=X_i} dw + O_P(h_x^m) .
\end{aligned} \tag{4.9.28}$$

where χ_1 is the compact support of X_1 , χ_{-1} is the compact support of $(X_2, \dots, X_d)^t$ and $x_{-1} = (x_2, \dots, x_d)^t$. To obtain the last equality, we have also used the facts that v_f is assumed to be continuous on χ and equals 0 for values of x on the boundary of χ_0 and $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \frac{\partial^\tau}{\partial x_1^\tau} D_{f,1}(u, x) \right| < \infty$ for $\tau \in \{0, 1\}$ and for all $\alpha_1, \dots, \alpha_d$ such that $\alpha_j \in \{0, \dots, m-1\}$ for $j = 1, \dots, d$ and where $|\alpha| = \sum_{j=1}^d \alpha_j$ is such that $|\alpha| \leq m-1$, which is a consequence of Condition (C7). Finally, combining (4.9.27) and (4.9.28), we obtain:

$$\begin{aligned}
& \int_{\chi} v_f(u_0, x) \nabla_f S_1(u, x) [\widehat{f}_X - f_X] dx \\
&= \int_{\chi} v_f(u_0, x) \{B_1(u, x) + B_2(u, x)\} dx \\
&= \frac{1}{n} \sum_{i=1}^n v_f(u_0, X_i) \int_0^u D_{f,0}(w, X_i) dw \\
&\quad - \frac{1}{n} \sum_{i=1}^n \int_0^u \frac{\partial}{\partial x_1} \left[v_f(u_0, x) D_{f,1}(w, x) \right] \Big|_{x=X_i} dw + o_P(n^{-1/2}) .
\end{aligned}$$

This concludes the proof. $\square$

4.9.2 Proofs of the main results

In this section, we give the proofs of the three main theorems introduced in Sections 4.3 and 4.5. The first one establishes how we can identify the transformation $\Gamma(u)$ and the second one and the third one establish respectively the limiting distributions of our new nonparametric estimators $\widehat{\Gamma}_{LS}(u)$ and $\widehat{\Gamma}_{LAD}(u)$ of the transformation $\Gamma(u)$ introduced in (4.4.7) and (4.4.8).

Proof of Theorem 4.3.1. From condition (C4), we choose ρ such that $\mathcal{A}_\rho$ is non empty and we take any point $x \in \mathcal{A}_\rho$. Since $\mathcal{U}_0$ is an interval of $\mathbb{R}$ that contains 0, we integrate from 0 to u both sides of the equation (4.3.3) and we obtain:

$$\int_0^u \Gamma'(w) dw = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) \int_0^u \frac{\varphi_u(w, x)}{\varphi_\rho(w, x)} dw = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) S_\rho(u, x) .$$

Hence,

$$\Gamma(u) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) S_\rho(u, x) , \quad (4.9.29)$$

since $\Gamma(0) = 0$ and $\int_0^u \Gamma'(w) dw = \Gamma(u) - \Gamma(0) = \Gamma(u)$. Using the normalization condition $\Gamma(1) = 1$, equation (4.9.29) becomes

$$1 = \Gamma(1) = - \left(\frac{\partial}{\partial x_\rho} m(x) \right) S_\rho(1, x) \iff - \left(\frac{\partial}{\partial x_\rho} m(x) \right) = \frac{1}{S_\rho(1, x)} . \quad (4.9.30)$$

Finally, combining equations (4.9.29) and (4.9.30), we obtain $\Gamma(u) = S_\rho(u, x)/S_\rho(1, x)$. Second, we will prove that $S_\rho(u, x)/S_\rho(1, x)$ is independent of x and ρ . We assume that there exists another index ℓ , $\ell \in \{1, \dots, d\}$ such that $\mathcal{A}_\ell$ is non empty and take any point $\tilde{x} \in \mathcal{A}_\ell$. Moreover, we assume that $\ell \neq \rho$ and $\tilde{x} \neq x$. Note that $s_\rho(u, x) = \varphi_u(u, x)/\varphi_\rho(u, x)$ and $S_\rho(1, x) = -1/\frac{\partial}{\partial x_\rho} m(x)$ by (4.9.30). Hence, for any $u \in \mathcal{U}_0$, $\Gamma'(u) = s_\rho(u, x)/S_\rho(1, x)$ using equation (4.3.3). Similarly, $\Gamma'(u) = s_\ell(u, \tilde{x})/S_\ell(1, \tilde{x})$, which implies that

$$\frac{s_\rho(u, x)}{S_\rho(1, x)} = \frac{s_\ell(u, \tilde{x})}{S_\ell(1, \tilde{x})} .$$

Finally,

$$\Gamma(u) = \frac{S_\rho(u, x)}{S_\rho(1, x)} = \int_0^u \frac{s_\rho(w, x)}{S_\rho(1, x)} dw = \int_0^u \frac{s_\ell(w, \tilde{x})}{S_\ell(1, \tilde{x})} dw .$$

This concludes the proof. $\square$

Proof of Theorem 4.5.1. First, using successively Theorem 4.3.1, equation

(4.4.6) and Proposition 4.9.1, we have

$$\begin{aligned}
& \widehat{\lambda}_1(u, x) - \Gamma(u) \\
&= \widehat{\lambda}_1(u, x) - \lambda_1(u, x) \\
&= \frac{\widehat{S}_1(u, x)}{\widehat{S}_1(1, x)} - \frac{S_1(u, x)}{S_1(1, x)} \\
&= \frac{1}{S_1(1, x)} \left\{ \widehat{S}_1(u, x) - S_1(u, x) \right\} - \frac{S_1(u, x)}{S_1^2(1, x)} \left\{ \widehat{S}_1(1, x) - S_1(1, x) \right\} \\
&\quad - \frac{1}{\widehat{S}_1(1, x)S_1(1, x)} \left\{ \widehat{S}_1(1, x) - S_1(1, x) \right\} \left\{ \widehat{S}_1(u, x) - S_1(u, x) \right\} \\
&\quad + \frac{S_1(u, x)}{\widehat{S}_1(1, x)S_1^2(1, x)} \left\{ \widehat{S}_1(1, x) - S_1(1, x) \right\}^2 . \tag{4.9.31}
\end{aligned}$$

Next, using the facts that

$$\widehat{S}_1(1, x) - S_1(1, x) \leq \sup_{(u, x) \in \mathcal{U}_0 \times \chi_0} |\widehat{S}_1(u, x) - S_1(u, x)| = \|\widehat{S}_1 - S_1\|_\infty ,$$

$$\widehat{S}_1(u, x) - S_1(u, x) \leq \sup_{(u, x) \in \mathcal{U}_0 \times \chi_0} |\widehat{S}_1(u, x) - S_1(u, x)| = \|\widehat{S}_1 - S_1\|_\infty ,$$

and $\sup_{(u, x) \in \mathcal{U}_0 \times \chi_0} |S_1(u, x)| < \infty$ and $\inf_{x \in \chi_0} |S_1(1, x)| > 0$ by Conditions (C7), (C8) and the fact that $\|\widehat{S}_1 - S_1\|_\infty = o_P(1)$, we can conclude that the third and the fourth terms on the right hand side of (4.9.31) are $O(\|\widehat{S}_1 - S_1\|_\infty^2)$. These two terms are also $o_P(n^{-1/2})$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$ using successively Propositions 4.9.5 and 4.9.4(ii).

Moreover, using Proposition 4.9.5 and the facts that $\sup_{x \in \chi_0} |v(x)| < \infty$ and $\inf_{x \in \chi_0} |S_1(1, x)| > 0$ by Conditions (C8) and (C9), we have

$$\begin{aligned}
& \int_{\chi} \frac{v(x)}{S_1(1, x)} \left\{ \widehat{S}_1(u, x) - S_1(u, x) \right\} dx \\
&= \int_{\chi} \frac{v(x)}{S_1(1, x)} \left\{ \nabla_p S_1(u, x)[\widehat{p} - p] + \nabla_f S_1(u, x)[\widehat{f}_X - f_X] \right\} dx + o_P(n^{-1/2}) .
\end{aligned}$$

Hence, we can apply Propositions 4.9.6 and 4.9.7 with $u_0 = 1$ and where $v_1(u_0, x) = \frac{v(x)}{S_1(u_0, x)}$ is the common function both for $v_p(\cdot)$ and $v_f(\cdot)$. We obtain:

$$\int_{\chi} \frac{v(x)}{S_1(1, x)} \left\{ \widehat{S}_1(u, x) - S_1(u, x) \right\} dx = \frac{1}{n} \sum_{i=1}^n \delta_i^{v_1}(1, u) + o_P(n^{-1/2}) , \tag{4.9.32}$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, where

$$\begin{aligned}
& \delta_i^{\tilde{v}}(u_0, u) \\
&= \int_{\max(0, U_i)}^u \left\{ \tilde{v}(u_0, X_i) D_{p,0}(w, X_i) - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=X_i} \right\} dw \\
&+ \int_0^u \left\{ \tilde{v}(u_0, X_i) D_{f,0}(w, X_i) - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{f,1}(w, x) \right] \Big|_{x=X_i} \right\} dw \\
&+ (1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) \tilde{v}(u_0, X_i) D_{p,u}(U_i, X_i) \\
&+ \int_0^u \left[\frac{1_{\{U_i \leq w\}} - 1_{\{U_i \leq 0\}}}{F_Y(1) - F_Y(0)} - w \right] \int_{\chi} \left(\left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(w, x)] \right\} f_{U,X}(w, x) \right. \\
&\quad \left. + \tilde{v}(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw \\
&- \left(\frac{1_{\{Y_i \leq 1\}} - 1_{\{Y_i \leq 0\}}}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\chi} \left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) \right. \\
&\quad \left. - \tilde{v}(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(u, x)] \right\} f_{U,X}(w, x) dx dw .
\end{aligned} \tag{4.9.33}$$

Similarly, using successively Proposition 4.9.5, the facts that $\sup_{(u,x) \in \mathcal{U}_0 \times \chi_0} |S_1(u, x)| < \infty$, $\inf_{x \in \chi_0} |S_1(1, x)| > 0$ and $\sup_{x \in \chi_0} |v(x)| < \infty$ as consequences of Conditions (C7), (C8) and (C9) and Propositions 4.9.6 and 4.9.7 with $u_0 = u$ and where $v_2(u_0, x) = \frac{v(x)S_1(u_0, x)}{S_1^2(1, x)}$ is the common function both for $v_p(\cdot)$ and $v_f(\cdot)$, we obtain

$$\begin{aligned}
& \int_{\chi} \frac{v(x)S_1(u, x)}{S_1^2(1, x)} \left\{ \widehat{S}_1(1, x) - S_1(1, x) \right\} dx \\
&= \int_{\chi} \frac{v(x)S_1(u, x)}{S_1^2(1, x)} \left\{ \nabla_p S_1(1, x) [\widehat{p} - p] \right. \\
&\quad \left. + \nabla_f S_1(1, x) [\widehat{f}_X - f_X] \right\} dx + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \delta_i^{v_2}(u, 1) + o_P(n^{-1/2}) ,
\end{aligned} \tag{4.9.34}$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$, where $\delta_i^{v_2}(u_0, u)$ is defined in (4.9.33). Next, using the definition of $\widehat{\Gamma}_{LS}(u)$ in (4.4.7) and the equations (4.9.31), (4.9.32)

and (4.9.34), we have

$$\begin{aligned}
\sqrt{n}(\widehat{\Gamma}_{LS}(u) - \Gamma(u)) &= \sqrt{n} \int_{\mathcal{X}} v(x) \{\widehat{\lambda}_1(u, x) - \Gamma(u)\} dx \\
&= \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \delta_i^{v_1}(1, u) - \frac{1}{n} \sum_{i=1}^n \delta_i^{v_2}(u, 1) + o_P(n^{-1/2}) \right\} \\
&= n^{-1/2} \sum_{i=1}^n \delta_i^v(u) + o_P(1),
\end{aligned}$$

where $\delta_i^v(u) = \delta_i^{v_1}(1, u) - \delta_i^{v_2}(u, 1)$. The pointwise weak convergence of $\sqrt{n}(\widehat{\Gamma}_{LS}(u) - \Gamma(u))$ to a Gaussian distribution is obtained using the central limit theorem. Next, to extend this result to weak functional convergence, we will prove that the class

$$\begin{aligned}
\mathcal{F} &= \{(a, b, c) \rightarrow \delta_{a,b}^{\tilde{v}}(u_0, u), u \in \mathcal{U}_0, u_0 \in \mathcal{U}_0\} \\
&= \left\{ (a, b, c) \rightarrow \int_{\max(0,a)}^u \left\{ \tilde{v}(u_0, b) D_{p,0}(w, b) \right. \right. \\
&\quad \left. \left. - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{p,1}(w, x) \right] \right|_{x=b} \right\} dw \\
&\quad + \int_0^u \left\{ \tilde{v}(u_0, b) D_{f,0}(w, b) - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{f,1}(w, x) \right] \right|_{x=b} \right\} dw \\
&\quad + (1_{\{a \leq u\}} - 1_{\{a \leq 0\}}) \tilde{v}(u_0, b) D_{p,u}(a, b) \\
&\quad + \int_0^u \left[\frac{1_{\{a \leq w\}} - 1_{\{a \leq 0\}}}{F_Y(1) - F_Y(0)} - w \right] \int_{\mathcal{X}} \left(\left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(w, x)] \right\} f_{U,X}(w, x) \right. \\
&\quad \left. + \tilde{v}(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw \\
&\quad - \left(\frac{1_{\{c \leq 1\}} - 1_{\{c \leq 0\}}}{F_Y(1) - F_Y(0)} - 1 \right) \int_0^u w \int_{\mathcal{X}} \left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) \right. \\
&\quad \left. - \tilde{v}(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) \right. \\
&\quad \left. + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(u, x)] \right\} f_{U,X}(w, x) dx dw : u, u_0 \in \mathcal{U}_0 \},
\end{aligned}$$

where $\tilde{v}$ will represent either v_1 or v_2 , is Donsker. To obtain this result, it

suffices to prove that the classes

$$\begin{aligned}
\mathcal{F}_1 &= \left\{ (a, b) \rightarrow \int_{\max(0, a)}^u \left\{ \tilde{v}(u_0, b) D_{p,0}(w, b) - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{p,1}(w, x) \right] \Big|_{x=b} \right\} dw \right. \\
&\quad \left. + \int_0^u \left\{ \tilde{v}(u_0, b) D_{f,0}(w, b) - \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, x) D_{f,1}(w, x) \right] \Big|_{x=b} \right\} dw : u, u_0 \in \mathcal{U}_0 \right\}, \\
\mathcal{F}_2 &= \{ a \rightarrow 1_{\{a \leq u\}} - 1_{\{a \leq 0\}} : u \in \mathcal{U}_0 \}, \\
\mathcal{F}_3 &= \{ (a, b) \rightarrow \tilde{v}(u_0, b) D_{p,u}(a, b) : u, u_0 \in \mathcal{U}_0 \}, \\
\mathcal{F}_4 &= \left\{ a \rightarrow \frac{1}{F_Y(1) - F_Y(0)} \int_{\max(0, a)}^u \int_{\mathcal{X}} \left(\left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) \right. \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(w, x)] \right\} f_{U,X}(w, x) \right. \\
&\quad \left. \left. + \tilde{v}(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw : u, u_0 \in \mathcal{U}_0 \right\}, \\
\mathcal{F}_5 &= \left\{ a \rightarrow \frac{1_{\{a \leq 0\}}}{F_Y(1) - F_Y(0)} \right\}, \\
\mathcal{F}_6 &= \left\{ - \int_0^u \int_{\mathcal{X}} \left(\left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(w, x)] \right\} \right. \right. \\
&\quad \left. \left. \times f_{U,X}(w, x) + \tilde{v}(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw : u, u_0 \in \mathcal{U}_0 \right\}, \\
\mathcal{F}_7 &= \left\{ - \int_0^u w \int_{\mathcal{X}} \left(\left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(w, x)] \right\} \right. \right. \\
&\quad \left. \left. \times f_{U,X}(w, x) + \tilde{v}(u_0, x) D_{p,u}(w, x) \frac{\partial}{\partial w} f_{U,X}(w, x) \right) dx dw : u, u_0 \in \mathcal{U}_0 \right\}, \\
\mathcal{F}_8 &= \left\{ c \rightarrow - \frac{1_{\{c \leq 1\}} - 1_{\{c \leq 0\}}}{F_Y(1) - F_Y(0)} + 1 \right\}, \\
\mathcal{F}_9 &= \left\{ \int_0^u w \int_{\mathcal{X}} \left\{ \tilde{v}(u_0, x) D_{p,0}(w, x) - \tilde{v}(u_0, x) \frac{\partial}{\partial w} D_{p,u}(w, x) \right. \right. \\
&\quad \left. \left. + \frac{\partial}{\partial x_1} [\tilde{v}(u_0, x) D_{p,1}(u, x)] \right\} f_{U,X}(w, x) dx dw : u, u_0 \in \mathcal{U}_0 \right\},
\end{aligned}$$

are Donsker and to apply Examples 2.10.7 and 2.10.8 in Van der Vaart and Wellner (1996) since

$$\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2 \mathcal{F}_3 + \mathcal{F}_4 + \mathcal{F}_5 \mathcal{F}_6 + \mathcal{F}_7 + \mathcal{F}_8 \mathcal{F}_9.$$

The reasoning that follows holds both for $\tilde{v} = v_1$ and $\tilde{v} = v_2$. First, note that all the functions that compose $\mathcal{F}_1, \mathcal{F}_3, \mathcal{F}_4, \mathcal{F}_6, \mathcal{F}_7$ and $\mathcal{F}_9$ are at least m times continuously differentiable by Conditions (C7) and (C9) which implies that m is the smoothness of these classes. Hence, applying Corollary 2.7.2 in Van der Vaart and Wellner (1996), we get that $\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P_1)) \leq C_1 \tilde{\varepsilon}^{-(d+1)/m}$, where C_1 is some positive constant, $N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P_1))$ is the $\tilde{\varepsilon}$ -bracketing number of the class $\mathcal{F}_1$, P_1 is the probability measure corresponding to the distribution of (U, X) , and $L_2(P_1)$ is the L_2 -norm. Consequently,

$$\begin{aligned} \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P_1))} d\tilde{\varepsilon} &= \int_0^{2C} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_1, L_2(P_1))} d\tilde{\varepsilon} \\ &\leq \int_0^{2C} \sqrt{C_1 \tilde{\varepsilon}^{-(d+1)/m}} d\tilde{\varepsilon} \\ &= \frac{\sqrt{C_1} (2C)^{-\frac{d-1}{2m}+1}}{-\frac{d-1}{2m}+1} \\ &< \infty, \end{aligned}$$

provided $m > (d+1)/2$, which is the case since $m > d+2$ by condition (C6), and where C is a uniform upper bound for f_1 when $f_1 \in \mathcal{F}_1$. We obtained the first equality since only one $\tilde{\varepsilon}$ -bracket suffices to cover $\mathcal{F}_1$ if $\tilde{\varepsilon} > 2C$. Using exactly the same reasoning, we can prove for $j = 3, 4, 6, 7, 9$ that $\int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_j, L_2(P_j))} d\tilde{\varepsilon} < \infty$ where $P_j \equiv P_1$.

Moreover, $\mathcal{F}_2$ is a class of increasing functions with image in $[0, 1]$. Hence, applying Theorem 2.7.5 in Van der Vaart and Wellner (1996), we get that $\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_2, L_2(P_2)) \leq C_2 \tilde{\varepsilon}^{-1}$ where C_2 is some positive constant and P_2 is the probability measure corresponding to the distribution of U . Hence,

$$\begin{aligned} \int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_2, L_2(P_2))} d\tilde{\varepsilon} &= \int_0^1 \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_2, L_2(P_2))} d\tilde{\varepsilon} \\ &\leq \sqrt{C_2} \int_0^1 \tilde{\varepsilon}^{-1/2} d\tilde{\varepsilon} \\ &= 2\sqrt{C_2} \\ &< \infty, \end{aligned}$$

where the first equality was obtained since only one $\tilde{\varepsilon}$ -bracket suffices to cover $\mathcal{F}_2$ if $\tilde{\varepsilon} > 1$. Finally, only one bracket suffices to cover $\mathcal{F}_5$ and $\mathcal{F}_8$ since these classes of functions do not depend on u and u_0 . Then, $\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_5, L_2(P_5)) = 0$ and $\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_8, L_2(P_8)) = 0$ where P_5 and P_8 are the probability measures corresponding to the distribution of Y .

Moreover, for $j = 1, \dots, 9$, we will now prove that the envelope functions F_j of the classes $\mathcal{F}_j$ possess a weak second moment, i.e., $P(F_j > w) = o(w^{-2})$ as $w \rightarrow \infty$. The envelope function of $\mathcal{F}_j$ is some function F_j such that $|f_j(u, x)| \leq F_j(u, x)$ for all $(u, x) \in \mathcal{U}_0 \times \chi_0$ and $f_j \in \mathcal{F}_j$.

We start with $\mathcal{F}_2$ which is the easier case. Indeed, for all functions $f_2 \in \mathcal{F}_2$, we have $|f_2| \leq F_2 = 1$ uniformly in $(u, x) \in \mathcal{U}_0 \times \chi_0$. Hence, using Markov's inequality, $P(F_2 > w) \leq \frac{E(F_2^3)}{w^3} = \frac{1}{w^3}$, which is also $O(w^{-3})$ and $o(w^{-2})$. Next, if $f_5 \in \mathcal{F}_5$ and $f_8 \in \mathcal{F}_8$, we necessarily have $|f_5| \leq F_5 = \frac{1}{F_Y(1) - F_Y(0)}$, and $|f_8| \leq F_8 = \frac{1}{F_Y(1) - F_Y(0)} + 1$, which are constants. Then, we have again $P(F_5 > w) = o(w^{-2})$ and $P(F_8 > w) = o(w^{-2})$.

Finally, the reasoning that follows is the same for the remaining classes and we explain it in detail only for $\mathcal{F}_1$. If $f_1 \in \mathcal{F}_1$, we have $|f_1| \leq F_1$ where F_1 is such that

$$\begin{aligned} F_1^3 \leq C_4 & \left\{ \sup_{(u, u_0, x) \in \mathcal{U}_0^2 \times \chi_0} \left| \tilde{v}(u_0, x) D_{p,0}(u, x) \right|^3 \right. \\ & + \sup_{(u, u_0, x) \in \mathcal{U}_0^2 \times \chi_0} \left| \frac{\partial}{\partial x_1} (\tilde{v}(u_0, x) D_{p,1}(u, x)) \right|^3 \\ & + \sup_{(u, u_0, x) \in \mathcal{U}_0^2 \times \chi_0} \left| \tilde{v}(u_0, x) D_{f,0}(u, x) \right|^3 \\ & \left. + \sup_{(u, u_0, x) \in \mathcal{U}_0^2 \times \chi_0} \left| \frac{\partial}{\partial x_1} (\tilde{v}(u_0, x) D_{f,1}(u, x)) \right|^3 \right\}, \quad (4.9.35) \end{aligned}$$

where C_4 is some positive constant. All the functions on the right hand side of (4.9.35) are continuous and the set $\mathcal{U}_0$ and χ_0 are compacts. Hence, $E(F_1^3) < \infty$ which implies that $P(F_1 > w)$ is $o(w^{-2})$ using Markov's inequality.

Consequently, for $j = 1, \dots, 9$, we have proved that

$$\int_0^{+\infty} \sqrt{\log N_{[]}(\tilde{\varepsilon}, \mathcal{F}_2, L_2(P_j))} d\tilde{\varepsilon} < \infty,$$

and the envelope functions F_j of $\mathcal{F}_j$ possess a weak second moment, which implies that the classes $\mathcal{F}_j$ are Donsker by Theorem 2.5.6 in Van der Vaart and Wellner (1996). In particular, the classes $\{(a, b) \rightarrow \delta_{a,b}^{v_1}(1, u), u \in \mathcal{U}_0\}$ and $\{(a, b) \rightarrow \delta_{a,b}^{v_2}(u, 1), u \in \mathcal{U}_0\}$ are Donsker and thus also the class $\{(a, b) \rightarrow \delta_{a,b}^v(u), u \in \mathcal{U}_0\}$. As a consequence, the process $\sqrt{n}(\hat{\Gamma}_{LS}(u) - \Gamma(u))$ converges to a Gaussian process.

Finally, we prove that $E\{\delta_i^{\tilde{v}}(u_0, u)\} = 0$, where $\delta_i^{\tilde{v}}(u_0, u)$ is defined in (4.9.33), which implies that $E\{\delta_i^v(u)\} = 0$. First, we calculate the expectation of the first part of the first term on the right hand side of (4.9.33):

$$\begin{aligned} & E\left(\int_{\max(0, U_i)}^u \tilde{v}(u_0, X_i) D_{p,0}(w, X_i) dw\right) \\ &= \int_{\chi} \int_{\mathcal{U}} \left[\int_{\max(0, t)}^u \tilde{v}(u_0, x) D_{p,0}(w, x) dw \right] f_{U,X}(t, x) dt dx \\ &= \int_{\chi} \tilde{v}(u_0, x) \int_{\mathcal{U}} \left[\int_0^u D_{p,0}(w, x) 1_{\{t \leq w\}} dw \right] f_{U,X}(t, x) dt dx. \end{aligned}$$

Rearranging the different integrals, the last expression is equal to

$$\begin{aligned} & \int_{\chi} \tilde{v}(u_0, x) \int_0^u D_{p,0}(w, x) \left[\int_{\mathcal{U}} f_{U,X}(t, x) 1_{\{t \leq w\}} dt \right] dw dx \\ &= \int_{\chi} \tilde{v}(u_0, x) \int_0^u D_{p,0}(w, x) \left[\int_{-\infty}^w f_{U,X}(t, x) dt \right] dw dx \end{aligned}$$

Using the definition of $p(u, x)$ given in (4.4.2), we obtain

$$\begin{aligned} & E \left(\int_{\max(0, U_i)}^u \tilde{v}(u_0, X_i) D_{p,0}(w, X_i) dw \right) \\ &= \int_{\chi} \tilde{v}(u_0, x) \int_0^u D_{p,0}(w, x) p(w, x) dw dx . \end{aligned} \quad (4.9.36)$$

Similarly, the expectation of the second part of the first term on the right hand side of (4.9.33) is equal to

$$\begin{aligned} & E \left(- \int_{\max(0, U_i)}^u \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, \tilde{x}) D_{p,1}(w, \tilde{x}) \right] \Big|_{\tilde{x}=X_i} dw \right) \\ &= - \int_{\chi} \int_0^u \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, \tilde{x}) D_{p,1}(w, \tilde{x}) \right] \Big|_{\tilde{x}=x} p(w, x) dw dx \\ &= - \int_{\chi_{-1}} \int_0^u \left[\tilde{v}(u_0, x) D_{p,1}(w, x) p(w, x) \right]_{\chi_1} dw dx_{-1} \\ &\quad + \int_{\chi} \int_0^u \tilde{v}(u_0, x) D_{p,1}(w, x) p_1(w, x) dw dx . \end{aligned} \quad (4.9.37)$$

where χ_1 is the compact support of X_1 , χ_{-1} is the compact support of $(X_2, \dots, X_d)^t$, $x_{-1} = (x_2, \dots, x_d)^t$, and the last expression was obtained using integration by parts. Using the same reasoning, the expectations of the first and second parts of the second term on the right hand side of (4.9.33) are respectively equal to

$$E \left(\int_0^u \tilde{v}(u_0, X_i) D_{f,0}(w, X_i) dw \right) = \int_{\chi} \tilde{v}(u_0, x) \int_0^u D_{f,0}(w, x) f_X(x) dw dx , \quad (4.9.38)$$

and

$$\begin{aligned} & E \left(- \int_0^u \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, \tilde{x}) D_{f,1}(w, \tilde{x}) \right] \Big|_{\tilde{x}=X_i} dw \right) \\ &= - \int_{\chi} \int_0^u \frac{\partial}{\partial x_1} \left[\tilde{v}(u_0, \tilde{x}) D_{f,1}(w, \tilde{x}) \right] \Big|_{\tilde{x}=x} f_X(x) dw dx \\ &= - \int_{\chi_{-1}} \int_0^u \left[\tilde{v}(u_0, x) D_{f,1}(w, x) f_X(x) \right]_{\chi_1} dw dx_{-1} \\ &\quad + \int_{\chi} \tilde{v}(u_0, x) \int_0^u D_{f,1}(w, x) f_{X,1}(x) dw dx , \end{aligned} \quad (4.9.39)$$

where the last equality was obtained using integration by parts. Moreover, the expectation of the third term on the right hand side of (4.9.33) is equal to

$$\begin{aligned}
& E \left((1_{\{U_i \leq u\}} - 1_{\{U_i \leq 0\}}) \tilde{v}(u_0, X_i) D_{p,u}(U_i, X_i) \right) \\
&= \int_{\mathcal{X}} \int_{\mathcal{U}} (1_{\{t \leq u\}} - 1_{\{t \leq 0\}}) \tilde{v}(u_0, x) D_{p,u}(t, x) f_{U,X}(t, x) dt dx \\
&= \int_{\mathcal{X}} \tilde{v}(u_0, x) \int_0^u D_{p,u}(w, x) p_u(w, x) dw dx, \tag{4.9.40}
\end{aligned}$$

where the last equality was obtained because $p_u(u, x) = f_{U,X}(u, x)$. Next, the expectation of the fourth term on the right hand side of (4.9.33) is equal to 0 since

$$\begin{aligned}
E \left[\frac{1_{\{U_i \leq w\}} - 1_{\{U_i \leq 0\}}}{F_Y(1) - F_Y(0)} - w \right] &= \frac{F_U(w) - F_U(0)}{F_Y(1) - F_Y(0)} - w \\
&= \frac{w\{F_Y(1) - F_Y(0)\} + F_Y(0) - F_U(0)}{F_Y(1) - F_Y(0)} - w \\
&= 0.
\end{aligned}$$

We have used the fact that $U \sim \text{Un} \left[\frac{-F_Y(0)}{F_Y(1) - F_Y(0)}, \frac{1 - F_Y(0)}{F_Y(1) - F_Y(0)} \right]$ which implies that $F_U(w) = w\{F_Y(1) - F_Y(0)\} + F_Y(0)$ and also $F_U(0) = F_Y(0)$. Similarly, the expectation of the fifth term on the right hand side of (4.9.33) is also equal to 0. Indeed, it suffices to use the fact that $E(\hat{F}_Y(1) - \hat{F}_Y(0)) = F_Y(1) - F_Y(0)$. Finally, combining (4.9.36) to (4.9.40), we obtain

$$\begin{aligned}
& E(\delta_i^{\tilde{v}}(u_0, u)) \\
&= \int_{\mathcal{X}} \tilde{v}(u_0, x) \int_0^u \left\{ D_{f,0}(w, x) f_X(x) + D_{f,1}(w, x) f_{X,1}(x) \right. \\
&\quad \left. + D_{p,0}(w, x) p(w, x) + D_{p,u}(w, x) p_u(w, x) + D_{p,1}(w, x) p_1(w, x) \right\} dw dx \\
&\quad - \int \int_0^u \left[\tilde{v}(u_0, x) \{ D_{p,1}(w, x) p(w, x) + D_{f,1}(w, x) f_X(x) \} \right]_{\mathcal{X}_1} dw dx_{-1} \\
&= \int_{\mathcal{X}} \tilde{v}(u_0, x) \left\{ \nabla_p S_1(u, x) [f_X] + \nabla_p S_1(u, x) [p] \right\} dx \\
&= 0.
\end{aligned}$$

Indeed, $\varphi(w, x) = \frac{p(w, x)}{f_X(x)}$ which implies that

$$\begin{aligned} & D_{p,1}(w, x)p(w, x) + D_{f,1}(w, x)f_X(x) \\ &= \frac{-\varphi_u(w, x)}{f_X(x)\varphi_1^2(w, x)} \cdot p(w, x) + \frac{\varphi_u(w, x)\varphi(w, x)}{\varphi_1^2(w, x)f_X(x)} \cdot f_X(x) \\ &= \frac{-\varphi_u(w, x)\varphi(w, x)}{\varphi_1^2(w, x)} + \frac{\varphi_u(w, x)\varphi(w, x)}{\varphi_1^2(w, x)} \\ &= 0, \end{aligned}$$

and we have also showed that $\nabla_p S_1(u, x)[p] = 0$ and $\nabla_p S_1(u, x)[f_X] = 0$ at the beginning of the proofs of Propositions 4.9.6 and 4.9.7 respectively. $\square$

Proof of Theorem 4.5.2. First, using the definition of the estimator $\hat{\Gamma}_{LAD}(u)$ in (4.4.8), we have

$$\begin{aligned} \hat{\Gamma}_{LAD}(u) &= \arg \min_{q_m \in \mathbb{R}} \int_{\mathcal{X}} v(x) \left| \hat{\lambda}_1(u, x) - q_m \right| dx \\ &= \arg \min_{q_m \in \mathbb{R}} \int_{\mathcal{X}} v(x) \left(\hat{\lambda}_1(u, x) - q_m \right) \left[2 \cdot 1_{\{\hat{\lambda}_1(u, x) - q_m > 0\}} - 1 \right] dx. \end{aligned}$$

Note that the function we have to minimize is not differentiable with respect to q_m . Hence, we follow Horowitz (1998) by introducing the following smoothed version of $\hat{\Gamma}_{LAD}(u)$:

$$\hat{\Gamma}_{LAD,b}(u) = \arg \min_{q_m \in \mathbb{R}} Q_b(q_m, \hat{\lambda}_1(u, \cdot)),$$

where

$$Q_b(q_m, \lambda_1(u, \cdot)) = \int_{\mathcal{X}} v(x) (\lambda_1(u, x) - q_m) [2L_b(\lambda_1(u, x) - q_m) - 1] dx, \quad (4.9.41)$$

where $L_b(\cdot) = L(\cdot/b)$, L is some twice continuously differentiable distribution function with uniformly bounded derivatives, with median at zero and $b > 0$ is some bandwidth. We also assume that $L_b(\cdot)$ satisfies Assumption (A5) in Horowitz (1998). Moreover, by very similar arguments as in Theorem 2.1 in Horowitz (1998), we can prove that $\sqrt{n}(\hat{\Gamma}_{LAD,b}(u) - \hat{\Gamma}_{LAD}(u)) = o_P(1)$. Consequently,

$$\sqrt{n}(\hat{\Gamma}_{LAD}(u) - \Gamma(u)) = \sqrt{n}(\hat{\Gamma}_{LAD,b}(u) - \Gamma(u)) + o_P(1),$$

and it suffices to establish the limiting distribution of $\sqrt{n}(\hat{\Gamma}_{LAD,b}(u) - \Gamma(u))$. This approach was also considered in Chiappori et al. (2015). We will now prove that $\hat{\Gamma}_{LAD,b}(u) = \hat{\Gamma}_{LS}(u) + o_P(n^{-1/2})$. Using the mean-value theorem

with some $\tilde{\Gamma}(u)$ between $\Gamma(u)$ and $\hat{\Gamma}_{LAD,b}(u)$, we have

$$\begin{aligned} & \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\Gamma}_{LAD,b}(u)} - \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} \\ &= \left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \tilde{\Gamma}(u)} \times \left\{ \hat{\Gamma}_{LAD,b}(u) - \Gamma(u) \right\} . \end{aligned}$$

By definition,

$$\left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\Gamma}_{LAD,b}(u)} = 0 .$$

Hence,

$$\begin{aligned} & \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} \\ &= - \left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \tilde{\Gamma}(u)} \times \left\{ \hat{\Gamma}_{LAD,b}(u) - \Gamma(u) \right\} . \quad (4.9.42) \end{aligned}$$

Moreover, by a Taylor expansion, we have

$$\begin{aligned} & \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} \\ &= \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\lambda}_1(u, x)} \\ &+ \left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \hat{\lambda}_1(u, x)} \times \left\{ \Gamma(u) - \hat{\lambda}_1(u, x) \right\} \\ &+ O(\|\Gamma(u) - \hat{\lambda}_1(u, x)\|_\infty^2) . \quad (4.9.43) \end{aligned}$$

Recall the definition of the function Q_b in (4.9.41). The derivative of Q_b with respect to q_m is equal to

$$\begin{aligned} \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} &= - \int_{\chi} v(x) \left\{ 2L_b(\hat{\lambda}_1(u, x) - q_m) - 1 \right\} dx \\ &- 2 \int_{\chi} v(x) \left\{ \hat{\lambda}_1(u, x) - q_m \right\} L'_b(\hat{\lambda}_1(u, x) - q_m) dx , \end{aligned}$$

where $L'_b(\cdot) = L'(\cdot/b)/b$. Consequently,

$$\begin{aligned} \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\lambda}_1(u, x)} &= - (2L_b(0) - 1) \int_{\chi} v(x) dx \\ &= 0 , \quad (4.9.44) \end{aligned}$$

because $\int_{\chi} v(x) dx = 1$, $L_b(0) = L(0) = \frac{1}{2}$ since L is some distribution function with median at zero and $L'_b(0) < \infty$. Similarly, the second derivative of Q_b with respect to q_m is equal to

$$\begin{aligned} \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} &= 4 \int_{\chi} v(x) L'_b(\hat{\lambda}_1(u, x) - q_m) dx \\ &\quad + 2 \int_{\chi} v(x) \left\{ \hat{\lambda}_1(u, x) - q_m \right\} L''_b(\hat{\lambda}_1(u, x) - q_m) dx, \end{aligned}$$

where $L''_b(\cdot) = L''(\cdot/b)/b^2$. Consequently,

$$\left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \hat{\lambda}_1(u, x)} = 4L'_b(0), \quad (4.9.45)$$

because $\int_{\chi} v(x) dx = 1$ and $L'_b(0) < \infty$. Inserting (4.9.44) and (4.9.45) in (4.9.43), we obtain,

$$\left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} = 4L'_b(0) \left\{ \Gamma(u) - \hat{\lambda}_1(u, \cdot) \right\} + O(\|\Gamma(u) - \hat{\lambda}_1(u, x)\|_{\infty}^2).$$

Using Theorem 4.5.1, we know that $\hat{\Gamma}_{LS}(u) - \Gamma(u) = \int_{\chi} v(x) \{\hat{\lambda}_1(u, x) - \Gamma(u)\} dx$ is $O_P(n^{-1/2})$ which implies that $\|\hat{\lambda}_1(u, x) - \Gamma(u)\|_{\infty}$ is also $O_P(n^{-1/2})$ since $\int_{\chi} v(x) dx = 1$. Hence, $\|\hat{\lambda}_1(u, x) - \Gamma(u)\|_{\infty}^2$ is $O_P(n^{-1})$ and also $o_P(n^{-1/2})$. Then,

$$\left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} = 4L'_b(0) \left\{ \Gamma(u) - \hat{\lambda}_1(u, x) \right\} + o_P(n^{-1/2}). \quad (4.9.46)$$

Next, using a Taylor expansion, we have

$$\begin{aligned} &\left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \tilde{\Gamma}(u)} \\ &= \left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \hat{\lambda}_1(u, x)} + O(\|\tilde{\Gamma}(u) - \hat{\lambda}_1(u, \cdot)\|_{\infty}). \end{aligned} \quad (4.9.47)$$

Moreover, note that

$$\Gamma(u) = \arg \min_{q_m \in \mathbb{R}} Q_b(q_m, \lambda_1(u, \cdot)).$$

Indeed, $\lambda_1(u, \cdot)$ is independent of x and equals $\Gamma(u)$ by Theorem 4.3.1 which implies that

$$Q_b(q_m, \lambda_1(u, \cdot)) = (\Gamma(u) - q_m) (2L_b(\Gamma(u) - q_m) - 1),$$

using the expression (4.9.41) and the fact that $\int_{\mathcal{X}} v(x) dx = 1$. By construction, $Q_b(q_m, \lambda_1(u, \cdot))$ is a positive function. Consequently, the value of q_m that minimizes $Q_b(q_m, \lambda_1(u, \cdot))$ is the values of q_m such that $Q_b(q_m, \lambda_1(u, \cdot)) = 0$, i.e., $\Gamma(u)$.

Using this result and the fact that $\hat{\Gamma}_{LAD,b}(u) = \arg \min_{q_m \in \mathbb{R}} Q_b(q_m, \hat{\lambda}_1(u, \cdot))$, we have

$$\begin{aligned} 0 &= \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\Gamma}_{LAD,b}(u)} - \left. \frac{\partial Q_b(q_m, \Gamma(u))}{\partial q_m} \right|_{q_m = \Gamma(u)} \\ &= \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \hat{\Gamma}_{LAD,b}(u)} - \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} \\ &\quad + \left. \frac{\partial Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m} \right|_{q_m = \Gamma(u)} - \left. \frac{\partial Q_b(q_m, \Gamma(u))}{\partial q_m} \right|_{q_m = \Gamma(u)}. \end{aligned}$$

Next, using twice the mean value theorem with some $\tilde{\Gamma}(u)$ between $\Gamma(u)$ and $\hat{\Gamma}_{LAD,b}(u)$ and with some $\xi(u)$ between $\Gamma(u)$ and $\hat{\lambda}_1(u, \cdot)$, the last expression becomes

$$\begin{aligned} 0 &= \left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \tilde{\Gamma}(u)} (\hat{\Gamma}_{LAD,b}(u) - \Gamma(u)) \\ &\quad + \left. \frac{\partial^2 Q_b(q_m, \lambda_1(u, \cdot))}{\partial q_m \partial \lambda_1} \right|_{\substack{q_m = \Gamma(u) \\ \lambda_1 = \xi(u)}} (\hat{\lambda}_1(u, \cdot) - \Gamma(u)). \end{aligned}$$

Note that the function Q_b is twice continuously differentiable with respect to each component and with uniformly bounded derivatives. Since $\|\hat{\lambda}_1(u, x) - \Gamma(u)\|_{\infty}$ is $o_P(n^{-1/4})$, we can conclude that $\|\hat{\Gamma}_{LAD,b}(u) - \Gamma(u)\|_{\infty}$ is $o_P(n^{-1/4})$. Then, the expression $\|\tilde{\Gamma}(u) - \hat{\lambda}_1(u, \cdot)\|_{\infty}$ is also $o_P(n^{-1/4})$ since

$$\|\tilde{\Gamma}(u) - \hat{\lambda}_1(u, \cdot)\|_{\infty} \leq \|\tilde{\Gamma}(u) - \Gamma(u)\|_{\infty} + \|\Gamma(u) - \hat{\lambda}_1(u, \cdot)\|_{\infty},$$

and $\tilde{\Gamma}(u)$ is between $\Gamma(u)$ and $\hat{\Gamma}_{LAD,b}(u)$. Next, using (4.9.45), the expression (4.9.47) becomes

$$\left. \frac{\partial^2 Q_b(q_m, \hat{\lambda}_1(u, \cdot))}{\partial q_m^2} \right|_{q_m = \tilde{\Gamma}(u)} = 4L'_b(0) + o_P(n^{-1/4}). \quad (4.9.48)$$

Finally, inserting (4.9.46) and (4.9.48) in (4.9.42), we obtain

$$\begin{aligned} &-4L'_b(0) \left\{ \hat{\lambda}_1(u, \cdot) - \Gamma(u) \right\} + o_P(n^{-1/2}) \\ &= - \left\{ 4L'_b(0) + o_P(n^{-1/4}) \right\} \left\{ \hat{\Gamma}_{LAD,b}(u) - \Gamma(u) \right\} \\ &= -4L'_b(0) \left\{ \hat{\Gamma}_{LAD,b}(u) - \Gamma(u) \right\} + o_P(n^{-1/2}), \end{aligned}$$

since $\|\widehat{\Gamma}_{LAD,b}(u) - \Gamma(u)\|_\infty$ is $o_P(n^{-1/4})$. Consequently, $\widehat{\Gamma}_{LAD,b}(u) = \widehat{\lambda}_1(u, \cdot) + o_P(n^{-1/2})$ and thus also $\widehat{\Gamma}_{LAD,b}(u) = \widehat{\Gamma}_{LS}(u) + o_P(n^{-1/2})$ by definition of $\widehat{\Gamma}_{LS}(u)$. Moreover,

$$\sqrt{n} \left(\widehat{\Gamma}_{LAD,b}(u) - \Gamma(u) \right) = \sqrt{n} \left(\widehat{\Gamma}_{LS}(u) - \Gamma(u) \right) + o_P(1) ,$$

and it suffices to apply Theorem 4.5.1 to conclude the proof. $\square$

Chapter 5

Discussion

5.1 General conclusions

In Chapter 1, we explained first how we can identify a semiparametric transformation model, i.e., a model where the transformation of the response is parametric and the regression function is nonparametric. To estimate the transformation parameter in that model, we introduced the profile likelihood and the mean squared distance from independence estimators proposed by Linton et al. (2008). Moreover, again for the same model, we presented the estimators of the error density and the error distribution functions defined in Colling et al. (2015) and Heuchenne et al. (2015) and explained their usefulness. Next, we explained how Chiappori et al. (2015) constructed a fully nonparametric estimator of the transformation in a model where both the transformation and the regression function are nonparametric. Finally, we focused on the problem of testing the parametric form of the regression function. We presented a literature review of existing goodness-of-fit tests in regression models without transformation of the response. More precisely, we explained among others the approach of Van Keilegom et al. (2008a) using the error distribution function, the approach of Bierens (1982), Stute (1997) and Escanciano (2006a) using the integrated regression function, the approach of Härdle and Mammen (1993) using the regression function, the approach of Hušková and Meintanis (2009) using the characteristic function and the approach of Van Keilegom et al. (2008b) using empirical likelihood ratio tests. We also mentioned the other aspects of a regression model that can be tested.

In Chapter 2, we constructed a test for the parametric form of the regression function in a semiparametric transformation model. 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 . These test statistics were based on a profile likelihood estimator for the transformation parameter and on a local

polynomial estimator for the regression function. 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, both using a bootstrap procedure to approximate the critical values of the test statistics under the null hypothesis. We also carried out a simulation study to illustrate the performance of the estimators of the error density function and the error distribution function that are useful to construct the test statistics. Under the null hypothesis, we concluded that the estimations of the nominal level were globally good when $d = 1$ and that the method suffers from curse-of-dimensionality problems when d increases, which can be reduced by increasing the sample size. Under the alternative, we concluded as expected that the power increases with the sample size and that the power decreases when θ_0 increases from 0 to 1. Finally, we concluded that the use of the profile likelihood estimator of θ_0 was more appropriate than the use of the mean squared distance from independence estimator of θ_0 since the corresponding obtained powers were considerably bigger.

In Chapter 3, we used the integrated approach to construct a new test for the parametric form of the regression function in a semiparametric transformation model. We used the same profile likelihood estimator for the transformation parameter and the same local polynomial estimator for the regression function as in Chapter 2. The main idea of our new test was to compare the integrated regression function estimated in a semiparametric way to the one estimated under H_0 . We defined a Kolmogorov-Smirnov and a Cramér-von Mises test statistic, both based on an empirical residual process depending on a general weighting function that allowed us to apply several approaches we found in the literature. We established the limiting distribution of these two test statistics under the null hypothesis and under a local alternative and we noticed that our tests have power against local alternatives that converge to the null model at parametric rate, like the tests constructed in Chapter 2. We compared the performance of this new test to the previous test developed in Chapter 2 by means of a large simulation study and using the same bootstrap procedure as the one in Chapter 2 to approximate the critical values of the test statistics under the null hypothesis. Finally, we applied our test on a real data set. Under the null hypothesis, we concluded that the estimations of the nominal level were globally good when $d = 1$ and that the new method suffers from the same curse-of-dimensionality problems as the test constructed in Chapter 2 when d increases. Under the alternative, we again observed that the power decreases when θ_0 increases from 0 to 1 and also when the variance of ε increases. In conclusion, we recommend to use either the weighting function $\exp(ix^t X)$ when the deviation from the null hypothesis is monotone and to use the test constructed in Chapter 2 when the deviation from the null hypothesis is non monotone, especially when the dimension d increases. Moreover, it is not a good idea to use indicator weights $1_{\{X \leq x\}}$ and $1_{\{\gamma^t X \leq x\}}$ since they give a

very poor power, except when the model is easier to estimate, i.e., when $d = 1$ and $\theta_0 = 0$ for example. Finally, we recommend to use the weighting functions $\exp(x^t X)$, $(1 + \exp(-x^t X))^{-1}$ and $\sin(x^t X)$ in the particular case where the deviation from the null hypothesis is monotone and $d = 1$.

In Chapter 4, we constructed new nonparametric estimators for the transformation in a nonparametric transformation model, the first one based on the least squares loss and the second one based on the least absolute deviation loss. The motivation was to improve the estimators developed by Chiappori et al. (2015) that can work very badly in practice since they are constructed using kernel smoothing on Y and the distribution of Y can be very asymmetric and have very long tails. To overcome this problem, our main idea was to construct estimators using kernel smoothing on $U = T(Y)$ where U is uniformly distributed. Our new nonparametric estimators for the transformation were defined on the basis of kernel estimators of the conditional distribution function of U given X and of its derivatives with respect to u and x_1 . We established the asymptotic normality of these two estimators. Finally, a simulation study was carried out to illustrate the performance of our estimators, to compare these new estimators to the previous one and to see under which model conditions which estimator behaves the best. We concluded that our estimators clearly outperform the ones defined in Chiappori et al. (2015) when the transformation of the response is not the identity and when we consider values of y that are far away from the points that define the identification conditions whereas the estimators given in Chiappori et al. (2015) only outperforms our estimators when we consider values of y close to these identification points or when we consider a nonparametric regression model without transformation of the response.

5.2 Future research

5.2.1 Related extensions

In Chapter 2, we focused on the problem of testing the parametric form of the regression function in a semiparametric transformation model. As mentioned in Chapter 1, several other aspects of that model could also be tested using the approach of the error distribution function. For instance, we could carry over to a semiparametric transformation model the works of Neumeyer and Dette (2007), Pardo-Fernández et al. (2007), Neumeyer and Van Keilegom (2010) and Heuchenne and Van Keilegom (2010) who constructed respectively tests for the symmetry of the residual distribution, the equality of regression curves, the additivity of the regression function and a specific parametric form of the error distribution for a location-scale model of the form $Y = m(X) + \sigma(X)\varepsilon$, see Section 1.4.2 for more details. The spirit of these new tests would be the same as the ones described in these papers except that the Kolmogorov-Smirnov and Cramér-von Mises test statistics would be built using the empirical distribution function of the semiparametric residuals $\hat{\varepsilon} = \Lambda_{\hat{\theta}}(Y) - \hat{m}(X)$ instead

of $\widehat{\varepsilon} = \frac{Y - \widehat{m}(X)}{\widehat{\sigma}(X)}$.

Moreover, Neumeyer et al. (2016) already extended the works of Linton et al. (2008) and Einmahl and Van Keilegom (2008) in the sense that they constructed a profile likelihood estimator for the transformation parameter θ_0 in a heteroscedastic transformation model of the form $\Lambda_\theta(Y) = m(X) + \sigma(X)\varepsilon$ and also constructed a test for the validity of that model whereas Einmahl and Van Keilegom (2008) did the same for the location-scale model. However, other possibilities of future research could be for example the construction of tests for the parametric form of the variance function and for the equality of variance functions in a heteroscedastic transformation model, which would be extensions of the works of Dette et al. (2007) and Pardo-Fernández et al. (2015b) respectively.

In Chapters 2 and 3, we constructed tests for the parametric form of the regression function in a semiparametric transformation model using the approaches of the error distribution function and the integrated regression function respectively. However, as mentioned in Section 1.4.4, other ways could be considered to construct such tests, which would give useful alternatives for the tests developed in this thesis. First, we could extend the work of Härdle and Mammen (1993) presented in Section 1.4.4.1 to a semiparametric transformation model and define test statistics based on the difference between $\widehat{m}(x)$, the local polynomial estimator of m defined in Section 2.2.1, and $\widehat{m}_{\widehat{\beta}}(x)$ the smoothed parametric estimator of m obtained under H_0 and defined in Section 2.2.2. Second, we could extend the test constructed by Hušková and Meintanis (2009) and presented in Section 1.4.4.2 and define Kolmogorov-Smirnov and Cramér-von Mises test statistics measuring the distance between the characteristic function of ε estimated in a general context by

$$\widehat{\phi}_{\widehat{\varepsilon}}(y) = n^{-1} \sum_{j=1}^n e^{iy\widehat{\varepsilon}_j} ,$$

and the characteristic function of ε estimated under H_0 by

$$\widehat{\phi}_{\varepsilon_0}(y) = n^{-1} \sum_{j=1}^n e^{iy\widehat{\varepsilon}_{j0}} ,$$

where $\widehat{\varepsilon}_j = \Lambda_{\widehat{\theta}}(Y_j) - \widehat{m}(X_j)$ and $\widehat{\varepsilon}_{j0} = \Lambda_{\widehat{\theta}}(Y_j) - \widehat{m}_{\widehat{\beta}}(X_j)$. Finally, we could also extend the work of Van Keilegom et al. (2008b) presented in Section 1.4.4.3 and base our test on the following empirical likelihood ratio:

$$\sup \left\{ n^n \prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i w(X_i, x, \gamma) [\Lambda_{\widehat{\theta}}(Y_i) - m_{\widehat{\beta}}(X_i)] = 0 \right\} ,$$

where w is some weighting function that depends on some parameter $\gamma \in \mathbb{R}^{d_\gamma}$, like in Chapter 3.

The fully nonparametric estimators $\hat{\Gamma}_{LS}(u)$ and $\hat{\Gamma}_{LAD}(u)$ of the transformation developed in Chapter 4 could be very useful to construct new estimators for the transformation parameter θ_0 in semiparametric transformation models. Indeed, the main idea of these estimators would be a least squares minimization, with respect to $\theta \in \Theta$, of the distance between a nonparametric estimator of the transformation, either $\hat{\Gamma}_{LS}(\hat{T}(y))$ or $\hat{\Gamma}_{LAD}(\hat{T}(y))$, and the parametric transformation $\Lambda_\theta(y)$. It would be interesting to prove the consistency of these estimators, to establish their asymptotic distributions and to perform a simulation study to conclude under which model conditions these new estimators of θ_0 behave better in practice than the profile likelihood estimator introduced by Linton et al. (2008). Note that these new least squares estimators of the parameter θ_0 could be used to construct new tests for the parametric form of the regression function in semiparametric transformation models. Indeed, the theory of Chapters 2 and 3 could be adapted using these new estimators instead of the profile likelihood estimator of θ_0 .

Moreover, the tests in Chapters 2 and 3 could be extended to the context of fully nonparametric transformation models. Indeed, the construction of the test statistics would follow the same way as in Chapters 2 and 3 but using the nonparametric estimators $\hat{\Gamma}_{LS}(\hat{T}(y))$ and $\hat{\Gamma}_{LAD}(\hat{T}(y))$ of the transformation instead of a parametric estimator $\Lambda_{\hat{\theta}}(y)$ of the transformation.

Finally, another application of the nonparametric estimators defined in Chapter 4 would be the construction of a test for the parametric form of the transformation, i.e., we could test $H_0 : \Lambda \in \mathcal{L}$ where $\mathcal{L} = \{\Lambda_\theta : \theta \in \Theta\}$ and Θ is a compact subset of $\mathbb{R}^q$. The main idea of the corresponding Kolmogorov-Smirnov and Cramér-von Mises test statistics would be to compare a nonparametric estimator of the transformation, i.e., either $\hat{\Gamma}_{LS}(\hat{T}(y))$ or $\hat{\Gamma}_{LAD}(\hat{T}(y))$, to an estimator of the transformation obtained under H_0 , i.e., $\Lambda_{\hat{\theta}}(y)$ where $\hat{\theta}$ is either the profile likelihood estimator defined by Linton et al. (2008) or the least squares estimator presented in the previous paragraph.

5.2.2 Transformation models in other contexts

It is also important to notice that transformation models can be encountered in other various contexts, like in survival analysis and in quantile regression for example. First, consider the context of survival analysis and let T be the failure time and the response variable and let X be a covariate vector. Suppose that we want to analyze the effect of X on the response T . If there are censored observations in the data, we can consider among others the Cox proportional hazards model, see Cox (1972), or the proportional odds model, see Bennett (1983), to examine the covariate effect. The Cox model can be written as

$$\log [-\log \{S_{T|X}(t)\}] = h(t) + X^t \beta ,$$

where $S_{T|X}(t) = P(T > t | X = x)$ is the survival function of T given X , $h(\cdot)$ is a completely unspecified strictly increasing function and β is a vector of

unknown regression coefficients. Moreover, the proportional odds model can be written as

$$-\text{logit}\{S_{T|X}(t)\} = h(t) + X^t\beta ,$$

where $\text{logit}(x) = \log\{x/(1-x)\}$. A natural generalisation of these two models is given by $g(S_{T|X}(t)) = h(t) + X^t\beta$, where $g(\cdot)$ is a known decreasing function, which is also equivalent to the following linear transformation model:

$$h(T) = -X^t\beta + \varepsilon , \quad (5.2.1)$$

where the random error ε has a known distribution function $F_\varepsilon(e) = 1 - g^{-1}(e)$. Model (5.2.1) is the Cox proportional hazards model when $F_\varepsilon(e) = 1 - \exp\{-\exp(s)\}$ is the extreme value distribution and model (5.2.1) is the proportional odds model when $F_\varepsilon(e) = \frac{1}{1+\exp(-y)}$ is the standard logistic distribution. We also observe that model (5.2.1) is a linear transformation model and is semiparametric since the transformation h of the response is unknown and the regression function is linear and is known up to the parameter β . Cheng et al. (1995) proposed a class of simple estimating functions for β in model (5.2.1) with possibly censored observations and showed under suitable conditions that these estimators are consistent and asymptotically normal. However, model (5.2.1) is less general than the semiparametric and nonparametric transformation models considered in this thesis since the regression function is linear and the distribution of the error term is known. Consequently, another possibility of future research could be the extension of the tests and estimators developed in this thesis to the case of censored data.

Second, if we are interested in estimating the conditional median, or more generally other quantiles, of the response variable given certain values of the predictor variables instead of the conditional mean, we have to use another type of regression analysis which is called the quantile regression. More formally, the quantile of order $0 < \tau < 1$ of a random variable Y is defined by $q_\tau = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$ where $F(y) = P(Y \leq y)$ is the distribution function of Y . The quantile regression model was introduced by Koenker and Bassett (1978), see also the book of Koenker (2005), and assumes that:

$$Q_\tau(Y|X = x) = x^t\beta_\tau ,$$

where X is a vector of regressors, Y is a scalar dependent variable, β is an unknown vector of parameters and $Q_\tau(Y|X = x)$ denotes the τ -th conditional quantile of Y on the regressor vector X . In that case, the parameter β can be estimated by solving the following minimization problem:

$$\hat{\beta}_\tau = \arg \min_{b \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - X_i^t b) ,$$

where $\rho_\tau(u) = |u| - (2\tau - 1)u$ and (X_i, Y_i) is a random sample from some population for $i = 1, \dots, n$. For instance, if we are interested in the conditional

median of Y , i.e., $\tau = 0.5$ and $\rho_\tau(u) = |u|$, this is equivalent to minimizing the sum of absolute errors, which corresponds to the well-known least absolute deviation estimator. This study can be extended by transforming parametrically the response variable Y and thus by considering the model:

$$Q_\tau(\Lambda_{\theta_\tau}(Y)|X = x) = x^t \beta_\tau ,$$

which is also equivalent to

$$\Lambda_{\theta_\tau}(Y) = X^t \beta_\tau + \varepsilon \quad \text{with} \quad Q_\tau(\varepsilon|X = x) = 0 ,$$

where θ_τ is an unknown transformation parameter and $\Lambda_{\theta_\tau}(\cdot)$ is a family of strictly increasing transformations. Hence, this model is similar to the semi-parametric transformation model considered in Chapter 2 except that we assume now that a certain quantile of ε is equal to 0 instead of its mean. Note also that transformations on quantile models are natural due to the equivariance property of the quantiles under monotone transformations, i.e., the quantiles of a transformed distribution are the transformed quantiles of the original distribution. In the literature, the most common used transformation in that context is the Box and Cox (1964) transform. The corresponding model is called the “Box-Cox quantile regression” and we can find several important contributions for estimating the transformation parameter θ_τ in this particular linear quantile regression model, see Buchinsky (1995), Machado and Mata (2000), Mu and He (2007) and Fitzenberger et al. (2010) among others. Consequently, these works could also be generalized first by considering an unknown regression function m instead of the linear term $X^t \beta_\tau$ and second by constructing tests for the parametric form of this function m .

Bibliography

- Ahmad, I. A. and Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *Journal of Nonparametric Statistics*, 7(3):279–293.
- Akritis, M. G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, 28(3):549–567.
- Alcalá, J., Cristóbal, J., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statistics & Probability Letters*, 42(1):39–46.
- Bellver, C. (1987). Influence of particulate pollution on the positions of neutral points in the sky at seville (spain). *Atmospheric Environment (1967)*, 21(3):699–702.
- Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in medicine*, 2(2):273–277.
- Beverton, R. J. and Holt, S. J. (1957). *On The Dynamics Of Exploited Fish Populations*. Her Majestys Stationery Office; London.
- Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):296–311.
- Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134.
- Bierens, H. J. (1984). Model specification testing of time series regressions. *Journal of Econometrics*, 26(3):323–353.
- Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58(6):1443–1458.
- Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65(5):1129–1151.
- Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.

- Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80(391):580–598.
- Buchinsky, M. (1995). Quantile regression, box-cox transformation model, and the us wage structure, 1963–1987. *Journal of Econometrics*, 65(1):109–154.
- Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, volume 30. CRC Press.
- Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70(4):1683–1697.
- Chen, S. X. and Cui, H. (2006). On bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 93(1):215–220.
- Cheng, F. and Sun, S. (2008). A goodness-of-fit test of the errors in nonlinear autoregressive time series models. *Statistics & Probability Letters*, 78(1):50–59.
- Cheng, S., Wei, L., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845.
- Chiappori, P.-A., Komunjer, I., and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188(1):22–39.
- Cleveland, W. S. (1993). *Visualizing Data*. Hobart Press.
- Colling, B., Heuchenne, C., Samb, R., and Van Keilegom, I. (2015). Estimation of the error density in a semiparametric transformation model. *Annals of the Institute of Statistical Mathematics*, 67(1):1–18.
- Colling, B. and Van Keilegom, I. (2016). Goodness-of-fit tests in semiparametric transformation models. *Test*, 25(2):291–308.
- Colling, B. and Van Keilegom, I. (2017a). Estimation of fully nonparametric transformation models. (*submitted*).
- Colling, B. and Van Keilegom, I. (2017b). Goodness-of-fit tests in semiparametric transformation models using the integrated regression function. *Journal of Multivariate Analysis*, 160:10–30.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34:187–220.
- Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics*, 27(3):1012–1040.

- Dette, H., Kusi-Appiah, S., and Neumeyer, N. (2002). Testing symmetry in nonparametric regression models. *Journal of Nonparametric Statistics*, 14(5):477–494.
- Dette, H. and Munk, A. (1998). Validation of linear regression models. *The Annals of Statistics*, 26(2):778–800.
- Dette, H., Munk, A., and Wagner, T. (2000). Testing model assumptions in multivariate linear regression models. *Journal of Nonparametric Statistics*, 12(3):309–342.
- Dette, H., Neumeyer, N., and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490.
- 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 (Statistical Methodology)*, 69(5):903–917.
- Einmahl, J. H. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, 143(1):88–102.
- Escanciano, J. C. (2006a). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051.
- Escanciano, J. C. (2006b). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101(474):531–541.
- Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica*, 17(1):115–138.
- Escanciano, J. C. and Jacho-Chávez, D. T. (2012). $\sqrt{n}$ -uniformly consistent density estimation in nonparametric regression models. *Journal of Econometrics*, 167(2):305–316.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, volume 66. CRC Press.
- Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test*, 16(3):409–444.
- Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics*, 29(1):153–193.
- Fitzenberger, B., Wilke, R. A., and Zhang, X. (2010). Implementing box–cox quantile regression. *Econometric Reviews*, 29(2):158–181.

- González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, 22(3):361–411.
- Hall, P. and Yatchew, A. (2005). Unified approach to testing functional hypotheses in semiparametric contexts. *Journal of Econometrics*, 127(2):225–252.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947.
- Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit Tests*. Springer.
- Heuchenne, C., Samb, R., and Van Keilegom, I. (2015). Estimating the residual distribution in semiparametric transformation models. *Electronic Journal of Statistics*, 9:2391–2419.
- Heuchenne, C. and Van Keilegom, I. (2010). Goodness-of-fit tests for the error distribution in nonparametric regression. *Computational Statistics & Data Analysis*, 54(8):1942–1951.
- Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, 37(3):1079–1111.
- Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–137.
- Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 66(6):1327–1351.
- Horowitz, J. L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, 69(2):499–513.
- Hušková, M. and Meintanis, S. G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. *Kybernetika*, 45(6):960–971.
- Hušková, M. and Meintanis, S. G. (2010). Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *Test*, 19(1):92–112.
- Jacho-Chavez, D., Lewbel, A., and Linton, O. (2010). Identification and nonparametric estimation of a transformation additively separable model. *Journal of Econometrics*, 156(2):392–407.
- Jennrich, R. I. (1969). Asymptotic properties of non linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643.

- John, J. and Draper, N. (1980). An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):190–197.
- Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *The Annals of Statistics*, 27(1):204–236.
- Linton, O., Sperlich, S., and Van Keilegom, I. (2008). Estimation of a semi-parametric transformation model. *The Annals of Statistics*, 36(2):686–718.
- Machado, J. A. and Mata, J. (2000). Box-cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics*, 15(3):253–274.
- MacKinnon, J. G. and Magee, L. (1990). Transforming the dependent variable in regression models. *International Economic Review*, 31(2):315–339.
- Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.
- Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 102(477):269–279.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.
- Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scandinavian Journal of Statistics*, 36(2):204–228.
- Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametric regression models. *Statistica Sinica*, 17(2):775–795.
- Neumeyer, N., Noh, H., and Van Keilegom, I. (2016). Heteroscedastic semiparametric transformation models: estimation and testing for validity. *Statistica Sinica*, 26:925–954.
- 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(5):1067–1078.
- Owen, A. B. (1998). *Empirical Likelihood*. Wiley Online Library.

- Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057.
- Pardo-Fernández, J. C., Jiménez-Gamero, M. D., and El Ghouch, A. (2015a). A non-parametric anova-type test for regression curves based on characteristic functions. *Scandinavian Journal of Statistics*, 42(1):197–213.
- Pardo-Fernández, J. C., Jiménez-Gamero, M. D., and El Ghouch, A. (2015b). Tests for the equality of conditional variance functions in nonparametric regression. *Electronic Journal of Statistics*, 9(2):1826–1851.
- Pardo-Fernández, J. C., Van Keilegom, I., and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, 17(3):1115–1137.
- Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430.
- Ricker, W. and Smith, H. (1975). A revised interpretation of the history of the skeena river sockeye salmon. *Journal of the Fisheries Board of Canada*, 32(8):1369–1381.
- Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada*, 11(5):559–623.
- Sakia, R. (1992). The Box-Cox transformation technique: a review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(2):169–178.
- Sherman, R. P. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *The Annals of Statistics*, 22(1):439–459.
- Silverman, B. and Young, G. (1987). The bootstrap: to smooth or not to smooth? *Biometrika*, 74(3):469–479.
- Stinchcombe, M. B. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(3):295–325.
- Stute, W. (1982). The oscillation behavior of empirical processes. *The Annals of Probability*, 10(1):86–107.
- Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25(2):613–641.
- Stute, W. and Zhu, L.-X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*, 29(3):535–545.
- Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge university press.

- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and Empirical Processes*. Springer.
- Van Keilegom, I., González Manteiga, W., and Sánchez Sellero, C. (2008a). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, 17(2):401–415.
- Van Keilegom, I., Sellero, C. S., and Manteiga, W. G. (2008b). Empirical likelihood based testing for regression. *Electronic Journal of Statistics*, 2:581–604.
- Vanhems, A. and Van Keilegom, I. (2017). Semiparametric transformation model with endogeneity: a control function approach. *Econometric Theory (under revision)*.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372.
- Wang, Y.-J. (2000). Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics*, 98(1):27–46.
- Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959.
- Zellner, A. and Revankar, N. S. (1969). Generalized production functions. *The Review of Economic Studies*, 36(2):241–250.
- Zhang, C. M. (2003). Adaptive tests of regression functions via multiscale generalized likelihood ratios. *Canadian Journal of Statistics*, 31(2):151–171.