Non-Linear Mean Impact Analysis

Dem Fachbereich 03: Mathematik/Informatik der
Universität Bremen
zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

eingereichte Dissertation

von
Herrn Dipl.-Math. **Martin Scharpenberg**
geb. am 18.03.1988 in Warendorf

Erstgutachter: Prof. Dr. Werner Brannath
Zweitgutachter: Univ.-Prof. Mag. Dr. Andreas Futschik

Einreichung am: 29.04.2015
Verteidigung am: 30.06.2015

Acknowledgments

I would like to thank Prof. Dr. Werner Brannath for enabling me to write this thesis. I am grateful for his constructive words of advice, as well as his constant support during my studies.
I also would like to thank Univ.-Prof. Mag. Dr. Andreas Futschik for being the second reviewer of this thesis.

Special thanks go to my colleague Svenja for always having an open ear and time for fruitful discussions which considerably helped in the progress of this thesis.

Last but not least I would like to thank my wife Janina for supporting me in every condition of life.

Contents

Introduction

The interpretation and the validity of the results from linear regression rely on strong modeling assumptions (e.g. linearity of the conditional mean of Y given $X_1, ..., X_k$) which are known not to be satisfied in many cases. In order to overcome the problems in the interpretation of regression results Scharpenberg (2012) and Brannath and Scharpenberg (2014) introduced a new, population-based and generally non-linear measure of association called *mean impact*. The mean impact of an independent variable X on a target variable Y is defined as the maximum possible change in the mean of Y, when changing the density of X (in the population) in a suitably standardized way. Based on the mean impact further parameters, one of which is a non-linear measure for determination, were defined. There is also a natural extension to the case of multiple independent variables $X_1, ..., X_k$, where we are interested in quantifying the association between Y and X_1 corrected for possible associations driven by $X_2, ..., X_k$ (corresponding to multiple regression). However, Scharpenberg (2012) and Brannath and Scharpenberg (2014) point out that a restriction of the possible distributional disturbances is needed when estimating the mean impact in order to avoid overfitting problems. Therefore, they restrict themselves to functions linear in X. Doing so, they obtain conservative estimates for the mean impact and build conservative confidence intervals on their basis. Additionally, it is shown that this procedure leads to a new interpretation of linear regression coefficients under mean model miss specification.

The restriction to linear distributional disturbances seems very strict and the resulting estimates are often very conservative. The goal of this thesis is to move from linear distributional disturbances to non-linear ones. Doing so we expect to obtain less conservative estimates of the mean impact. Estimates as well as confidence intervals for the mean impact based on different non-linear regression techniques will be derived and their (asymptotical) behavior will be investigated in the course of this thesis. We will do this for the single independent variable case, as well as for the case of multiple independent variables.

The thesis is organized as follows: In the first section we present the theoretical foundations of the mean impact analysis. The main results of Scharpenberg (2012), including the theory for the (partial) linear mean impact (which is the mean impact where we restrict the set of distributional disturbances to linear functions), are presented as well as major improvements of the asymptotic normality results for the signed (partial) linear mean impact. Furthermore, the common mean impact of several variables $X_1, .., X_k$ on

a target variable Y is defined. Again restriction to linear disturbances is made resulting in the linear common mean impact. Also presented is the partial common mean impact which serves to quantify the common influence of a set of variables $X_1, .., X_k$ on a target variable Y which goes beyond the possible influence of a second set of variables $Q_1, ..., Q_l$. Again a restriction to linear functions is made. In a further step second order accurate bootstrap intervals are derived for the newly defined parameters. Furthermore, an alternative approach to the quantification of the influence of X_1 which goes beyond the possible influence of other covariates $X_2, ..., X_k$ is also introduced. In this approach this influence is defined as the difference of the common mean impact of all variables $X_1, ..., X_k$ and the common mean impact of $X_2, ..., X_k$. This difference can then be seen as the excess of dependence when adding X_1 to the set of covariates considered.

The second section deals with the relaxation of the restriction to linear functions in the single covariate case. We derive conservative estimates of the mean impact based on non-linear regression techniques like polynomial regression and kernel smoothers. Higher order local regression is also considered. Confidence intervals based on asymptotic normality results as well as bootstrap confidence intervals, for the mean impact based on non-linear regression techniques are derived.

In Section 3 we define partial mean impacts based on non-linear regression techniques, which allows us to quantify the influence of a single covariate X_1 on Y which goes beyond the possible influence of other covariates $X_2, ..., X_k$ in a more flexible way than in the linear partial mean impact setup. The non-linear regression techniques used include again polynomial regression and kernel smoothing. We extend the alternative approach to the quantification of partial influences of Section 2 to non-linear regression techniques.

In the last Section we present results from a simulation study in which we consider the coverage probability of the confidence intervals derived in this thesis. We also investigate the probability of exclusion of zero (i.e. the power) in cases where the mean impact is not equal to zero. The results of the non-linear mean impact analyses are compared to the linear mean impact analysis in order to evaluate the benefit (or the possible drawback) when moving from linear to non-linear impact analysis.

In the appendix a brief overview of the regression techniques and the bootstrap techniques which are used in this thesis as well as proofs which are left out in the course of the thesis are given.

1. Theoretical foundations - Impact analysis

In this section the main results of the impact analysis derived in Scharpenberg (2012) are given.

In classical regression analysis one tries to describe the dependency of a target variable Y from independent variables $X_1, ..., X_k$ (which we will call covariates in the sequel) by a probabilistic model. Since one usually interprets the results of regression analysis on an individual basis the regression model describes the distribution of Y of an individual in dependence on its covariate values. Interpreting the results in this individual-based manner implies that they depend only on the conditional distribution of Y given $X_1, ..., X_k$ and are independent of the marginal distribution of the covariates in the underlying population. Assumptions like linearity of the conditional mean of Y given $X_1, ..., X_k$ in the covariates or that no other covariates have an influence, which justify the individual-based way of interpretation do not generally hold. This means that the results of regression analysis may often depend on the marginal distribution of the covariates which can make the individual-based approach misleading.

In order to avoid this type of misinterpretation, Scharpenberg (2012) and Brannath and Scharpenberg (2014) introduce an approach in which one looks at changes in the distribution of the target variable across the population when the marginal distribution of the covariates is perturbed. The dependence of the results on the specific population and the way the population is perturbed are thereby acknowledged.

Scharpenberg (2012) first investigates the scenario of one covariate whose influence on the target variable is described. Later this approach is generalized to the case of several observed covariates where it is aimed to investigate the influence of one covariate on the target variable which goes beyond the possible influence of the other covariates. In this thesis we will explain the main idea in the context of the special case of one covariate. The results derived in Scharpenberg (2012) are only given for the general case, where the special case is carried along as an example since large parts of this thesis are constructed for this special case.

1.1. Mathematical presentation

As mentioned before, in order to introduce the idea of the new approach, we take a look at the influence of a single real valued covariate X onto a real valued target variable Y, where we assume that $Y, X \in L^2_{\mathbf{P}}$ and the distribution of (X, Y) has a density on $\mathbb{R}^2$ with respect to the Lebesgue-measure.

In contrast to classical regression analysis we are not directly looking at the influence

of a covariate X on the conditional mean $E(Y|X)$ of a target variable Y. We investigate how $E(Y)$ the marginal population mean of Y changes when the marginal distribution of X in the population is changed, instead. Let f and h be the marginal densities of X and Y. Let $h(Y|X)$ be the conditional density of Y given X. Since X and Y are independent if and only if $h(y|x) = h(y)$ for all x we obtain in the case of independence of X and Y that

$$\begin{aligned} E_f(Y) &= \iint h(y|x)f(x)y\,dxdy \overset{\substack{X \text{ and } Y \\ \text{independent}}}{=} \iint h(y)f(x)y\,dxdy \\ &= \int h(y)y\,dy. \end{aligned}$$

The last expression is independent of the marginal density f of X. Hence, the mean of Y does not depend on the density f of X which means that the question "Has X got an influence on Y?" leads to the question "Does the mean of Y change when the density of X is changed (in the population)?". These considerations suggest that the change of the mean of Y when changing the density of X in the population is a good indicator for the influence of X on Y. Define

$$E_{f_i}(Y) = \iint yh(y|x)f_i(x)\,dx\,dy$$

where f_i, $i = 1, 2$ are densities of X. Then the change of the mean of Y when the density of X is changed from f_1 to f_2 can be written as

$$\begin{aligned} \Delta E(Y) = E_{f_2}(Y) - E_{f_1}(Y) &= \iint yh(y|x)\{f_2(x) - f_1(x)\}\,dxdy \\ &= \iint yh(y|x)\delta(x)f_1(x)\,dxdy = E(Y\delta(x)) \end{aligned}$$

where $\delta(x) = \frac{f_2(x)-f_1(x)}{f_1(x)} = \frac{f_2(x)}{f_1(x)} - 1$. Such δ exists, according to the Radon-Nikodym theorem, if P_{f_2} is absolutely continuous with respect to P_{f_1}, where P_{f_i} is the measure with Lebesgue-density f_i, $i = 1, 2$ (cf. Klenke, 2008, p159).

The key quantity of the new approach, which is called "Mean Impact Analysis (MImA)" in Scharpenberg (2012), is the *mean impact* of a covariate X on Y

$$\iota_Y(X) = \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R}):\, E_{\mathbf{P}}\{\delta(X)\}=0, E_{\mathbf{P}}\{\delta^2(X)\}=1} E_{\mathbf{P}}\{Y\delta(X)\}.$$

It "describes the maximum change in the mean of Y when the density f of X (in the population) is changed to $(1 + \delta(x))f(x)$ in a way that δ is $L^2_{\mathbf{P}}(\mathbb{R})$-integrable with

norm equal to 1" (Scharpenberg, 2012, p. 20). One can see with the help of Cauchy´s inequality, that the mean impact is bounded by the standard deviation $\sqrt{Var_{\mathbf{P}}(Y)}$ of Y. Note that the name *mean impact* might be misleading, since we do not describe causal influences. The mean impact is rather a measure of association.

1.2. Partial mean impact

In all considerations of this section we assume $Y, X_1, ..., X_k \in L^2_{\mathbf{P}}$. One can generalize the concept of the mean impact analysis to the case were we consider more than one covariate in order to investigate the influence of X_1 on the target variable Y which goes beyond the influence of the other covariates $X_2, ..., X_k$. Similar to the univariate case perturbations of the distribution of the covariates in the population are considered and one has a look at the change of the mean of Y. One only regards perturbations that leave the means of $X_2, ..., X_k$ unchanged in order to account for the potential influence of other covariates than X_1.

1.2.1. General approach

The k regarded covariates are denoted by $X_1, ..., X_k$ and $\mathbf{X} = (X_1, ..., X_k)$ is the vector of the covariates. Given this set of covariates one is interested in the question if a covariate e.g. X_1 has influence on Y beyond the (potential) influence of $X_2, ..., X_k$. This question is answered by estimating the regression coefficient for X_1 of the multiple regression model in the theory of linear models. The regression coefficient shows how the conditional expectation $E_{\mathbf{P}}(Y|\mathbf{X})$ changes, when X_1 is changed and the other covariates are fixed.

In the new, population-based approach Scharpenberg (2012) defines another quantity to characterize the influence of X_1 on Y going beyond the influence of $X_2, ..., X_k$. This quantity is called the *partial mean impact* of X_1 on Y and is defined as

$$\iota_{X_1}(Y|X_2, ..., X_k) = \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k): \delta(\mathbf{X}) \in \mathcal{H}_2^{\perp}, E_{\mathbf{P}}\{\delta^2(\mathbf{X})\}=1} E_{\mathbf{P}}\left\{Y\delta(\mathbf{X})\right\}, \tag{1.1}$$

where $\mathcal{H}_2 = span(1, X_2, X_3, ..., X_k) \subseteq L^2_{\mathbf{P}}$.

"The partial mean impact describes the maximum change in the mean of Y when the density f of $X_1, ..., X_k$ (in the population) is changed to $(1 + \delta)f$ in a way that δ is $L^2_{\mathbf{P}}(\mathbb{R}^k)$-integrable with norm equal to one and the means of the other covariates $X_2, ..., X_k$ are not changed" (Scharpenberg, 2012, p. 54).

With $P_{\mathcal{H}_2^\perp}$ being the orthogonal projection onto $\mathcal{H}_2^\perp$ we obtain for $\delta(\mathbf{X}) \in \mathcal{H}_2^\perp$ with $E_{\mathbf{P}}(\delta^2(\mathbf{X})) = 1$ that:

$$E_{\mathbf{P}}(Y\delta(\mathbf{X})) = E_{\mathbf{P}}(Z\delta(\mathbf{X})) \overset{\text{Cauchy}}{\leq} \sqrt{E_{\mathbf{P}}(Z^2)}\sqrt{E_{\mathbf{P}}(\delta^2(\mathbf{X}))} = \sqrt{Var_{\mathbf{P}}(Z)}$$

where $Z = P_{\mathcal{H}_2^\perp} Y$ and the last equation follows from $E_{\mathbf{P}}(Z) = E_{\mathbf{P}}(Z \cdot 1) = 0$. In the single-covariate case we have $\mathcal{H}_2 = \text{span}(1)$ and $P_{\mathcal{H}_2^\perp} X = X - E_{\mathbf{P}}(X)$.

The following theorem implies that the supremum in (1.1) is always attained. Therefore, we could write max instead of sup in (1.1).

Theorem 1.1. *Let $Y \in L^2_{\mathbf{P}}$. The partial mean impact $\iota_{X_1}(Y|X_2, ..., X_k)$ of X_1 on Y is equal to*

(1) the upper bound $\sqrt{Var_{\mathbf{P}}(Z)}$ if and only if $Y = g(\boldsymbol{X})$ for a measurable function $g : \mathbb{R}^k \to \mathbb{R}$,

(2) $\sqrt{Var_{\mathbf{P}}\{P_{\mathcal{H}_2^\perp} g(X)\}}$ if $Y = g(X) + \epsilon$ where ϵ is a square integrable random variable with mean $E_{\mathbf{P}}(\epsilon) = 0$ which is independent of $\boldsymbol{X}$,

(3) 0 if and only if $E_{\mathbf{P}}(Y|\boldsymbol{X}) \in \mathcal{H}_2$,

(4) if $\iota_{X_1}(Y|X_2, ..., X_k) \neq 0$, then $\iota_{X_1}(Y|X_2, ..., X_k) = E\{Y\hat{\delta}(\boldsymbol{X})\}$ where

$$\hat{\delta}(\boldsymbol{X}) = P_{\mathcal{H}_2^\perp} E_{\mathbf{P}}(Y|\boldsymbol{X}) / \sqrt{Var_{\mathbf{P}}\{P_{\mathcal{H}_2^\perp} E_{\mathbf{P}}(Y|\boldsymbol{X})\}}.$$

In the singe-covariate case this theorem simplifies in the following way.

Theorem 1.2. *Let $Y \in L^2_{\mathbf{P}}$. The mean impact $\iota_X(Y)$ of X on Y is equal to*

(1) the upper bound $\sqrt{Var_{\mathbf{P}}(Y)}$ if and only if $Y = g(X)$ for a measurable function $g : \mathbb{R} \to \mathbb{R}$, i.e., Y depends on X in a deterministic way.

(2) $\sqrt{Var_{\mathbf{P}}\{g(X)\}}$ if $Y = g(X) + \epsilon$ where ϵ is a square integrable random variable with mean $E_{\mathbf{P}}(\epsilon) = 0$ which is independent of X.

(3) 0 if and only if $E_{\mathbf{P}}(Y|X) = E_{\mathbf{P}}(Y)$ almost surely.

(4) if $\iota_X(Y) \neq 0$, then $\iota_X(Y) = E_{\mathbf{P}}\{Y\hat{\delta}(X)\}$ where

$$\hat{\delta}(X) = \left[E_{\mathbf{P}}(Y|X) - E_{\mathbf{P}}(Y)\right] / \sqrt{Var_{\mathbf{P}}\{E_{\mathbf{P}}(Y|X)\}}.$$

and the sign of $s\iota_X(Y)$ is the sign of $Cor\{X, E_{\mathbf{P}}(Y|X)\}$.

Assume that the covariates have Lebesgue-density f. In the definition of the *partial mean impact* (1.1), $1+\delta$ is the factor which we have to multiply to the density f of $\mathbf{X}$ in the population to obtain the "new" density to which f is changed. By maximizing over all $\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k)$ it is possible that the resulting density $f(1+\delta)$ becomes negative at some points. Since a density has to be non-negative one should only regard those $\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k)$, for which $f(\mathbf{X})(1+\delta(\mathbf{X})) \geq 0$. In Scharpenberg (2012) not exaclty this result is shown, but it is shown that there is a sequence δ_n of measurable functions that are asymptotically orthogonal to $\mathcal{H}_2$ and for which $E_{\mathbf{P}}(Y\delta_n(\mathbf{X}))/\sqrt{E_{\mathbf{P}}(\delta_n^2(\mathbf{X}))} \underset{n\to\infty}{\to} \iota_{X_1}(Y|X_2,...,X_k)$.

Theorem 1.3. *There is a sequence $\delta_n(\boldsymbol{X})$ with $(1+\delta_n(\boldsymbol{X}))f(\boldsymbol{X}) \geq 0$ and $E_{\mathbf{P}}(\delta_n(\boldsymbol{X})) = 0$ for all n, $E_{\mathbf{P}}(X_j\delta_n(\boldsymbol{X}))/\sqrt{E_{\mathbf{P}}(\delta_n^2(\boldsymbol{X}))} \underset{n\to\infty}{\to} 0$ for all $j = 2,...,k$ and $E_{\mathbf{P}}(Y\delta_n(\boldsymbol{X}))/\sqrt{E_{\mathbf{P}}(\delta_n^2(\boldsymbol{X}))} \underset{n\to\infty}{\to} \iota_{X_1}(Y|X_2,...,X_k)$.*

In the single-covariate case the desired stronger version of this theorem holds. This means that we have:

Theorem 1.4. *We have that*

$$\iota_X(Y) = \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R}): E_{\mathbf{P}}(\delta(X))=0, f(X)(1+\delta(X))\geq 0} E_{\mathbf{P}}(Y\delta(X))/\sqrt{E_{\mathbf{P}}(\delta^2(X))}.$$

From the definition of the partial mean impact (1.1) follows that it only accounts for linear influences of the covariates $X_2,...,X_k$. Due to this, it is possible that the partial mean impact is positive although Y does not depend on X_1. The following expample illustrates this.

Example 1.5. *Let $Y = \theta_0 + \theta_1 X_2 + \theta_2 X_2^2 + \epsilon$ where $X_2 \sim N(0,1)$ and $\epsilon \sim N(0,1)$ are stochastically independent and $\theta_l \neq 0$ for $l = 0,1,2$. Then we have according to Theorem 1.1 with $\mathcal{H}_2 = span(1, X_2)$*

$$\begin{aligned}\iota_{X_1}(Y|X_2) &= \sqrt{Var_{\mathbf{P}}\{P_{\mathcal{H}_2^\perp}(\theta_0 + \theta_1 X_2 + \theta_2 X_2^2)\}} = \sqrt{Var_{\mathbf{P}}\{P_{\mathcal{H}_2^\perp}(\theta_2 X_2^2)\}} \\ &= \sqrt{Var_{\mathbf{P}}\{(\theta_2\{X_2^2 - E_{\mathbf{P}}(X_2^2)\})\}} \\ &= |\theta_2|\sqrt{Var_{\mathbf{P}}(X_2^2)} > 0.\end{aligned}$$

Hence, $\iota_{X_1}(Y|X_2) \neq 0$ although X_1 and Y are independent.

One possible way to account for non linear influences of the covariates is to add X_j^2 to the set of covariates for all $j = 2,...,k$ (this procedure accounts for quadratic influences). To account for the influences of all measurable transformations of the covariates

$X_2, ..., X_k$ one would have to demand $E_{\mathbf{P}}(\delta(\mathbf{X})g(X_j)) = 0$ for all measurable g and all $j = 2, ..., k$ in the definition of the partial mean impact. This approach leads to a complex statistical problem and is not followed up by Scharpenberg (2012).

It can be shown that the perturbation δ leading to the impact is almost surely uniquely determined.

Theorem 1.6. *If $\iota_{X_1}(Y|X_2, ..., X_k) > 0$ then the perturbation $\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k)$ for which $E_{\mathbf{P}}(\delta(\boldsymbol{X})) = 0$, $E_{\mathbf{P}}(\delta^2(\boldsymbol{X})) = 1$ and $E_{\mathbf{P}}(Y\delta(\boldsymbol{X})) = \iota_{X_1}(Y|X_2, ..., X_k)$ is $\mathbf{P}$ almost surely uniquely determined.*

Note, that the partial mean impact (as well as the mean impact in the single-covariate case) is by definition always non-negative. Hence, the partial mean impact does not give any hint in which direction the change in the distribution of X changes the mean of Y. In order to be able to indicate the direction of the change the so called *signed partial mean impact* is defined by

$$s\iota_{X_1}(Y|X_2, ..., X_k) = \text{sign}(E_{\mathbf{P}}\{X_1\delta_0(\mathbf{X})\})\iota_{X_1}(Y|X_2, ..., X_k)$$

where $\delta_0 \in L^2_{\mathbf{P}}(\mathbb{R}^k)$ is such that $\delta_0(\mathbf{X}) \in \mathcal{H}_2^{\perp}$, $E_{\mathbf{P}}\left\{\delta_0^2(\mathbf{X})\right\} = 1$ and

$$\iota_{X_1}(Y|X_2, ..., X_k) = E_{\mathbf{P}}\{Y\delta_0(\mathbf{X})\}.$$

It is possible that the signed partial mean impact equals zero, although the partial mean impact is non-negative which happens when $E_{\mathbf{P}}(X_1\delta_0(\mathbf{X})) = 0$. Since this hints to a non-linear relationship between Y and X_1 one could consider a non-linear transformation $T(X_1)$ of X_1 and regard the signed partial mean impact for $T(X_1)$ in order to describe the influence of X_1 in a better way. Analogous to the signed partial mean impact, the signed mean impact is given by

$$s\iota_X(Y) = \text{sign}\left(E_{\mathbf{P}}\{X\delta_0(X)\}\right)\iota_X(Y)$$

in the single covariate case. Note that $(E_{\mathbf{P}}\{X_1\delta_0(\mathbf{X})\})$ indicates by which amount the mean of X_1 is changed with the disturbance $\delta_0(\mathbf{X})$ that maximizes the change of the mean of Y.

Another quantity which is based on the partial mean impact is the *partial mean slope*.

It is given by

$$\theta_{X_1}(Y|X_2,...,X_k) = \iota_{X_1}(Y|X_2,...,X_k)/E_{\mathbf{P}}\{X_1\delta_0(\mathbf{X})\}$$

if $\iota_{X_1}(Y|X_2,...,X_k) > 0$ and $E_{\mathbf{P}}(X_1\delta_0(X)) \neq 0$, where $\delta_0 \in L^2_{\mathbf{P}}(\mathbb{R}^k)$, $\delta_0(\mathbf{X}) \in \mathcal{H}_2^{\perp}$, $E_{\mathbf{P}}(\delta_0(\mathbf{X})^2) = 1$ and $\iota_{X_1}(Y|X_2,...,X_k) = E_{\mathbf{P}}(Y\delta_0(\mathbf{X}))$. Note that for $\iota_{X_1}(Y|X_2,...,X_k) > 0$ with δ_0 the partial mean slope is also uniquely determined. If $\iota_{X_1}(Y|X_2,...,X_k) = 0$ it is defined to be zero. It gives the amount the mean of Y changes if the mean of X_1 is changed (without changing the mean of the other covariates) by one unit.

Theorem 1.7. *If $Y = \theta_0 + \sum_{j=1}^k \theta_j X_j + \epsilon$ where $(X_1,...,X_k)$ and ϵ are independent and $E_{\mathbf{P}}(\epsilon) = 0$ then the partial mean slope is $\theta_{X_1}(Y|X_2,...,X_k) = \theta_1$ and the partial mean impact is $\iota_{X_1}(Y|X_2,...,X_k) = |\theta_1|\sqrt{E_{\mathbf{P}}\{(P_{\mathcal{H}_2^{\perp}}X_1)^2\}}$. The signed partial mean impact is $s\iota_{X_1}(Y|X_2,...,X_k) = \theta_1\sqrt{E_{\mathbf{P}}\{(P_{\mathcal{H}_2^{\perp}}X_1)^2\}}$.*

In the single-covariate case this theorem reduces to:

Theorem 1.8. *If $Y = \theta_0 + \theta_1 X + \epsilon$ where X and ϵ are independent and $E_{\mathbf{P}}(\epsilon) = 0$ then the mean slope is $\theta_X(Y) = \theta_1$ and the mean impact is $\iota_X(Y) = |\theta_1|\sqrt{Var_{\mathbf{P}}(X)}$. The signed mean impact is $s\iota_X(Y) = \theta_1\sqrt{Var_{\mathbf{P}}(X)}$.*

Hence, in the case of an underlying linear model the new parameters, (partial) mean slope and signed (partial) mean impact are closely related to the coefficients of this model. This relationship between impact analysis and linear regression in the case of the regression model to be true is faced again when considering the asymptotic distribution of the estimators which will be derived later. In the single covariate case we define, additionally to the new parameters above, the *population coefficient for determination*, which is given by

$$R^2_{\mathbf{P}}(X) = \frac{\iota^2_X(Y)}{Var_{\mathbf{P}}(Y)}. \tag{1.2}$$

Note that the population coefficient for determination is equal to Pearson´s correlation ratio given in Doksum and Samarov (1995). A partial population coefficient for determination will be introduced in Section 1.9.4.

1.2.2. Restricted and linear partial mean impact

There may be reasons to restrict the set of perturbations δ of the density $f(X)$ in definition (1.1) of the partial mean impact of X_1 on Y. We will see later that estimation and testing will require restrictions, otherwise we obtain meaningless results due to the

problem of overfitting. This leads to the following general definition of the *restricted partial mean impact*. Let $\mathcal{R}$ be a closed subset of $L^2_{\mathbf{P}}(\mathbb{R}^k)$. We define for $\mathcal{R}$ the restricted partial mean impact as

$$\iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k) = \sup_{\delta\in\mathcal{R}:\delta(\mathbf{X})\in\mathcal{H}_2^{\perp},E_{\mathbf{P}}\{\delta^2(\mathbf{X})\}=1} E_{\mathbf{P}}\{Y\delta(\mathbf{X})\}$$

where $\mathcal{H}_2 = \text{span}(1, X_2, X_3, ..., X_k)$. Restriction to a linear subspace $\mathcal{R}$ leads again always to a non-negative number because with δ also $-\delta$ belongs to $\mathcal{R}$.

When regarding the special set of perturbations $\mathcal{R}_X = \{h(\mathbf{X}) = a_0 + \sum_{j=1}^k a_j X_j : a_i \in \mathbb{R}\} \subseteq L^2_{\mathbf{P}}(\mathbb{R}^k)$ one obtains the so called *linear partial mean impact*

$$\iota^{lin}_{X_1}(Y|X_2,...,X_k) = \iota^{\mathcal{R}_X}_{X_1}(Y|X_2,...,X_k).$$

It "describes the maximum change in the mean of Y when the density f of $X_1, ..., X_k$ (in the population) is to $(1 + \delta)f$ a way that δ is linear in $(1, X_1, ..., X_k)$, $L^2_{\mathbf{P}}(\mathbb{R}^k)$-integrable with norm equal to one and the means of the other covariates $X_2, ..., X_k$ are not changed" (cf. Scharpenberg, 2012, p. 61). Since the partial mean impact is defined as the supremum over all perturbations of the density of the covariates, every restriction of the set of perturbation leads to a smaller (restricted) impact than the unrestricted impact (1.1). Consequently, $\iota^{lin}_{X_1}(Y|X_2,...,X_k)$ is a lower bound for the unrestricted partial impact $\iota_{X_1}(Y|X_2,...,X_k)$ and consistent estimates and one-sided tests for $\iota^{lin}_{X_1}(Y|X_2,...,X_k)$ with control of the type I error rate will be conservative with regard to the unrestricted partial impact $\iota_{X_1}(Y|X_2,...,X_k)$.

Proposition 1.9. *We have* $\iota^{lin}_{X_1}(Y|X_2,...,X_k) = \left| E_{\mathbf{P}}\left(Y \frac{P_{\mathcal{H}_2^{\perp}} X_1}{\sqrt{E_{\mathbf{P}}((P_{\mathcal{H}_2^{\perp}} X_1)^2)}} \right) \right|$.

Similar to the unrestricted partial mean impact, a signed version for the restricted partial mean impact can be defined

$$s\iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k) = \text{sign}(E_{\mathbf{P}}\{X_1\delta_0(\mathbf{X})\})\iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k)$$

where $\delta_0 \in \mathcal{R}$ with $\delta_0(\mathbf{X}) \in \mathcal{H}_2$ and $E_{\mathbf{P}}(\delta_0^2(\mathbf{X})) = 1$ is the unique disturbance with $E_{\mathbf{P}}(Y\delta_0(\mathbf{X})) = \iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k)$.

Lemma 1.10. *We have*

$$s\iota^{lin}_{X_1}(Y|X_2,...,X_k) = E_{\mathbf{P}}\left(Y \frac{P_{\mathcal{H}_2^{\perp}} X_1}{\sqrt{E_{\mathbf{P}}((P_{\mathcal{H}_2^{\perp}} X_1)^2)}} \right).$$

In the previous section we mentioned that it can be desirable, in order to account for the influence of all measurable transformations of the covariates $X_2, ..., X_k$, to demand that

$$E_{\mathbf{P}}(\delta(\mathbf{X})g(X_j)) = 0$$

for all measurable functions g and all $j = 2, ..., k$ in the definition of the partial mean impact. For the δ from the linear mean impact we can show the following result.

Proposition 1.11. *If* $E_{\mathbf{P}}(X_1|X_2, ..., X_k) = \xi_1 + \sum_{j=2}^{k} \xi_j X_j$ *for suitable* $\xi_j \in \mathbb{R}$ *and* $\delta(\boldsymbol{X}) = \frac{P_{\mathcal{H}_2^\perp} X_1}{\sqrt{E_{\mathbf{P}}((P_{\mathcal{H}_2^\perp} X_1)^2)}}$ *then we have*

$$E_{\mathbf{P}}\{\delta(\boldsymbol{X})g(X_j)\} = 0$$

for all measurable functions g *and all* $j = 2, ..., k$.

Hence, when the conditional mean of X_1 given the other covariates is a linear function of those covariates, the (signed) linear partial mean impact accounts for the influence of all measurable transformations of $X_2, ..., X_k$.

The *linear partial mean slope* is defined as

$$\theta_{X_1}^{lin}(Y|X_2, ..., X_k) = \iota_{X_1}^{lin}(Y|X_2, ..., X_k)/E_{\mathbf{P}}(\delta_0(\mathbf{X})X_1)$$

where $\delta_0 \in L^2_{\mathbf{P}}(\mathbb{R}^k)$ with $\delta_0(\mathbf{X}) \in \mathcal{H}_2^\perp$, $E_{\mathbf{P}}(\delta_0^2(\mathbf{X})) = 1$ and

$$\iota_{X_1}^{lin}(Y|X_2, ..., X_k) = E_{\mathbf{P}}(\delta_0(\mathbf{X})Y).$$

Proposition 1.12. $\theta_{X_1}^{lin}(Y|X_2, ..., X_k)$ *equals the coefficient for* X_1 *in the orthogonal projection of* Y *onto* $\mathcal{H} = \text{span}(1, X_1, X_2, ..., X_k)$, *i.e. when* $P_{\mathcal{H}}Y = \theta_0 + \sum_{j=1}^{k} \theta_j X_j$ *then* $\theta_{X_1}^{lin}(Y|X_2, ..., X_k) = \theta_1$.

Hence, in the case of a linear model $E_{\mathbf{P}}(Y|X_1, ..., X_k) = \theta_0 + \sum_{j=1}^{k} \theta_j X_j$ the linear partial mean slope is the regression coefficient θ_1.

By Theorem 1.1 we have $\delta_0(\mathbf{X}) = P_{\mathcal{H}_2^\perp} X_1 / \sqrt{Var_{\mathbf{P}} P_{\mathcal{H}_2^\perp} X_1}$. Together with the fact that the linear partial mean impact $\iota_{X_1}^{lin}(Y|X_2, ..., X_k)$ is a lower bound for the unrestricted partial mean impact $\iota_{X_1}(Y|X_2, ..., X_k)$ we obtain that $|\theta_{X_1}^{lin}(Y|X_2, ..., X_k)|$ is a lower bound for the absolute value of the unrestricted partial mean impact $|\theta_{X_1}(Y|X_2, ..., X_k)|$. Therefore, tests for the hypothesis $H_0 : |\theta_{X_1}^{lin}(Y|X_2, ..., X_k)| \leq v$ for $v > 0$ with control

of the type I error rate will be conservative for $H_0' : |\theta_{X_1}(Y|X_2, ..., X_k)| \leq v$.

In the single-covariate case we can write the linear versions of the parameters as

$$\iota_X^{lin}(Y) = \left| E_{\mathbf{P}} \left(Y \frac{X - E_{\mathbf{P}}(X)}{\sqrt{Var_{\mathbf{P}}(X)}} \right) \right|, \; s\iota_X^{lin}(Y) = E_{\mathbf{P}} \left(Y \frac{X - E_{\mathbf{P}}(X)}{\sqrt{Var_{\mathbf{P}}(X)}} \right)$$

and

$$\theta_X^{lin}(Y) = E_{\mathbf{P}} \left(Y \frac{X - E_{\mathbf{P}}(X)}{Var_{\mathbf{P}}(X)} \right).$$

1.3. Examples

In this section we give the values of $\iota_X(Y)$, $s\iota_X(Y)$ and $\theta_X(Y)$ in the case where $Y = g(X) + \epsilon$ for a square integrable random variable ϵ with mean $E(\epsilon) = 0$ which is independent of X. Obviously $\iota_X(Y)$ and $\theta_X(Y)$ depend on $g(X)$ and the distribution of X, $\mathcal{L}(X)$. In the following we consider a specific $g(X)$ and $\mathcal{L}(X)$ and compute the resulting $\iota_X(Y)$ and $\theta_X(Y)$. The example presented here originates from Scharpenberg (2012), more examples can be found there. Let $\mathcal{L}(X) = N(\mu, \sigma^2)$ and $g(X) = ae^X$ for $a \neq 0$. Then we have

$$\iota_X(Y) = |a| \sqrt{e^{2(\mu+\sigma^2)} - e^{2(\mu+\frac{\sigma^2}{2})}} = |a| e^{\mu} e^{\sigma^2/2} \sqrt{e^{\sigma^2} - 1}.$$

Furthermore, it can be shown that

$$\theta_X(Y) = a e^{\mu} e^{\sigma^2/2} (e^{\sigma^2} - 1)/\sigma^2$$

and

$$s\iota_X(Y) = \text{sign}(E_{\mathbf{P}}\{X\delta(X)\}) \iota_X(Y) = a e^{\mu} e^{\sigma^2/2} \sqrt{e^{\sigma^2} - 1}.$$

For the linear versions of the parameters we obtain

$$s\iota_X^{lin}(Y) = a\sigma e^{\mu + \frac{\sigma^2}{2}},$$

which implies

$$\iota_X^{lin}(Y) = |s\iota_X^{lin}(Y)| = |a| \sigma e^{\mu + \frac{\sigma^2}{2}}$$

and

$$\theta_X^{lin}(Y) = |a| e^{\mu + \frac{\sigma^2}{2}}.$$

The following table presents values of $\iota_X(Y)$, $\iota_X^{lin}(Y)$, $s\iota_X(Y)$, $s\iota_X^{lin}(Y)$, $\theta_X(Y)$ and $\theta_X^{lin}(Y)$ for $a = 1$ and different μ and σ^2:

μ	σ^2	$\iota_X(Y)$	$\iota_X^{lin}(Y)$	$s\iota_X(Y)$	$s\iota_X^{lin}(Y)$	$\theta_X(Y)$	$\theta_X^{lin}(Y)$
0	1	2.161	1.649	2.161	1.649	2.833	1.649
-1	1	0.795	0.607	0.795	0.607	1.042	0.607
1	1	5.875	4.482	5.875	4.482	7.701	4.482
1	0.25	1.642	1.540	1.642	1.540	3.499	3.080

Table 1: Parameter values which are used in Figure 1

We can see that the absolute mean slope as well as the linear absolute mean slope are less dependent on the variance σ^2 than their mean impact counterparts. Figure 1 presents the graph of $g(X) = e^X$, the densities of different normal distributions and a straight line with slope $\theta_X(Y)$ which crosses the point
$(E_{\mathbf{P}}(X), g(E_{\mathbf{P}}(X)))$.

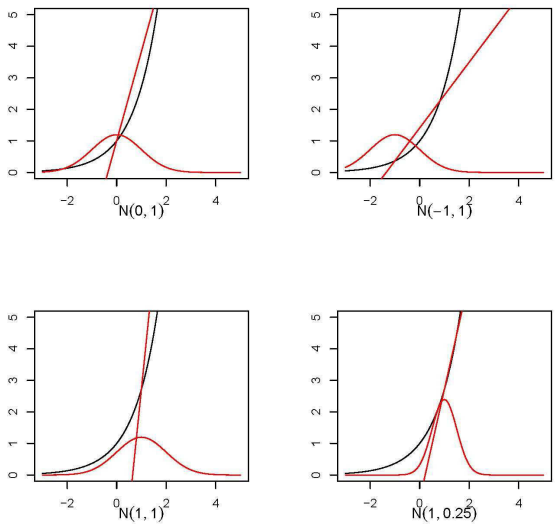

Figure 1: Behavior of the absolute mean slope for $g(X) = e^X$ and different normal distributions for X.

Figure 1 suggests, that for $\sigma^2 \to 0$ the mean slope $\theta_X(Y)$ will converge to the derivative

of $g(X)$ in the point $x = \mu = E_{\mathbf{P}}(X)$ which indeed is the case.

1.4. Estimation of the partial mean impact

We now deal with the estimation of the partial mean impact and the other new parameters. To this end we consider observations $(Y_i, X_{i1}, ..., X_{ik})$, $i = 1, ..., n$. The most intuitive way of estimating $\iota_{X_1}(Y|X_2, ..., X_k)$ is using the estimator

$$\hat{\iota}_{X_1}(Y|X_2, ..., X_k) = \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k): \delta(\mathbf{x}) \perp \mathbf{X}_i \, i=2,...,n, \frac{1}{n}\|\delta(\mathbf{x})\|^2_{\mathbb{R}^n}=1, \delta(\mathbf{x}) \perp \mathbf{1}} \frac{1}{n} \mathbf{Y}^T \delta(\mathbf{X}),$$

where $\|\mathbf{a}\|_{\mathbb{R}^n} = \sqrt{\sum_{i=1}^n a_i^2}$ for $a \in \mathbb{R}^n$ is the euclidean norm on $\mathbb{R}^n$, $\mathbf{X}_j = (x_{1j}, ..., x_{nj})^T$, $\mathbf{x} = (\mathbf{X}_1, ..., \mathbf{X}_k)$, $\delta(\mathbf{x}) = (\delta(X_{11}, ..., X_{1k}), ..., \delta(X_{n1}, ..., X_{nk}))^T$, $\mathbf{Y} = (y_1, ..., y_n)^T$ and $\mathbf{1}$=$(1, ..., 1)^T$. Here $\perp$ means orthogonality in $\mathbb{R}^n$, hence $\delta(\mathbf{x}) \perp \mathbf{X}_i \Leftrightarrow \sum_{j=1}^n (\delta(\mathbf{x}))_j (\mathbf{X}_i)_j = 0$. As we will show next this way of estimating the impact leads to overfitting. With $\mathcal{M}_2 = span(\mathbf{1}, \mathbf{X}_2, ...\mathbf{X}_k)$ the linear subspace of $\mathbb{R}^n$ spanned by the observation vectors $\mathbf{X}_2, ..., \mathbf{X}_k$ and the assumption that the observation vector Y does not belong to $\mathcal{M}_2$ we obtain for

$$\delta(\mathbf{x}) = \frac{\hat{\mathbf{Z}}}{\frac{1}{\sqrt{n}}\|\hat{\mathbf{Z}}\|_{\mathbb{R}^n}}$$

where $P_{\mathcal{M}_2^\perp} \mathbf{Y} = \hat{\mathbf{Z}} = (\hat{Z}_1, ..., \hat{Z}_n)$ that $\delta \in L^2_{\mathbf{P}}(\mathbb{R}^k)$. Furthermore,

$$\begin{aligned} \frac{1}{n}\|\delta(\mathbf{x})\|^2_{\mathbb{R}^n} &= \frac{\frac{1}{n}\|\hat{\mathbf{Z}}\|^2_{\mathbb{R}^n}}{\frac{1}{n}\|\hat{\mathbf{Z}}\|^2_{\mathbb{R}^n}} = 1 \\ \delta(\mathbf{x}) &= \frac{\hat{\mathbf{Z}}}{\frac{1}{\sqrt{n}}\|\hat{\mathbf{Z}}\|_{\mathbb{R}^n}} \in \mathcal{M}_2^\perp \\ \Rightarrow \delta(\mathbf{x}) &\perp \mathbf{1}, \delta(\mathbf{x}) \perp \mathbf{X}_i \, i = 2, ..., n. \end{aligned}$$

Therefore,

$$\hat{\iota}_{X_1}(Y|X_2, ..., X_k) \geq \frac{1}{n} \mathbf{Y}^T \frac{\hat{\mathbf{Z}}}{\frac{1}{\sqrt{n}}\|\hat{\mathbf{Z}}\|_{\mathbb{R}^n}} = \frac{1}{\sqrt{n}}\|\hat{\mathbf{Z}}\|_{\mathbb{R}^n}.$$

Since $\|\hat{\mathbf{Z}}\|_{\mathbb{R}^n} > 0$ for $\mathbf{Y} \notin \mathcal{M}_2$, a positive impact of X_1 on Y could always be found by using the estimator $\hat{\iota}_{X_1}(Y|X_2, ..., X_k)$, even when $\iota_{X_1}(Y|X_2, ..., X_k) = 0$. Therefore, using this estimator leads to meaningless results.

One can avoid the problem of overfitting by restricting the set of functions for δ and estimate restricted partial mean impacts. We consider the special case of linear functions and use the estimator

$$\hat{\iota}^{lin}_{X_1}(Y|X_2,...,X_k) = \sup_{\delta(\mathbf{x})=a_0\mathbf{1}+a_1\mathbf{X}_1+...+a_k\mathbf{X}_k, \delta(\mathbf{x})\in\mathcal{M}_2^\perp, \frac{1}{n}\|\delta(\mathbf{x})\|^2_{\mathbb{R}^n}=1} \frac{1}{n}\mathbf{Y}^T\delta(\mathbf{x}).$$

One can show that, with $P_{\mathcal{M}_2^\perp}\mathbf{X}_1 = \hat{\mathbf{U}} = (\hat{U}_1,...,\hat{U}_n)$, the estimator for the linear partial mean impact can be written as

$$\begin{aligned}\hat{\iota}^{lin}_{X_1}(Y|X_2,...,X_k) &= \left|\frac{1}{n}\mathbf{Y}^T\frac{\hat{\mathbf{U}}}{\sqrt{\frac{1}{n}\|\hat{\mathbf{U}}\|^2_{\mathbb{R}^n}}}\right| \\ &= |\hat{\theta}_1|\frac{1}{\sqrt{n}}\|\hat{\mathbf{U}}\|_{\mathbb{R}^n}\end{aligned}$$

where $\hat{\theta}_1$ is the least squares estimator of the coefficient θ_1 in the multivariate linear regression model. The second equation is valid due to the fact that the least squares estimator of the regression coefficient θ_1 can be estimated by the least squares estimator of a simple linear regression model with Y as dependent and the residual vector $P_{\mathcal{M}_2^\perp}\mathbf{X}_1$ as independent variable.

Analogously to this the signed linear partial mean impact can be estimated by

$$\begin{aligned}\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k) &= \frac{1}{n}\mathbf{Y}^T\frac{\hat{\mathbf{U}}}{\sqrt{\frac{1}{n}\|\hat{\mathbf{U}}\|^2_{\mathbb{R}^n}}} \\ &= \hat{\theta}_1\frac{1}{\sqrt{n}}\|\hat{\mathbf{U}}\|_{\mathbb{R}^n}.\end{aligned}$$

Hence, estimating the linear signed impact of X_1 on Y leads to a scaled version of the coefficient from a multiple linear regression. For the estimators of the parameters from the single-covariate setup we have

$$\hat{\iota}^{lin}_X(Y) = \left|\frac{1}{n}\sum_{i=1}^n \frac{Y_i(X_i-\bar{X})}{\sqrt{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2}}\right| = |\hat{\theta}_1|\sqrt{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2},$$

$$\hat{s\iota}^{lin}_X(Y) = \frac{1}{n}\sum_{i=1}^n \frac{Y_i(X_i-\bar{X})}{\sqrt{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2}} = \hat{\theta}_1\sqrt{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2},$$

and

$$\hat{\theta}_X^{lin}(Y) = \sum_{i=1}^{n}(X_i - \bar{X})Y_i / \sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Here $\hat{\theta}_1$ is the regression coefficient from the univariate regression model.

1.4.1. Asymptotic normality and hypothesis testing

Let $(Y_i, X_{i1}, ..., X_{ik})$, $i = 1, ..., n$, be i.i.d. observations with the same multivariate distribution as the real random variables $Y, X_1, ..., X_k \in L^2_{\mathbf{P}}$. In order to simplify the notation we write $\mathbf{Y} = (Y_1, ..., Y_n)$ and $\mathbf{X}_j = (X_{1j}, ..., X_{nj})$ for $j = 1, ..., k$. In application of the theory derived before, one surely will be interested in testing for $v \in \mathbb{R}$ the one-sided hypothesis

$$H_0 : s\iota_{X_1}^{lin}(Y|X_2, ..., X_k) \leq v \quad vs. \quad H_1 : s\iota_{X_1}^{lin}(Y|X_2, ..., X_k) > v \tag{1.3}$$

or for $v \geq 0$ the hypothesis

$$H_0 : \iota_{X_1}^{lin}(Y|X_2, ..., X_k) \leq v \quad vs. \quad H_1 : \iota_{X_1}^{lin}(Y|X_2, ..., X_k) > v. \tag{1.4}$$

Furthermore, confidence intervals for the parameters are of great interest. Since we have $\iota_{X_1}^{lin}(Y|X_2, ..., X_k) = |s\iota_{X_1}^{lin}(Y|X_2, ..., X_k)|$ the one-sided null hypothesis (1.4) coincides with the null hypothesis $H_0 : -v \leq s\iota_{X_1}^{lin}(Y|X_2, ..., X_k) \leq v$. Hence, we start constructing a test for (1.3) and build from this a test for (1.4). We know that a level α test for $H_0 : \iota_{X_1}^{lin}(Y|X_2, ..., X_k) \leq v$ is a conservative level α test for $H_0 : \iota_{X_1}(Y|X_2, ..., X_k) \leq v$. We are also interested in testing for $v \in \mathbb{R}$ the one-sided hypothesis

$$H_0 : \theta_{X_1}^{lin}(Y|X_2, ..., X_k) \leq v \quad vs. \quad H_1 : \theta_{X_1}^{lin}(Y|X_2, ..., X_k) > v. \tag{1.5}$$

Asymptotic normality

Remember the subspace

$$\mathcal{H}_2 = \text{span}(1, X_2, ..., X_k) = \{\beta_1 + \sum_{j=2}^{k} X_j\beta_j : \beta = (\beta_1, ..., \beta_k)^T \in \mathbb{R}^k\}$$

and its orthogonal complement $\mathcal{H}_1 = \mathcal{H}_2^{\perp}$ in $L^2_{\mathbf{P}}$. We consider the decomposition of X_1

$$X_1 = U + \tilde{X}_1 \quad \text{with} \quad \tilde{X}_1 = \xi_1 + \sum_{j=2}^{k} X_j\xi_j \in \mathcal{H}_2 \quad \text{and} \quad U \in \mathcal{H}_1.$$

Hence, $\tilde{X}_1$ is the orthogonal projection of X_1 onto $\mathcal{H}_2$. The same decomposition can be made for X_{i1}, namely $X_{i1} = U_i + \tilde{X}_{i1}$ where $\tilde{X}_{i1} = \xi_1 + \sum_{j=2}^{k} X_{ij}\xi_j \in \mathcal{H}_2$. To establish asymptotic results we need the assumption

$$E_{\mathbf{P}}\{U^2Y^2\} < \infty \tag{1.6}$$

which implies that the random variable UY has finite variance. This assumption follows, for instance, if $Y, X_1, ..., X_k \in L^2_{\mathbf{P}}$, $Y = g(X_1, ..., X_k) + \epsilon$ where ϵ is a random variable independent of $X_1, ..., X_k$ with $\epsilon \in L^2_{\mathbf{P}}$ and $g(X_1, ..., X_k)$ is bounded or $g(X_1, ..., X_k)$ and U are stochastically independent. (1.6) also follows if

$$E_{\mathbf{P}}(Y^4) < \infty \qquad \text{and} \qquad E_{\mathbf{P}}(X_j^4) < \infty \quad \text{for all } j = 1, \ldots, k.$$

A similar decomposition of $\mathbf{X}_1$ as vector in $\mathbb{R}^n$ can be considered. With the random matrix $D_n = (\mathbf{1}, \mathbf{X}_2, ..., \mathbf{X}_k)$ where $\mathbf{1} = (1, ..., 1)^T \in \mathbb{R}^n$ and the assumption that $\operatorname{rank}(D_n) = k$ we can define $\hat{\xi} = (D_n^T D_n)^{-1} D_n^T \mathbf{X}_1$, the least squares estimate of $\xi = (\xi_1, \ldots, \xi_k)$. Obviously we have

$$\mathcal{M}_2 = \operatorname{span}(D_n) = \{D_n\beta : \beta \in \mathbb{R}^k\} \subseteq \mathbb{R}^n.$$

Therefore, the definition $\hat{X}_{i1} = (D_n\hat{\xi})_i = (\hat{\xi}_1 + \sum_{j=2}^{k} X_{ij}\hat{\xi}_j)_{i=1}^n$ leads to the conclusion $\hat{X}_{i1} = (P_{\mathcal{M}_2}\mathbf{X}_1)_i$ which implies $\hat{U}_i = (\hat{U})_i = X_{i1} - \hat{X}_{i1}$.

Lemma 1.13. *We have* $\hat{\xi} \xrightarrow{p} \xi$.

Lemma 1.14. *Let* $V_1, V_2, ...$ *be an i.i.d. sequence of random variables in* $L^2_{\mathbf{P}}$. *Then, for* U_i *and* $\hat{U}_i$ *defined above the following statements are true.*

(a) $\sum_{i=1}^{n}(U_i - \hat{U}_i)^2 = \sum_{i=1}^{n}(\tilde{X}_{i1} - \hat{X}_{i1})^2$ *is bounded in probability.*

(b) If $E_{\mathbf{P}}(V_i) = E_{\mathbf{P}}(V_i X_{ij}) = 0$ *for* $j = 2, ..., k$ *then*

$$(1/\sqrt{n})\sum_{i=1}^{n}(U_i - \hat{U}_i)V_i \xrightarrow{p} 0.$$

(c) If $E_{\mathbf{P}}(|V_i X_{ij}|) < \infty$ *and* $E_{\mathbf{P}}(|V_i X_{ij} X_{il}|) < \infty$ *for all* $2 \le j, l \le k$ *then*

$$(1/n)\sum_{i=1}^{n}(U_i - \hat{U}_i)^2 V_i \xrightarrow{p} 0.$$

In the following let again $Z = P_{\mathcal{H}_2^\perp} Y$. In order to show the asymptotic normality of the linear signed partial mean impact we first show a proposition.

Proposition 1.15. *We have that*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^2 - \frac{1}{n}\sum_{i=1}^{n}U_i^2\right) \xrightarrow{p} 0.$$

Proof. We have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^2 - \frac{1}{n}\sum_{i=1}^{n}U_i^2\right) = \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{U}_i - U_i)^2}_{(1)} + \underbrace{\frac{2}{\sqrt{n}}\sum_{i=1}^{n}(\hat{U}_i - U_i)U_i}_{(2)},$$

where both (1) and (2) converge to 0 in probability by Lemma 1.14. □

With the help of Proposition 1.15 we are able to show the asymptotic normality of $s\iota_{X_1}^{lin}(Y|X_2,...,X_k)$ stated in the following theorem (In Scharpenberg (2012) this result was not shown).

Theorem 1.16. *We have*

$$\sqrt{n}(\hat{s\iota}_{X_1}^{lin}(Y|X_2,...,X_k) - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)) \xrightarrow{\mathcal{L}} N\left(0, \frac{\varphi}{\eta^2}\right),$$

where

$$\begin{aligned}\varphi =& \kappa^2 - \frac{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}{\eta}\left(E_{\mathbf{P}}(U_i^3 Z_i) - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta^3\right) \\ &+ \left(\frac{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}{2\eta}\right)^2 Var_{\mathbf{P}}(U_i^2),\end{aligned}$$

with $\eta^2 = E_{\mathbf{P}}(U^2)$ *and* $\kappa^2 = E_{\mathbf{P}}(UZ)$.

Proof. We have

$$\begin{aligned}& \sqrt{n}(\hat{s\iota}_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta}) \\ =\ & \sqrt{n}(\hat{s\iota}_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta) \\ & -\sqrt{n}(s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta).\end{aligned}$$

We regard the random vector $\binom{U_iZ_i}{U_i^2}$ which has mean

$$E_{\mathbf{P}}\begin{pmatrix} U_iZ_i \\ U_i^2 \end{pmatrix} = \begin{pmatrix} s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \\ \eta^2 \end{pmatrix}$$

and covariance matrix

$$Cov_{\mathbf{P}}\begin{pmatrix} U_iZ_i \\ U_i^2 \end{pmatrix} = \begin{pmatrix} \kappa^2 & \rho \\ \rho & \gamma^2 \end{pmatrix} = \Sigma,$$

where $\kappa^2 = Var_{\mathbf{P}}(UZ)$, $\gamma^2 = Var_{\mathbf{P}}(U^2)$ and $\rho = Cov_{\mathbf{P}}(U_iZ_i, U_i^2) = E_{\mathbf{P}}(U_i^3Z_i) - E_{\mathbf{P}}(U_iZ_i)E_{\mathbf{P}}(U_i^2) = E_{\mathbf{P}}(U_i^3Z_i) - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta^3$.
By the strong law of large numbers ((cf. van der Vaart, 2000, p. 16)) we obtain

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} U_iZ_i \\ U_i^2 \end{pmatrix} - \begin{pmatrix} s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \\ \eta^2 \end{pmatrix}\right) \xrightarrow{\mathcal{L}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right). \tag{1.10}$$

From Proposition 1.15 we know that

$$\sqrt{n}\begin{pmatrix} \hat{s\iota}_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - \frac{1}{n}\sum_{i=1}^{n} U_iZ_i \\ \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^2 - \frac{1}{n}\sum_{i=1}^{n}U_i^2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.11}$$

Additionally we have

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}{\eta+\hat{\eta}} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 & 0 \\ 0 & \frac{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}{2\eta} \end{pmatrix} =: A. \tag{1.12}$$

Therefore, it follows, when adding (1.10) and (1.11), by (van der Vaart, 2000, p.11) that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \hat{U}_iZ_i \\ \hat{U}_i^2 \end{pmatrix} - \begin{pmatrix} s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \\ \eta^2 \end{pmatrix}\right) \xrightarrow{\mathcal{L}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

and in conclusion by multiplying with (1.12)

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}{\eta+\hat{\eta}} \end{pmatrix} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \hat{U}_iZ_i \\ \hat{U}_i^2 \end{pmatrix} - \begin{pmatrix} s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \\ \eta^2 \end{pmatrix}\right)$$

$$= \sqrt{n}\begin{pmatrix} \hat{s\iota}_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \\ s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\hat{\eta} - s\iota_{X_1}^{lin}(Y|X_2,...,X_k)\eta \end{pmatrix} \xrightarrow{\mathcal{L}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma'\right),$$

where

$$\Sigma' = A\Sigma A^T = \begin{pmatrix} \kappa^2 & \frac{s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\rho}{2\eta} \\ \frac{s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\rho}{2\eta} & \left(\frac{s\iota^{lin}_{X_1}(Y|X_2,...,X_k)}{2\eta}\right)^2 \gamma^2 \end{pmatrix}.$$

From this it follows by the Cramér-Wold Device and the properties of the normal distribution that

$$\sqrt{n}(\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta} - s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta})$$

$$= \begin{pmatrix} 1 & -1 \end{pmatrix} \sqrt{n} \begin{pmatrix} \hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta} - s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta \\ s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta} - s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta \end{pmatrix}$$

$$\xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix} 1 & -1 \end{pmatrix} \Sigma' \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right) = N(0, \varphi),$$

and therefore

$$\sqrt{n}(\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k) - s\iota^{lin}_{X_1}(Y|X_2,...,X_k)) \xrightarrow{\mathcal{L}} N\left(0, \frac{\varphi}{\eta^2}\right).$$

□

In order to estimate the asymptotic normal distribution of $\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)$ we need to estimate the variance φ/η^2. The next theorem shows how φ/η^2 can be consistently estimated.

Theorem 1.17. *We have that*

$$\hat{\varphi}/\hat{\eta}^2 \xrightarrow{p} \varphi/\eta^2,$$

with

$$\hat{\varphi} = \hat{\kappa}^2 - \frac{\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)}{\hat{\eta}}\hat{\rho} + \left(\frac{\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)}{2\hat{\eta}}\right)^2 \hat{\gamma}^2,$$

where $\hat{\rho} = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^3\hat{Z}_i - \hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta}^3$, $\hat{\gamma}^2 = \frac{1}{n}\sum_{i=1}^n(\hat{U}_i^2 - \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2)^2$, $\hat{\kappa}^2 = \frac{1}{n}\sum_{i=1}^n\{\hat{U}_i\hat{Z}_i - \hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta}\}^2$ *and* $\hat{\eta}^2 = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2$.

Proof. From Scharpenberg (2012) we know that $\hat{\kappa}^2$, $\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)$ and $\hat{\eta}$ are consistent estimators of κ^2, $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)$ and η. This implies that we only have to show the consistency of $\hat{\rho}$ and $\hat{\gamma}^2$ for ρ and γ^2. This follows directly from the assumptions (e.g. existing means, i.i.d. random variables, ...) and the fact that $\hat{\xi} \xrightarrow{p} \xi$. □

Note that in Scharpenberg (2012) it was only shown that

$$\sqrt{n}\{\hat{s\iota}^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta} - s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta\} \xrightarrow{\mathcal{L}} N(0,\kappa^2)$$

where $\kappa^2 = Var_{\mathbf{P}}(UZ)$ and $\hat{\eta}^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{U}_i^2 \xrightarrow{p} \eta^2 = E_{\mathbf{P}}(U^2)$. This result is less satisfactory than Theorem 1.16 since it only allows the derivation of a confidence interval for $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta$ instead of $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)$. The asymptotic normality of $\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k)$ can also be shown. According to Proposition 1.9 we have that

$$\theta^{lin}_{X_1}(Y|X_2,...,X_k) = E_{\mathbf{P}}(UY)/E_{\mathbf{P}}(U^2)$$

which leads to the estimator

$$\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k) = \sum_{i=1}^{n}\hat{U}_iY_i/\sum_{i=1}^{n}\hat{U}_i^2.$$

$\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k)$ is identical to the least squares estimate of the regression coefficient from the linear model with Y as dependent variable and $X_1,...,X_k$ as independent covariables. It can be shown that

$$\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k) = \sum_{i=1}^{n}\hat{U}_iY_i/\sum_{i=1}^{n}\hat{U}_i^2 = \sum_{i=1}^{n}\hat{U}_iZ_i/\sum_{i=1}^{n}\hat{U}_i^2.$$

In order to show the asymptotic normality of the estimate $\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k)$ we need the assumption

$$E_{\mathbf{P}}(|X_iX_jX_lX_m|) < \infty \text{ for all } 1 \le i,j,l,m \le k \tag{1.13}$$

which follows, for instance, if all X_j are bounded. Conclusions of (1.13) are e.g. that $E_{\mathbf{P}}\{U_i^2Z_i^2\} < \infty$ and $E_{\mathbf{P}}(U_i^4) < \infty$.

Theorem 1.18. *If $(Y_i, X_{i1},...,X_{ik})$, $i = 1,...,n$, are i.i.d. and satisfy assumption* (1.13) *then*

$$\sqrt{n}\{\hat{\theta}^{lin}_{X_1}(Y|X_2,...,X_k) - \theta^{lin}_{X_1}(Y|X_2,...,X_k)\} \xrightarrow{\mathcal{L}} N\left(0,\frac{\tau^2}{\eta^4}\right)$$

where $\tau^2 = E_{\mathbf{P}}[U^2\{Z - U\theta^{lin}_{X_1}(Y|X_2,...,X_k)\}^2]$ and $\eta^2 = E_{\mathbf{P}}(U^2)$.

We already know how to estimate η^2. The following theorem gives a consistent estimate for τ^2.

Theorem 1.19. *Using the same assumptions as for Theorem 1.18 we can state that*

$$\hat{\tau}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 \hat{\epsilon}_i^2 \xrightarrow{p} \tau^2$$

where $\hat{\epsilon}_i$ are the residuals from a linear regression analysis with dependent variable Y_i and independent variables $X_{i1}, ..., X_{ik}$.

If ϵ and the covariates $X_1, ..., X_k$ are independent we obtain

$$\tau^2 = E_{\mathbf{P}}(U_i^2 \epsilon_i^2) = E_{\mathbf{P}}(U_i^2) E_{\mathbf{P}}(\epsilon_i^2) = \eta^2 \sigma^2,$$

which implies

$$\frac{\tau^2}{\eta^4} = \frac{\sigma^2}{\eta^2}.$$

Hence, $\sqrt{n}\{\hat{\theta}_{X_1}^{lin}(Y|X_2, ..., X_k) - \theta_{X_1}^{lin}(Y|X_2, ..., X_k)\}$ converges to the same normal distribution as $\sqrt{n}\{\hat{\theta}_1 - \theta_1\}$ where $\hat{\theta}_1$ is the least squares estimate for the regression coefficient θ_1 from a linear regression analysis with dependent variable Y and independent variables $X_1, ..., X_k$. Additionally τ^2/η^4 would be estimated by $\hat{\sigma}^2/\hat{\eta}^2$ where $\hat{\sigma}^2$ is the estimate for the residual variance form the linear model.

Transferring these results into the case of a linear regression model $Y = \theta_1 X_1 + ... + \theta_k X_k + \epsilon$ with $E_{\mathbf{P}}(\epsilon) = 0$ and ϵ uncorrelated to the covariates, Theorems 1.18 and 1.19 are similar to the results in White (1980a) and White (1980b) for the regression coefficient θ_1.

The single-covariate versions of Theorems 1.18 and 1.19 are

Theorem 1.20. *Under the setup of this section we have that*

$$\sqrt{n}\{\hat{\theta}_X^{lin}(Y) - \theta_X^{lin}(Y)\} \xrightarrow{\mathcal{L}} N\left(0, \frac{\tau^2}{\eta^4}\right)$$

where $\tau^2 = E_{\mathbf{P}}\{(X - E_{\mathbf{P}}(X))^2[(Y - E_{\mathbf{P}}(Y)) - (X - E_{\mathbf{P}}(X))\theta_X^{lin}(Y)]^2\}$ and $\eta^2 = E_{\mathbf{P}}(\{X - E_{\mathbf{P}}(X)\}^2) = Var_{\mathbf{P}}(X)$

and

Theorem 1.21. *Under the same assumptions as in Theorem 1.20 we have that*

$$\hat{\tau}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 \hat{\epsilon}_i^2 \xrightarrow{p} \tau^2$$

where $\hat{\epsilon}_i$ are the residuals from a linear regression analysis with target variable Y_i and covariates $1, X_i$.

Hypothesis testing and confidence intervals for the partial linear mean slope

A direct consequence of Theorems 1.18 and 1.19 is that with

$$T_v = \sqrt{n}\{\hat{\theta}^{lin}_{X_1}(Y|X_2, ..., X_k) - v\}(\hat{\eta}^2/\hat{\tau})$$

the rejection rule $T_v \geq \Phi^{-1}(1-\alpha)$ provides a test with significance level close to α for the hypothesis (1.5). Analogous

$$CI^{\theta}_{\alpha} = [\hat{\theta}^{lin}_{X_1}(Y|X_2, ..., X_k) - (\hat{\tau}/\hat{\eta}^2)\Phi^{-1}(1-\alpha)/\sqrt{n}, \infty)$$

is expected to have coverage probability close to $1-\alpha$ for $\theta^{lin}_{X_1}(Y|X_2, ..., X_k)$.

Similar to this the rejection rule $|T_v| \geq \Phi^{-1}(1-\frac{\alpha}{2})$ is expected to provide an approximate level α test for the two-sided hypothesis

$$H_0 : \theta^{lin}_{X_1}(Y|X_2, ..., X_k) = v \quad vs. \quad H_1 : \theta^{lin}_{X_1}(Y|X_2, ..., X_k) \neq v$$

where $v \in \mathbb{R}$. An approximate two-sided confidence interval for the linear partial mean impact is then given by

$$CI^{\theta}_{\alpha,\text{2-sided}} = (\hat{\theta}^{lin}_{X_1}(Y|X_2, ..., X_k) \pm (\hat{\tau}/\hat{\eta}^2)\Phi^{-1}(1-\frac{\alpha}{2})/\sqrt{n}).$$

In order to improve the type one error rate of the tests and the coverage probability of the confidence intervals one could follow the heuristic approach replace the quantile $\Phi^{-1}(1-\frac{\alpha}{2})$ of the normal distribution by the $(1-\frac{\alpha}{2})$-quantile of the t-distribution with $n-(k+1)$ degrees of freedom.

Hypothesis testing and confidence intervals for the partial linear signed mean impact

Since in Scharpenberg (2012) it was only shown that

$$\sqrt{n}\{\hat{s\iota}^{lin}_{X_1}(Y|X_2, ..., X_k)\hat{\eta} - s\iota^{(n)}_{X_1}(Y|X_2, ..., X_k)\eta\} \xrightarrow{\mathcal{L}} N(0, \kappa^2)$$

only confidence intervals for $s\iota^{lin}_{X_1}(Y|X_2, ..., X_k)\eta$ could be constructed. A heuristic approach to the construction of a confidence interval for $s\iota^{lin}_{X_1}(Y|X_2, ..., X_k)$ from those of

$s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta$ and $\theta^{lin}_{X_1}(Y|X_2,...,X_k)$ is given there. We have that

$$CI^{s\iota\eta}_{\alpha} = [\hat{s}\iota^{lin}_{X_1}(Y|X_2,...,X_k)\hat{\eta} - \hat{\kappa}\Phi^{-1}(1-\alpha)/\sqrt{n}, \infty)$$

is expected to have coverage probability close to $1-\alpha$ for $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)\eta$ for sufficiently large sample sizes. Again one could improve the coverage probability of this interval by replacing the normal quantile by the quantile of the t-distribution with $n-(k+1)$ degrees of freedom. Note, that this consideration is only heuristic.

To construct the confidence interval for $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)$ we note that both, $CI^{s\iota\eta}_{\alpha}/\eta$ and $CI^{\theta}_{\alpha}\eta$, are approximate one-sided $(1-\alpha)$ confidence intervals for the linear signed impact. We can rewrite them as

$$CI^{s\iota\eta}_{\alpha}/\eta = (CI^{s\iota\eta}_{\alpha}/\hat{\eta})\left(\frac{\hat{\eta}}{\eta}\right) \quad \text{and} \quad CI^{\theta}_{\alpha}\eta = CI^{\theta}_{\alpha}\hat{\eta}\frac{\eta}{\hat{\eta}} \tag{1.14}$$

where one of the terms $\frac{\hat{\eta}}{\eta}$ and $\frac{\eta}{\hat{\eta}}$ is always smaller than 1 while the other one is greater than one. Therefore, we choose our confidence interval for the linear signed mean impact to be

$$CI^{s\iota,old}_{\alpha} = CI^{s\iota\eta}_{\alpha}/\hat{\eta} \cup CI^{\theta}_{\alpha}\hat{\eta} = [\hat{s}\iota^{lin}_{X_1}(Y|X_2,...,X_k) - c, \infty)$$

where $c = (\Phi^{-1}(1-\alpha)/\sqrt{n})\max\{\hat{\kappa}/\hat{\eta}, \hat{\tau}/\hat{\eta}\}$. Since this interval always contains at least one of the two intervals in (1.14). Hence, we expect this interval to have asymptotic coverage probability of $1-\alpha$. Similarly, the rejection rule $v \notin CI^{s\iota,old}_{\alpha}$ is expected to provide an approximate level α test for (1.3). A test for the two-sided hypothesis

$$H_0 : s\iota^{lin}_{X_1}(Y|X_2,...,X_k) = v \quad vs. \quad H_1 : s\iota^{lin}_{X_1}(Y|X_2,...,X_k) \neq v$$

can be derived from the two-sided confidence interval

$$CI^{s\iota,old}_{\alpha,\text{2-sided}} = (\hat{s}\iota^{lin}_{X_1}(Y|X_2,...,X_k) - c, \hat{s}\iota^{lin}_{X_1}(Y|X_2,...,X_k) + c)$$

where $c = (\Phi^{-1}(1-\alpha/2)/\sqrt{n})\max\{\hat{\kappa}/\hat{\eta}, \hat{\tau}/\hat{\eta}\}$.

However, this approach to the construction of a confidence interval for the linear singed mean impact is only heuristic. Theorem 1.16 implies that

$$CI^{s\iota^{lin}_{X_1}(Y|X_2,...,X_k)}_{\alpha} = [\hat{s}\iota^{lin}_{X_1}(Y|X_2,...,X_k) - \frac{\sqrt{\hat{\varphi}}}{\hat{\eta}}\Phi^{-1}(1-\alpha)/\sqrt{n}, \infty).$$

is a one-sided asymptotic $(1-\alpha)\%$ confidence interval for $s\iota^{lin}_{X_1}(Y|X_2,...,X_k)$ Therefore,

the rejection rule $v \notin CI_\alpha^{s\iota_{X_1}^{lin}(Y|X_2,...,X_k)}$ is expected to provide an approximate level α test for the null hypothesis $H_0 : s\iota_{X_1}^{lin}(Y|X_2,...,X_k) \leq v$ for $v \in \mathbb{R}$.

As a next step we want to construct a test for (1.4). For $v \geq 0$ we have

$$\iota_{X_1}^{lin}(Y|X_2,...,X_k) \leq v \Leftrightarrow -v \leq s\iota_{X_1}^{lin}(Y|X_2,...,X_k) \leq v.$$

This implies that the rejection rule

$$v < min\{|a| : a \in CI_{\alpha,\text{2-sided}}^{s\iota}\}$$

provides an approximate level α test for (1.4). Thus, an approximate level α confidence interval is given by

$$CI_\alpha^\iota = [min\{|a| : a \in CI_{\alpha,\text{2-sided}}^{s\iota}\}, \infty). \tag{1.15}$$

1.4.2. Simulations

In order to investigate if the derived confidence interval for $s\iota_{X_1}^{lin}(Y|X_2,...,X_k)$ really improves the old one we make some simulations and compare the two intervals with respect to the coverage probability and the probability of not covering zero. For the comparison of the intervals we choose the scenarios (1) and (2) of Section 5.1 Scharpenberg (2012). All simulations used $n = 100$ observations and 1000 repetitions.

(1) We assume that $Y = \frac{1}{8}e^X + \epsilon$ where $X \sim N(\mu, \sigma^2)$ and $\epsilon \sim N(0, 1)$ are independent. Table 2 gives the power of the test of $H_0 : \iota_Y^{lin}(X) \leq v$ with $v = 0$ and the power of the z-test from linear regression for the one-sided null hypothesis that the first regression coefficient is less or equal zero ($H_0 : \theta_1 \leq 0$) assuming that $X \sim N(\mu, \sigma^2)$. The tables also give the linear mean slope and the mean slope.

μ	σ^2	$\theta_X^{lin}(Y)$	$\theta_X(Y)$	Power linear-slope-test	Power z-test
0	1	0.206	0.354	0.6041	0.6216
-1	1	0.076	0.130	0.1896	0.1807
1	1	0.560	0.963	0.9942	0.9971
1	0.25	0.385	0.437	0.6067	0.5980

Table 2: Power of the test for $\theta_X^{lin}(Y)$ and the z-test from linear regression.

One can see that the linear mean slope test may suffer a slight loss in power compared to the z-test but it can also be more powerful in some cases.

μ	σ^2	$s\iota_X^{lin}(Y)$	$s\iota_X(Y)$	Power new impact test	Power old test
0	1	0.206	0.270	0.6018	0.5695
-1	1	0.076	0.099	0.1904	0.1723
1	1	0.560	0.734	0.9929	0.9878
1	0.25	0.193	0.205	0.6073	0.5822

Table 3: Power of the new test for $s\iota_X^{lin}(Y)$ and the old test.

One can see, that using the new confidence interval increases the power by up to 3% compared to the use of the interval from Scharpenberg (2012). The power of the new test is now near to the power of the test for the linear slope.

μ	σ^2	$s\iota_X^{lin}(Y)$	$s\iota_X(Y)$	$CI_\alpha^{s\iota,new} \ni s\iota_X^{lin}(Y)$	$CI_\alpha^{s\iota,old} \ni s\iota_X^{lin}(Y)$
0	1	0.206	0.270	0.9427	0.9692
-1	1	0.076	0.099	0.9417	0.9552
1	1	0.560	0.734	0.9600	0.9938
1	0.25	0.193	0.205	0.9660	0.9385

Table 4: Coverage probabilities of the two confidence intervals for $s\iota$ for different normal distributions of X.

In some cases the new confidence interval tends to undercoverage although it improves the old interval in terms of coverage probability in the last case.

(2) We now let $Y = \frac{1}{2}e^X + \epsilon$ where $X \sim Exp(\lambda)$ is independent from $\epsilon \sim N(0,1)$. The simulations gave the following results.

λ	$\theta_X^{lin}(Y)$	$\theta_X(Y)$	Power linear-slope-test	Power z-test
3	1.125	1.500	0.9059	0.9050
5	0.781	0.833	0.4540	0.4634

Table 5: Power of the test for $\theta_X^{lin}(Y)$ and the z-test from linear regression.

λ	$s\iota_X^{lin}(Y)$	$s\iota_X(Y)$	Power new impact test	Power old test
3	0.375	0.433	0.8873	0.8027
5	0.156	0.161	0.4634	0.3842

Table 6: Power of the new test for $s\iota_X^{lin}(Y)$ and the old test.

In these scenarios the power of the tests could be improved by approximately 8% by using the new confidence intervals. In this case as well the power of the test for the signed linear mean impact is now close to the one of the linear slope. One

λ	$s\iota_X^{lin}(Y)$	$s\iota_X(Y)$	$CI_\alpha^{s\iota new} \ni s\iota_X^{lin}(Y)$	$CI_\alpha^{s\iota old} \ni s\iota_X^{lin}(Y)$
3	0.375	0.433	0.9667	0.9983
5	0.156	0.161	0.9424	0.9824

Table 7: Coverage probabilities of the two confidence intervals for $s\iota$ for different normal distributions of X.

can see, that similar to the first simulations the use of the new confidence intervals reduces the coverage probability. Nevertheless the new coverage probabilities are much closer to stated level than the old ones.

1.5. Absolute mean slope

Up to this point the mean slope was defined by

$$\theta_{X_1}(Y|X_2,...,X_k) = \frac{\iota_{X_1}(Y|X_2,...,X_k)}{E_{\mathbf{P}}(X_1\delta_0(\mathbf{X}))}.$$

Here, δ_0 is the almost surely uniquely defined perturbation for which we have that $\iota_{X_1}(Y|X_2,...,X_k) = E_{\mathbf{P}}(Y\delta_0(\mathbf{X}))$. It describes the maximum change in the mean of Y when changing the distribution of the covariates in a way that the mean of X_1 is changed by one unit with the same distributional change. However, such a statement is only useful if there is a linear relationship between Y and X_1. When moving to non-linear and therefore possibly non-monotonous relationships the mean slope becomes meaningless. For example when regarding quadratic influences of X_1 on Y (say $Y = X_1^2 + \epsilon$) the term $E_{\mathbf{P}}(X_1\delta_0(\mathbf{X}))$ could become very small or zero. Therefore, we suggest a new measure of association which we call *partial absolute mean slope*. It is defined as the maximum

change in the mean of Y relative to the maximum possible change in the mean of X_1 when changing the density of the covariates. This can be formalized as follows:

$$\theta_{X_1}(Y|X_2,...,X_k) = \frac{\iota_{X_1}(Y|X_2,...,X_k)}{\iota_{X_1}(X_1|X_2,...,X_k)} = \frac{\iota_{X_1}(Y|X_2,...,X_k)}{\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}_2^\perp}X_1)}}, \tag{1.16}$$

where $\mathcal{H}_2 = \text{span}(X_2,...,X_k)$. With the definition as the ratio of maximum possible changes in the means of Y and X_1 under distributional changes of the covariates, the absolute mean slope becomes meaningful again. Note that the mean impact depends strongly on the distribution of X_1. The mean slope is not completely but more invariant with respect to this distribution (see also Brannath and Scharpenberg (2014)).
In the single covariate case the absolute mean slope simplifies to

$$\theta_X(Y) = \frac{\iota_X(Y)}{\iota_X(X)} = \frac{\iota_X(Y)}{SD_{\mathbf{P}}(X)}. \tag{1.17}$$

As already pointed out there may be reasons to regard restricted versions of the partial absolute mean slope (1.16) (e.g. to avoid overfitting). Let $\mathcal{R}$ be a closed subset of $L^2_{\mathbf{P}}(\mathbb{R})$. We define the *restricted partial absolute mean slope* as

$$\theta^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k) = \frac{\iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k)}{\iota^{\mathcal{R}}_{X_1}(X_1|X_2,...,X_k)}$$

where $\iota^{\mathcal{R}}_{X_1}(Y|X_2,...,X_k)$ is the restricted partial mean impact. In the special case of restriction to linear subspaces we obtain that the linear partial absolute mean slope is the absolute value of the linear partial mean slope. Hence, when we restrict to linear functions δ the absolute mean slope has still the interpretation of the maximum change in the mean of Y when we change X_1 by one unit, which has a simple interpretation in the linear setup. In the course of this thesis we will regard the absolute mean slope instead of the mean slope.

1.6. Common mean impact of several variables

In generalization to the mean impact we can define the *common mean impact* of a set of covariates $\mathbf{X} = (X^{(1)},...,X^{(k)})$. It is given by

$$\iota_{X^{(1)},...,X^{(k)}}(Y) = \sup_{\delta(\mathbf{X})\in L^2_{\mathbf{P}}(\mathbf{R}),\ E_{\mathbf{P}}[\delta(\mathbf{X})]=0,\ E_{\mathbf{P}}[\delta^2(\mathbf{X})]=1} E_{\mathbf{P}}[Y\delta(\mathbf{X})]. \tag{1.18}$$

The common mean impact quantifies the maximum change in the mean of the target variable Y, when the common density f of $X^{(1)},...,X^{(k)}$ is changed to $f(1+\delta)$, where δ has mean zero and variance equal to one. Hence, the common mean impact is a measure of the multivariate association between Y and $X^{(1)},...,X^{(k)}$.

Theorem 1.22. *Let $X^{(1)},...,X^{(k)}$ and Y be square integrable. Then*

(a) $\iota_{\mathbf{X}}(Y) = \sqrt{Var_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})]}$

(b) $\iota_{\mathbf{X}}(Y) = 0$ *if and only if* $E_{\mathbf{P}}(Y|\mathbf{X}) = E_{\mathbf{P}}(Y)$ *is independent from* $\mathbf{X}$.

(c) $0 \leq \iota_{\mathbf{X}}(Y) \leq \iota_Y(Y) = SD_{\mathbf{P}}(Y)$ *where* $SD_{\mathbf{P}}(Y) = \sqrt{Var_{\mathbf{P}}(Y)}$.

(d) $\iota_{\mathbf{X}}(Y) = \iota_Y(Y)$ *if and only if* Y *depends on* $\mathbf{X}$ *deterministically, i.e.,* $Y = g(\mathbf{X})$ *for a measurable function* $g : \mathbb{R}^{m+1} \to \mathbb{R}$.

(e) if $Y = g(\mathbf{X}) + U$, *where* $g : \mathbb{R}^{m+1} \to \mathbb{R}$ *is measurable and* U *and* $\mathbf{X}$ *are stochastically independent, then* $\iota_{\mathbf{X}}(Y) = \iota_{\mathbf{X}}[g(\mathbf{X})] = SD[g(\mathbf{X})]$.

Proof. (a) follows from Cauchy-Schwartz´s inequality in $L^2(\mathbb{R})$, which implies

$$\begin{aligned} E_{\mathbf{P}}[Y\delta(\mathbf{X})] &= E_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})\delta(\mathbf{X})] = E_{\mathbf{P}}[\{E_{\mathbf{P}}(Y|\mathbf{X}) - E_{\mathbf{P}}(Y)\}\delta(\mathbf{X})] \\ &\leq SD_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})]. \end{aligned}$$

For $\delta(\mathbf{X}) = \{E_{\mathbf{P}}(Y|\mathbf{X}) - E_{\mathbf{P}}(Y)\}/SD_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})]$ we obtain $E_{\mathbf{P}}[\delta(\mathbf{X})] = 0$, $E_{\mathbf{P}}[\delta^2(\mathbf{X})] = 1$ and $E_{\mathbf{P}}[Y\delta(\mathbf{X})] = SD_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})]$. This implies $\iota_{\mathbf{X}}(Y) = SD_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})]$. Statements (b) to (e) follow from (a) and $Var_{\mathbf{P}}(Y) = Var_{\mathbf{P}}[E_{\mathbf{P}}(Y|\mathbf{X})] + E_{\mathbf{P}}[Var_{\mathbf{P}}(Y|\mathbf{X})]$. □

1.7. Common linear mean impact of several variables

Similar to the case of the mean impact we would run into overfitting problems, when trying to estimate the common mean impact (1.18). As a solution to this, we restrict the set of allowed perturbations δ to the set of functions linear in random variables $X^{(1)},...,X^{(k)}$ (we write $\mathbf{X} = (X^{(1)},...,X^{(k)})$), where we assume that $X^{(1)} = 1$. This means we have

$$\iota^{lin}_{\mathbf{X}}(Y) = \sup_{\delta(\mathbf{X})\in\mathcal{H}; E_{\mathbf{P}}\{\delta(\mathbf{X})\}=0; E_{\mathbf{P}}\{\delta^2(\mathbf{X})\}=1} E_{\mathbf{P}}\{Y\delta(\mathbf{X})\}$$

where $\mathcal{H} = \text{span}(X^{(1)},...,X^{(k)}) \subseteq L^2_{\mathbf{P}}$. This *common linear mean impact* is clearly a lower bound for the common mean impact (1.18). Applications of this scenario will

cover polynomial fits or fitting natural splines and are further described in later sections. However, the common linear impact can be used to describe non-linear associations between the target variable Y and one (ore more) independent variables.

As a next step, we show that the common linear impact of $\mathbf{X}$ equals $\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}}Y)}$. By Cauchy´s inequality we obtain for all $\delta \in \mathcal{H}$ with $E_{\mathbf{P}}(\delta(\mathbf{X})) = 0$ and $E_{\mathbf{P}}(\delta^2(\mathbf{X})) = 1$ that

$$\begin{aligned} E_{\mathbf{P}}(Y\delta(\mathbf{X})) =& E_{\mathbf{P}}(P_{\mathcal{H}}Y\delta(\mathbf{X})) = E_{\mathbf{P}}(\{P_{\mathcal{H}}Y - E_{\mathbf{P}}(P_{\mathcal{H}}Y)\}\delta(\mathbf{X})) \\ \leq& \sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}}Y)}. \end{aligned}$$

Hence, if $\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}}Y)} = 0$ then $\iota^{lin}_{\mathbf{X}}(Y) = 0$, otherwise chose $\delta(\mathbf{X}) = \{P_{\mathcal{H}}Y - E_{\mathbf{P}}(P_{\mathcal{H}}Y)\}/\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}}Y)}$ and obtain $\iota^{lin}_{\mathbf{X}}(Y) = \sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}}Y)}$. Note that $Var_{\mathbf{P}}(P_{\mathcal{H}}Y) = E_{\mathbf{P}}\{(P_{\mathcal{H}_1}Y)^2\}$, where $\mathcal{H}_1 = \mathcal{H} \cap \text{span}(1)^{\perp} = \mathcal{H} - \text{span}(1)$.

By these arguments the linear mean impact can be estimated by

$$\begin{aligned} \hat{\iota}^{lin}_{\mathbf{X}}(Y) =& \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(P_{\mathcal{M}}\mathbf{Y})_i - \frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i\right]^2} \\ =& \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i\right)^2} \end{aligned} \tag{1.19}$$

where $\mathbf{Y} = (Y_1, ..., Y_n)^T$, $\mathcal{M} = \text{span}(\mathbf{X}^{(1)}, ..., \mathbf{X}^{(k)}) \subseteq \mathbb{R}^n$ and $\mathbf{X}^{(j)} = \left(X_1^{(j)}, ..., X_n^{(j)}\right)^T$ is the vector of observations of $X^{(j)}$. Consistency of this estimator can be shown as follows. Let $\hat{\xi}_1, ..., \hat{\xi}_k$ be the coefficients of the projection of $\mathbf{Y}$ onto $\mathcal{M}$ in $\mathbb{R}^n$ and $\xi_1, ..., \xi_k$ the coefficients of the projection of Y onto $\mathcal{H}$ in $L^2_{\mathbf{P}}$. We know that $(\hat{\xi}_1, ..., \hat{\xi}_k) \xrightarrow{p} (\xi_1, ..., \xi_k)$. Therefore we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i &= \frac{1}{n}\sum_{i=1}^{n}(\hat{\xi}_1 X_i^{(1)} + ... + \hat{\xi}_k X_i^{(k)}) \\ &= \hat{\xi}_1\frac{1}{n}\sum_{i=1}^{n}X_i^{(1)} + ... + \hat{\xi}_k\frac{1}{n}\sum_{i=1}^{n}X_i^{(k)} \\ &\xrightarrow{p} \xi_1 E_{\mathbf{P}}(X^{(1)}) + ... + \xi_k E_{\mathbf{P}}(X^{(k)}) \\ &= E_{\mathbf{P}}(P_{\mathcal{H}}Y). \end{aligned}$$

Analogously it can be shown that $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2 \xrightarrow{p} E_{\mathbf{P}}([P_{\mathcal{H}}Y]^2)$ and therefore $\hat{\iota}_{\mathbf{X}}^{lin}(Y) \xrightarrow{p} \iota_{\mathbf{X}}^{lin}(Y)$.

1.7.1. A test for the linear common mean impact being zero

As a next step we want to derive a test for

$$H_0 : \iota_{\mathbf{X}}^{lin}(Y) = 0 \quad \text{vs.} \quad H_1 : \iota_{\mathbf{X}}^{lin}(Y) \neq 0 \Leftrightarrow H_1 : \iota_{\mathbf{X}}^{lin}(Y) > 0.$$

We make the assumption

Assumption 1.23. *There exists no $\xi \in \mathbb{R}^k$ with $\xi \neq \mathbf{0}$ so that the linear combination $\xi_1 X^{(1)} + ... + \xi_k X^{(k)}$ is almost surely constant.*

With this assumption we obtain

$$\iota_{\mathbf{X}}^{lin}(Y) = 0 \Leftrightarrow P_{\mathcal{H}}Y = \text{const. almost surely} \Leftrightarrow R\xi = 0,$$

where $R = \begin{pmatrix} 0 & \\ \vdots & I_{k-1} \\ 0 & \end{pmatrix}$, with I_{k-1} being the $(k-1)$ dimensional identity matrix and $\xi = (\xi_1, ..., \xi_k)^T$ the vector of coefficients of the orthogonal projection of Y on $\mathcal{H}$. Therefore,

$$H_0 : \iota_{\mathbf{X}}^{lin}(Y) = 0 \Leftrightarrow H_0' : R\xi = 0.$$

To construct a test for H_0' we make the following assumptions which originate from White (1980b).

Assumption 1.24. *The true model is*

$$Y_i = g(W_i) + \epsilon_i, \quad i = 1, ..., n$$

where g is an unknown measurable function and (W_i, ϵ_i) are i.i.d. random $(p+1)$ vectors $(p \geq 1)$ such that $E(W_i) = 0$, $E(W_i^T W_i) = M_{WW}$ finite and non-singular, $E(\epsilon_i) = 0$, $E(\epsilon_i^2) = \sigma_\epsilon^2 < \infty$, $E(W_i^T \epsilon_i) = 0$ and $E(g(W_i)^2) = \sigma_g^2 < \infty$.

Assumption 1.25. $\mathbf{X} = (X^{(1)}, ..., X^{(k)})$ *is a measurable function of W.*

Assumption 1.25 means that the elements of $\mathbf{X}_i$ are functions of W_i, but not necessarily functions of every element of W_i, some variables may be omitted. We also need to assume

Assumption 1.26. $E_{\mathbf{P}}(g(W_i)\epsilon_i) = 0$, $E_{\mathbf{P}}(\boldsymbol{X}_i^T\epsilon_i) = 0$, $E_{\mathbf{P}}(\boldsymbol{X}_i^T\boldsymbol{X}_i) = M_{XX}$ *is finite and nonsingular.*

White (1980b) shows that under assumptions 1.24, 1.25 and 1.26 the following asymptotic result holds:

$$\sqrt{n}(\hat{\xi} - \xi) \xrightarrow{\mathcal{L}} N_k(\mathbf{0}, \Sigma)$$

where Σ can be consistently estimated by $(\mathbf{X}^T\mathbf{X}/n)^{-1}\hat{V}(\mathbf{X}^T\mathbf{X}/n)^{-1}$ with $\hat{V} = n^{-1}\sum_{i=1}^n(Y_i - \mathbf{X}_i\hat{\xi})^2\mathbf{X}_i^T\mathbf{X}_i$ and $\hat{\xi}$ is the vector of estimated coefficients from a linear regression with target variable Y and covariates $X^{(1)}, ..., X^{(k)}$. Since R has rank $k-1$ we obtain

$$\sqrt{n}R(\hat{\xi} - \xi) \xrightarrow{\mathcal{L}} N_{k-1}(\mathbf{0}, R\Sigma R^T)$$

which implies

$$n[R(\hat{\xi} - \xi)]^T[R(\mathbf{X}^T\mathbf{X}/n)^{-1}\hat{V}(\mathbf{X}^T\mathbf{X}/n)^{-1}R^T]^{-1}[R(\hat{\xi} - \xi)] \xrightarrow{\mathcal{L}} \chi^2_{k-1},$$

thus under $H_0' : R\xi = 0$

$$n[R\hat{\xi}]^T[R(\mathbf{X}^T\mathbf{X}/n)^{-1}\hat{V}(\mathbf{X}^T\mathbf{X}/n)^{-1}R^T]^{-1}[R\hat{\xi}] \xrightarrow{\mathcal{L}} \chi^2_{k-1}.$$

This implies that we can reject H_0' at an asymptotic significance level α if

$$T = n[R\hat{\xi}]^T[R(\mathbf{X}^T\mathbf{X}/n)^{-1}\hat{V}(\mathbf{X}^T\mathbf{X}/n)^{-1}R^T]^{-1}[R\hat{\xi}] \geq Q^{\chi^2}_{k-1}(1-\alpha),$$

where $Q^{\chi^2}_{k-1}(1-\alpha)$ is the $(1-\alpha)-$quantile of the χ^2_{k-1} distribution.

1.7.2. A shrinkage-like approach to the construction of confidence intervals for the linear common mean impact

In this section we want to derive lower confidence intervals for $\iota^{lin}_{\mathbf{X}}(Y)$. We will start by constructing confidence intervals for the squared impact, from which one can easily obtain the desired confidence bounds for the unsquared restricted impact. First of all we assume that the assumptions 1.23, 1.24, 1.25 and 1.26 hold. It was shown in the previous section that these assumptions imply

$$n(\hat{\xi} - \xi)^T\hat{\Sigma}^{-1}(\hat{\xi} - \xi) \xrightarrow{\mathcal{L}} \chi^2_k,$$

with $\hat{\Sigma} = (\mathbf{X}^T\mathbf{X}/n)^{-1}\hat{V}_n(\mathbf{X}^T\mathbf{X}/n)^{-1}$ from above. When testing the squared impact via the coefficient vector ξ of the orthogonal projection of Y onto $\mathcal{H}$ one has to keep in mind

that multiple ξ can lead to the same squared impact. Thus, to be able to reject a certain impact we have to be able to reject all coefficient vectors leading to this impact. To this end we note that

$$\hat{\iota}_{\mathbf{X}}^{lin^2}(Y) = \hat{\xi}^T \underbrace{\left(\mathbf{X}^T\mathbf{X}/n - \mathbf{X}^T \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \mathbf{X}/n^2 \right)}_{=:\mathbf{U}} \hat{\xi}$$

and choose a shrinkage-like approach. The idea is to find for $\lambda > 0$

$$\text{argmin}_{\xi}\, n(\hat{\xi} - \xi)^T \hat{\Sigma}^{-1}(\hat{\xi} - \xi) + \lambda \xi^T \mathbf{U} \xi. \tag{1.20}$$

In this approach we penalize the χ^2-test for H_0' by the estimated squared impact obtained by ξ. In the following we will show that this minimization problem is equivalent to finding the minimum of

$$n(\hat{\xi} - \xi)^T \hat{\Sigma}^{-1}(\hat{\xi} - \xi)$$

under the constraint $\xi^T \mathbf{U} \xi \leq s(\lambda)$ for $s(\lambda) = \xi_\lambda^T \mathbf{U} \xi_\lambda$, where ξ_λ is the unique solution to (1.20). This means that by testing the coefficient ξ_λ which solves (1.20), we essentially test the hypotheses

$$H_0 : \iota_{\mathbf{X}}^{lin^2}(Y) \leq s(\lambda) \quad \text{vs.} \quad H_1 : \iota_{\mathbf{X}}^{lin^2}(Y) > s(\lambda). \tag{1.21}$$

By testing these hypotheses for all $s(\lambda)$ in a decreasing manner we will be able to find the desired asymptotic lower confidence interval (the last $s(\lambda)$ which cannot be rejected). In order to understand the behavior of ξ_λ and $s(\lambda)$ when λ changes we make the following considerations.

Proposition 1.27. $\xi_\lambda = [n\hat{\Sigma}^{-1} + \lambda \boldsymbol{U}]^{-1} n\hat{\Sigma}^{-1}\hat{\xi}$ *is the unique solution to* (1.20).

Proof. First of all we show that $n\hat{\Sigma}^{-1} + \lambda \mathbf{U}$ is non-singular. We have

$$\begin{aligned} \hat{\Sigma} &= (\mathbf{X}^T\mathbf{X}/n)^{-1} \hat{V}_n (\mathbf{X}^T\mathbf{X}/n)^{-1} \\ &= (\mathbf{X}^T\mathbf{X}/n)^{-1} \mathbf{X}^T \Omega / n \mathbf{X} (\mathbf{X}^T\mathbf{X}/n)^{-1}, \end{aligned}$$

with $\Omega = \text{diag}(\hat{\epsilon}_1^2, ..., \hat{\epsilon}_n^2)$ and $\hat{\epsilon}_i^2 = (Y_i - \mathbf{X}_i\hat{\xi})^2 \geq 0$. Therefore, for $c \neq 0$, $c \in \mathbb{R}^k$ we obtain

$$c^T \hat{\Sigma} c = c^T (\mathbf{X}^T\mathbf{X}/n)^{-1} \mathbf{X}^T \Omega / n \mathbf{X} (\mathbf{X}^T\mathbf{X}/n)^{-1} c = \tilde{c}^T \Omega / n \tilde{c} \geq 0$$

with $\tilde{c} = \mathbf{X}(\mathbf{X}^T\mathbf{X}/n)^{-1}c \neq 0$ since $\mathbf{X}$ has full rank. Hence $\hat{\Sigma}$ is positive-semidefinite and since it is non-singular it is positive-definite. This implies that $n\hat{\Sigma}^{-1}$ is also positive-definite. Furthermore, we have for $c \neq 0$, $c \in \mathbb{R}^k$ and $\mathbb{1} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$

$$\begin{aligned} c^T\mathbf{U}c =& c^T\mathbf{X}^T\mathbf{X}/nc - c^T\mathbf{X}^T\mathbb{1}\mathbf{X}/n^2c \\ =& \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{k} c_j X_{ij}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} c_j X_{ij}\right)^2 \geq 0, \end{aligned}$$

where "=" in $\geq$ only holds if $\text{rank}(\mathbf{X}) < k$. Hence we have > 0 instead of ≥ 0. In consequence we obtain for $c \neq 0$, $c \in \mathbb{R}^k$

$$c^T(n\hat{\Sigma}^{-1} + \lambda\mathbf{U})c = \underbrace{c^T n\hat{\Sigma}^{-1}c}_{>0} + \underbrace{\lambda c^T\mathbf{U}c}_{>0} > 0.$$

Hence $n\hat{\Sigma}^{-1} + \lambda\mathbf{U}$ is positive-definite and thereby invertible.
Now we show the statement of the proposition. To this end we consider

$$\begin{aligned} &\frac{\partial}{\partial\xi} n(\hat{\xi} - \xi)^T\hat{\Sigma}^{-1}(\hat{\xi} - \xi) + \lambda\xi^T\mathbf{U}\xi = 0 \\ \Leftrightarrow& 2[n\hat{\Sigma}^{-1}(\xi - \hat{\xi}) + \lambda\mathbf{U}\xi] = 0 \\ \Leftrightarrow& n\hat{\Sigma}^{-1}\xi - n\hat{\Sigma}^{-1}\hat{\xi} + \lambda\mathbf{U}\xi = 0 \\ \Leftrightarrow& [n\hat{\Sigma}^{-1} + \lambda\mathbf{U}]\xi = n\hat{\Sigma}^{-1}\hat{\xi} \\ \Leftrightarrow& \xi = \xi_\lambda = [n\hat{\Sigma}^{-1} + \lambda\mathbf{U}]^{-1}n\Sigma^{-1}\hat{\xi}. \end{aligned}$$

Hence ξ_λ is the unique solution to (1.20). □

Proposition 1.28. *Let $\lambda > 0$. ξ_λ is also a minimizer of the expression*

$$n(\hat{\xi} - \xi)^T\hat{\Sigma}^{-1}(\hat{\xi} - \xi)$$

under the constraint $\xi^T\mathbf{U}\xi \leq s(\lambda) = \xi_\lambda^T\mathbf{U}\xi_\lambda$.

Proof. Let $\tilde{\xi}\mathbf{U}\tilde{\xi} \leq s(\lambda)$. Then we have

$$\begin{aligned} 0 &\leq n(\hat{\xi}-\tilde{\xi})^T\hat{\Sigma}^{-1}(\hat{\xi}-\tilde{\xi}) - n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda) + \underbrace{\lambda\tilde{\xi}^T\mathbf{U}\tilde{\xi} - \lambda s(\lambda)}_{\leq 0} \\ &\leq n(\hat{\xi}-\tilde{\xi})^T\hat{\Sigma}^{-1}(\hat{\xi}-\tilde{\xi}) - n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda). \end{aligned}$$

Therefore, ξ_λ is also a minimizer of the expression

$$n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi)$$

under the constraint $\xi^T\mathbf{U}\xi \leq s(\lambda)$. □

Proposition 1.29. *Let $\lambda > 0$. A minimizer of $n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi)$ under the constraint $\xi^T\mathbf{U}\xi \leq s(\lambda)$ is also a minimizer of $n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi) + \lambda\xi^T\mathbf{U}\xi$.*

Proof. Let $\tilde{\xi}$ be a minimizer of $n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi)$ under the constraint $\xi^T\mathbf{U}\xi \leq s(\lambda)$. Since ξ_λ obviously fulfills the constraint we have

$$n(\hat{\xi}-\tilde{\xi})^T\hat{\Sigma}^{-1}(\hat{\xi}-\tilde{\xi}) \leq n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda)$$

which implies

$$\begin{aligned} n(\hat{\xi}-\tilde{\xi})^T\hat{\Sigma}^{-1}(\hat{\xi}-\tilde{\xi}) + \lambda\tilde{\xi}^T\mathbf{U}\tilde{\xi} &\leq n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda) + \lambda\tilde{\xi}^T\mathbf{U}\tilde{\xi} \\ &\leq n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda) + \lambda\xi_\lambda^T\mathbf{U}\xi_\lambda. \end{aligned}$$

Hence the required minimization. □

The latter two propositions imply that the minimization problem (1.20) is equivalent to the minimization of $n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi)$ under the constraint $\xi^T\mathbf{U}\xi \leq s(\lambda) = \xi_\lambda^T\mathbf{U}\xi_\lambda$.

Proposition 1.30. *$s(\lambda)$ is decreasing in $\lambda > 0$.*

Proof. Let $0 < \lambda_1 \leq \lambda_2$. Due to the minimization property of ξ_{λ_2} we have

$$n(\hat{\xi}-\xi_{\lambda 2})^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_{\lambda 2}) + \lambda_2 s(\lambda_2) \leq n(\hat{\xi}-\xi_{\lambda 1})^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_{\lambda 1}) + \lambda_2 s(\lambda_1)$$

which implies

$$\lambda_2\left(s(\lambda_2) - s(\lambda_1)\right) \leq n(\hat{\xi}-\xi_{\lambda 1})^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_{\lambda 1}) - n(\hat{\xi}-\xi_{\lambda 2})^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_{\lambda 2}).$$

Analogously we obtain

$$\lambda_1(s(\lambda_2) - s(\lambda_1)) \geq n(\hat{\xi} - \xi_{\lambda 1})^T \hat{\Sigma}^{-1}(\hat{\xi} - \xi_{\lambda 1}) - n(\hat{\xi} - \xi_{\lambda 2})^T \hat{\Sigma}^{-1}(\hat{\xi} - \xi_{\lambda 2}).$$

Hence

$$\lambda_2(s(\lambda_2) - s(\lambda_1)) \leq \lambda_1(s(\lambda_2) - s(\lambda_1)).$$

Since $0 < \lambda_1 \leq \lambda_2$ this implies that $s(\lambda_2) \leq s(\lambda_1)$. □

We have already observed that we have to reject all coefficient vectors ξ that lead to a specific squared impact to be able to reject this squared impact. We have shown that ξ_λ minimizes the test statistic $n(\hat{\xi}-\xi)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi)$ for all ξ with $\iota^{lin^2}_{\mathbf{X},\xi}(Y) = \xi^T\mathbf{U}\xi \leq \xi_\lambda^T\mathbf{U}\xi_\lambda = \iota^{lin^2}_{\mathbf{X},\xi_\lambda}(Y)$. Which means that by testing ξ_λ we essentially test the hypotheses (1.21). This implies that being able to reject ξ_λ means being able to reject $\iota^{lin^2}_{\mathbf{X},\xi_\lambda}(Y)$ and all smaller values. Since $s(\lambda)$ is monotonously decreasing in λ the expression $n(\hat{\xi}-\xi_\lambda)^T\hat{\Sigma}^{-1}(\hat{\xi}-\xi_\lambda)$ is monotonously increasing in λ. Therefore, we can test ξ_λ for all λ and construct a lower confidence interval out of all $\iota^{lin^2}_{\mathbf{X},\xi_\lambda}(Y)$ for which ξ_λ could not be rejected. Since it is impossible to test for all $\lambda > 0$ we choose a different procedure in applications. In a first step we perform the test for $\iota^{lin}_{\mathbf{X}}(Y) = 0$ from the previous section. If we cannot reject we choose our lower confidence limit to be 0. In the case where $\iota^{lin}_{\mathbf{X}}(Y) = 0$ can be rejected we test for an arbitrary increasing sequence of λ until ξ_λ can be rejected the first time. Let λ_l be the first λ of the sequence for which we can reject ξ_λ. Then we undertake a bisection search between λ_{l-1} and λ_l and stop when the difference between the smallest squared impact which could not be rejected and the largest squared impact which could be rejected is less or equal a pre-chosen margin ϵ. We then choose the lower bound of the confidence interval to be the largest squared impact which could be rejected. A lower bound for the $\iota^{lin}_{\mathbf{X}}(Y)$ is then just the square root of this bound. However, as will turn out in Section 4 these intervals have poor coverage probability, when the mean impact equals zero and the sample size is small (i.e. $n \approx 100$). This may be due to the fact that the use of the robust covariance estimate of White (1980b) leads to type-I error inflation. In order to overcome this lack of performance for small impacts and small sample sizes, we will derive bootstrap intervals for the mean impact in Section 1.7.5, which will have better coverage properties. Simulations with larger sample sizes ($n = 200, n = 500$) showed that the coverage probabilities of the intervals derived here come close to the nominal level. Therefore, bootstrap methods are only necessary when the sample size is small.

1.7.3. Common population coefficient for determination

We can also define the population coefficient for determination in the given setup. Remember, that the unrestricted population coefficient for determination (1.2) for a single variable X_1 is defined by

$$R^2_{\mathbf{P}} = \iota^2_{X_1}(Y)/Var_{\mathbf{P}}(Y)$$

and quantifies how close $\iota_{X_1}(Y)$ is to its upper bound $\sqrt{Var_{\mathbf{P}}(Y)}$. In generalization to this we define the common population coefficient for determination of $X^{(1)}, ..., X^{(k)}$, which coincides with Pearson´s correlation ratio considered in Doksum and Samarov (1995), by

$$R^2_{\mathbf{P}} = \iota^2_{\mathbf{X}}(Y)/Var_{\mathbf{P}}(Y).$$

This expression quantifies how close $\iota_{\mathbf{X}}(Y)$ is to its upper bound $\sqrt{Var_{\mathbf{P}}(Y)}$. The natural restriction to linear perturbations leads to the linear common population coefficient for determination of $X^{(1)}, ..., X^{(k)}$

$$R^{lin^2}_{\mathbf{P}} = \iota^{lin^2}_{\mathbf{X}}(Y)/Var_{\mathbf{P}}(Y). \tag{1.22}$$

We can rewrite (1.22) as $R^{lin^2}_{\mathbf{P}} = Var_{\mathbf{P}}(P_{\mathcal{H}}Y)/Var_{\mathbf{P}}(Y)$ and estimate this by

$$\hat{R}^{lin^2}_{\mathbf{P}} = \hat{\iota}^{lin^2}_{\mathbf{X}}(Y)/\hat{\sigma}^2_Y, \tag{1.23}$$

where $\hat{\sigma}^2_Y = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$ and $\mathcal{H} = \text{span}(X^{(1)}, ..., X^{(k)})$.

1.7.4. Common absolute mean slope

We recall the definition (1.17) for the absolute mean slope of a singe covariate X_1:

$$\theta_{X_1}(Y) = \frac{\iota_{X_1}(Y)}{\iota_{X_1}(X_1)}.$$

When $X^{(1)}, ..., X^{(k)}$ are functions of X_1 we can use the common linear mean impact of $X^{(1)}, ..., X^{(k)}$ to approximate the the absolute mean slope of X_1, namely by the common linear absolute mean slope

$$\theta^{lin}_{\mathbf{X}}(Y) = \frac{\iota^{lin}_{\mathbf{X}}(Y)}{\iota_{X_1}(X_1)} = \frac{\iota^{lin}_{\mathbf{X}}(Y)}{\sqrt{Var_{\mathbf{P}}(X_1)}}, \tag{1.24}$$

which does not necessarily equal the linear slope of X_1 alone.

It can be estimated by

$$\hat{\theta}^{lin}_{\mathbf{X}}(Y) = \frac{\hat{\iota}^{lin}_{\mathbf{X}}(Y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2}}.$$

Note, that in theory we could also define a signed version of the polynomial based impact. Since the consideration of a signed impact is questionable in non-monotonous relationships this approach is not followed up further in this thesis.

The common absolute mean slope will be called "mean slope" in the sequel.

1.7.5. Bootstrap intervals for the common linear mean impact

As it was already said, the confidence intervals for the common linear mean impact based on the shrinkage-like approach of Section 1.7.2 tend to undercoverage, when the sample size is not sufficiently large and the mean impact is close (or equal) to zero. In the following we will give a theoretical justification for using bootstrap methods in the construction of confidence intervals for the common linear mean impact based on the variables $X^{(1)}, ..., X^{(k)}$ in order to overcome these undercoverage issues. Therefore, we will show that the conditions of the *"smooth function model"* of Hall (Hall (1988) and Hall (1992)) which is reviewed in Section A.3 are fulfilled.

Theorem 1.31. *If* $E_{\mathbf{P}}(|X^{(j)}X^{(l)}X^{(m)}X^{(o)}|) < \infty$ *and* $E_{\mathbf{P}}(|X^{(j)}X^{(l)}Y^2|) < \infty$ *for all* $1 \leq j, l, m, o \leq k$ *bootstrap* BC_a *and studentized bootstrap intervals for* $\iota^{lin^2}_{\mathbf{X}}(Y)$ *based on* $\hat{\iota}^{lin^2}_{\mathbf{X}}(Y)$ *are second order accurate.*

Proof. We show that $\hat{\iota}^{lin^2}_{\mathbf{X}}(Y)$ is a smooth function of arithmetic means of certain i.i.d. random vectors and $\iota^{lin^2}_{\mathbf{X}}(Y)$ is the same smooth function of the expectation of these variables. To show this we only need to show that the terms arising in (1.19), namely

$$\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i$$

are smooth functions of i.i.d. means. Since square and subtraction are smooth this would imply the smoothness of $\hat{\iota}^{lin^2}_{\mathbf{X}}(Y)$. We make the following considerations:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2 =& \frac{1}{n}\mathbf{Y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X}/n)^{-1}\frac{1}{n}\mathbf{X}^T\mathbf{Y} \\ =& V^T C^{-1} V,\end{aligned}$$

where

$$V = \frac{1}{n}\mathbf{X}^T\mathbf{Y} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)}Y_i, ..., \frac{1}{n}\sum_{i=1}^{n} X_i^{(k)}Y_i\right)^T$$

and

$$C^{-1} = (\mathbf{X}^T\mathbf{X}/n)^{-1} = \left[\left(\frac{1}{n}\sum_{l=1}^{n} X_l^{(i)}X_l^{(j)}\right)_{i,j=1,...,k}\right]^{-1}.$$

Hence, $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2$ is a polynomial in the entry of the mean of the i.i.d. vectors

$$Z_i = (X_i^{(1)}Y_i, ..., X_i^{(k)}Y_i, X_i^{(1)}X_i^{(1)}, ..., X_i^{(1)}X_i^{(k)}, ..., X_i^{(k)}X_i^{(k)})$$

for $i = 1, ..., n$. Since all partial derivatives of this quotient of polynomials are quotients of polynomials themselves they are also differentiable. Induction yields the smoothness of $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2$. Note that the denominator of $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i^2$, which is the determinant of C, is always greater than zero since it is the determinant of a (by assumption) non-singular matrix. It follows by similar arguments, that $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i$ is a smooth function of i.i.d. means. We have

$$\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i = \frac{1}{n}\mathbb{1}^T\mathbf{X}(\mathbf{X}^T\mathbf{X}/n)^{-1}\frac{1}{n}\mathbf{X}^T\mathbf{Y} = W^TC^{-1}V,$$

where C and V are defined as before and $\mathbb{1} = (1, ..., 1)^T$, consequently

$$W = \frac{1}{n}X^T\mathbb{1} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)}, ..., \frac{1}{n}\sum_{i=1}^{n} X_i^{(k)}\right)^T.$$

This implies that $\frac{1}{n}\sum_{i=1}^{n}(P_{\mathcal{M}}\mathbf{Y})_i$ also is a quotient of polynomials of means of i.i.d. random variables and therefore, by the argumentation above a smooth function in i.i.d. means. Note that the matrices W, C and V converge to the same matrices, but with expected values instead of arithmetic means as entries. Hence, since $\hat{\iota}_{\mathbf{X}}^{lin^2}(Y)$ is consistent for $\iota_{\mathbf{X}}^{lin^2}(Y)$, $\iota_{\mathbf{X}}^{lin^2}(Y)$ can be written as the same smooth function as $\hat{\iota}_{\mathbf{X}}^{lin^2}(Y)$ but of the corresponding expectations instead of arithmetic means. Thus, $\hat{\iota}_{\mathbf{X}}^{lin^2}(Y)$ fulfills the conditions of Halls smooth function model. This means that bootstrap BC_a and studentized bootstrap intervals for $\iota_{\mathbf{X}}^{lin^2}(Y)$ based on $\hat{\iota}_{\mathbf{X}}^{lin^2}(Y)$ are second order accurate. □

From these intervals we can derive second order accurate confidence intervals for

$\iota_{\mathbf{X}}^{lin}(Y)$ by choosing the lower bound

$$l_\alpha = \begin{cases} 0 & \text{if } l_\alpha^{boot} \leq 0 \\ \sqrt{l_\alpha^{boot}} & \text{if } l_\alpha^{boot} > 0 \end{cases},$$

where l_α^{boot} is the bootstrap confidence bound for $\iota_{\mathbf{X}}^{lin^2}(Y)$. The second order accuracy of this bounds is due to the second order accuracy of the bootstrap bound, the monotony of $\sqrt{\cdot}$ and the fact that the impact is non-negative.
Note, that if $\iota_{\mathbf{X}}^{lin}(Y) \neq 0$ the estimated linear common mean impact itself (instead of its squared version) fulfills the conditions of the smooth function model ($\sqrt{\cdot}$ is smooth on $\mathbb{R}^+$). Hence, when the common linear mean impact is strictly positive, bootstrap BC_a and studentized bootstrap intervals based on $\hat{\iota}_{\mathbf{X}}^{lin}(Y)$ are second order accurate for $\iota_{\mathbf{X}}^{lin}(Y) > 0$. However, when using bootstrap bounds based on the unsquared estimate $\hat{\iota}_{\mathbf{X}}^{lin}(Y)$ we have to make sure that $\iota_{\mathbf{X}}^{lin}(Y) > 0$. To ensure this, we can pre perform the test for the null-hypothesis $\iota_{\mathbf{X}}^{lin}(Y) = 0$ of Section 1.7.1 prior to the calculation of the confidence intervals. In Section 4 we compare the method of computing the confidence bounds via the squared estimate to the approach where we bootstrap the unsquared estimate with pre-performed test for $\iota_{\mathbf{X}}^{lin}(Y) = 0$ (Set the confidence bound to zero, when the test can not reject the null-hypothesis).

From the results above it follows that $\hat{\theta}_{\mathbf{X}}^{lin^2}(Y)$ also fulfills the smooth function model leading to second order accuracy of bootstrap BC_a and studentized bootstrap intervals. From these intervals we can obtain confidence bounds for $\theta_{\mathbf{X}}^{lin}(Y)$ by the same transformation as in Section 1.9.3 namely by

$$l_\alpha = \begin{cases} 0 & \text{if } l_\alpha^{boot} \leq 0 \\ \sqrt{l_\alpha^{boot}} & \text{if } l_\alpha^{boot} > 0 \end{cases},$$

where l_α^{boot} is the bootstrap confidence bound for $\theta_{\mathbf{X}}^{lin^2}(Y)$.

Note, that if $\theta_{\mathbf{X}}^{lin}(Y) > 0$ its estimate $\hat{\theta}_{\mathbf{X}}^{lin}(Y)$ fulfills the smooth function model. Hence, we could also compute bootstrap bounds based on the unsqared estimate $\hat{\theta}_{\mathbf{X}}^{lin}(Y)$ when pre-performing a test for $\iota_{\mathbf{X}}^{lin}(Y) = 0$ (which is essentially the same as to test for $\theta_{\mathbf{X}}^{lin}(Y) = 0$).
We can also show that in this setup the linear common population coefficient for determination (1.22) with its estimate (1.23) meets the conditions of the smooth function

model. To this end, we note that

$$\hat{R}^{lin^2}_{\mathbf{P}} = \hat{\iota}^{lin^2}_{\mathbf{X}}(Y)/\hat{\sigma}^2_Y,$$

with $\hat{\sigma}^2_Y = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$ contains the smooth function $\iota^{lin^2}_{\mathbf{X}}(Y)$. Obviously the expression $\hat{\sigma}^2_Y$ is a smooth function of means of i.i.d. random variables. Furthermore, it is consistent for $\sigma^2_Y = Var_{\mathbf{P}}(Y)$. Thus, the conditions of the smooth function model are met and bootstrap BC_a and studentized bootstrap intervals for $R^{lin^2}_{\mathbf{P}}$ are second order accurate.

1.8. Partial common mean impact

Again we consider the case where we want to quantify the association between a target variable Y and a set of covariates $\mathbf{X} = (X^{(1)}, ..., X^{(k)})$ which goes beyond the possible influence of further variables $\mathbf{Q} = (Q_1, ..., Q_l)$. We define the partial common mean impact by

$$\begin{aligned} &\iota_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l) \\ &= \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R}^{k+l}): E(\delta(\mathbf{X},\mathbf{Q}))=0,\ E(\delta^2(\mathbf{X},\mathbf{Q}))=1,\ E(\delta(\mathbf{X},\mathbf{Q})Q_j)=0\ \forall j=1,...,l} E_{\mathbf{P}}[Y\delta(\mathbf{X},\mathbf{Q})]. \quad (1.25) \end{aligned}$$

It describes the maximum change in the mean of Y, when the common density f of $(X^{(1)}, ..., X^{(k)}, Q_1, ..., Q_l)$ is changed to $f(1+\delta)$, where δ has mean zero and variance equal to one and the means of $(Q_1, ..., Q_l)$ remain unchanged.

When the variables $X^{(1)}, ..., X^{(k)}$ are functions of a single variable X_1 we can in generalization to the common mean slope (1.24) define a partial common mean slope by

$$\theta_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l) = \frac{\iota_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l)}{\iota_{X^{(1)},...,X^{(k)}}(X_1|Q_1, ..., Q_l)}.$$

1.9. Linear partial common impact analysis

1.9.1. Definition of the linear partial common mean impact

In order to avoid the problem of overfitting which arises in the estimation of the common partial mean impact (1.25), we regard the common linear influence of the set of covariates $X^{(1)}, ..., X^{(k)}$ which goes beyond the possible influence of $Q_1, ..., Q_l$. Hence, we regard

the linear partial common mean impact which is given by

$$\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \sup_{\delta\in\mathcal{H}\cap\mathcal{H}_2^{\perp},\, E_{\mathbf{P}}\{\delta^2(\mathbf{X},\mathbf{Q})\}=1} E_{\mathbf{P}}\{Y\delta(\mathbf{X},\mathbf{Q})\}, \tag{1.26}$$

where $\mathcal{H} = \text{span}(Q_1,...,Q_l,1,X^{(1)},...,X^{(k)})$ and $\mathcal{H}_2 = \text{span}(1,Q_1,...,Q_l)$. The linear partial mean impact $\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ describes in a sense the maximum change of the mean of Y when the density f of the variables $X^{(1)},...,X^{(k)},Q_1,...,Q_l$ in the population is changed to $(1+\delta)f$ in a way that δ is linear in $X^{(1)},...,X^{(k)},Q_1,...,Q_l$, $L^2(\mathbb{R}^{k+l})$-integrable with norm equal to one and the means of the covariates $Q_1,...,Q_l$ remain unchanged. Obviously the linear partial common mean impact (1.26) is a lower bound for the (unrestricted) partial common mean impact (1.25). In order to be able to calculate this impact, we make the following consideration.

Proposition 1.32. *With the above definitions of $\mathcal{H}$ and $\mathcal{H}_2$ we have that*

$$P_{\mathcal{H}\cap\mathcal{H}_2^{\perp}}Z = P_{\mathcal{H}}P_{\mathcal{H}_2^{\perp}}Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z,$$

for all $Z\in L^2(\mathbb{R}^{k+l})$.

Proof. First of all, since $\mathcal{H}_2\subset\mathcal{H}$ we have for all $Z\in L^2(\mathbb{R}^{k+l})$ that

$$\begin{aligned} P_{\mathcal{H}}P_{\mathcal{H}_2^{\perp}}Z =& P_{\mathcal{H}}Z - P_{\mathcal{H}}P_{\mathcal{H}_2}Z = P_{\mathcal{H}}Z - P_{\mathcal{H}_2}Z \\ &= P_{\mathcal{H}}Z - P_{\mathcal{H}_2}P_{\mathcal{H}}Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z. \end{aligned}$$

Thus, $P_{\mathcal{H}}P_{\mathcal{H}_2^{\perp}}Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z \in \mathcal{H}\cap\mathcal{H}_2^{\perp}$. This, together with the definition of the projection provides for all $U\in\mathcal{H}\cap\mathcal{H}_2^{\perp}$

$$\begin{aligned} E_{\mathbf{P}}(UP_{\mathcal{H}\cap\mathcal{H}_2^{\perp}}Z) =& E_{\mathbf{P}}(UZ) \\ =& E_{\mathbf{P}}(UP_{\mathcal{H}}Z) \\ =& E_{\mathbf{P}}(UP_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z) = E_{\mathbf{P}}(UP_{\mathcal{H}}P_{\mathcal{H}_2^{\perp}}Z). \end{aligned}$$

Thus, since the projection is uniquely defined, we obtain

$$P_{\mathcal{H}\cap\mathcal{H}_2^{\perp}}Z = P_{\mathcal{H}}P_{\mathcal{H}_2^{\perp}}Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z.$$

□

We can rewrite the impact (1.26) as

$$\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \sup_{\delta\in L^2_{\mathbf{P}},\ Var_{\mathbb{P}}(\delta(\mathbf{X},\mathbf{Q}))>0} E_{\mathbf{P}}\left(Y\frac{P_{\tilde{\mathcal{H}}}\delta(\mathbf{X},\mathbf{Q})}{\sqrt{E_{\mathbf{P}}\{(P_{\tilde{\mathcal{H}}}\delta(\mathbf{X},\mathbf{Q}))^2\}}}\right),$$

where $\tilde{\mathcal{H}} = \mathcal{H}\cap\mathcal{H}_2^{\perp}$. The same argumentation as in Section 1.7 gives

$$\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \sqrt{E_{\mathbf{P}}\{(P_{\tilde{\mathcal{H}}}Y)^2\}} = \sqrt{Var_{\mathbf{P}}(P_{\tilde{\mathcal{H}}}Y)}.$$

Proposition 1.33. *With the definitions of this section we have that*

$$\tilde{\mathcal{H}} = \mathcal{H}\cap\mathcal{H}_2^{\perp} = \mathcal{H}_1,$$

where $\mathcal{H}_1 = \text{span}(P_{\mathcal{H}_2^{\perp}}X^{(1)},...,P_{\mathcal{H}_2^{\perp}}X^{(k)})$.

Proof. We start by showing that $\tilde{\mathcal{H}}\subseteq\mathcal{H}_1$. Let $Z\in\tilde{\mathcal{H}}$, then we have trivially that $P_{\tilde{\mathcal{H}}}Z = Z$. By Proposition 1.32 we can rewrite this as $P_{\tilde{\mathcal{H}}}Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z = Z$. This means that we have

$$Z = P_{\mathcal{H}_2^{\perp}}P_{\mathcal{H}}Z = P_{\mathcal{H}_2^{\perp}}\left(\sum_{j=1}^{k}\eta_j X^{(j)} + \eta_0 + \sum_{m=1}^{l}\eta_{k+m}Q_l\right)$$

for some coefficients $\eta_0,...,\eta_{j+m}$,

$$= \sum_{j=1}^{k}\eta_j P_{\mathcal{H}_2^{\perp}}X^{(j)} \in \mathcal{H}_1.$$

Consequently, we obtain $\tilde{\mathcal{H}}\subseteq\mathcal{H}_1$. Next we show the reverse statement $\mathcal{H}_1\subseteq\tilde{\mathcal{H}}$. Since we obviously have that $\mathcal{H}_1\subseteq\mathcal{H}_2^{\perp}$ it suffices to show that $\mathcal{H}_1\subseteq\mathcal{H}$. To this end let $Z\in\mathcal{H}_1$, this means that we can write for some coefficients $\nu_1,...,\nu_j$ and $\zeta_0,...,\zeta_l$

$$\begin{aligned} Z &= \sum_{j=1}^{k}\nu_j P_{\mathcal{H}_2^{\perp}}X^{(j)} = \sum_{j=1}^{k}\nu_j(X^{(j)} - P_{\mathcal{H}_2}X^{(j)}) \\ &= \sum_{j=1}^{k}\nu_j(X^{(j)} - \zeta_0 - \sum_{m=1}^{l}\zeta_m Q_m) \\ &= \sum_{j=1}^{k}\nu_j X_j - \zeta_0\sum_{j=1}^{k}\nu_j - \sum_{j=1}^{k}\nu_j\sum_{m=1}^{l}\zeta_m Q_m \in \mathcal{H}. \end{aligned}$$

From this we obtain $\mathcal{H}_1 \subseteq \mathcal{H}$ and consequently $\mathcal{H}_1 \subseteq \tilde{\mathcal{H}}$. Together with $\tilde{\mathcal{H}} \subseteq \mathcal{H}_1$ the assertion follows. □

With the result of Proposition 1.33, we can also write

$$\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \sqrt{Var_{\mathbf{P}}\left\{P_{\mathcal{H}_1}Y\right\}},$$

where $\mathcal{H}_1 = \text{span}(P_{\mathcal{H}_2^\perp}X^{(1)},...,P_{\mathcal{H}_2^\perp}X^{(k)})$.

In the case of $X^{(1)},...,X^{(k)}$ being functions of a single variable X_1 (with $X^{(j)} = X_1$ for one j), the linear partial common mean slope is defined as

$$\theta^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \frac{\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)}{\iota^{lin}_{X^{(1)},...,X^{(k)}}(X_1|Q_1,...,Q_l)} = \frac{\sqrt{Var_{\mathbf{P}}\left(P_{\mathcal{H}_1}Y\right)}}{\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}_2^\perp}X_1)}}.$$

1.9.2. Estimation of the linear partial common mean impact

Given i.i.d. observations $\mathbf{Y} = (Y_1,...,Y_n)^T$, $\mathbf{Q}_j = (Q_{1j},...,Q_{nj})^T$, $\mathbf{X}^{(m)} = (X_1^{(m)},...,X_n^{(m)})^T$ of Y, Q_j and $X^{(m)}$ for $j = 1,...,l$ and $m = 1,...,k$ we estimate $\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ by

$$\hat{\iota}^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \sqrt{n^{-1}\|P_{\mathcal{M}_1}\mathbf{Y}\|^2} = \frac{1}{\sqrt{n}}\|P_{\mathcal{M}_1}\mathbf{Y}\|,$$

where $\mathcal{M}_1 = \text{span}\left(\hat{\mathbf{X}}^{(1)},...,\hat{\mathbf{X}}^{(k)}\right)$, with $\hat{\mathbf{X}}^{(j)} = P_{\mathcal{M}_2^\perp}\mathbf{X}^{(j)}$ and $M_2 = \text{span}\left(\mathbf{1},\mathbf{Q}_1,...,\mathbf{Q}_l\right)$.

When $X^{(1)},...,X^{(k)}$ are functions of a single variable X_1 (with $X^{(j)} = X_1$ for one j), the linear partial common mean slope is well defined and can be estimated by

$$\hat{\theta}^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \frac{\hat{\iota}^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)}{\sqrt{\frac{1}{n}\sum_{i=1}^n\left((P_{\mathcal{M}_2^\perp}\mathbf{X}_1)_i - \overline{P_{\mathcal{M}_2^\perp}\mathbf{X}_1}\right)^2}}.$$

Here, $\mathbf{X}_1$ is the vector of i.i.d. observations of the variable X_1.

1.9.3. Bootstrap confidence intervals in linear partial common impact analysis

We will show that $\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ meets the conditions of the smooth function model of Hall described in Section A.3.3 which implies that bootstrap-BC_a and studen-

tized bootstrap intervals are second order accurate. From these intervals one can easily derive confidence intervals for $\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$.

Theorem 1.34. *Bootstrap BC_a and studentized bootstrap confidence intervals for $\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ based on $\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ are second order accurate.*

Proof. With $\hat{\mathbf{X}}$ being the matrix with j-th column $\hat{\mathbf{X}}^{(j)}$ we have that

$$\begin{aligned}\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) =& \frac{1}{n}\mathbf{Y}^T \underbrace{\hat{\mathbf{X}}\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^T}_{P_{\mathcal{M}_1}}\hat{\mathbf{X}}\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^T\mathbf{Y} \\ =& \frac{1}{n}\mathbf{Y}^T\hat{\mathbf{X}}\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}/n\right)^{-1}\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{Y}.\end{aligned}$$

We can rewrite $\hat{\mathbf{X}}^{(j)}$ as

$$\hat{\mathbf{X}}^{(j)} = \mathbf{X}^{(j)} - \mathbf{Q}(\mathbf{Q}^T\mathbf{Q}/n)^{-1}\frac{1}{n}\mathbf{Q}^T\mathbf{X}^{(j)},$$

where $\mathbf{Q} = (\mathbf{1}, \mathbf{Q}_1, ..., \mathbf{Q}_l)$. We can see that

$$\frac{1}{n}\mathbf{Q}^T\mathbf{X}^{(j)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^{(j)} \\ \frac{1}{n}\sum_{i=1}^n X_i^{(j)}Q_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^{(j)}Q_{il} \end{pmatrix}$$

is a vector of means of i.i.d. random variables. Obviously, $(\mathbf{Q}^T\mathbf{Q}/n)^{-1}$ is a matrix whose entries are smooth functions of means of i.i.d. random variables (this was already shown in Section 1.7). Hence, we can write

$$(\mathbf{Q}^T\mathbf{Q}/n)^{-1}\frac{1}{n}\mathbf{Q}^T\mathbf{X}^{(j)} = \left(f_m^{(j)}\right)_{m=1,...,l+1},$$

where $f_m^{(j)}$ are smooth functions of means of i.i.d. random variables. Consequently $\hat{\mathbf{X}}^{(j)}$ is given by

$$\hat{\mathbf{X}}^{(j)} = \mathbf{X}^{(j)} - \mathbf{Q}\begin{pmatrix} f_1^{(j)} \\ \vdots \\ f_{l+1}^{(j)} \end{pmatrix} = \left\{ X_i^{(j)} - \left(f_1^{(j)} + \sum_{m=2}^{l+1} f_m^{(j)} Q_{i(m-1)} \right) \right\}_{i=1,...,n}.$$

This implies that

$$\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{Y} = \left\{\frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}Y_i - \left(f_1^{(j)}\frac{1}{n}Y_i + \sum_{m=2}^{l+1} f_m^{(j)}\frac{1}{n}\sum_{i=1}^{n} Y_i Q_{i(m-1)}\right)\right\}_{j=1,\ldots,k}$$

is a vector of smooth function of means of i.i.d. random variables. Analogously it can be shown that $\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}/n\right)^{-1}$ is a matrix with smooth functions of means i.i.d. random variables as entries. From this it follows that $\hat{\iota}^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l)$ meets the conditions of the smooth function model. If we can show that $\hat{\iota}^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l) \to \iota^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l)$ the considerations in Hall (1988) and Hall (1992) concerning the smooth function model give the second order accuracy of the BC_a and the studentized bootstrap interval for $\iota^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l)$. For the consistency we make the following considerations. A direct conclusion of the proof of Lemma 1.13 (which is also given in Appendix B) is that

$$\begin{aligned}
&\iota^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l)\\
&= \left\{E_{\mathbf{P}}\left(P_{\mathcal{H}_2^\perp}X^{(j)}\right)_{j=1,\ldots,k}\right\}\left\{E_{\mathbf{P}}\left(P_{\mathcal{H}_2^\perp}X^{(a)}P_{\mathcal{H}_2^\perp}X^{(b)}\right)_{a,b}\right\}^{-1}\left\{E_{\mathbf{P}}\left(P_{\mathcal{H}_2^\perp}X^{(j)}\right)\right\}^T_{j=1,\ldots,k}.
\end{aligned} \tag{1.27}$$

As mentioned before we have

$$\hat{\iota}^{lin^2}_{X^{(1)},\ldots,X^{(k)}}(Y|Q_1,\ldots,Q_l) = \frac{1}{n}\mathbf{Y}^T\hat{\mathbf{X}}\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}/n\right)^{-1}\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{Y},$$

with

$$\hat{\mathbf{X}} = \mathbf{X} - \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{X}, \quad \text{where } \mathbf{X} = (\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(k)}).$$

Using this notation we obtain

$$\begin{aligned}
\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{Y} =& \frac{1}{n}\left(\langle\mathbf{X}^{(j)},\mathbf{Y}\rangle\right)_{j=1,\ldots,k} - \frac{1}{n}\left(\langle\mathbf{X}^{(j)},P_{\mathcal{M}_2}\mathbf{Y}\rangle\right)_{j=1,\ldots,k}\\
=& \frac{1}{n}\left(\langle\mathbf{X}^{(j)},\mathbf{Y}\rangle\right)_{j=1,\ldots,k} - \frac{1}{n}\left(\langle P_{\mathcal{M}_2}\mathbf{X}^{(j)},\mathbf{Y}\rangle\right)_{j=1,\ldots,k}\\
=& \frac{1}{n}\left(\langle P_{\mathcal{M}_2^\perp}\mathbf{X}^{(j)},\mathbf{Y}\rangle\right)_{j=1,\ldots,k}\\
\xrightarrow{p}& E_{\mathbf{P}}\left(P_{\mathcal{H}_2^\perp}X^{(j)}Y\right)_{j=1,\ldots,k},
\end{aligned} \tag{1.28}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathbb{R}^n$. The convergence in (1.28) follows instantly

from the convergence of the coefficients of the projection in $\mathbb{R}^n$ to those of the projection in L^2 (see for example Lemma 1.13). Let $(\hat{\xi}_1, ..., \hat{\xi}_{l+1})$ be the coefficients of the projection of $\mathbf{X}^{(j)}$ onto $\mathcal{M}_2$ and $(\xi_1, ..., \xi_{l+1})$ the coefficients of the projection of $X^{(j)}$ onto $\mathcal{H}_2$. Then we have

$$\frac{1}{n}\langle P_{\mathcal{M}_2}\mathbf{X}^{(j)}, \mathbf{Y}\rangle = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i\hat{\xi}_1 + \sum_{j=2}^{l+1}\hat{\xi}_j Q_{i(j-1)}Y_i\right) = \hat{\xi}_1\frac{1}{n}\sum_{i=1}^{n}Y_i + \sum_{j=2}^{l+1}\hat{\xi}_j\frac{1}{n}\sum_{i=1}^{n}Q_{i(j-1)}Y_i$$

$$\xrightarrow{p}\xi_1 E_{\mathbf{P}}(Y) + E_{\mathbf{P}}\left(Y\sum_{j=2}^{l+1}\xi_j Q_{j-1}\right) = E_{\mathbf{P}}\left(YP_{\mathcal{H}_2}X^{(j)}\right),$$

thus, (1.28) holds. For $\frac{1}{n}\hat{\mathbf{X}}^T\hat{\mathbf{X}}$ we obtain

$$\begin{aligned}
\frac{1}{n}\hat{\mathbf{X}}^T\hat{\mathbf{X}} =& \frac{1}{n}\left(\mathbf{X}^T - \mathbf{X}^T\mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\right)\left(\mathbf{X} - \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{X}\right)\\
=& \frac{1}{n}\mathbf{X}^T\mathbf{X} - \frac{1}{n}\mathbf{X}^T\mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{X}\\
=& \left(\frac{1}{n}\langle\mathbf{X}^{(a)}, \mathbf{X}^{(b)}\rangle\right)_{a,b} - \left(\frac{1}{n}\langle\mathbf{X}^{(a)}, P_{\mathcal{M}_2}\mathbf{X}^{(b)}\rangle\right)_{a,b}\\
\xrightarrow{p}& \left\{E_{\mathbf{P}}\left(X^{(a)}X^{(b)}\right)\right\}_{a,b} - \left\{E_{\mathbf{P}}\left(X^{(a)}P_{\mathcal{H}_2}X^{(b)}\right)\right\}_{a,b} \qquad (1.29)\\
=& \left\{E_{\mathbf{P}}\left(X^{(a)}P_{\mathcal{H}_2^\perp}X^{(b)}\right)\right\}_{a,b} = \left\{E_{\mathbf{P}}\left(P_{\mathcal{H}_2^\perp}X^{(a)}P_{\mathcal{H}_2^\perp}X^{(b)}\right)\right\}_{a,b}, \qquad (1.30)
\end{aligned}$$

where the convergence in (1.29) follows analogous to that in (1.28). Combining (1.27), (1.28) and (1.30) gives

$$\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l) \xrightarrow{p} \iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l).$$

Thus, bootstrap BC_a and studentized bootstrap confidence intervals for $\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l)$ based on $\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l)$ are second order accurate. □

From these intervals we can derive second order accurate confidence intervals for $\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1, ..., Q_l)$ by choosing the lower bound

$$l_\alpha = \begin{cases} 0 & \text{if } l_\alpha^{boot} \leq 0 \\ \sqrt{l_\alpha^{boot}} & \text{if } l_\alpha^{boot} > 0 \end{cases}, \qquad (1.31)$$

where l_α^{boot} is the bootstrap confidence bound for $\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$. The second order accuracy of this bounds is due to the second order accuracy of the bootstrap bound, the monotony of $\sqrt{\cdot}$ and the fact that the impact is non-negative. Analogously to Section 1.7.5 one can see that the unsquared estimate $\hat{\iota}^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ also fulfills the smooth function model, if $\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) \neq 0$. Thus, we could also use bootstrap bounds based on $\hat{\iota}^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ when performing a test for the null-hypothesis $\iota^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = 0$ prior to the calculation of the confidence bounds (set the bound to be zero if the test does not reject).

In the same manner it can be shown that bootstrap BC_a and studentized bootstrap intervals for $\theta^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ based on $\hat{\theta}^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$ are second order accurate. Using the same transformation as in (1.31) we obtain confidence intervals for $\theta^{lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)$.

1.9.4. Alternative Approach

Alternative partial mean impact

Another approach to quantify the influence of covariates $X^{(1)},...,X^{(k)}$ on Y which goes beyond the possible influence of the other covariates $Q_1,...,Q_l$ is to regard the difference in the common influence of $X^{(1)},...,X^{(k)},Q_1,...,Q_l$ and the common influence of $Q_1,...,Q_l$. Hence we look at

$$\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota_{Q_1,...,Q_l}(Y). \tag{1.32}$$

This difference describes the additional maximum change in the mean of Y, when changing the distribution of $X^{(1)},...,X^{(k)},Q_1,...,Q_l$ instead of $Q_1,...,Q_l$ in the population. It is therefore a measure of the influence of $X^{(1)},...,X^{(k)}$ which goes beyond the influence of $Q_1,...,Q_l$. Moving to the linear versions of the common impacts leads to

$$\iota^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin}_{Q_1,...,Q_l}(Y).$$

Note that this difference is not necessarily a lower bound for the difference of the unrestricted impacts (1.32). To obtain a lower bound for (1.32) one would need to have a conservative estimate of $\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y)$ (which can be done by using the linear impact) and a consistent or anticonservative estimate of $\iota_{Q_1,...,Q_l}(Y)$. Moreover, to be able to make use of the smooth function model asymptotic, we need to look at the

difference of the squared impacts

$$\iota^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin^2}_{Q_1,...,Q_l}(Y).$$

Since $\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y)$ and $\hat{\iota}^{lin^2}_{Q_1,...,Q_l}(Y)$ both fulfill the smooth function model of Hall (1988) and Hall (1992), their difference does so as well. This implies that bootstrap BC_a and studentized bootstrap intervals for $\iota^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin^2}_{Q_1,...,Q_l}(Y)$ based on $\hat{\iota}^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \hat{\iota}^{lin^2}_{Q_1,...,Q_l}(Y)$ are second order accurate. When regarding the difference of the squared impacts it follows from orthogonality that

$$\iota^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin^2}_{Q_1,...,Q_l}(Y) = \iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l). \tag{1.33}$$

Similar to the previous cases the smooth function model also applies to the difference of the unsquared impacts, when both $\iota^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y)$ and $\iota^{lin}_{Q_1,...,Q_l}(Y)$ are strictly positive.

Alternative partial mean slope

In the setup where $X^{(1)},...,X^{(k)}$ are functions of a single covariate X_1 (assuming that $X^{(j)} = X_1$ for one j), the partial absolute common mean slope was defined as

$$\theta_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \frac{\iota_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)}{\iota_{X^{(1)},...,X^{(k)}}(X_1|Q_1,...,Q_l)}.$$

A straightforward application of the alternative approach to is then given by

$$\begin{aligned}\theta^{alt}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) &= \frac{\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota_{Q_1,...,Q_l}(Y)}{\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(X_1) - \iota_{Q_1,...,Q_l}(X_1)} \\ &= \frac{\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota_{Q_1,...,Q_l}(Y)}{SD_{\mathbf{P}}(X_1) - \iota_{Q_1,...,Q_l}(X_1)}.\end{aligned}$$

It shows how much more the mean of Y can be changed by adding $X^{(1)},...,X^{(k)}$ to the set of covariates relative to the excess in the maximum change of the mean of X_1 when adding those covariates. A linear version of this parameter is given by

$$\begin{aligned}\theta^{alt,\ lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) &= \frac{\iota^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin}_{Q_1,...,Q_l}(Y)}{\iota^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(X_1) - \iota^{lin}_{Q_1,...,Q_l}(X_1)} \\ &= \frac{\iota^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin}_{Q_1,...,Q_l}(Y)}{SD_{\mathbf{P}}(X_1) - \iota^{lin}_{Q_1,...,Q_l}(X_1)}.\end{aligned}$$

This can be estimated by

$$\hat{\theta}^{alt,\ lin}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l) = \frac{\hat{\iota}^{lin}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \hat{\iota}^{lin}_{Q_1,...,Q_l}(Y)}{\widehat{SD}_{\mathbf{P}}(X_1) - \hat{\iota}^{lin}_{Q_1,...,Q_l}(X_1)},$$

where $\widehat{SD}_{\mathbf{P}}(X_1) = n^{-1}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2$. If we want to use the smooth function model we need look at the squares of the impacts again. Thus, we consider the parameter

$$\frac{\iota^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) - \iota^{lin^2}_{Q_1,...,Q_l}(Y)}{Var_{\mathbf{P}}(X_1) - \iota^{lin^2}_{Q_1,...,Q_l}(X_1)}.$$

Orthogonality provides that this equals

$$\frac{\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l)}{\iota^{lin^2}_{X^{(1)},...,X^{(k)}}(X_1|Q_1,...,Q_l)} = \theta^{lin^2}_{X^{(1)},...,X^{(k)}}(Y|Q_1,...,Q_l).$$

Thus, in the linear case, the alternative approach leads to similar results as the classical approach. However, when moving away from linear restrictions, as will be done in Section 3.2 this is no longer the case.

Partial coefficient for determination

The alternative approach enables us to define a partial measure of determination, namely by

$$\frac{R^2_{X^{(1)},...,X^{(k)},Q_1,...,Q_l} - R^2_{Q_1,...,Q_l}}{1 - R^2_{Q_1,...,Q_l}}. \tag{1.34}$$

Note, that this quantity is not defined for $R^2_{Q_1,...,Q_l} = 1$, however, in this case all variation of $E_{\mathbf{P}}(Y|X)$ can be explained by $Q_1,...,Q_l$ and one would not need to add further variables to the set of explanatory variables. Hence, without loss of generality we will assume $R^2_{Q_1,...,Q_l} < 1$ in the sequel. Furthermore, this definition is very similar to the partial coefficient for determination of the linear regression which is for example defined in Paulson (2007, Ch. 5). This quantity explains how much of the variation of Y that could not be explained by $Q_1,...,Q_l$ can be explained by adding $X^{(1)},...,X^{(k)}$ to the model. It lies between zero and one and is zero if and only if $\iota_{X^{(1)},...,X^{(k)},Q_1,...,Q_l}(Y) = \iota_{Q_1,...,Q_l}(Y)$, which can be interpreted as that $X^{(1)},...,X^{(k)}$ do not have an effect on Y which goes beyond the effect of $Q_1,...,Q_l$. It is one if $R^2_{X^{(1)},...,X^{(k)},Q_1,...,Q_l} = 1$, which means that adding $X^{(1)},...,X^{(k)}$ to the set of explanatory variables explains all variation of $E_{\mathbf{P}}(Y|X)$ that could not already be explained by $Q_1,...,Q_l$ alone. The linear version

of (1.34) is given by

$$\frac{R^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l} - R^{lin^2}_{Q_1,...,Q_l}}{1 - R^{lin^2}_{Q_1,...,Q_l}}$$

and can be estimated by its natural estimate

$$\frac{\hat{R}^{lin^2}_{X^{(1)},...,X^{(k)},Q_1,...,Q_l} - \hat{R}^{lin^2}_{Q_1,...,Q_l}}{1 - \hat{R}^{lin^2}_{Q_1,...,Q_l}}.$$

In this setup as well, the smooth function model holds, leading to second order accurate bootstrap BC_a and studentized bootstrap intervals.

1.9.5. Example

In this section we want to investigate the difference between the two approaches to quantify the influence of X_1 on Y which goes beyond the possible influence of $X_2, ..., X_k$ of the Sections 1.9.1 and 1.9.4. In this example we examine the case of two covariates X_1 and X_2. For simplicity we only look at linear influences. Let

$$Y = a_1 X_1 + a_2 X_2 + \epsilon,$$

where ϵ has mean zero and is independent of X_1 and X_2. Furthermore, assume that $E_{\mathbf{P}}(X_i) = 0$ and $Var_{\mathbf{P}}(X_i) = 1$ for $i = 1, 2$ as well as $Corr_{\mathbf{P}}(X_1, X_2) = \rho$. The common linear mean impact of X_1 and X_2 on Y is then given as

$$\iota^{lin}_Y(X_1, X_2) = \sqrt{Var_{\mathbf{P}}\left(a_1 X_1 + a_2 X_2\right)} = \sqrt{a_1^2 + 2a_1 a_2 \rho + a_2^2}.$$

The (linear) mean impact of X_2 on Y can be seen to equal

$$\iota^{lin}_Y(X_2) = \left| E_{\mathbf{P}}\left(Y \frac{X_2 - E_{\mathbf{P}}(X_2)}{\sqrt{Var_{\mathbf{P}}(X_2)}} \right) \right| = \left| a_1 \rho + a_2 \right|.$$

Hence the alternative approach to the quantification of the influence of X_1 on Y which goes beyond the possible influence of X_2 gives

$$\iota^{lin}_Y(X_1, X_2) - \iota^{lin}_Y(X_2) = \sqrt{a_1^2 + 2a_1 a_2 \rho + a_2^2} - \left| a_1 \rho + a_2 \right|.$$

Note that, if $a_1 = 0$ also $\iota^{lin}_Y(X_1, X_2) - \iota^{lin}_Y(X_2) = 0$. However, if $a_1 \neq 0$ this expression depends on the coefficient a_2 of X_2.

The partial mean impact of interest is in this setup given by

$$\iota^{lin}_{Y:X_2}(X_1) = \left| E_{\mathbf{P}} \left(Y \frac{P_{\mathcal{H}_2^\perp} X_1}{\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}_2^\perp} X_1)}} \right) \right| = |a_1| \sqrt{1-\rho^2},$$

where $\mathcal{H}_2 = \text{span}(1, X_2)$. One can see that the partial mean impact is independent of a_2 and is therefore preferred to the alternative approach.

In simulations no substantial differences between the alternative approach and the common partial mean impact approach could be discovered for the scenario of this example. Therefore, the simulation results are not shown here. However, in Section 4 additional simulation results for other scenarios are provided.

1.10. Application of Impact analysis to data with a zero-inflated covariate

Another example for the application of the impact analysis and the smooth function model is given, when the variable X has a compound distribution with a probability mass at zero and an otherwise continuous distribution. We restrict ourselves to the case where we have a metric target variable Y and a single independent variable X. In this case an ordinary linear regression is questionable, because the part of the data for which X is zero has a strong influence on the fit for the part of the data where $X \neq 0$. Hence, the results of a linear regression are difficult to interpret. With the help of the impact analysis we can overcome this problem. We do so by estimating the mean impact (conservatively) by

$$\hat{\iota}_X(Y) = \frac{1}{n} \sum_{i=1}^{n} Y_i \frac{\hat{Y}_i - \bar{\hat{Y}}}{\sqrt{n^{-1} \sum_{i=1}^{n} (\hat{Y}_i - \bar{\hat{Y}})^2}}, \tag{1.35}$$

where the prognoses $\hat{Y}_i$ are equal to the mean of all observations Y_i for which $X_i = 0$, if $X_i = 0$ and equal to the prognoses of a linear regression with all data-points for which $X \neq 0$, if $X \neq 0$. This means that we split the data in the parts where $X_i = 0$ and $X_i \neq 0$ and fit different models in each part. Formally, we can obtain these prognoses by a linear regression with covariates $\mathbb{1}_{\{X=0\}}, \mathbb{1}_{\{X\neq 0\}}, X$ without intercept. Obviously the estimate (1.35) is the estimate of the common linear mean impact of $\mathbb{1}_{\{X=0\}}, \mathbb{1}_{\{X\neq 0\}}, X$. Hence, by the argumentation of Section 1.7 we obtain that its square fulfills the smooth function model and that it is consistent for

$$\iota_{\mathbb{1}_{\{X=0\}}, \mathbb{1}_{\{X\neq 0\}}, X}(Y) = \sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}} Y)},$$

with $\mathcal{H} = \text{span}(\mathbb{1}_{\{X=0\}},\ \mathbb{1}_{\{X\neq 0\}},\ X)$. Hence Bootstrap BC_a and studentized bootstrap intervals are second order accurate for the squared common impact. Transformation of the confidence bounds yields confidence bounds for the common mean impact itself.

Note that the linear fit in the data where $X \neq 0$ can be generalized to non-linear fits of polynomial regression and spline-methods with fixed knots (or any other additive model, namely by adding the required basis terms to the model) without affecting the theoretical results. When doing so, one has to keep in mind, that it is necessary to multiply the resulting basis-terms by $\mathbb{1}_{\{X\neq 0\}}$ in order to perform the fitting only in the subset of the data, where $X \neq 0$. Another advantage of the splitting of the data set is that we can now use transformations to the data in one subset of the data which was not possible on the whole data set (e.g. we could use a log-transformation for $X \neq 0$ and then fit models in $\log(X)$). This means that the impact analysis gives us the ability to interpret models (namely by the estimated mean impact as the inner product of Y and the standardized prognoses) that were hardly interpretable before.

Table 8 gives the simulation results of 10,000 simulation runs with $n = 100$ observations. Compared are the classical linear regression with robust variance estimate and the application of the mean impact analysis described in this section (prognoses taken from the linear regression independent variables $\mathbb{1}_{\{X=0\}},\ \mathbb{1}_{\{X\neq 0\}},\ X$). The test for the impact analysis is based on basic bootstrap intervals with two pre-performed tests (to check if the impact is larger than zero). Performed were the test of Section 1.7.1 and a global F-test from linear regression. Further details on this tests can be found in Section 4.1.2, where different methods for the calculation of confidence intervals for the linear common mean impact are compared. One can see that in the models I-III (defined below) the impact analysis outperforms the linear regression in terms of rejection probability by $9\%(Model\ II) - 84\%(Model\ I)$. In Model IV the coefficient from the linear regression and the mean impact are both equal to zero. Therefore, the given rejection probability is equal to the type-I-error. One can see that linear regression suffers from slight type-I-error inflation, whereas the impact analysis does not.

Model	p_0	Rejection Probability		Parameter Estimate	
		Linear model	Impact analysis	Regression coefficient	Mean impact
I	0.3	0.106	1.000	0.066	0.648
I	0.5	0.159	1.000	-0.156	0.612
II	0.3	0.898	0.983	0.534	0.512
II	0.5	0.740	0.935	0.424	0.437
III	0.3	0.876	1.000	1.339	1.617
III	0.5	0.528	1.000	0.885	1.359
IV	0.3	0.060	0.039	0.001	0.126
IV	0.5	0.067	0.044	0.005	0.127

Table 8: Simulation results comparing classical linear regression with the robust variance estimate to impact analysis in four different set-ups (Model I: $Y = 2 \cdot \mathbb{1}_{\{X=0\}} + X + \epsilon$; Model II: $Y = \mathbb{1}_{\{X=0\}} + X + \epsilon$; Model III: $Y = 2 \cdot \mathbb{1}_{\{X=0\}} + X^2 + \epsilon$; Model IV: $Y = \epsilon$). In each scenario $X = 0$ with probability p_0 and follows a log-normal distribution otherwise. The error $\epsilon \sim N(0, 1)$ is independent from X. Given are the rejection probabilities (tests performed at significance level 0.05) of the hypothesis that the parameters (the regression coefficient respectively the mean impact) are zero.

2. Non-linear impact analysis

The theoretical framework derived in White (1980a), White (1980b), Scharpenberg (2012) and Brannath and Scharpenberg (2014) gives a justification for using linear regression techniques even when the assumptions of the linear model are not valid. In this case one estimates the linear impact, which is a lower bound for the mean impact. This implies that even when the assumptions of linear regressions are violated using linear regression techniques in estimation of the mean impact is a conservative method.

In the further course of this thesis we will use non-linear and non-parametric regression techniques such as Splines and Kernel Smoothers (cf. Sections A.1.2 and A.1.1) to estimate non-linear versions of the mean impact. We will restrict ourselves to the case of one observed covariate and, since according to (4) of Theorem 1.2, the mean impact is given by

$$E_{\mathbf{P}}\left(Y\frac{E_{\mathbf{P}}(Y|X)-E_{\mathbf{P}}(Y)}{\sqrt{Var_{\mathbf{P}}(E_{\mathbf{P}}(Y|X))}}\right),$$

estimate the mean impact by

$$\hat{\iota}_X^{\mathcal{R}}(Y)=\frac{1}{n}\sum_{i=1}^{n}Y_i\frac{\hat{\delta}(X_i)-n^{-1}\sum_{j=1}^{n}\hat{\delta}(X_j)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\delta}(X_i)-n^{-1}\sum_{j=1}^{n}\hat{\delta}(X_j)\right\}^2}},$$

where $\hat{\delta}(X_i)$ will be estimates of $E(Y|X)$ at the X_i obtained by methods like natural cubic splines, polynomial regression or kernel smoothers. Furthermore, we will derive the asymptotic distribution of these estimators so that confidence intervals for the mean impact can be derived.

Another advantage of using non-linear or non-parametric regression techniques to estimate the mean impact is that users of these methods can easily compute an estimate and a confidence bound for the mean impact when fitting curves to their data. Hence, whenever fitting curves to the data we get a parameter quantifying the influence of the covariate X on Y.

2.1. Impact analysis based on polynomials and splines

With the help of the restricted common mean impact defined in Section 1.7 we are able to investigate the common linear influence of two ore more covariates $X^{(1)},...,X^{(k)}$.

An application of this scenario involves polynomial fitting. In this case the variables $X^{(1)},...,X^{(k)}$ are taken to be $1, X, X^2,...,X^{(k-1)}$. With this set of variables we can detect

polynomial influences of X up to $(k-1)$st order. We expect a restricted impact based on polynomials to be closer to the true impact than the linear mean impact. Another way for covering non-linear influences of a covariate X on the target variable Y enabled by the setup provided here is fitting natural splines. For natural cubic splines with fixed knot sequence $\zeta_1, ..., \zeta_m$ we obtain according to (Hastie et al., 2001, p.121 f) that $X^{(1)}, ..., X^{(k)} = N_1(X), ..., N_{m+2}(X)$, where

$$N_1(X) = 1,\ N_2(X) = X,\ N_{l+2}(X) = d_l(X) - d_{m-1}(X),$$

for $l = 1, ..., m$ with

$$d_l(X) = \frac{(X-\zeta_l)_+^3 - (X-\zeta_m)_+^3}{\zeta_m - \zeta_l},$$

as mentioned in Appendix A.1.2. The problem of how to choose the knot sequence immediately arises. A data dependent choice of the knots, like empirical quantiles of X will usually cause a violation of the smooth function model, since then $N_{l+2}(X_i)$ are not i.i.d. anymore. This means that the theoretical results derived here are not valid anymore. However, simulations indicate that the spline based impact analysis still works (in terms of coverage probability of confidence intervals) when employing data dependent knot sequences. When choosing knot sequences one has to keep in mind that it must be guaranteed that all knots lie inside of the range of the observations X. Especially when using bootstrap methods this condition can easily fail when drawing from the observed data.

The setup provided above gives us not only the opportunity to detect non-linear influences of one covariate but also to investigate the common non-linear influence of two ore more covariates $X_1, ..., X_k$ on the target Variable Y. To this end we can fit polynomials in $X_1, ..., X_k$ and proceed in the same way as above. Another way to detect non-linear influences of several covariates is fitting multidimensional splines with pre-chosen knot-grid. The same knot-choosing problems as in the univariate case arise here too. One should be aware that the number of basis functions (when for example using the tensor product basis already introduced in Section A.1.2) grows exponentially fast with the number of covariates included (Hastie et al., 2001, p. 139).

All setups described here can trivially also be applied to the absolute mean slope.

2.2. Kernel-method-based impacts

In this section we write $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2), ...$ for the i.i.d. observations of the random variables X and Y. We are now interested in estimating the mean impact of a

single covariate X on Y (The case of a common impact of several covariates is a straightforward generalization of the theory derived here an will be discussed in Section 2.2.6). The definition of the mean impact $\iota_X(Y)$ suggests the use of an estimator which has the form

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{\delta}(X_i) - \bar{\delta}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\delta}(X_i) - \bar{\delta}\right)^2}}, \tag{2.1}$$

where $\hat{\delta}$ is a perturbation estimated from the data and $\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)$. In the following sections we will choose perturbations $\hat{\delta}$ based on kernel smoothers and other kernel methods. Application of the theory of U-statistics and the delta method will yield asymptotic normality of this estimators. Furthermore, as pointed out in Section A.3, it follows from Bickel and Freedman (1981) and an additional argument that the bootstrap is valid in these cases.

All cases which are considered below use kernels with fixed bandwidth. For practical application of the derived methods there is the difficulty to choose the bandwidth. For the argumentation for the asymptotic normality derived below the bandwidth must be chosen data independent. For the comparison of methods in Section 4 we will compute the restricted impacts for several fixed bandwidths h as well as for data dependent bandwidths.

2.2.1. Kernel-smoother-based impact analysis

We now use a perturbation $\hat{\delta}$ inspired by a Nadaraya-Watson kernel regression estimator. Let

$$\hat{\delta}(x) = \frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)Y_j, \tag{2.2}$$

where $K_h(u) = K(u/h)$ is a symmetric kernel weight function with fixed bandwidth $h > 0$. Note that we obtain $\hat{\delta}(x)$ from the Nadaraya-Watson kernel regression estimator

$$\hat{m}(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{K_h(x - X_j)Y_j}{n^{-1}\sum_{l=1}^{n} K_h(x - X_l)}$$

by multiplying to each value $\hat{m}(x)$ the kernel density estimator

$$\hat{f}(x) = \frac{1}{n}\sum_{l=1}^{n} K_h(x - X_l)$$

of the density f of X. This means that values of Y whose X values lead to higher values of $f(X)$ are given a higher weight than in the kernel regression estimator. A natural estimator for a restricted kernel smoother-based impact is then as mentioned before given by

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{\delta}(X_i)-\bar{\delta}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\delta}(X_i)-\bar{\delta}\right)^2}},$$

where $\bar{\delta}=\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)$. We name this estimator $\hat{\iota}_X^{ks}(Y)$ (ks for kernel smoother) and rewrite it in the following way:

$$\hat{\iota}_X^{ks}(Y)=\frac{\tilde{\iota}_1-\tilde{\iota}_2}{\sqrt{\tilde{\iota}_3-\tilde{\iota}_4^2}},$$

where $\tilde{\iota}_1=\frac{1}{n}\sum_{i=1}^{n} Y_i\hat{\delta}(X_i)$, $\tilde{\iota}_2=\frac{1}{n}\sum_{i=1}^{n} Y_i\bar{\delta}$, $\tilde{\iota}_3=\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)^2$ and $\tilde{\iota}_4=\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)$. We aim to show that $\tilde{\iota}=(\tilde{\iota}_1,...,\tilde{\iota}_4)^T$ is essentially a third order U-statistic and therefore asymptotically normal. Application of the delta method will yield asymptotic normality of $\hat{\iota}_X^{ks}(Y)$.

For the proof of the asymptotic normality we need the following lemma and assumption.

Lemma 2.1. *Let $w:\mathbb{R}^m\to\mathbb{R}^p$ be a permutation-symmetric function of the random vectors $Z_{j_1},...,Z_{j_m}$ such that $E_{\mathbf{P}}\left(w(Z_{j_1},...,Z_{j_m})\right)$ and $E_{\mathbf{P}}\left(w^2(Z_{j_1},...,Z_{j_m})\right)$ exist for all $(j_1,...,j_m)\in\{1,...,n\}^m$. Then we have that*

$$\sqrt{n}\frac{1}{n^m}\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n} w(Z_{j_1},...,Z_{j_m})=\sqrt{n}\frac{1}{n^m}\sum_{C(\{j_1,...,j_m\})} w(Z_{j_1},...,Z_{j_m})+o_p(1),$$

where $C(\{j_1,...,j_m\})$ is the set of all combinations which can be drawn without replacement from $\{1,...,n\}$ in m draws.

Proof. Because of the symmetry of w we can write

$$\begin{aligned}
&\frac{\sqrt{n}}{n^m}\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n} w(Z_{j_1},...,Z_{j_m})\\
=&\frac{\sqrt{n}}{n^m}\sum_{C(\{j_1,...,j_m\})} w(Z_{j_1},...,Z_{j_m})\\
&+\sum_{k=1}^{m-1}\sum_{a\in A(k)}\varphi(a)\frac{\sqrt{n}}{n^m}\sum_{C(\{j_1,...,j_k\})} w(\underbrace{Z_{j_1},...,Z_{j_1}}_{a_1},...,\underbrace{Z_{j_k},...,Z_{j_k}}_{a_k}),
\end{aligned}$$

where $A(k) = \{(a_1, ..., a_k) : \sum_{l=1}^{k} a_l = m, a_l \in \mathbb{N}_{>0}\}$, $\varphi(a)$ is independent of n, and Z_{j_l} appears a_l-times in the function w. Hence, we only need to show, that for given k and $a \in A(k)$ the term

$$\frac{\sqrt{n}}{n^m} \sum_{C(\{j_1,...,j_k\})} w(Z_{j_1}, ..., Z_{j_1}, ..., Z_{j_k}, ..., Z_{j_k})$$

converges to zero in probability. At first we rewrite this expression as

$$\frac{\sqrt{n}}{n^m} \sum_{C(\{j_1,...,j_k\})} w^*(Z_{j_1}, ..., Z_{j_k}),$$

where $w^*(Z_{j_1}, ..., Z_{j_k}) = w(Z_{j_1}, ..., Z_{j_1}, ..., Z_{j_k}, ..., Z_{j_k})$ with Z_{j_l} appearing a_l-times. Obviously, with $S(\{1, ..., k\})$ being the set off all permutations of $\{1, ..., k\}$ this equals

$$\begin{aligned}
&\frac{\sqrt{n}}{n^m} \sum_{C(\{j_1,...,j_k\})} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(k)}}) \\
=&\frac{\sqrt{n}}{n^m} k! \sum_{j_1 < ... < j_k} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(k)}}) \\
=&\frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)! n^k} \binom{n}{k}^{-1} \sum_{j_1 < ... < j_k} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(k)}}).
\end{aligned}$$

Since $m - k > 1/2$, the term $\frac{\sqrt{n}}{n^{m-k}}$ converges to zero. Furthermore, we have $\frac{n!}{(n-k)! n^k} \to 1$. With the assumptions of this lemma, Theorem A.7 gives that

$$\binom{n}{k}^{-1} \sum_{j_1 < ... < j_k} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(k)}})$$

converges to its (existing) mean and is therefore bounded in probability. We obtain

$$\frac{\sqrt{n}}{n^m} \sum_{C(\{j_1,...,j_k\})} w(Z_{j_1}, ..., Z_{j_1}, ..., Z_{j_k}, ..., Z_{j_k}) = o_p(1).$$

□

Assumption 2.2. *Let*

$$g_1(Z_i, Z_j, Z_l) = K_h(X_i - X_j) Y_i Y_j, \quad g_2(Z_i, Z_j, Z_l) = K_h(X_j - X_l) Y_i Y_l,$$

$$g_3(Z_i,Z_j,Z_l)=K_h(X_i-X_j)Y_jK_h(X_i-X_l)Y_l,\quad g_4(Z_i,Z_j,Z_l)=K_h(X_i-X_j)Y_j.$$

Furthermore, let $g=(g_1,...,g_4)$ *and*

$$\begin{aligned}w(Z_i,Z_j,Z_l)=&\frac{1}{6}\{g(Z_i,Z_j,Z_l)+g(Z_i,Z_l,Z_j)+g(Z_j,Z_i,Z_l)\\&+g(Z_j,Z_l,Z_i)+g(Z_l,Z_i,Z_j)+g(Z_l,Z_j,Z_i)\}.\end{aligned}\tag{2.3}$$

Assume that

$$E_{\mathbf{P}}(w(Z_i,Z_j,Z_l))\quad\text{and}\quad E_{\mathbf{P}}(w^2(Z_i,Z_j,Z_l))$$

exist for all $(i,j,l)\in\{1,...,n\}^3$ *and define* $\vartheta=E_{\mathbf{P}}(w(Z_i,Z_j,Z_l))$ *for* $i\neq j\neq l\neq i$.

Theorem 2.3. *Under Assumption 2.2 we have that*

$$\sqrt{n}(\hat{\iota}_X^{ks}(Y)-\iota_X^{ks}(Y))\xrightarrow{\mathcal{L}}N(0,\sigma^2),$$

with $\sigma^2=DF(\vartheta)^TVDF(\vartheta)$,

$$F\left((a_1,...,a_4)^T\right)\mapsto\frac{a_1-a_2}{\sqrt{a_3-a_4^2}},$$

$\iota_X^{ks}(Y)=F(\vartheta)$ *and*

$$V=9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right),$$

for $\tilde{w}(Z_i)=E_{\mathbf{P}}(w(Z_i,Z_j,Z_l)|Z_i)-\vartheta$.

Proof. With the notation of this section we have

$$\tilde{\iota}_1=\frac{1}{n}\sum_{i=1}^n Y_i\hat{\delta}(X_i)=\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\underbrace{K_h(X_i-X_j)Y_iY_j}_{=g_1(Z_i,Z_j,Z_l)}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n w_1(Z_i,Z_j,Z_l).$$

Similarly,

$$\tilde{\iota}_2=\frac{1}{n}\sum_{i=1}^n Y_i\bar{\delta}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\underbrace{K_h(X_j-X_l)Y_iY_l}_{=g_2(Z_i,Z_j,Z_l)}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n w_2(Z_i,Z_j,Z_l),$$

$$\tilde{\iota}_3=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\underbrace{K_h(X_i-X_j)Y_jK_h(X_i-X_l)Y_l}_{=g_3(Z_i,Z_j,Z_l)}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n w_3(Z_i,Z_j,Z_l)$$

and

$$\tilde{\iota}_4 = \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\underbrace{K_h(X_i - X_j)Y_j}_{=:g_4(Z_i,Z_j,Z_l)} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} w_4(Z_i, Z_j, Z_l).$$

Hence, we obtain for $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$ that

$$\sqrt{n}\tilde{\iota} = \frac{\sqrt{n}}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} w(Z_i, Z_j, Z_l).$$

By Lemma 2.1 and Assumption 2.2 we obtain

$$\begin{aligned}
\sqrt{n}\tilde{\iota} =& \frac{\sqrt{n}}{n^3}\sum\sum\sum_{i\neq j\neq l\neq i} w(Z_i, Z_j, Z_l) + o_p(1) \\
=& \frac{\sqrt{n}}{n^3} 6 \sum\sum\sum_{i<j<l} w(Z_i, Z_j, Z_l) + o_p(1) \\
=& \frac{n(n-1)(n-2)}{n^3}\sqrt{n}\underbrace{\binom{n}{3}^{-1}\sum\sum\sum_{i<j<l} w(Z_i, Z_j, Z_l)}_{=:U_n} + o_p(1)
\end{aligned}$$

where U_n is a third-order U-statistics. Consequently, we have

$$\begin{aligned}
\sqrt{n}(\tilde{\iota} - \vartheta) =& \frac{n(n-1)(n-2)}{n^3}\sqrt{n}U_n + o_p(1) - \sqrt{n}\vartheta \\
=& \underbrace{\frac{n(n-1)(n-2)}{n^3}}_{\to 1}\underbrace{\sqrt{n}(U_n - \vartheta)}_{\xrightarrow{\mathcal{L}} N(0,V)} + o_p(1) + \underbrace{\left(\frac{n(n-1)(n-2)}{n^3}\sqrt{n} - \sqrt{n}\right)\vartheta}_{\to 0} \\
\xrightarrow{\mathcal{L}}& N(0, V)
\end{aligned}$$

with

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right) \tag{2.4}$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$. Since the mapping

$$F\left((a_1, ..., a_4)^T\right) \mapsto \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}$$

is continuously differentiable application of the delta-method with $\iota_X^{ks}(Y) = F(\vartheta)$ yields

$$\sqrt{n}(\hat{\iota}_X^{ks}(Y) - \iota_X^{ks}(Y)) = \sqrt{n}\,(F(\tilde{\iota}) - F(\vartheta)) \xrightarrow{\mathcal{L}} DF(\vartheta)^T N(0, V) = N(0, \sigma^2), \tag{2.5}$$

where $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$. □

The next lemma shows how σ^2 can be consistently estimated.

Lemma 2.4. *σ^2 can be consistently estimated by*

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{V} DF(\tilde{\iota}),$$

where

$$\hat{V} = 9\left(\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \tilde{g}(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}, Z_{\pi(a)}, Z_{\pi(b)}) - \tilde{\iota}\tilde{\iota}^T\right)$$

and $\tilde{g}(Z_i, Z_j, Z_l, Z_a, Z_b) = w(Z_i, Z_j, Z_l) w^T(Z_i, Z_a, Z_b)$.

Proof. Since $\tilde{\iota}$ is consistent for ϑ, $DF(\vartheta)$ can be consistently estimated by $DF(\tilde{\iota})$. To find a consistent estimator for V we make the following considerations (which are similar to those in Kowalski and Tu (2008, p. 259))

$$\begin{aligned}
V/9 =& E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right) \\
=& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j, Z_l)|Z_i\right\} E_{\mathbf{P}}\left\{w^T(Z_i, Z_j, Z_l)|Z_i\right\}\right) \\
& - 2\vartheta E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j, Z_l)|Z_i\right\}\right) + \vartheta\vartheta^T \\
=& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j, Z_l)|Z_i\right\} E_{\mathbf{P}}\left\{w^T(Z_i, Z_j, Z_l)|Z_i\right\}\right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j, Z_l) w^T(Z_i, Z_a, Z_b)|Z_i\right\}\right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}}\underbrace{\left(w(Z_i, Z_j, Z_l) w^T(Z_i, Z_a, Z_b)\right)}_{=:\tilde{g}(Z_i, Z_j, Z_l, Z_a, Z_b)} - \vartheta\vartheta^T,
\end{aligned}$$

where $a \notin \{i, j, l\}$, $b \notin \{i, j, l, a\}$. By the theory of U-statistics (cf. Kowalski and Tu (2008)) $E_{\mathbf{P}}\left(\tilde{g}(Z_i, Z_j, Z_l, Z_a, Z_b)\right)$ can be consistently estimated by

$$\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \tilde{\tilde{g}}(Z_i, Z_j, Z_l, Z_a, Z_b),$$

where $\tilde{\tilde{g}}(Z_i, Z_j, Z_l, Z_a, Z_b)$ is a symmetric version of $\tilde{g}(Z_i, Z_j, Z_l, Z_a, Z_b)$ say

$$\tilde{\tilde{g}}(Z_i, Z_j, Z_l, Z_a, Z_b) = \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \tilde{g}(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}, Z_{\pi(a)}, Z_{\pi(b)})$$

with $S(\{i, j, l, a, b\})$ being the set of all permutations of $\{i, j, l, a, b\}$. Hence a consistent estimator for V is given by

$$\hat{V} = 9 \left(\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \tilde{g}(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}, Z_{\pi(a)}, Z_{\pi(b)}) - \tilde{\iota}\tilde{\iota}^T \right) \quad (2.6)$$

which leads to

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{V} DF(\tilde{\iota}) \quad (2.7)$$

as consistent estimator for σ. □

Combining (2.5) and (2.7) we obtain

$$\sqrt{n} \frac{\hat{\iota}_X^{ks}(Y) - \iota_X^{ks}(Y)}{\hat{\sigma}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Once we have an asymptotic normality result like that, an asymptotic level α confidence interval for $\iota_X^{ks}(Y)$ is given by

$$CI = [\hat{\iota}_X^{ks}(Y) - z_{1-\alpha}\hat{\sigma}/\sqrt{n}, \infty) \quad (2.8)$$

where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution.

It will turn out in the simulations of Section 4 that the estimator $\hat{\iota}_X^{ks}(Y)$ may be anti conservative especially when $\iota_X^{ks}(Y) = 0$. In a late state of this thesis we became aware of the work of Doksum and Samarov (1995), which deals with obtaining reliable estimators for a nonparametric coefficient of determination. Their results suggest that using a "leave-one-out" estimator where we replace $\hat{\delta}(X_i)$ in (2.2) by

$$\hat{\delta}'(X_i) = \frac{1}{n-1} \sum_{j \neq i} K_h(X_i - X_j) Y_j,$$

thus leaving the i-th observation of Y out, leads to more conservative estimates. A proof of asymptotic normality of the leave-one-out estimator in our case can be done in the same spirit as the proof of Theorem 2.3. In fact, Lemma 2.1 would only be needed to ensure that $\sqrt{n}\tilde{\iota}_3$ is a U-statistics plus a term which is $o_p(1)$. The other terms can easily

be seen to be U-Statistics without this lemma. The leave-one-out type of estimation can also be applied to all the following Kernel-method based estimators for the mean impact. The investigation of the leave-one-out estimator will not be subject of this thesis, but is clearly an interesting topic for further research.

We now come back to the estimator using all observations. We have to check, if $\iota_X^{ks}(Y)$ is a lower bound for the "true" impact $\iota_X(Y)$. This would imply that the confidence interval in (2.8) is conservative for $\iota_X(Y)$. To this end we make the following considerations. We have that

$$\vartheta = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)) = \begin{pmatrix} E_{\mathbf{P}}(K_h(X_i - X_j)Y_iY_j) \\ E_{\mathbf{P}}(K_h(X_j - X_l)Y_lY_i) \\ E_{\mathbf{P}}(K_h(X_i - X_j)K_h(X_i - X_l)Y_jY_l) \\ E_{\mathbf{P}}(K_h(X_i - X_j)Y_j) \end{pmatrix}.$$

From this it follows that

$$\begin{aligned}
&\vartheta_1 - \vartheta_2 \\
=&E_{\mathbf{P}}(Y_iK_h(X_i - X_j)Y_j) - E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}(K_h(X_j - X_l)Y_j) \\
=&E_{\mathbf{P}}\left\{E_{\mathbf{P}}[Y_iK_h(X_i - X_j)Y_j|X_i]\right\} - E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_j - X_l)Y_j|X_i]\right\} \\
=&E_{\mathbf{P}}\left\{E_{\mathbf{P}}(Y_i|X_i)E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]\right\} - E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_j - X_l)Y_j|X_i]\right\} \\
=&E_{\mathbf{P}}\left\{Y_iE_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]\right\} - E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_j - X_l)Y_j|X_i]\right\} \\
=&Cov_{\mathbf{P}}(Y_i, E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]) \\
=&E_{\mathbf{P}}\left\{Y_i\left[E_{\mathbf{P}}(K_h(X_i - X_j)Y_j|X_i) - E_{\mathbf{P}}(E_{\mathbf{P}}(K_h(X_i - X_j)Y_j|X_i))\right]\right\},
\end{aligned}$$

where the third equality follows from $E_{\mathbf{P}}(E_{\mathbf{P}}(Y|X)h(X)) = E_{\mathbf{P}}(E_{\mathbf{P}}[Yh(X)|X]) = E_{\mathbf{P}}(Yh(X))$, for every measurable function h (see for example Klenke (2008)). Additionally, we have

$$\begin{aligned}
&\vartheta_3 - \vartheta_4^2 \\
=&E_{\mathbf{P}}(K_h(X_i - X_j)Y_jK_h(X_i - X_l)Y_l) - E_{\mathbf{P}}(K_h(X_i - X_j)Y_j)^2 \\
=&E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]E_{\mathbf{P}}[K_h(X_i - X_l)Y_l|X_i]\right\} - E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]\right\}^2 \\
\\
=&E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]^2\right\} - E_{\mathbf{P}}\left\{E_{\mathbf{P}}[K_h(X_i - X_j)Y_j|X_i]\right\}^2 \\
=&Var_{\mathbf{P}}\left(E_{\mathbf{P}}\{K_h(X_i - X_j)Y_j|X_i\}\right).
\end{aligned}$$

Hence

$$\iota_X^{ks}(Y) = \frac{\vartheta_1 - \vartheta_2}{\sqrt{\vartheta_3 - \vartheta_4^2}} = E_{\mathbf{P}}\left\{ Y \frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(X)}} \right\},$$

where $\delta(X) = E_{\mathbf{P}}\left\{K_h(X - X_j)Y_j | X\right\}$,

$$\leq \sup_{\delta \in L^2_{\mathbb{P}}(\mathbb{R})} E_{\mathbf{P}}\left\{ Y \frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(\delta(X))}} \right\} = \iota_X(Y).$$

Thus $\iota_X^{ks}(Y) \leq \iota_X(Y)$ and (2.8) is a (potentially conservative) asymptotic level α confidence interval for $\iota_X(Y)$.

Since the computation of $\hat{\sigma}$ in (2.7) requires substantial computational effort for typical n it would be more convenient to use bootstrap methods to calculate a confidence interval for $\iota_X^{ks}(Y)$. In order to establish the consistency of the bootstrap in this setup we will need a "bootstrap version" of Lemma 2.1.

Lemma 2.5. *Let $w : \mathbb{R}^m \to \mathbb{R}^p$ be a permutation-symmetric function of the random vectors $Z_{j_1}, ..., Z_{j_m}$ such that $E_{\mathbf{P}}\left(w(Z_{j_1}, ..., Z_{j_m})\right)$ and $E_{\mathbf{P}}\left(w^2(Z_{j_1}, ..., Z_{j_m})\right)$ exist for all $(j_1, ..., j_m) \in \{1, ..., n\}^m$. Furthermore, let $Z_1^*, ..., Z_n^*$ be bootstrap repetitions of $Z_1, ..., Z_n$, i.e. i.i.d. random vectors which are distributed according to the empirical distribution function F_n of $Z_1, ..., Z_n$. Then we have for almost all sequences $Z_1, Z_2, ...$ that*

$$\sqrt{n}\frac{1}{n^m}\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n} w(Z_{j_1}^*, ..., Z_{j_m}^*) = \sqrt{n}\frac{1}{n^m}\sum_{C(\{j_1,...,j_m\})} w(Z_{j_1}^*, ..., Z_{j_m}^*) + o_{p|Z}(1),$$

where

$$A_n = o_{p|Z}(1) \Leftrightarrow \mathbb{P}(|A_n| \geq \epsilon | Z_1, ..., Z_n) \to 0 \ \forall \epsilon > 0, \ \textit{almost surely}$$

and $C\left(\{j_1, ..., j_m\}\right)$ is the set of all combinations which can be drawn without replacement from $\{1, ..., n\}$ in m draws.

Proof. By the same argumentation as in the proof of Lemma 2.1 it suffices to show that for a given integer k with $1 \leq k \leq m - 1$ and

$$a \in A(k) = \left\{ (a_1, ..., a_k) : \sum_{l=1}^{k} a_l = m, a_l \in \mathbb{N}_{>0} \right\}$$

the expression

$$\frac{\sqrt{n}}{n^m} \sum_{C(\{j_1,...,j_k\})} w(Z^*_{j_1},...,Z^*_{j_1},...,Z^*_{j_k},...,Z^*_{j_k}),$$

where $Z^*_{j_l}$ appears a_l times, converges to zero in probability given $Z_1,...,Z_n$ for almost all $Z_1, Z_2, ...$ We can rewrite this as

$$\frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)!n^k} \underbrace{\binom{n}{k}^{-1} \sum_{j_1<...<j_k} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z^*_{j_{\pi(1)}},...,Z^*_{j_{\pi(k)}})}_{h_a(G_n)}, \tag{2.9}$$

where G_n is the empirical distribution function of $Z^*_1,...,Z^*_n$ and $w^*(Z^*_{j_1},...,Z^*_{j_k}) = w(Z^*_{j_1},...,Z^*_{j_1},...,Z^*_{j_k},...,Z^*_{j_k})$ with $Z^*_{j_l}$ appearing a_l-times. With the notation

$$h_a(F_n) = \binom{n}{k}^{-1} \sum_{j_1<...<j_k} \frac{1}{k!} \sum_{\pi \in S(\{1,...,k\})} w^*(Z_{j_{\pi(1)}},...,Z_{j_{\pi(k)}})$$

we obtain that (2.9) equals

$$\frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)!n^k} (h_a(G_n) - h_a(F_n)) + \frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)!n^k} h_a(F_n).$$

One can see that $\frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)!n^k}$ converges to zero. According to Bickel and Freedman (1981) we have that the term $(h_a(G_n) - h_a(F_n))$ is $o_{p|Z}(1)$ for almost all $Z_1, Z_2, ...$. Lemma 2.1 implies that $\frac{\sqrt{n}}{n^{m-k}} \frac{n!}{(n-k)!n^k} h_a(F_n)$ is $o_{p|Z}(1)$ for almost all $Z_1, Z_2, ...$ too. Hence the statement of the Lemma is shown. □

Theorem 2.6. *Let $\hat{\iota}^{ks^*}_X(Y)$ be $\hat{\iota}^{ks}_X(Y)$ based on the bootstrap sample $Z^*_1,...,Z^*_n$. The conditional distribution of $\sqrt{n}(\hat{\iota}^{ks^*}_X(Y) - \hat{\iota}^{ks}_X(Y))$ (given $Z_1,...,Z_n$) and the distribution of $\sqrt{n}(\hat{\iota}^{ks}_X(Y) - \iota^{ks}_X(Y))$ converge to the same limit for almost all $Z_1, Z_2, ...$.*

Proof. Let $\tilde{\iota}^*$ be $\tilde{\iota}$ based on the bootstrap sample $Z^*_1,...,Z^*_n$. Performing the same calculations as in the proof of Theorem 2.3 and replacing Lemma 2.1 by Lemma 2.5 in this argumentation we obtain for almost all sequences $Z_1, Z_2, ...$ and all $a \in \mathbb{R}^4$

$$\mathbb{P}\left(\sqrt{n}(\tilde{\iota}^* - \tilde{\iota}) \le a | Z_1,...,Z_n\right)$$

$$= \mathbb{P}\left(\frac{n(n-1)(n-2)}{n^3} \sqrt{n} (h(G_n) - h(F_n)) + o_{p|Z}(1) \le a \,\middle|\, Z_1,...,Z_n \right) \to \Phi_{0,V}(a),$$

where

$$h(G_n) = \binom{n}{3}^{-1} \sum\sum\sum_{i<j<l} w(Z_i^*, Z_j^*, Z_l^*),\ h(F_n) = \binom{n}{3}^{-1} \sum\sum\sum_{i<j<l} w(Z_i, Z_j, Z_l),$$

with w from Assumption 2.2 and $\Phi_{0,V}(a)$ is the distribution function of $N(0,V)$ and V is as in (2.4). The convergence to the normal distribution follows from Bickel and Freedman (1981). Application of the delta-method yields that for almost all sequences $Z_1, Z_2, ...$ and all $b \in \mathbb{R}$

$$\mathbb{P}\left(\sqrt{n}(\hat{\iota}_X^{ks^*}(Y) - \hat{\iota}_X^{ks}(Y)) \leq b \middle| Z_1, ..., Z_n \right) \to \Phi_{0,\sigma^2}(b),$$

with Φ_{0,σ^2} being the distribution function of $N(0,\sigma^2)$, σ^2 from (2.5) and $\hat{\iota}_X^{ks^*}(Y)$ being the bootstrap version of $\hat{\iota}_X^{ks}(Y)$. Hence, the conditional distribution of $\sqrt{n}(\hat{\iota}_X^{ks^*}(Y) - \hat{\iota}_X^{ks}(Y))$ (given $Z_1, ..., Z_n$) and the distribution of $\sqrt{n}(\hat{\iota}_X^{ks}(Y) - \iota_X^{ks}(Y))$ converge to the same limit for almost all $Z_1, Z_2, ...$. □

Theorem 2.6 states that the bootstrap is consistent for the kernel smoother based impact. This consistency of the bootstrap justifies the use of bootstrap confidence intervals.

Another possibility to reduce the computation time for the variance estimator (2.7) is to modify the covariance estimator $\hat{V}$. We will show that it is sufficient to take the first sum in (2.6) not over all $i < j < l < a < b$ but to sample from these combinations of indices in a suitable manner for the estimator to remain consistent. To this end we state the following lemma.

Lemma 2.7. *Let*

$$\tilde{\Sigma} = \frac{1}{B} \sum_{(i,j,k,a,b) \in W} \underbrace{\frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \tilde{g}(Z_{\pi(i)}, ..., Z_{\pi(b)})}_{v(i,...,b)},$$

where $W = \{W_1, ..., W_B\}$ and the W_i are sampled without replacement (using a uniform distribution) from $C_n^5 := \{i < j < l < a < b : i, ..., b \in \{1, ..., n\}\}$. Let further $\binom{n}{5}/B \to 1$ for $n \to \infty$. Then we have

$$\hat{\Sigma} - \tilde{\Sigma} \xrightarrow{p} 0,$$

where

$$\hat{\Sigma} = \binom{n}{5}^{-1} \sum_{i<j<l<a<b} v(i,...,b)$$

is as in (2.6)*.*

Proof. We have that

$$\begin{aligned}
\hat{\Sigma} - \tilde{\Sigma} =& \binom{n}{5}^{-1} \sum_{(i,...,b)\in C_n^5} v(i,...,b) - \frac{1}{B} \sum_{(i,...,b)\in W} v(i,...,b) \\
=& \left(1 - \frac{\binom{n}{5}}{B}\right) \binom{n}{5}^{-1} \sum_{(i,...,b)\in C_n^5} v(i,...,b) \\
&+ \frac{1}{B} \left[\sum_{(i,...,b)\in C_n^5} v(i,...,b) - \sum_{(i,...,b)\in W} v(i,...,b) \right] \\
=& \left(1 - \frac{\binom{n}{5}}{B}\right) \hat{\Sigma} + \frac{1}{B} \sum_{(i,...,b)\in C_n^5 \setminus W} v(i,...,b).
\end{aligned}$$

The term $1 - \binom{n}{5}/B$ converges to zero, $\hat{\Sigma}$ is as seen above consistent for some finite matrix and $\sum_{(i,...,b)\in C_n^5\setminus W} v(i,...,b)$ converges to zero, since $|C_n^5 \setminus W| = \binom{n}{5} - B \to 0$. Hence the statement of the lemma holds. □

From this lemma it follows that we can replace $\hat{\Sigma}$ by $\tilde{\Sigma}$ in the estimation of the variance $\hat{\sigma}^2$ without losing the consistency of the estimator. Choosing this procedure leads to a reduction of computation time but requires to choose the sample size B. Since there is no intuitive way for doing so we will not pursue this approach any further.

2.2.2. Population coefficient for determination based on kernel smoothers

We can also give an estimate for the population coefficient for determination which is based on kernel smoothers. We estimate the parameter

$$R_{\mathbf{P}}^{ks^2} = \iota_X^{ks^2}(Y)/Var_{\mathbf{P}}(Y)$$

by

$$\hat{R}_{\mathbf{P}}^{ks^2} = \hat{\iota}_X^{ks^2}(Y)/\hat{\sigma}_Y^2, \tag{2.10}$$

where $\hat{\sigma}_Y^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$. Our aim is to show that $\hat{R}_{\mathbf{P}}^{ks^2}$ converges to a normal distribution and that the bootstrap is consistent in this setup. We do so by applying

the the theory of U-statistics similar to Section 2.2.1, where we examined $\hat{\iota}_X^{ks}(Y)$.

Assumption 2.8. *Let $g_1, ..., g_4$ as in Assumption 2.2 and additionally $g_5(Z_i, Z_j, Z_l) = \frac{1}{2}(Y_i - Y_j)^2$. Furthermore, let w be defined analogous to (2.3) and assume that*

$$E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)) \quad \text{and} \quad E_{\mathbf{P}}(w^2(Z_i, Z_j, Z_l))$$

exist for all $(i, j, l) \in \{1, ..., n\}^3$ and define $\vartheta = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l))$ for $i \neq j \neq l \neq i$. Assume that additionally $\vartheta_5 \neq 0$, which is equivalent to $Var_{\mathbf{P}}(Y) \neq 0$.

Theorem 2.9. *Under Assumption 2.8 we have that*

$$\sqrt{n}\left(\hat{R}_{\mathbf{P}}^{ks^2} - R_{\mathbf{P}}^{ks^2}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

with $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$,

$$F\left((a_1, ..., a_5)^T\right) = \left[\frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}\right]^2 / a_5 \quad \text{and} \quad V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

as well as $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$.

Proof. We define $\tilde{\iota}_5 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2 = \hat{\sigma}_Y^2$. We have that

$$\tilde{\iota}_5 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \frac{1}{2}(Y_i - Y_j)^2 = \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n \underbrace{\frac{1}{2}(Y_i - Y_j)^2}_{g_5(Z_i, Z_j, Z_l)} = \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n w_5(Z_i, Z_j, Z_l).$$

We now regard the vectors $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_5)^T$ and ϑ. With the same argumentation as in the proof of Theorem 2.3, but with $w = (w_1, ..., w_5)^T$ instead of $w = (w_1, ..., w_4)^T$, we obtain

$$\begin{aligned}
\sqrt{n}(\tilde{\iota} - \vartheta) =& \frac{n(n-1)(n-2)}{n^3}\sqrt{n}U_n + o_p(1) - \sqrt{n}\vartheta \\
=& \underbrace{\frac{n(n-1)(n-2)}{n^3}}_{\to 1}\underbrace{\sqrt{n}(U_n - \vartheta)}_{\xrightarrow{\mathcal{L}} N(0,V)} + o_p(1) + \underbrace{\left(\frac{n(n-1)(n-2)}{n^3}\sqrt{n} - \sqrt{n}\right)\vartheta}_{\to 0} \\
\xrightarrow{\mathcal{L}}& N(0, V)
\end{aligned}$$

with

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$. The mapping

$$F\left((a_1, ..., a_5)^T\right) = \left[\frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}\right]^2 / a_5$$

is continuously differentiable with $F(\tilde{\iota}) = \hat{R}_{\mathbf{P}}^{ks^2}$ and $F(\vartheta) = R_{\mathbf{P}}^{ks^2}$. Hence, by applying the delta method we obtain

$$\sqrt{n}\left(\hat{R}_{\mathbf{P}}^{ks^2} - R_{\mathbf{P}}^{ks^2}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

with $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$. □

For the estimation of σ^2 we can use the estimator (2.7) or the same estimator but with the modified estimator for V from Lemma 2.7 (in both cases with the new w, F and $\tilde{\iota}$).

It can be shown analogously to Section 2.2.1 that the bootstrap is consistent in this case, which implies that we can also compute bootstrap confidence intervals for $R_{\mathbf{P}}^{ks^2}$. Every confidence interval for $R_{\mathbf{P}}^{ks^2}$ will be a conservative confidence interval for $R_{\mathbf{P}}^2$.

Note that the estimator (2.10) is very similar to (and in a sense a special case of) one of the estimates Doksum and Samarov (1995) propose for Person´s correlation ratio

$$\eta^2 = \frac{Var_{\mathbf{P}}(E_{\mathbf{P}}(Y|X))}{Var_{\mathbf{P}}(Y)}.$$

However, our results differ from those of Doksum and Samarov (1995) as they use a leave-one-out type estimator (which can also be employed here, as seen in Section 2.2.1) and do not derive a result of bootstrap consistency, as we do in this thesis.

2.2.3. Mean slope based on kernel-smoothers

In a similar spirit we can show that the canonical estimate of the mean slope based on kernel-smoothers is also a differentiable function of a U-statistics under suitable additional assumptions. This implies that the construction of confidence intervals with and without the usage of bootstrap-methods works in the same manner as for the mean impact and the population coefficient for determination based on kernel-smoothing. We estimate the mean slope based on kernel smoothing by

$$\hat{\theta}_X^{ks}(Y) = \frac{\hat{\iota}_Y^{ks}(X)}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}}.$$

Assumption 2.10. *Let $g_1, ..., g_4$ as in Assumption 2.2 and additionally $g_5(Z_i, Z_j, Z_l) = \frac{1}{2}(X_i - X_j)^2$. Furthermore, let w be defined analogous to* (2.3) *and assume that*

$$E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)) \quad \text{and} \quad E_{\mathbf{P}}(w^2(Z_i, Z_j, Z_l))$$

exist for all $(i, j, l) \in \{1, ..., n\}^3$ and define $\vartheta = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l))$ for $i \neq j \neq l \neq i$. Assume that additionally $\vartheta_5 \neq 0$, which is equivalent to $Var_{\mathbf{P}}(X) \neq 0$.

Theorem 2.11. *Under Assumption 2.10 we have that*

$$\sqrt{n}(\hat{\theta}_Y^{ks}(X) - \theta_X^{ks}(Y)) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

with $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$, $\vartheta = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l))$,

$$F\left((a_1, ..., a_5)^T\right) = \frac{(a_1 - a_2)/\sqrt{a_5}}{\sqrt{a_3 - a_4^2}} \quad \text{and} \quad V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right),$$

as well as $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$.

Proof. The proof is similar to that of Theorem 2.9. One only has to replace Y by X in the definition of $\tilde{\iota}_5$ and to use the new function F in the application of the delta method. □

Similar to the previous sections bootstrap and non-bootstrap methods can be applied in order to construct asymptotic confidence intervals.

2.2.4. Loess-based impact analysis

In this section we choose a perturbation $\hat{\delta}$ inspired by a local linear regression estimator. We choose

$$\hat{\delta}(x) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j - x)(X_j - X_l)K_h(x - X_j)K_h(x - X_l)Y_l,$$

where $K_h(u) = K(u/h)$ is a symmetric kernel weight function with fixed bandwidth $h > 0$. Note that we obtain $\hat{\delta}(x)$ from the local linear regression regression estimator

$$\frac{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j - x)(X_j - X_l)K_h(x - X_j)K_h(x - X_l)Y_l}{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j^2 - X_jX_l)K_h(x - X_j)K_h(x - X_l)},$$

derived in (A.3) by dropping the denominator. Again, we use the natural estimator

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{\delta}(X_i) - \bar{\delta}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\delta}(X_i) - \bar{\delta}\right)^2}},$$

where $\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)$, and rewrite it as

$$\hat{\iota}_X^{loess}(Y) = \frac{\tilde{\iota}_1 - \tilde{\iota}_2}{\sqrt{\tilde{\iota}_3 - \tilde{\iota}_4^2}},$$

where $\tilde{\iota}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i\hat{\delta}(X_i)$, $\tilde{\iota}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i\bar{\delta}$, $\tilde{\iota}_3 = \frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)^2$ and $\tilde{\iota}_4 = \frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)$. We define $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$. Analogous to the case where $\hat{\delta}$ is based on a kernel smoother (see Section 2.2.1) we will prove that $\hat{\iota}_X^{loess}(Y)$ is asymptotically normal by showing that it is essentially a function of a fifth order U-statistic. All steps in the following examination are very similar to those in the kernel smoother case. We aim to show that $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$ is essentially a fifth order U-statistic and therefore asymptotic normal. Application of the delta method yields asymptotic normality of $\hat{\iota}_X^{loess}(Y)$.

Assumption 2.12. *Let*

$$g_1(Z_i, Z_j, Z_l, Z_k, Z_m) = (X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_lY_i,$$

$$g_2(Z_i, Z_j, Z_l, Z_k, Z_m) = (X_j - X_k)(X_j - X_l)K_h(X_k - X_j)K_h(X_k - X_l)Y_lY_i,$$

$$\begin{aligned} g_3(Z_i, Z_j, Z_l, Z_k, Z_m) =& \{(X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l \\ & (X_k - X_i)(X_k - X_m)K_h(X_i - X_k)K_h(X_i - X_m)Y_m\}, \end{aligned}$$

and

$$g_4(Z_i, Z_j, Z_l, Z_k, Z_m) = (X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l.$$

Furthermore let $g = (g_1, ..., g_4)$ *and*

$$w(Z_i, Z_j, Z_l, Z_k, Z_m) = \frac{1}{5!}\sum_{\pi\in S(\{i,j,l,k,m\})} g(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}, Z_{\pi(k)}, Z_{\pi(m)})$$

with $S(\{i,j,l,k,m\})$ being the set of all permutations of $\{i,j,l,k,m\}$. Assume that

$$E_{\mathbf{P}}(w(Z_i,Z_j,Z_l,Z_k,Z_m)) \quad \text{and} \quad E_{\mathbf{P}}(w^2(Z_i,Z_j,Z_l,Z_k,Z_m))$$

exist for all $(i,j,l,k,m) \in \{1,...,n\}^5$ and define $\vartheta = E_{\mathbf{P}}(w(Z_i,Z_j,Z_l,Z_k,Z_m))$ for $i < j < ... < m$.

Theorem 2.13. *Under Assumption 2.12 we have that*

$$\sqrt{n}(\iota_X^{loess}(Y) - \iota_X^{loess}(Y)) \xrightarrow{\mathcal{L}} N(0,\sigma^2), \tag{2.11}$$

where $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$,

$$F\left((a_1,...,a_4)^T\right) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}} \quad V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right),$$

as well as $\iota_X^{loess}(Y) = F(\vartheta)$ and $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i,Z_j,Z_l,Z_k,Z_m)|Z_i) - \vartheta$.

Proof. The proof is similar to the one of Theorem 2.3 for the kernel smoother based mean impact can be found in Appendix B. □

Our next task is finding a consistent estimator for σ^2. The following lemma gives such an estimator.

Lemma 2.14. *A consistent estimator for σ^2 is given by*

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{V} DF(\tilde{\iota}), \tag{2.12}$$

where

$$\hat{V} = 25\left(\binom{n}{9}^{-1} \sum_{i<...<d} \frac{1}{9!} \sum_{\pi \in S(\{i,...,d\})} \tilde{g}(Z_{\pi(i)},...,Z_{\pi(d)}) - \tilde{\iota}\tilde{\iota}^T\right)$$

and

$$\tilde{g}(Z_i,Z_j,Z_l,Z_k,Z_m,Z_a,Z_b,Z_c,Z_d) = w(Z_i,Z_j,Z_l,Z_k,Z_m)w^T(Z_i,Z_a,Z_b,Z_c,Z_d).$$

Proof. This proof can also be found in Appendix B. □

Combining (2.11) and (2.12) we obtain

$$\sqrt{n}\frac{\hat{\iota}_X^{loess}(Y) - \iota_X^{loess}(Y)}{\hat{\sigma}} \xrightarrow{\mathcal{L}} N(0,1).$$

For the confidence interval based on the above asymptotic result to be a (potentially conservative) asymptotic level α confidence interval for $\iota_X(Y)$ it is crucial to know that $\iota_X^{loess}(Y) \leq \iota_X(Y)$. This can be shown as follows. Let $a(Z_i, Z_j, Z_l) = (X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l$, then we can rewrite ϑ as

$$\begin{aligned}
\vartheta_1 =& E_{\mathbf{P}}\left\{a(Z_i, Z_j, Z_l)Y_i\right\} = E_{\mathbf{P}}\left\{E_{\mathbf{P}}[Y_i a(Z_i, Z_j, Z_l)|X_i]\right\} \\
=& E_{\mathbf{P}}\left\{E_{\mathbf{P}}(Y_i|X_i)E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)|X_i]\right\} = E_{\mathbf{P}}\left\{Y_i E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)|X_i]\right\}, \\
\vartheta_2 =& E_{\mathbf{P}}\left\{a(Z_k, Z_j, Z_l)Y_i\right\} = E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}\left\{a(Z_k, Z_j, Z_l)\right\} \\
=& E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}\left\{E_{\mathbf{P}}[a(Z_k, Z_j, Z_l)|X_i]\right\},
\end{aligned}$$

hence,

$$\vartheta_1 - \vartheta_2 = Cov_{\mathbf{P}}(Y_i, \delta(X_i)) = E_{\mathbf{P}}\left\{Y_i[\delta(X_i) - E_{\mathbf{P}}(\delta(X_i))]\right\},$$

where $\delta(X_i) = E_{\mathbf{P}}[(X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l|X_i]$. Furthermore, we have

$$\begin{aligned}
\vartheta_3 =& E_{\mathbf{P}}\left\{a(Z_i, Z_j, Z_l)a(Z_i, Z_k, Z_m)\right\} = E_{\mathbf{P}}\left\{E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)a(Z_i, Z_k, Z_m)|X_i]\right\} \\
=& E_{\mathbf{P}}\left\{E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)|X_i]E_{\mathbf{P}}[a(Z_i, Z_k, Z_m)|X_i]\right\} = E_{\mathbf{P}}\left\{E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)|X_i]^2\right\} \\
=& E_{\mathbf{P}}(\delta(X_i)^2),
\end{aligned}$$

where the third equality follows from the structure of a (given X_i all variables occurring in $a(Z_i, Z_j, Z_l)$ are independent of those occurring in $a(Z_i, Z_k, Z_m)$). Additionally, we have

$$\vartheta_4 = E_{\mathbf{P}}\left\{a(Z_i, Z_j, Z_l)\right\} = E_{\mathbf{P}}\left\{E_{\mathbf{P}}[a(Z_i, Z_j, Z_l)|X_i]\right\} = E_{\mathbf{P}}(\delta(X_i)).$$

Therefore, we obtain

$$\iota_X^{loess}(Y) = \frac{\vartheta_1 - \vartheta_2}{\sqrt{\vartheta_3 - \vartheta_4^2}} = E_{\mathbf{P}}\left\{Y\frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(X)}}\right\},$$

where $\delta(X) = E_{\mathbf{P}}[(X_j - X)(X_j - X_l)K_h(X - X_j)K_h(X - X_l)Y_l|X_i]$,

$$\leq \sup_{\delta \in L^2_{\mathbb{P}}(\mathbb{R})} E_{\mathbf{P}}\left\{Y\frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(\delta(X))}}\right\} = \iota_X(Y).$$

Thus $\iota_X^{loess}(Y) \leq \iota_X(Y)$ and an asymptotic level α confidence interval for $\iota_X^{loess}(Y)$ is a (potentially conservative) asymptotic level α confidence interval for $\iota_X(Y)$.

Since the computation of $\hat{\sigma}$ in (2.12) requires similar to the case in Section 2.2.1 great computational effort if n is not very small it may be convenient to use bootstrap methods to calculate a confidence interval for $\iota_Y^{loess}(X)$ rather than to compute the variance estimator. Using the same arguments as in the previous section we obtain that the bootstrap is consistent in this setup, i.e. when we denote the estimator $\hat{\iota}_X^{loess}(Y)$ computed based on a bootstrap sample $Z_1^*, ..., Z_n^*$ by $\hat{\iota}_X^{loess^*}(Y)$ we have that for almost all sequences $Z_1, Z_2, ...$ the conditional distribution of $\sqrt{n}(\hat{\iota}_X^{loess^*}(Y) - \hat{\iota}_X^{loess}(Y))$ (given $Z_1, ..., Z_n$) and the distribution of $\sqrt{n}(\hat{\iota}_X^{loess}(Y) - \iota_X^{loess}(Y))$ converge to the same limiting distribution.

2.2.5. Impact analysis based on local polynomials

In generalization to Sections 2.2.1 and 2.2.4 we now choose a perturbation $\hat{\delta}$ based on the predictions of a degree k local polynomial fit. The case $k = 1$ then gives the ordinary kernel smoother, and $k = 2$ leads to local linear regression. The local polynomial regression estimator of degree k at x is, as also presented in Section A.1.1 given by

$$\hat{m}(x) = (1, x, ..., x^k) \left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)^{-1} \frac{1}{n}\mathbf{B}^T\mathbf{W}(x)\mathbf{Y},$$

where $\mathbf{B}$ is the matrix with $(1, X_i, ..., X_i^k)$ as i-th row, $W(x) = \text{diag}\left\{(K_h(X_i - x))_{i=1,..,n}\right\}$ for a symmetric kernel $K_h(u) = K(u/h)$ with bandwidth h and $\mathbf{Y} = (Y_1, ..., Y_n)^T$ is the vector of observations of the target variable. For the estimation of the mean impact we will use

$$\hat{\delta}(x) = (1, x, ..., x^k)\, \text{cof}\left\{\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)\right\} \frac{1}{n}\mathbf{B}^T\mathbf{W}(x)\mathbf{Y},$$

in (2.1). Note that $\hat{\delta}$ arises from $\hat{m}$ by replacing

$$\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)^{-1}$$

by

$$\text{cof}\left\{\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)\right\},$$

where cof denotes the matrix of cofactors. The matrix of cofactors of a quadratic $k \times k$-matrix A is given by

$$\{\text{cof}(A)\}_{ij} = (-1)^{i+j} M_{ij},$$

where M_{ij} is the determinant of the sub-matrix arising from A by canceling the i-th row

and the j-column. Additionally, we have according to Fischer (2005) that

$$\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)^{-1} = \det\left\{\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)\right\}^{-1}\left\{\text{cof}\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)\right\}^T.$$

Since $\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)$ is symmetric its matrix of cofactors is symmetric too. Hence, by replacing $\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)^{-1}$ by $\text{cof}\left\{\left(n^{-1}\mathbf{B}^T\mathbf{W}(x)\mathbf{B}\right)\right\}$ we only drop the determinant in the denomiator. In the following we will show that the estimator

$$\hat{\iota}_X^{locpol}(Y) = \frac{\tilde{\iota}_1 - \tilde{\iota}_2}{\sqrt{\tilde{\iota}_3 - \tilde{\iota}_4^2}},$$

with $\tilde{\iota}_1 = \frac{1}{n}\sum_{i=1}^n Y_i\hat{\delta}(X_i)$, $\tilde{\iota}_2 = \frac{1}{n}\sum_{i=1}^n Y_i\bar{\delta}$, $\tilde{\iota}_3 = \frac{1}{n}\sum_{i=1}^n \hat{\delta}(X_i)^2$ and $\tilde{\iota}_4 = \frac{1}{n}\sum_{i=1}^n \hat{\delta}(X_i)$ is asymptotically normal. We do so by showing that $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$ is essentially a (2k+3)rd-order U-statistics. Application of the delta-method yields the desired asymptotic normality of the estimator. As in the special cases above Bickel and Freedman (1981) provide the validity of the bootstrap in this scenario.

As a first step we note that

$$n^{-1}\mathbf{B}^T\mathbf{W}(X_i)\mathbf{B} = \left\{\frac{1}{n}\sum_{j=1}^n K_h(X_j - X_i)X_j^{(m+l-1)}\right\}_{l,m=1,...,k+1}.$$

Since the elements of the matrix of cofactors of this matrix are the signed determinants of $k \times k$ sub-matrices they are of the form

$$\frac{1}{n^k}\sum_{j_1=1}^n \cdots \sum_{j_k=1}^n h(Z_{j_1}, ..., Z_{j_k}),$$

where $h(Z_{j_1}, ..., Z_{j_k})$ is a sum respectively a difference of products of terms of the form $K_h(X_{j_a} - X_i)X_{j_a}^v$ for suitable vs. Hence, we can rewrite the matrix of cofactors as

$$\text{cof}\left\{n^{-1}\mathbf{B}^T\mathbf{W}(X_i)\mathbf{B}\right\} = \left\{\frac{1}{n^k}\sum_{j_1=1}^n \cdots \sum_{j_k=1}^n h_{lm}(Z_{j_1}, ..., Z_{j_k})\right\}_{l,m=1,...,k+1}.$$

Assumption 2.15. *Let*

$$g_1(Z_{j_1}, ..., Z_{j_{2k+3}}) = Y_{j_{k+2}}\tilde{w}(Z_{j_1}, ..., Z_{j_{k+1}}, Z_{j_{k+2}}),$$

$$g_2(Z_{j_1}, ..., Z_{j_{2k+3}}) = Y_{j_{k+3}}\tilde{w}(Z_{j_1}, ..., Z_{j_{k+2}}),$$

$$g_3(Z_{j_1},...,Z_{j_{2k+3}}) = \tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_{j_{2k+3}})\tilde{w}(Z_{j_{k+2}},...,Z_{j_{2k+2}},Z_{j_{2k+3}}),$$

$$g_4(Z_{j_1},...,Z_{j_{2k+3}}) = \tilde{w}(Z_{j_1},...,Z_{j_{k+2}}),$$

with

$$\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i) = \sum_{l=1}^{k+1}\sum_{m=1}^{k+1} h_{lm}(Z_{j_1},...,Z_{j_k})K_h(X_{j_{k+1}}-X_i)Y_{j_{k+1}}X_{j_{k+1}}^{m-1}X_i^{m-1}.$$

Furthermore, define $g = (g_1,...,g_4)$ *and let*

$$w(Z_{j_1},...,Z_{j_{2k+3}}) = \frac{1}{(2k+3)!}\sum_{\pi\in S(\{1,...,2k+3\})} g(Z_{j_{\pi(1)}},...,Z_{j_{\pi(2k+3)}}).$$

Additionally, assume that

$$E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{2k+3}})) \quad \textit{and} \quad E_{\mathbf{P}}(w^2(Z_{j_1},...,Z_{j_{2k+3}}))$$

exist for all $(j_1,...,j_{2k+3}) \in \{1,...,n\}^{2k+3}$ *and define* $\vartheta = E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{2k+3}}))$ *for* $j_1 < ... < j_{k+3}$.

Theorem 2.16. *Under Assumption 2.15 we have that*

$$\sqrt{n}\left\{\hat{\iota}_X^{locpol}(Y) - \iota_X^{locpol}(Y)\right\} \xrightarrow{\mathcal{L}} N(0,\sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T\Sigma DF(\vartheta)$, $F\left((a_1,...,a_4)^T\right) = \frac{a_1-a_2}{\sqrt{a_3-a_4^2}}$, $\iota_X^{locpol}(Y) = F(\vartheta)$ *and*

$$\Sigma = (2k+3)^2 E_{\mathbf{P}}\left\{\left(E_{\mathbf{P}}\left[w(Z_{j_1},...,Z_{j_{2k+3}})|Z_{j_1}\right] - \vartheta\right)\left(E_{\mathbf{P}}\left[w^T(Z_{j_1},...,Z_{j_{2k+3}})|Z_{j_1}\right] - \vartheta^T\right)\right\}.$$

Proof. The proof of this theorem can be found in Appendix B. □

Lemma 2.17. *The variance* σ^2 *can be consistently estimated by*

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T\hat{\Sigma}DF(\tilde{\iota}),$$

where

$$\hat{\Sigma} = (2k+3)^2\left[\binom{n}{4k+5}^{-1}\sum_{j_1<...<j_{4k+5}}\frac{1}{(4k+5)!}\sum_{\pi\in S(\{1,...,4k+5\})}\tilde{g}(Z_{j_{\pi(1)}},...,Z_{j_{\pi(4k+5)}}) - \tilde{\iota}\tilde{\iota}^T\right]$$

and

$$\tilde{g}(Z_{j_1},...,Z_{j_{4k+5}}) = w(Z_{j_1},...,Z_{j_{2k+3}})w^T(Z_{j_1},Z_{j_{2k+4}},...,Z_{j_{4k+5}}).$$

Proof. For the proof see Appendix B. □

From this result we obtain

$$\sqrt{n}\frac{\hat{\iota}_X^{locpol}(Y) - \iota_X^{locpol}(Y)}{\hat{\sigma}} \xrightarrow{\mathcal{L}} N(0,1).$$

For polynomial regression the argumentation of the previous sections, ensuring the consistency of the bootstrap and thus justifying the use of bootstrap confidence intervals apply as well. Finally, we will show that $\iota_X^{locpol}(Y) \leq \iota_X(Y)$. To this end we note that we can rewrite the elements of ϑ as

$$\begin{aligned}
\vartheta_1 &= E_{\mathbf{P}}\left(Y_i\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)\right)\\
\vartheta_2 &= E_{\mathbf{P}}\left(Y_i\tilde{w}(Z_{j_1},...,Z_{j_{k+2}})\right)\\
\vartheta_3 &= E_{\mathbf{P}}\left(\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)\tilde{w}(Z_{j_{k+2}},...,Z_{j_{2k+2}},Z_i)\right)\\
\vartheta_4 &= E_{\mathbf{P}}\left(\tilde{w}(Z_{j_1},...,Z_{j_{k+2}})\right).
\end{aligned}$$

From this it follows that

$$\begin{aligned}
\vartheta_1 &= E_{\mathbf{P}}\left\{E_{\mathbf{P}}\left(Y_i\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i\right)\right\}\\
&= E_{\mathbf{P}}\{E_{\mathbf{P}}(Y_i|X_i)E_{\mathbf{P}}[\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i]\}\\
&= E_{\mathbf{P}}\{Y_i \underbrace{E_{\mathbf{P}}[\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i]}_{=:\delta(X_i)}\}
\end{aligned}$$

and

$$\vartheta_2 = E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}(\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)) = E_{\mathbf{P}}(Y_i)E_{\mathbf{P}}(\delta(X_i)),$$

hence, $\vartheta_1 - \vartheta_2 = Cov_{\mathbf{P}}(Y_i,\delta(X_i)) = E_{\mathbf{P}}\left\{Y_i[\delta(X_i) - E_{\mathbf{P}}(\delta(X_i))]\right\}$. Furthermore, we have

$$\begin{aligned}
\vartheta_3 =& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)\tilde{w}(Z_{j_{k+2}},...,Z_{j_{2k+2}},Z_i)|X_i\right\}\right)\\
=& E_{\mathbf{P}}\left\{E_{\mathbf{P}}\left[\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i\right]E_{\mathbf{P}}\left[\tilde{w}(Z_{j_{k+2}},...,Z_{j_{2k+2}},Z_i)|X_i\right]\right\}\\
=& E_{\mathbf{P}}\left\{E_{\mathbf{P}}\left[\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i\right]^2\right\} = E_{\mathbf{P}}(\delta(X_i)^2),
\end{aligned}$$

where the second equality holds because of the structure of $\tilde{w}$ ($\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)$ does not use Y_i which implies that the conditional independence needed for the second equality

holds). Additionally we have

$$\vartheta_4 = E_{\mathbf{P}}\left\{E_{\mathbf{P}}\left(\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)|X_i\right)\right\} = E_{\mathbf{P}}(\delta(X_i)).$$

Consequently we obtain

$$\begin{aligned}\iota_X^{locpol}(Y) =& \frac{\vartheta_1 - \vartheta_2}{\sqrt{\vartheta_3 - \vartheta_4^2}} = E_{\mathbf{P}}\left\{Y\frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(\delta(X))}}\right\} \\ \leq & \sup_{\delta \in L^2_{\mathbf{P}}(\mathbb{R})} E_{\mathbf{P}}\left\{Y\frac{\delta(X) - E_{\mathbf{P}}(\delta(X))}{\sqrt{Var_{\mathbf{P}}(\delta(X))}}\right\} = \iota_X(Y).\end{aligned}$$

2.2.6. Common impact based on kernel-smoothing

In this section we generalize the approach of the kernel method based impact to the case where we want to quantify the common influence of a set of covariates $X_1,...,X_k$ onto the target variable Y. Therefore, we estimate the impact of $\mathbf{X} = (X_1,...,X_k)$ on Y based on kernel smoothing by

$$\hat{\iota}_{\mathbf{X}}^{ks}(Y) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{\delta}_1(\mathbf{X}_i) - \bar{\delta}_1(\mathbf{X})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\delta}_1(\mathbf{X}_j) - \bar{\delta}_1(\mathbf{X})\right)^2}},$$

where $\mathbf{X}_i = (X_{i1},...,X_{ik})$ and

$$\hat{\delta}_1(\mathbf{X}_i) = \frac{1}{n}\sum_{j=1}^{n} K_h(\mathbf{X}_i - \mathbf{X}_j)Y_j,$$

for a kernel

$$K_h(\mathbf{X}_i - \mathbf{X}_j) = D\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{h}\right)$$

and D from Section A.1.1. With this choice of the estimated impact we are essentially in the situation of Section 2.2.1. The same argumentation applies here and leads to asymptotic normality of the estimator for the restricted impact. Furthermore, it follows, that the bootstrap is consistent in this case.

A generalization to higher order local regression in the variables $X_1,...,X_k$ is straight forward and leads to the setup of Section 2.2.5.

2.2.7. Modification of the Kernel-smoother-based impact

In Section 2.2.1 we dealt with an impact based on the predictions of kernel smoothing where the denominator of the kernel-smoother was left out. We now consider a kernel-smoother based impact without dropping the denominator. Since the denominator is a function of all observations and the estimator of the impact is therefore not a U-statistics anymore the methods to derive the asymptotic distribution of the estimator used in 2.2.1 do not apply here. Nevertheless we can deal with this problem by modifying the way we estimate the impact. We will use all observations to compute the denominator of the kernel smoother but only $m_n < n$ observations to compute the estimator for the impact. Choosing a suitable m_n we can show the asymptotic normality of the estimator. This procedure is particularly useful and interesting in cases where we do not have observations of the target variable Y for all individuals. This might occur in large health insurance data sets, where the covariates (e.g. sex, age) are present for all insured people but the target variable is only observed for a subset of the insured people.

Note, that the results of Doksum and Samarov (1995) may be used to derive an asymptotic theory for an estimator for the kernel smoother based impact without dropping the denominator using all observations by using a leave-one-out type estimator for $\hat{\delta}$. However, this approach is not pursued in this thesis.

Again we need to assume that the bandwidth $h > 0$ of the symmetric kernel $K_h(u) = K(|u/h|)$ is fixed. We now need the additional assumptions

Assumption 2.18. *1. K is non-negative and bounded;*

2. $\int K(x)dx = 1$;

3. $K(x) = p(|x|)$, where p is a monotone decreasing function on $[0, \infty)$.

Note that for example the Gaussian kernel or the Epanechnikov quadratic kernel meet these assumptions. We now choose the perturbation

$$\hat{\delta}(x) = \frac{1}{m_n} \sum_{j=1}^{m_n} \frac{K_h(x - X_j)Y_j}{\frac{1}{n}\sum_{l=1}^{n} K_h(x - X_l)},$$

where $m_n \to \infty$ and the following assumption holds.

Assumption 2.19. *Let $m_n \to \infty$ such that $\frac{n \log(n)^{-1}}{m_n} \to \infty$.*

We then estimate the impact by

$$\hat{\iota}_X^{ks,mod}(Y) = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i \frac{\hat{\delta}(X_i) - \bar{\hat{\delta}}}{\sqrt{\frac{1}{m_n}\sum_{i=1}^{m_n}\left(\hat{\delta}(X_i) - \bar{\hat{\delta}}\right)^2}},$$

where $\bar{\hat{\delta}} = \frac{1}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)$. Similar to the cases before we decompose the estimator to

$$\tilde{\iota}_1 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i\hat{\delta}(X_i),\ \tilde{\iota}_2 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i\bar{\hat{\delta}},\ \tilde{\iota}_3 = \frac{1}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)^2,\ \tilde{\iota}_4 = \frac{1}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i),$$

with $\hat{\iota}_X^{ks,mod}(Y) = \frac{\tilde{\iota}_1 - \tilde{\iota}_2}{\sqrt{\tilde{\iota}_3 - \tilde{\iota}_4^2}}$. Let $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$ and

$$\tilde{f}(u) = E_{\mathbf{P}}\left(K_h(u - X)\right),$$

assuming this exists.

Assumption 2.20. *Let*

$$g_1(Z_i, Z_j, Z_l) = \frac{K_h(X_i - X_j)}{\tilde{f}(X_i)}Y_iY_j, \quad g_2(Z_i, Z_j, Z_l) = \frac{K_h(X_j - X_l)}{\tilde{f}(X_j)}Y_iY_j,$$

$$g_3(Z_i, Z_j, Z_l) = \frac{K_h(X_i - X_j)K_h(X_i - X_l)}{\tilde{f}(X_i)^2}Y_jY_l, \quad \textit{and} \quad g_4(Z_i, Z_j, Z_l) = \frac{K_h(X_i - X_j)}{\tilde{f}(X_i)}Y_j$$

and define $g = (g_1, ..., g_4)$. *Let* w *be defined as in* (2.3). *We assume that the following quantities exist for all* $(i, j, l) \in \{1, ..., n\}^3$*:*

$$E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)), \quad E_{\mathbf{P}}(w^2(Z_i, Z_j, Z_l)).$$

Definition 2.21. *Let* $i \neq j \neq l \neq i$ *the following parameters exist by Assumption 2.20*

$$\vartheta_k := E_{\mathbf{P}}\left(g_k(Z_i, Z_j, Z_l)\right), \quad \textit{for } k \in \{1, ..., 4\}.$$

Assumption 2.22. *There exsists* $\eta > 0$ *such that* $\tilde{f}(x) \geq \eta\ \forall x \in \text{supp}(X)$, *where* $\text{supp}(X)$ *denotes the support of* X. *Furthermore let the density* f *of* X *be bounded.*

With this assumption we can state the following lemma.

Lemma 2.23. *We have that*

$$\sqrt{m_n}\sup_{x\in\mathbb{R}}|\tilde{f}(x)-\hat{f}(x)|\xrightarrow{a.s.}0,$$

where $\hat{f}(x)=\frac{1}{n}\sum_{j=1}^{n}K_h(x-X_j)$.

Proof. Adopting the notation of Pollard (1984) we write $\hat{f}(x)=\mathbb{P}_nK_{h,x}$ and $\tilde{f}(x)=\mathbb{P}K_{h,x}$ and obtain

$$\sup_{x\in\mathbb{R}}|\hat{f}(x)-\tilde{f}(x)|=\sup_{x\in\mathbb{R}}|\mathbb{P}_nK_{h,x}-\mathbb{P}K_{h,x}|.$$

Let $\tilde{K}_h(u)=K_h(u)/c$, where c is an upper bound of K which exists according to Assumption 2.18. Note that we have $0\leq\tilde{K}\leq 1$. We then have

$$\sup_{x\in\mathbb{R}}|\mathbb{P}_nK_{h,x}-\mathbb{P}K_{h,x}|=c\sup_{x\in\mathbb{R}}|\mathbb{P}_n\tilde{K}_{h,x}-\mathbb{P}\tilde{K}_{h,x}|.$$

According to Pollard (1984, p.36) and his Lemma 25 all kernels that meet Assumption 2.18 also meet the assumptions of his Theorem 37 which we will apply here. Denoting the density of X by f we obtain

$$\begin{aligned}\mathbb{P}\tilde{K}_{h,x}^2&=\int\tilde{K}^2\left(\frac{x-y}{h}\right)f(y)dy\leq\int\tilde{K}\left(\frac{x-y}{h}\right)f(y)dy\\&=h\int\tilde{K}(t)f(x+ht)dt\leq qh\int\tilde{K}(t)dt=qh\frac{1}{c},\end{aligned}$$

where q is an upper bound of f, which exists by Assumption 2.22. We choose $\delta=qh/c$ and $\alpha_n=m_n^{-1/2}$ and obtain, since $\frac{n\log(n)^{-1}}{m_n}\to\infty$, that $n\delta^2\alpha_n^2\log(n)^{-1}\to\infty$. Hence we can apply Theorem 37 of Pollard (1984) which gives

$$\delta^{-2}\sqrt{m_n}\sup_{x\in\mathbb{R}}|\mathbb{P}_n\tilde{K}_{h,x}-\mathbb{P}\tilde{K}_{h,x}|\xrightarrow{a.s.}0,$$

which implies

$$\sqrt{m_n}\sup_{x\in\mathbb{R}}|\tilde{f}(x)-\hat{f}(x)|\xrightarrow{a.s.}0.$$

□

We now show that we can replace $\hat{f}$ by $\tilde{f}$ in $\hat{\delta}$ without affecting the asymptotic behavior of $\tilde{\iota}_1,...,\tilde{\iota}_4$.

Lemma 2.24. *Under the assumptions of this section have that*

$$\sqrt{m_n}\tilde{\iota} = \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n} w(Z_i, Z_j, Z_l) + o_p(1).$$

Proof. To this end let

$$L_{n,i} = Y_i\hat{\delta}(X_i)\frac{\hat{f}(X_i)}{\tilde{f}(X_i)}$$

and

$$A_{n,i} = Y_i\hat{\delta}(X_i)\frac{\tilde{f}(X_i) - \hat{f}(X_i)}{\tilde{f}(X_i)}.$$

We then obtain

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_1 =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}(L_{n,i} + A_{n,i})\\ =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} L_{n,i} + \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} L_{n,i}\frac{A_{n,i}}{L_{n,i}}\end{aligned}$$

For the second term of this expression we make the observation that

$$\begin{aligned}\left|\frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} L_{n,i}\frac{A_{n,i}}{L_{n,i}}\right| \leq& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}|L_{n,i}|\left|\frac{\tilde{f}(X_i) - \hat{f}(X_i)}{\hat{f}(X_i)}\right|\\ \leq& \sqrt{m_n}\sup_{x\in\mathbb{R}}\left|\frac{\tilde{f}(x) - \hat{f}(x)}{\hat{f}(x)}\right|\frac{1}{m_n}\sum_{i=1}^{m_n}|L_{n,i}|.\end{aligned}$$

Since $\tilde{f}$ is bounded away from zero on its support and $\sup_{x\in\mathbb{R}}|\tilde{f}(x) - \hat{f}(x)| \overset{a.s.}{\to} 0$ (by Lemma 2.23) we have that $\hat{f}(x) \geq \tilde{\eta}$ for all $x \in \operatorname{supp}(X)$ for suitable large n. Hence, for those n we can write

$$\sqrt{m_n}\sup_{x\in\mathbb{R}}\left|\frac{\tilde{f}(x) - \hat{f}(x)}{\hat{f}(x)}\right| \leq \frac{1}{\tilde{\eta}}\sqrt{m_n}\sup_{x\in\mathbb{R}}\left|\tilde{f}(x) - \hat{f}(x)\right| \overset{a.s.}{\to} 0.$$

Furthermore, we have

$$0 \leq \frac{1}{m_n}\sum_{i=1}^{m_n}|L_{n,i}| \leq \frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\underbrace{\left|\frac{K_h(X_i - X_j)Y_iY_j}{\tilde{f}(X_i)}\right|}_{=:\gamma(Z_i,Z_j)}$$

$$
\begin{aligned}
&= \frac{1}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \lambda(Z_i, Z_j) \\
&= \frac{1}{m_n^2} \sum_{i \neq j}^{m_n} \lambda(Z_i, Z_j) + \frac{1}{m_n^2} \sum_{i=1}^{m_n} \lambda(Z_i, Z_i),
\end{aligned}
$$

where $\lambda(Z_i, Z_j) = 1/2\{\gamma(Z_i, Z_j) + \gamma(Z_j, Z_i)\}$. Since by Assumption 2.20 $E(\lambda(Z_i, Z_j))$ exists for all $(i, j) \in \{1, ..., n\}^2$ this equals

$$
\begin{aligned}
&= \frac{1}{m_n^2} \sum_{i<j} 2\lambda(Z_i, Z_j) + o_p(1) \\
&= \frac{m_n(m_n - 1)}{m_n^2} \binom{m_n}{2}^{-1} \sum_{i<j} \lambda(Z_i, Z_j) + o_p(1) \\
&\xrightarrow{p} E\{\lambda(Z_i, Z_j)\} < \infty
\end{aligned}
$$

according to the theory of U-statistics (cf. Theorem A.7). Hence,

$$
\begin{aligned}
\sqrt{m_n}\tilde{\iota}_1 &= \frac{\sqrt{m_n}}{m_n} \sum_{i=1}^{m_n} L_{n,i} + o_p(1) \\
&= \sqrt{m_n} \frac{1}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \underbrace{\frac{K_h(X_i - X_j) Y_i Y_j}{\tilde{f}(X_i)}}_{=: g_1(Z_i, Z_j, Z_l)} + o_p(1) \\
&= \sqrt{m_n} \frac{1}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} w_1(Z_i, Z_j, Z_l) + o_p(1),
\end{aligned}
$$

where $w_1(Z_i, Z_j, Z_l) = \frac{1}{6} \sum_{\pi \in S(\{i,j,l\})} g_1(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)})$. In the following we will show, that we can replace the denominators in the expressions $\sqrt{m_n}\tilde{\iota}_2, \sqrt{m_n}\tilde{\iota}_3, \sqrt{m_n}\tilde{\iota}_4$ similarly. We can decompose $\sqrt{m_n}\tilde{\iota}_2$ to

$$
\begin{aligned}
\sqrt{m_n}\tilde{\iota}_2 &= \frac{\sqrt{m_n}}{m_n} \sum_{l=1}^{m_n} \hat{\delta}(X_l) \bar{Y} \\
&= \frac{\sqrt{m_n}}{m_n} \sum_{l=1}^{m_n} (\tilde{L}_{n,l} + \tilde{A}_{n,l}) \\
&= \frac{\sqrt{m_n}}{m_n} \sum_{l=1}^{m_n} \tilde{L}_{n,l} + \frac{\sqrt{m_n}}{m_n} \sum_{l=1}^{m_n} \tilde{L}_{n,l} \frac{\tilde{f}(X_l) - \hat{f}(X_l)}{\hat{f}(X_l)},
\end{aligned}
$$

where $\tilde{L}_{n,l} = \hat{\delta}(X_l)\bar{Y}\frac{\hat{f}(X_l)}{\tilde{f}(X_l)}$ and $\tilde{A}_{n,l} = \hat{\delta}(X_l)\bar{Y}\frac{\tilde{f}(X_l)-\hat{f}(X_l)}{\tilde{f}(X_l)}$. Again, we can show that the second term converges to zero in probability. We do so by observing

$$\left|\frac{\sqrt{m_n}}{m_n}\sum_{l=1}^{m_n}\tilde{L}_{n,l}\frac{\tilde{f}(X_l)-\hat{f}(X_l)}{\hat{f}(X_l)}\right| \leq \sqrt{m_n}\sup_{x\in\mathbb{R}}\left|\frac{\tilde{f}(x)-\hat{f}(x)}{\hat{f}(x)}\right|\frac{1}{m_n}\sum_{l=1}^{m_n}|\tilde{L}_{n,l}|.$$

The first term of this expression converges to zero by Lemma 2.23. For the second term we have

$$\begin{aligned}\frac{1}{m_n}\sum_{l=1}^{m_n}|\tilde{L}_{n,l}| \leq& \frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\underbrace{\left|\frac{K_h(X_l-X_j)Y_iY_j}{\tilde{f}(X_l)}\right|}_{\gamma_2(Z_i,Z_j,Z_l)}\\ =&\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\lambda_2(Z_i,Z_j,Z_l),\end{aligned}$$

where $\lambda_2(Z_i,Z_j,Z_l) = 1/3!\sum_{\pi\in S(\{i,j,l\})}\gamma_2(Z_{\pi(i)},Z_{\pi(j)},Z_{\pi(l)})$. With Assumption 2.20 and Lemma 2.1 it follows that

$$\begin{aligned}=&\frac{1}{m_n^3}\sum_{C(\{i,j,l\})}\lambda_2(Z_i,Z_j,Z_l)+o_p(m_n^{-1/2})\\ =&\frac{m_n^3-3m_n^2+2m_n}{m_n^3}\binom{m_n}{3}^{-1}\sum_{i<j<l}\lambda_2(Z_i,Z_j,Z_l)+o_p(m_n^{-1/2})\\ \xrightarrow{p}&E\{\lambda_2(Z_i,Z_j)\}<\infty\end{aligned}$$

by Theorem A.7. Concluding, we have that

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_2 =&\frac{\sqrt{m_n}}{m_n}\sum_{l=1}^{m_n}\tilde{L}_{n,l}+o_p(1)\\ =&\sqrt{m_n}\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\underbrace{\frac{K_h(X_i-X_j)Y_jY_l}{\tilde{f}(X_i)}}_{=:g_2(Z_i,Z_j,Z_l)}+o_p(1)\\ =&\sqrt{m_n}\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}w_2(Z_i,Z_j,Z_l)+o_p(1),\end{aligned}$$

where $w_2(Z_i,Z_j,Z_l) = \frac{1}{6}\sum_{\pi\in S(\{i,j,l\})}g_2(Z_{\pi(i)},Z_{\pi(j)},Z_{\pi(l)})$. By decomposing analogously

to the previous two cases and by application of Lemma 2.23 it follows that we have

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_3 =& \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\frac{K_h(X_i-X_l)K_h(X_i-X_j)Y_jY_l}{\hat{f}(X_i)^2}\\ =& \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}(L_i+A_i)\\ =& \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}L_i+\frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}L_i\frac{A_i}{L_i},\end{aligned}$$

where

$$L_i=\frac{1}{m_n^2}\frac{\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}K_h(X_i-X_j)K_h(X_i-X_l)Y_jY_l}{\hat{f}(X_i)\tilde{f}(X_i)}$$

and

$$A_i=\frac{1}{m_n^2}\frac{\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}K_h(X_i-X_j)K_h(X_i-X_l)Y_jY_l}{\hat{f}(X_i)^2}\frac{\tilde{f}(X_i)-\hat{f}(X_i)}{\tilde{f}(X_i)}.$$

Thus,

$$\sqrt{m_n}\tilde{\iota}_3=\frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}L_i+\frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}L_i\frac{\tilde{f}(X_i)-\hat{f}(X_i)}{\hat{f}(X_i)}.$$

For the last term we obtain,

$$\left|\frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}L_i\frac{\tilde{f}(X_i)-\hat{f}(X_i)}{\hat{f}(X_i)}\right|$$

$$\leq \underbrace{\sqrt{m_n}\sup_{x\in\mathbb{R}}\left|\frac{\tilde{f}(x)-\hat{f}(x)}{\hat{f}(x)}\right|}_{=o_p(1)}\underbrace{\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\left|\frac{K_h(X_i-X_j)K_h(X_i-X_l)Y_jY_l}{\hat{f}(X_i)\tilde{f}(X_i)}\right|}_{(a)}.$$

As a next step we examine the term (a).

$$\begin{aligned}(a) =& \frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\frac{|K_h(X_i-X_j)K_h(X_i-X_l)Y_jY_l|}{\tilde{f}(X_i)^2}\\ &+\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\frac{|K_h(X_i-X_j)K_h(X_i-X_l)Y_jY_l|}{\tilde{f}(X_i)^2}\frac{\tilde{f}(X_i)-\hat{f}(X_i)}{\hat{f}(X_i)}\end{aligned}$$

$$\leq \frac{1}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \frac{|K_h(X_i - X_j) K_h(X_i - X_l) Y_j Y_l|}{\tilde{f}(X_i)^2}$$
$$+ \sup_{x \in \mathbb{R}} \left| \frac{\tilde{f}(x) - \hat{f}(x)}{\hat{f}(x)} \right| \frac{1}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \frac{|K_h(X_i - X_j) K_h(X_i - X_l) Y_j Y_l|}{\tilde{f}(X_i)^2}.$$

By application of the theory of U-statistics together with Assumption 2.20 it can be shown that the expression

$$\frac{1}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \frac{|K_h(X_i - X_j) K_h(X_i - X_l) Y_j Y_l|}{\tilde{f}(X_i)^2}$$

converges to

$$E_{\mathbf{P}} \left(\frac{|K_h(X_i - X_j) K_h(X_i - X_l) Y_j Y_l|}{\tilde{f}(X_i)^2} \right) < \infty.$$

Furthermore, $\sup_{x \in \mathbb{R}} \left| \frac{\tilde{f}(x) - \hat{f}(x)}{\hat{f}(x)} \right| = o_p(1)$ by Lemma 2.23. Thus, we obtain that

$$(a) \xrightarrow{p} E_{\mathbf{P}} \left(\frac{|K_h(X_i - X_j) K_h(X_i - X_l) Y_j Y_l|}{\tilde{f}(X_i)^2} \right) < \infty.$$

Thereby

$$\left| \frac{\sqrt{m_n}}{m_n^3} \sum_{i=1}^{m_n} L_i \frac{\tilde{f}(X_i) - \hat{f}(X_i)}{\hat{f}(X_i)} \right| = o_p(1).$$

This implies

$$\sqrt{m_n} \tilde{\iota}_3 = \frac{\sqrt{m_n}}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \frac{K_h(X_i - X_l) K_h(X_i - X_j) Y_j Y_l}{\hat{f}(X_i)^2}$$
$$= \frac{\sqrt{m_n}}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \frac{K_h(X_i - X_l) K_h(X_i - X_j) Y_j Y_l}{\hat{f}(X_i) \tilde{f}(X_i)} + o_p(1).$$

Repeated application of Lemma 2.23 gives

$$= \frac{\sqrt{m_n}}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \underbrace{\frac{K_h(X_i - X_l) K_h(X_i - X_j) Y_j Y_l}{\tilde{f}(X_i)^2}}_{=: g_3(Z_i, Z_j, Z_l)} + o_p(1)$$
$$= \frac{\sqrt{m_n}}{m_n^3} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} w_3(Z_i, Z_j, Z_l) + o_p(1),$$

where $w_3(Z_i, Z_j, Z_l) = \frac{1}{6}\sum_{\pi\in S(\{i,j,l\})} g_3(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)})$. Similar to the cases of $\tilde{\iota}_1, ..., \tilde{\iota}_3$ we obtain that

$$\sqrt{m_n}\tilde{\iota}_4 = \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)$$

which equals by suitable decomposition and application of Lemma 2.23

$$\begin{aligned}
&= \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)\frac{\hat{f}(X_i)}{\tilde{f}(X_i)} + o_p(1)\\
&= \frac{\sqrt{m_n}}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\underbrace{\frac{K_h(X_i - X_j)Y_j}{\tilde{f}(X_i)}}_{=:g_4(Z_i,Z_j,Z_l)} + o_p(1)\\
&= \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n} w_4(Z_i, Z_j, Z_l) + o_p(1),
\end{aligned}$$

where $w_4(Z_i, Z_j, Z_l) = \frac{1}{6}\sum_{\pi\in S(\{i,j,l\})} g_4(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)})$. Thus, summarizing the results we obtain:

$$\sqrt{m_n}\tilde{\iota} = \frac{\sqrt{m_n}}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n} w(Z_i, Z_j, Z_l) + o_p(1)$$

□

With the preceding results we are able to prove the asymptotic normality of the modified version of the kernel smoother based impact estimator.

Theorem 2.25. *Under the assumptions of this section we have*

$$\sqrt{m_n}(\hat{\iota}_X^{ks,mod}(Y) - \iota_X^{ks,mod}(Y)) \xrightarrow{\mathcal{L}} N(0, \underbrace{DF(\vartheta)^T V DF(\vartheta)}_{\sigma^2}),$$

where $F(a,b,c,d) = \frac{a-b}{\sqrt{c-d^2}}$, $\hat{\iota}_X^{ks,mod}(Y) = F(\tilde{\iota})$ *and* $\iota_X^{ks,mod}(Y) = F(\vartheta)$ *(ks,mod stands for kernel smoother modified). In this case the covariance matrix* V *is given by* $V = 9E_{\mathbf{P}}\left\{\tilde{w}(Z_i)\tilde{w}(Z_i)^T\right\}$ *with* $\tilde{w}(Z_i) = E_{\mathbf{P}}\left(w(Z_i, Z_j, Z_l)|Z_i\right) - \vartheta$.

Proof. We obtain

$$\sqrt{m_n}(\tilde{\iota} - \vartheta) = \sqrt{m_n}\left(\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n} w(Z_i, Z_j, Z_l) - \vartheta\right) + o_p(1).$$

According to Lemma 2.1 we have

$$\frac{1}{m_n^3}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\sum_{l=1}^{m_n} w(Z_i,Z_j,Z_l) = \frac{1}{m_n^3}\sum_{C(\{i,j,l\})} w(Z_i,Z_j,Z_l) + o_p(m_n^{-1/2}).$$

Furthermore, we can rewrite this as

$$\begin{aligned}&\frac{1}{m_n^3}\sum_{C(\{i,j,l\})} w(Z_i,Z_j,Z_l) + o_p(m_n^{-1/2})\\ =&\underbrace{\frac{m_n^3-3m_n^2+2m_n}{m_n^3}}_{c_n}\underbrace{\binom{m_n}{3}^{-1}\sum_{i<j<l} w(Z_i,Z_j,Z_l)}_{U_n} + o_p(m_n^{-1/2}).\end{aligned}$$

This means that we have

$$\sqrt{m_n}(\tilde{\iota}-\vartheta) = c_n\sqrt{m_n}(U_n-\vartheta) + o_p(1) + (c_n-1)\sqrt{m_n}\vartheta,$$

where $(c_n-1)\sqrt{m_n}\to 0$ and $c_n\to 1$. Theorem A.7 together with Assumption 2.20 gives that

$$\sqrt{m_n}(U_n-\vartheta)\xrightarrow{\mathcal{L}} N(0,V)$$

and consequently

$$\sqrt{m_n}(\tilde{\iota}-\vartheta)\xrightarrow{\mathcal{L}} N(0,V),$$

where $V = 9E_{\mathbf{P}}\left\{\tilde{w}(Z_i)\tilde{w}(Z_i)^T\right\}$ with $\tilde{w}(Z_i) = E_{\mathbf{P}}\left(w(Z_i,Z_j,Z_l)|Z_i\right) - \vartheta$. Application of the delta method with $F(a,b,c,d) = \frac{a-b}{\sqrt{c-d^2}}$ yields the asymptotic normality of the estimator for the impact

$$\sqrt{m_n}(\hat{\iota}_X^{ks,mod}(Y) - \iota_X^{ks,mod}(Y))\xrightarrow{\mathcal{L}} N(0,\underbrace{DF(\vartheta)^T V DF(\vartheta)}_{\sigma^2}),$$

where $\hat{\iota}_X^{ks,mod}(Y) = F(\tilde{\iota})$ and $\iota_X^{ks,mod}(Y) = F(\vartheta)$. □

Similarly to the other kernel method based impacts it can be shown that $\iota_X^{ks,mod}(Y)\leq \iota_X(Y)$ (the proof is omitted here). We are now interested in finding a consistent estimator for σ^2.

Lemma 2.26. *The variance σ^2 can be consistently estimated by $DF(\tilde{\iota})^T 9(\hat{g}^* - \tilde{\iota}\tilde{\iota}^T)DF(\tilde{\iota})$,*

where

$$\hat{g}^* = \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \hat{w}(Z_i, Z_j, Z_l)\hat{w}^T(Z_i, Z_a, Z_b)$$

and $\hat{w}(Z_i, Z_j, Z_l)$ *is obtained from* $w(Z_i, Z_j, Z_l)$ *by replacing all* $\tilde{f}s$ *by* $\hat{f}s$.

Proof. The proof makes use of Lemma 2.23 and is otherwise very similar to the proofs of the analogous results of the preceding sections but with more complex sums. It can be found in Appendix B. □

Similar to the cases before, the computation of that variance estimate is very computer intensive for usual sample sizes. However, we do not yet have a justification for bootstrap methods in this scenario. This could be the subject of future research.

2.2.8. Another modification of the Kernel-smoother-based impact

In the previous section we introduced a modification to the kernel-smoother-based impact, where we did not need to drop the denominator of the kernel smoother at the cost that we could not use all data in the estimation of the impact. Furthermore, in all considerations up to this point we needed the bandwidth of the kernel smoother to be fixed. In this section we present an alternative modification to the estimator based on kernel smoothing, which enables us to keep the denominator of the kernel smoother as well as to use a bandwidth h that decreases with the sample size. However, the price to be paid for these advantages is that again we can not use all available data in the estimation of the mean impact and that we need to assume that X has bounded support. The new modification makes use of the results of Mack and Silverman (1982) and uses the estimator

$$\hat{\iota}_X^{ks,mod2}(Y) = \frac{1}{m_n} \sum_{i=1}^{m_n} Y_i \frac{\hat{\delta}(X_i) - \overline{\hat{\delta}(X)}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{\delta}(X_i) - \bar{\hat{\delta}}(X)\right)^2}}, \tag{2.13}$$

where

$$\hat{\delta}(x) = \frac{\sum_{j=1}^{n} K_h(x - X_j)Y_j}{\sum_{j=1}^{n} K_h(x - X_j)},$$

the ordinary kernel smoother. Note, that we use all data in the kernel-smoothing while we only use the first $m_n < n$ observations for the estimation of the mean impact. Furthermore, we do no longer need to assume that the bandwidth h is fixed. In order for

the asymptotic to work we need the following assumptions which originate from Mack and Silverman (1982).

Assumption 2.27.
- *K is uniformly continuous with modulus of continuity w_K, i.e. $|K(x) - K(y)| \leq w_K(|x - y|)$ for all $x, y \in \text{supp}(K)$ and $w_K : [0, \infty] \to [0, \infty]$ is continuous at zero with $w_K(0) = 0$. Furthermore K is of bounded variation $V(K)$;*
- *K is absolutely integrable with respect to the Lebesgue measure on the line;*
- *$K(x) \to 0$ as $|x| \to \infty$;*
- *$\int |x \log |x||^{\frac{1}{2}} |dK(x)| < \infty$,*

Assumption 2.28.
- *$E_{\mathbf{P}}|Y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)dy < \infty$, $s \geq 2$;*
- *f, g and l are continuous on an open interval containing the bounded interval J, where f is the joint density of X and Y, g is the marginal density of X and $l(x) = \int yf(x, y)dy$*

Assumption 2.29. *$m_n \to \infty$ and $h \to 0$ as $n \to \infty$ in a way that*

$$\Big(\frac{m_n}{nh} log(1/h)\Big)^{1/2} \to 0$$

and $m_n < n$ for all $n \in \mathbb{N}$.

Assumption 2.30. *(a) The support J of X is a bounded interval, on which its density is bounded away from zero.*

(b) The density of X and $\int f(x, y)y\, dy$ have bounded second derivatives, where $f(x, y)$ is the joint density of X and Y.

(c) $h^2 = o(a_n)$, where

$$a_n = \left(\frac{1}{nh} \log \frac{1}{h}\right)^{1/2}.$$

(d) $n^{2\eta-1}h \to \infty$ for some $\eta < 1 - s^{-1}$ and $s > 2$.

(e) The kernel function K is symmetric.

With this we can state the following theorem.

Theorem 2.31. *Assume that Assumptions 2.27, 2.28 and 2.30 hold. Then we have that*

$$a_n^{-1} \sup_{x \in J} \left| \hat{\delta}(x) - E_{\mathbf{P}}(Y|x) \right| = O_p(1).$$

Proof. For the proof see Theorem B from Mack and Silverman (1982). □

The following Corollary follows immediately.

Corollary 2.32. *Assume that the Assumption 2.29 as well as conditions of Theorem 2.31 are met. Then we have that*

$$\sqrt{m_n} \sup_{x \in J} \left| \hat{\delta}(x) - E_{\mathbf{P}}(Y|x) \right| = o_p(1).$$

To show the asymptotic normality of the estimate $\hat{\iota}_X^{ks,mod2}(Y)$ from (2.13) we need an additional assumption.

Assumption 2.33. *We define*

$$g_1(Z_i, Z_j) = Y_i E_{\mathbf{P}}(Y|X_i) \quad g_2(Z_i, Z_j) = Y_i E_{\mathbf{P}}(Y|X_j)$$

$$g_3(Z_i, Z_j) = E_{\mathbf{P}}(Y|X_i) E_{\mathbf{P}}(Y|X_j) \quad g_4(Z_i, Z_j) = E_{\mathbf{P}}(Y|X_i)$$

as well as $g = (g_1, ..., g_4)^T$ *and* $w(Z_i, Z_j) = \frac{1}{2}\left(g(Z_i, Z_j) + g(Z_j, Z_i)\right)$ *and assume that*

$$E_{\mathbf{P}}\left(w(Z_i, Z_j)\right) \quad \text{and} \quad E_{\mathbf{P}}\left(w^2(Z_i, Z_j)\right)$$

exist for all $(i, j) \in \{1, ..., m_n\}^2$. *Additionally, define* $\vartheta = E(w(Z_i, Z_j))$ *for* $i \neq j$.

Theorem 2.34. *Under the assumptions of this section we have that*

$$\sqrt{m_n}\left(\hat{\iota}_X^{ks,mod2}(Y) - \iota_X(Y)\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T \Sigma DF(\vartheta)$, $\Sigma = 2E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$, *where* $\tilde{w}(Z_i) = E_{\mathbf{P}}\left(w(Z_i, Z_j)|Z_i\right) - \vartheta$ *and* $F(a_1, ..., a_4) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}$.

Proof. For the proof see Appendix B. □

Note that this is the first asymptotic result we obtained directly for the (unrestricted) mean impact, and not for a lower bound of it. This means that, under the comparatively strong conditions imposed in this setup, we are able to make statements for the mean impact itself.

In order to be able to give an asymptotic confidence interval for the impact, we need to find a consistent estimate of the variance σ^2. The following lemma shows how this can be achieved.

Lemma 2.35. *Under the setup of this section a consistent estimate for* σ^2 *is given by*

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{\Sigma} DF(\tilde{\iota}),$$

where $\hat{\Sigma} = \hat{\hat{g}} - \tilde{\iota}\tilde{\iota}^T$ *with* $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$, *where* $\tilde{\iota}_1 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i \hat{\delta}(X_i)$, $\tilde{\iota}_2 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i \overline{\hat{\delta}(X)}$, $\tilde{\iota}_3 = \frac{1}{m_n}\sum_{i=1}^{m_n} \hat{\delta}(X_i)^2$ *and* $\tilde{\iota}_4 = \frac{1}{m_n}\sum_{i=1}^{m_n} \hat{\delta}(X_i)$ *and*

$$\hat{\hat{g}}_{u,v} = \binom{m_n}{3}^{-1} \sum_{i<j<l} \frac{1}{24} \sum_{\pi \in S(\{i,j,l\})} \sum_{\substack{\psi \in S(\{\pi(i),\pi(j)\}) \\ \rho \in S(\{\pi(i),\pi(l)\})}} \check{g}_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) \check{g}_v(Z_{\rho(\pi(i))}, Z_{\rho(\pi(l))}).$$

$\check{g}$ *is obtained from* g *by replacing* $E_{\mathbf{P}}(Y|x)$ *with* $\hat{\delta}(x)$.

Proof. For the proof see Appendix B. □

A consequence of this lemma is that a confidence interval for $\iota_X(Y)$ is given by

$$CI = [\hat{\iota}_X^{ks,mod2}(Y) - \frac{\hat{\sigma}}{\sqrt{n}} z_{1-\alpha}, \infty).$$

Since we have an asymptotic result directly for the mean impact $\iota_X(Y)$ we can also give a two-sided asymptotic confidence interval

$$CI_{2-sided} = \left[\hat{\iota}_X^{ks,mod2}(Y) - \frac{\hat{\sigma}}{\sqrt{n}} z_{1-\alpha/2}, \hat{\iota}_X^{ks,mod2}(Y) + \frac{\hat{\sigma}}{\sqrt{n}} z_{1-\alpha/2}\right].$$

One disadvantage of this alternative modification of the kernel smoother based estimation of the mean impact is that we do not have a theoretical justification to use bootstrap methods in the calculation of the confidence intervals for the mean impact. Such a result would need a bootstrap version of Corollary 2.32 which we do not have at hand. Finding such a corollary could be the subject of future research.

3. Partial non-linear impact analysis

So far we focused on finding methods for non-linear impact estimation. It might as well be desirable to find methods which allow us to describe non-linear partial mean impacts, that means non-linear influences of one covariate X_1 on the target variable Y, which go beyond the influence of other covariates $Q_1, ..., Q_l$. In the following we will discuss two approaches to this problem. The first one is a direct generalization of the procedures in Section 1.7 where we fitted functions linear in a set of covariates $X^{(1)}, ..., X^{(k)}$ to the data. The second approach tries to answer the question of a non-linear influence of one covariate which goes beyond the possible influence of other covariates via the application of kernel smoothers.

3.1. Partial non-linear impact based on polynomials and splines

In Section 1.9.1 we derived the theory to quantify the dependence of Y on a set of covariates $X^{(1)}, ..., X^{(k)}$ which goes beyond the possible influence of another set of random variables $Q_1, ..., Q_l$. An application of this framework includes the fitting of polynomials and splines in a single variable X_1 (analogous to Section 2.1). In that case one would simply choose $X^{(1)}, ..., X^{(k)}$ to be the respective basis terms. Furthermore, we could choose $Q_1, ..., Q_l$ to be polynomial or spline basis terms in other covariates, allowing us to characterize the non-linear influence of X_1 on Y while correcting for non-linear influences of other variables.

The scenarios outlined here can also be applied to the partial common absolute mean slope.

3.2. Partial non-linear impact based on kernel smoothers

We will derive the theory for a partial mean impact based on kernel smoothers for the scenario where we consider the influence of a single covariate X_1 on Y which goes beyond the influence of other covariates $X_2, ..., X_k$. An extension to non-linear influence of more than one variable is straight forward. One simply replaces the one-dimensional kernel-fit in the following by the higher-dimensional fit (as it was already done in Section 2.2.6).

3.2.1. Direct approach via density-changes

In the definition of the partial mean impact (1.1), we focused on perturbations δ of the common density of $X_1, ..., X_k$ which leave the means of the covariates $X_2, ..., X_k$ unchanged. In order to construct an estimate for the partial mean impact which is

based on kernel smoothers we choose the perturbation

$$\hat{\delta}(\mathbf{X}) = \frac{P_{\mathcal{M}^\perp}\hat{m}(\mathbf{X}_1)}{\sqrt{n^{-1}\sum_{j=1}^n \{P_{\mathcal{M}^\perp}\hat{m}(\mathbf{X}_1)\}_j^2}}, \tag{3.1}$$

where

$$\hat{m}(\mathbf{X}_1)_i = \frac{1}{n}\sum_{j=1}^n K_h(X_{i1} - X_{j1})Y_j,$$

$\mathbf{X}_j = (X_{1j}, ..., X_{nj})$ are the observations of X_j and $\mathcal{M} = \text{span}(\mathbf{1}, \mathbf{X}_2, ..., \mathbf{X}_k)$. This leads to the following estimate for the partial kernel-smoother-based impact

$$\begin{aligned}
\hat{\iota}^{ks}_{X_1}(Y|X_2, ..., X_k) =& \frac{1}{n}\mathbf{Y}^T\hat{\delta}(\mathbf{X}) \\
=& \frac{\frac{1}{n}\mathbf{Y}^T\hat{m}(\mathbf{X}_1) - \frac{1}{n}\mathbf{Y}^T P_{\mathcal{M}}\hat{m}(\mathbf{X}_1)}{\sqrt{n^{-1}\|\hat{m}(\mathbf{X}_1)\|_2^2 - n^{-1}\|P_{\mathcal{M}}\hat{m}(\mathbf{X}_1)\|_2^2}} \\
=& \frac{\frac{1}{n}\mathbf{Y}^T\hat{m}(\mathbf{X}_1) - \frac{1}{n}\mathbf{Y}^T\tilde{m}/d}{\sqrt{n^{-1}\|\hat{m}(\mathbf{X}_1)\|_2^2 - n^{-1}\|\tilde{m}\|_2^2/d^2}},
\end{aligned}$$

where

$$\tilde{m} = \check{\mathbf{X}}\,\text{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)\frac{1}{n}\check{\mathbf{X}}^T\hat{m}(\mathbf{X}_1),$$

$d = \det(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)$ and $\check{\mathbf{X}} = (\check{\mathbf{X}}_1, ..., \check{\mathbf{X}}_k)$ with $\check{\mathbf{X}}_1 = \mathbf{1}$ and $\check{\mathbf{X}}_j = \mathbf{X}_j$ for j=2,...,k. With the notation

$$s(\mathbf{X}) = \frac{1}{n}\hat{m}(\mathbf{X}_1)^T\check{\mathbf{X}}\,\text{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)\frac{1}{n}\check{\mathbf{X}}^T\hat{m}(\mathbf{X}_1)$$

(note that $n^{-1}\|\tilde{m}\|_2^2/d^2 = s(\mathbf{X})/d$) we aim to show that

$$\tilde{\iota} = \left(\frac{1}{n}\mathbf{Y}^T\hat{m}(\mathbf{X}_1), \frac{1}{n}\mathbf{Y}^T\tilde{m}(\mathbf{X}_1), d, \frac{1}{n}\|\hat{m}(\mathbf{X}_1)\|_2^2, s(\mathbf{X})\right)^T$$

is essentially a five-dimensional U-statistics. With the same argumentation as in Section 2.2.1, where we considered the ordinary impact based on kernel smoothers, we can then deduce asymptotic normality of our estimate. Furthermore, the bootstrap can then be shown to be consistent.

As a first step we investigate the expression $\text{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)$. We have that

$$(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)_{ab} = \frac{1}{n}\sum_{i=1}^n \check{X}_{ia}\check{X}_{ib},$$

where $\check{X}_{ib}$ is the i-th element of $\check{\mathbf{X}}_b$ (analogous for a). Since the entries of $\text{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)$

are signed determinants of $(k-1)\times(k-1)$-sub-matrices of $\check{\mathbf{X}}^T\check{\mathbf{X}}/n$ they can be written as

$$\frac{1}{n^{k-1}}\sum_{j_1=1}^{n}...\sum_{j_{k-1}=1}^{n} g(Z_{j_1},...,Z_{j_{k-1}}), \tag{3.2}$$

where $Z_i=(Y_i,X_{i1},...,X_{ik})^T$ and $g(Z_{j_1},...,Z_{j_{k-1}})$ is a sum respectively a difference of products and quotients of terms of the form $\check{X}_{j_o a}\check{X}_{j_p b}$, $o,p=1,...,k-1$ (the specific form of g depends on which element of $\mathrm{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)$ is expressed). Hence

$$\mathrm{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)=\left\{\frac{1}{n^{k-1}}\sum_{j_1=1}^{n}...\sum_{j_{k-1}=1}^{n} g_{lm}(Z_{j_1},...,Z_{j_{k-1}})\right\}_{l,m=1,...,k},$$

where g_{lm} is that function g from (3.2) which leads to the l,m-th entry of $\mathrm{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)$. Furthermore, we have

$$\frac{1}{n}\mathbf{Y}^T\check{\mathbf{X}}=\left\{\frac{1}{n}\sum_{i=1}^{n}\check{X}_{il}Y_i\right\}^T_{l=1,...,k}$$

and

$$\frac{1}{n}\check{\mathbf{X}}^T\hat{m}(\mathbf{X}_1)=\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(X_{i1}-X_{j1})Y_j\check{X}_{il}\right\}_{l=1,...,k}.$$

Defining $f_l(Z_{j_k},Z_{j_{k+1}})=K_h(X_{j_k1}-X_{j_{k+1}1})Y_{j_{k+1}}\check{X}_{j_k l}$ we can write

$$\begin{aligned}
\tilde{\iota}_2=&\frac{1}{n}\mathbf{Y}^T\check{\mathbf{X}}\,\mathrm{cof}(\check{\mathbf{X}}^T\check{\mathbf{X}}/n)\frac{1}{n}\check{\mathbf{X}}^T\tilde{m}\\
=&\sum_{l=1}^{k}\sum_{j=1}^{k}\frac{1}{n^{k+2}}\sum_{j_1=1}^{n}...\sum_{j_{k+2}=1}^{n} g_{lm}(Z_{j_1},...,Z_{j_{k-1}})f_l(Z_{j_k},Z_{j_{k+1}})\check{X}_{j_{k+2}m}Y_{j_{k+2}}\\
=&\frac{1}{n^{k+2}}\sum_{j_1=1}^{n}...\sum_{j_{k+2}=1}^{n}\underbrace{\sum_{l=1}^{k}\sum_{j=1}^{k} g_{lm}(Z_{j_1},...,Z_{j_{k-1}})f_l(Z_{j_k},Z_{j_{k+1}})\check{X}_{j_{k+2}m}Y_{j_{k+2}}}_{\tilde{v}_2(Z_{j_1},...,Z_{j_{k+2}})}\\
=&\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v_2(Z_{j_1},...,Z_{j_{k+3}}),
\end{aligned}$$

where $v_2(Z_{j_1},...,Z_{j_{k+3}})=\tilde{v}_2(Z_{j_1},...,Z_{j_{k+2}})$. Analogously it can be seen that

$$\tilde{\iota}_5=\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v_5(Z_{j_1},...,Z_{j_{k+3}}),$$

with

$$v_5(Z_{j_1},...,Z_{j_{k+3}}) = \sum_{l=1}^{l}\sum_{m=1}^{k} g_{lm}(Z_{j_1},...,Z_{j_{k-1}})f_l(Z_{j_k},Z_{j_{k+1}})f_m(Z_{j_{k+2}},Z_{j_{k+3}}).$$

For $\tilde{\iota}_3$ we obtain

$$\tilde{\iota}_3 = \frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v_3(Z_{j_1},...,Z_{j_{k+3}}),$$

where $v_3(Z_{j_1},...,Z_{j_{k+3}})$ contains sums and products of terms of the form $\check{X}_{ia}\check{X}_{ib}$, $a,b = 1,...,k$, $i = 1,...,n$. Furthermore, we have

$$\tilde{\iota}_1 = \frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v_1(Z_{j_1},...,Z_{j_{k+3}})$$

where $v_1(Z_{j_1},...,Z_{j_{k+3}}) = K_h(X_{j_1 1} - X_{j_2 1})Y_{j_1}Y_{j_2}$ and

$$\tilde{\iota}_4 = \frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v_4(Z_{j_1},...,Z_{j_{k+3}})$$

with $v_4(Z_{j_1},...,Z_{j_{k+3}}) = K_h(X_{j_1 1}-X_{j_2 1})^2 Y_{j_1}^2$. Thereby, with the definition $v = (v_1,...,v_5)^T$, we obtain

$$\begin{aligned}\tilde{\iota} =& \frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} v(Z_{j_1},...,Z_{j_{k+3}})\\ =& \frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n} w(Z_{j_1},...,Z_{j_{k+3}}),\end{aligned}$$

where

$$w(Z_{j_1},...,Z_{j_{k+3}}) = \frac{1}{(k+1)!}\sum_{\pi\in S(\{1,...,k+3\})} v(Z_{j_{\pi(1)}},...,Z_{j_{\pi(k+3)}}).$$

Assumption 3.1. *Assume that*

$$E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{k+3}})) \quad \textit{and} \quad E_{\mathbf{P}}(w^2(Z_{j_1},...,Z_{j_{k+3}}))$$

exist for all $(j_1,...,j_{k+3}) \in \{1,...,n\}^{k+3}$ *and define* $\vartheta = E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{k+3}}))$ *for* $j_1 < ... < j_{k+3}$.

With Assumption 3.1 we obtain that $\tilde{\iota} = U_n + o_p(n^{-1/2})$, where U_n is a five-dimensional $k+3$rd order U-statistics. Hence, we can use the argumentation of Section 2.2.1 to infer, that

$$\sqrt{n}(\tilde{\iota} - \vartheta) \xrightarrow{\mathcal{L}} N_5(0, V),$$

with $V = (k+3)! E_{\mathbf{P}}\left(\tilde{h}(Z_{j_1})\tilde{h}^T(Z_{j_1})\right)$, where $\tilde{h}(Z_{j_1}) = E_{\mathbf{P}}\left(w(Z_{j_1}, ..., Z_{j_{k+3}})|Z_{j_1}\right)$. Thus, application of the delta-method yields

$$\sqrt{n}(\hat{\iota}^{ks}_{X_1}(Y|X_2, ..., X_k) - \iota^{ks}_{X_1}(Y|X_2, ..., X_k)) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\iota^{ks}_{X_1}(Y|X_2, ..., X_k) = F(\vartheta)$, with $F(a_1, ..., a_5) = \frac{a_1 - a_2/a_3}{\sqrt{a_4 - a_5/a_3}}$ and the variance is given by $\sigma^2 = DF(\vartheta)VDF(\vartheta)^T$. Furthermore, it follows that the bootstrap is consistent in this scenario.

In the above considerations we built the estimator for the partial mean impact based on a kernel fit of Y in X_1. We could also use all covariates $X_1, ..., X_k$ in the kernel fitting, which would result in a different $\hat{m}(\mathbf{X})$ in (3.1), namely

$$\hat{m}(\mathbf{X})_i = \mathbf{X}_i \operatorname{cof}(\mathbf{X}^T\mathbf{W}(\mathbf{X}_i)\mathbf{X})\mathbf{X}^T\mathbf{W}(\mathbf{X}_i)\mathbf{Y},$$

where $\mathbf{W}(\mathbf{X}_i) = \operatorname{diag}\left(K_h(\|\mathbf{X}_1 - \mathbf{X}_i\|), ..., K_h(\|\mathbf{X}_n - \mathbf{X}_i\|)\right)$. With this choice of $\hat{m}$ we would also obtain that our estimate for the kernel smoother-based partial mean impact is essentially a U-statistics (the order of the U-statistics increases). Concluding, we also obtain asymptotic normality of the estimate and consistency of the bootstrap.

3.2.2. An alternative approach

Similar to Section 1.9.4 we can regard an alternative approach to decide whether there is an influence of the covariate X_1 which goes beyond the possible influence of other covariates $X_2, ..., X_k$. We do no longer regard changes of the density of the covariates which leave the means of $X_2, ..., X_k$ unchanged. Instead of this we estimate the impacts based on kernel smoothing in $X_1, ..., X_k$ and in $X_2, ..., X_k$ and say that X_1 has an influence on Y which goes beyond that of $X_2, ..., X_k$, if the estimated impact based on kernel smoothing in $X_1, ..., X_k$ is significantly larger than the that of $X_2, ..., X_k$. Remember that we estimate the impact of $\mathbf{X} = (X_1, ..., X_k)$ on Y based on kernel smoothing by

$$\hat{\iota}^{ks}_{\mathbf{X}}(Y) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{\delta}_1(\mathbf{X}_i) - \bar{\delta}_1(\mathbf{X})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\delta}_1(\mathbf{X}_j) - \bar{\delta}_1(\mathbf{X})\right)^2}},$$

where $\mathbf{X}_i = (X_{i1}, ..., X_{ik})$ and

$$\hat{\delta}_1(\mathbf{X}_i) = \frac{1}{n}\sum_{j=1}^{n} K_h(\mathbf{X}_i - \mathbf{X}_j)Y_j,$$

for a kernel

$$K_h(\mathbf{X}_i - \mathbf{X}_j) = D\left(\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{h}\right)$$

and D from Section A.1.1. Let the estimator $\hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y)$ of the impact of $\tilde{\mathbf{X}} = (X_2, ..., X_k)$ on Y be analogously defined as $\hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y)$ but with

$$\hat{\delta}_2(\tilde{\mathbf{X}}_i) = \frac{1}{n}\sum_{j=1}^{n} \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)Y_j \quad \text{and} \quad \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j) = D\left(\frac{\|\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j\|}{h}\right),$$

instead of $\hat{\delta}_1$ and K_h. As described above we now regard the difference between these two estimated impacts $\hat{\iota}^{ks}_{\mathbf{X}}(Y) - \hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y)$.

Assumption 3.2. *Let*

$$g_1(Z_i, Z_j, Z_l) = K_h(\mathbf{X}_i - \mathbf{X}_j)Y_iY_j, \quad g_2(Z_i, Z_j, Z_l) = K_h(\mathbf{X}_j - \mathbf{X}_l)Y_iY_l,$$

$$g_3(Z_i, Z_j, Z_l) = K_h(\mathbf{X}_i - \mathbf{X}_j)Y_iK_h(\mathbf{X}_i - \mathbf{X}_l)Y_l, \quad g_4(Z_i, Z_j, Z_l) = K_h(\mathbf{X}_i - \mathbf{X}_j)Y_j,$$

$$g_5(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)Y_iY_j, \quad g_6(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_j - \tilde{\mathbf{X}}_l)Y_iY_l,$$

and

$$g_7(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)\tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_l)Y_l, \quad g_8(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)Y_j.$$

We use the notation $g = (g_1, ..., g_8)$ and define w analogously to (2.3)*. Furthermore, we let $\vartheta = E(w(Z_i, Z_j, Z_l))$, for $i \neq j \neq l \neq i$.*

Theorem 3.3. *Under Assumption 3.2 we can conclude that*

$$\hat{\iota}^{ks,alt}_{X_1}(Y|X_2, ..., X_k) - \iota^{ks,alt}_{X_1}(Y|X_2, ..., X_k) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\hat{\iota}^{ks,alt}_{X_1}(Y|X_2, ..., X_k) = \hat{\iota}^{ks}_{\mathbf{X}}(Y) - \hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y)$ and $\iota^{ks,alt}_{X_1}(Y|X_2, ..., X_k) = \iota^{ks}_{\mathbf{X}}(Y) - \iota^{ks}_{\tilde{\mathbf{X}}}(Y) = F(\vartheta)$, $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$, $F((a_1, ..., a_8)^T) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}} - \frac{a_5 - a_6}{\sqrt{a_7 - a_8^2}}$ and

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$.

Proof. The key to determining the asymptotic distribution of this difference is once again the theory of U-statistics. We can rewrite the difference as

$$\hat{\iota}^{ks}_{\mathbf{X}}(Y) - \hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y) = \frac{\tilde{\iota}_1 - \tilde{\iota}_2}{\sqrt{\tilde{\iota}_3 - \tilde{\iota}_4^2}} - \frac{\tilde{\iota}_5 - \tilde{\iota}_6}{\sqrt{\tilde{\iota}_7 - \tilde{\iota}_8^2}},$$

with

$$\tilde{\iota}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i \hat{\delta}_1(\mathbf{X}_i), \quad \tilde{\iota}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i \bar{\delta}_1(\mathbf{X}), \quad \tilde{\iota}_3 = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_1(\mathbf{X}_i)^2, \quad \tilde{\iota}_4 = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_1(\mathbf{X}_i),$$

$$\tilde{\iota}_5 = \frac{1}{n}\sum_{i=1}^{n} Y_i \hat{\delta}_2(\tilde{\mathbf{X}}_i) \quad \tilde{\iota}_6 = \frac{1}{n}\sum_{i=1}^{n} Y_i \bar{\delta}_2(\tilde{\mathbf{X}}), \quad \tilde{\iota}_7 = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_2(\tilde{\mathbf{X}}_i)^2 \quad \tilde{\iota}_8 = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_2(\tilde{\mathbf{X}}_i).$$

We obtain in analogy to the proof of Theorem 2.3

$$\tilde{\iota} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} w(Z_i, Z_j, Z_l).$$

With Assumption 3.2 we can follow the argumentation of the proof of Theorem 2.3 to obtain

$$\sqrt{n}(\tilde{\iota} - \vartheta) \xrightarrow{\mathcal{L}} N(0, V)$$

and

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$. Application of the delta method with

$$F((a_1, ..., a_8)^T) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}} - \frac{a_5 - a_6}{\sqrt{a_7 - a_8^2}}$$

yields

$$\hat{\iota}^{ks,alt}_{X_1}(Y|X_2, ..., X_k) - \iota^{ks,alt}_{X_1}(Y|X_2, ..., X_k) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\hat{\iota}^{ks,alt}_{X_1}(Y|X_2, ..., X_k) = \hat{\iota}^{ks}_{\mathbf{X}}(Y) - \hat{\iota}^{ks}_{\tilde{\mathbf{X}}}(Y)$ and $\iota^{ks,alt}_{X_1}(Y|X_2, ..., X_k) = \iota^{ks}_{\mathbf{X}}(Y) - \iota^{ks}_{\tilde{\mathbf{X}}}(Y) = F(\vartheta)$ as well as $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$. □

By replacing F and $\tilde{\iota}$ in Section 2.2.1 by our new F and $\tilde{\iota}$ we can also apply the results of Section 2.2.1 concerning estimation of σ^2. The same considerations as in Section

2.2.1 show that the bootstrap is consistent in this case as well which is a justification for computing bootstrap confidence intervals in this case (e.g. in order to save computational time).

In the case of the linear common mean impact orthogonality provided (1.33), which means

$$\iota^{lin^2}_{X_1,...,X_k}(Y) - \iota^{lin^2}_{X_2,...,X_k}(Y) = \iota^{lin^2}_{X_1}(Y|X_2,...,X_k).$$

However, due to the lack of orthogonality, in the case of kernel smoothing such a decomposition does not hold in general.

3.2.3. Partial mean slope based on kernel smoothing

Direct approach

The partial mean slope based on kernel smoothing is defined as

$$\theta^{ks}_{X_1}(Y|X_2,...,X_k) = \frac{\iota^{ks}_{X_1}(Y|X_2,...,X_k)}{\iota_{X_1}(X_1|X_2,...,X_k)} = \frac{\iota^{ks}_{X_1}(Y|X_2,...,X_k)}{\sqrt{Var_{\mathbf{P}}(P_{\mathcal{H}_2^\perp}X_1)}},$$

where $\mathcal{H}_2 = \text{span}(1, X_2, ..., X_k)$. It is estimated by

$$\hat{\theta}^{ks}_{X_1}(Y|X_2,...,X_k) = \frac{\hat{\iota}^{ks}_{X_1}(Y|X_2,...,X_k)}{\sqrt{\frac{1}{n}\sum_{i=1}^n \left((P_{\mathcal{H}_2^\perp}\mathbf{X}_1)_i - \overline{P_{\mathcal{H}_2^\perp}\mathbf{X}_1}\right)^2}},$$

where $\mathcal{M}_2 = \text{span}(\mathbf{1}, \mathbf{X}_2, ..., \mathbf{X}_k)$. We use the notation from Section 3.2.1 and define

$$\tilde{\iota}_6 = \frac{1}{n}\sum_{i=1}^n X_{i1}^2 = \frac{1}{n^{k+3}}\sum_{j_1=1}^n ... \sum_{j_{k+3}=1}^n v_6(Z_{j_1},...,Z_{j_{k+3}}),$$

where $v_6(Z_{j_1},...,Z_{j_{k+3}}) = X_{i1}^2$ and

$$\tilde{\iota}_7 = \frac{d}{n}P_{\mathcal{M}}\mathbf{X}_1^T P_{\mathcal{M}}\mathbf{X}_1 = \frac{1}{n}\mathbf{X}_1^T\mathbf{X}\,\text{cof}(\mathbf{X}^T\mathbf{X}/n)\frac{1}{n}\mathbf{X}^T\mathbf{X}_1.$$

Analogous to the discussion of $\tilde{\iota}_2$ one can see that we have

$$\tilde{\iota}_7 = \frac{1}{n^{k+2}}\sum_{j_1=1}^n ... \sum_{j_{k+1}=1}^n \sum_{l=1}^k \sum_{j=1}^k g_{lm}(Z_{j_1},...,Z_{j_{k-1}})\tilde{X}_{j_k l}X_{j_k 1}\tilde{X}_{j_{k+1}m}X_{j_{k+1}1}$$

$$=\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n}v_7(Z_{j_1},...,Z_{j_{k+3}}),$$

where $v_7(Z_{j_1},...,Z_{j_{k+3}})=\sum_{l=1}^{k}\sum_{j=1}^{k}g_{lm}(Z_{j_1},...,Z_{j_{k-1}})\tilde{X}_{j_kl}X_{j_k1}\tilde{X}_{j_{k+1}m}X_{j_{k+1}1}$. Let $v=(v_1,...,v_7)$ and as before

$$w(Z_{j_1},...,Z_{j_{k+3}})=\frac{1}{(k+1)!}\sum_{\pi\in S(\{1,...,k+3\})}v(Z_{j_\pi(1)},...,Z_{j_\pi(k+3)}).$$

Then we obtain for $\tilde{\iota}=(\tilde{\iota}_1,...,\tilde{\iota}_7)^T$

$$\tilde{\iota}=\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n}v(Z_{j_1},...,Z_{j_{k+3}})=\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}...\sum_{j_{k+3}=1}^{n}w(Z_{j_1},...,Z_{j_{k+3}}),$$

where $w(Z_{j_1},...,Z_{j_{k+3}})=(k+3)!^{-1}\sum_{\pi\in S(\{j_1,...,j_{k+3}\})}v(Z_{j_1},...,Z_{j_{k+3}})$. Assume that the following holds:

Assumption 3.4. *Assume that*

$$\mathbf{P}(w(Z_{j_1},...,Z_{j_{k+3}}))\quad\textit{and}\quad E_{\mathbf{P}}(w^2(Z_{j_1},...,Z_{j_{k+3}}))$$

exist for all $(j_1,...,j_{k+3})\in\{1,...,n\}^{k+3}$ *and define* $\vartheta=E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{k+3}}))$ *for* $j_1<...<j_{k+3}$.

It follows from Lemma 2.1 and the theory of U-statistics that

$$\sqrt{n}\,(\tilde{\iota}-\vartheta)\xrightarrow{\mathcal{L}}N_7(\mathbf{0},\Sigma),$$

with $V=(k+3)!E_{\mathbf{P}}\left(\tilde{h}(Z_{j_1})\tilde{h}^T(Z_{j_1})\right)$, where $\tilde{w}(Z_{j_1})=E_{\mathbf{P}}\left(w(Z_{j_1},...,Z_{j_{k+3}})|Z_{j_1}\right)-\vartheta$ and $\vartheta=E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_{k+3}}))$. Thus, application of the delta-method with $F(a_1,...,a_7)=\frac{(a_1-a_2/a_3)}{\sqrt{a_4-a_5/a_3}\sqrt{a_6-a_7/a_3}}$ yields

$$\sqrt{n}\left(\hat{\theta}^{ks}_{X_1}(Y|X_2,...,X_k)-\theta^{ks}_{X_1}(Y|X_2,...,X_k)\right)\xrightarrow{\mathcal{L}}N(0,\sigma^2),$$

where $\theta^{ks}_{X_1}(Y|X_2,...,X_k)=F(\vartheta)$ and

$$\sigma^2=DF(\vartheta)^T\Sigma DF(\vartheta).$$

By the same argumentation as in Section 2.2.1 we can find a consistent estimate $\hat{\sigma}^2$ of

σ^2. Application of Lemma 2.5 together with the results of Bickel and Freedman (1981) gives consistency of the bootstrap in this case.

Alternative approach

We now adopt the notation of Section 3.2.2 where we considered the alternative approach for the partial mean impact. The extension to the partial mean slope can be done by regarding the difference

$$\theta_{X_1}^{ks,alt}(Y|X_2,...,X_k) = \frac{\iota_{\mathbf{X}}^{ks}(Y) - \iota_{\tilde{\mathbf{X}}}^{ks}(Y)}{\iota_{\mathbf{X}}(X_1) - \iota_{\tilde{\mathbf{X}}}^{ks}(X_1)}$$

which can be estimated by

$$\hat{\theta}_{X_1}^{ks,alt}(Y|X_2,...,X_k) = \frac{\hat{\iota}_{\mathbf{X}}^{ks}(Y) - \hat{\iota}_{\tilde{\mathbf{X}}}^{ks}(Y)}{\sqrt{\frac{1}{n}\sum_{i=1}^n \left(X_{i1} - \bar{\mathbf{X}}_1\right)^2} - \hat{\iota}_{\tilde{\mathbf{X}}}^{ks}(X_1)}.$$

We define $g_9(Z_i, Z_j, Z_l) = \frac{1}{n}(X_{i1} - X_{j1})^2$ as well as

$$g_{10}(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)X_{i1}X_{j1}, \quad g_{11}(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_j - \tilde{\mathbf{X}}_l)X_{i1}X_{l1},$$

and

$$g_{12}(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)\tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_l)X_{l1}, \quad g_{13}(Z_i, Z_j, Z_l) = \tilde{K}_h(\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j)X_{j1}.$$

Furthermore, let

$$\tilde{\iota} := \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n w(Z_i, Z_j, Z_l),$$

where still $w(Z_i, Z_j, Z_l) = \frac{1}{6}\{g(Z_i, Z_j, Z_l) + g(Z_i, Z_l, Z_j) + g(Z_j, Z_i, Z_l) + g(Z_j, Z_l, Z_i)$ $+g(Z_l, Z_i, Z_j) + g(Z_l, Z_j, Z_i)\}$, with $g = (g_1, ..., g_{13})$. Assume, that Assumption 3.2 still holds for this w. With this assumption we can follow the argumentation of Section 2.2.1 to obtain

$$\sqrt{n}(\tilde{\iota} - \vartheta) \xrightarrow{\mathcal{L}} N(0, V)$$

where ϑ is defined as in Assumption 3.2 but with the new w derived in this section and

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$. Application of the delta method with

$$F((a_1, ..., a_{13})^T) = \left(\frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}} - \frac{a_5 - a_6}{\sqrt{a_7 - a_8^2}} \right) / \left(a_9^{1/2} - \frac{a_{10} - a_{11}}{\sqrt{a_{12} - a_{13}^2}} \right)$$

yields

$$\hat{\theta}_{X_1}^{ks,alt}(Y|X_2, ..., X_k) - \theta_{X_1}^{ks,alt}(Y|X_2, ..., X_k) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\hat{\theta}_{X_1}^{ks,alt}(Y|X_2, ..., X_k) = \hat{\theta}_{\mathbf{X}}^{ks}(Y) - \hat{\theta}_{\tilde{\mathbf{X}}}^{ks}(Y)$ and $\theta_{X_1}^{ks,alt}(Y|X_2, ..., X_k) = \theta_{\mathbf{X}}^{ks}(Y) - \theta_{\tilde{\mathbf{X}}}^{ks}(Y) = F(\vartheta)$ as well as $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$. By replacing the function F in Section 2.2.1 by our new F we can also apply the results of that section concerning estimation of σ^2 (including Lemma 2.7). The same considerations as in Section 2.2.1 show that the bootstrap is consistent in this case as well which is a justification for computing bootstrap confidence intervals in this case (e.g. in order to save computational time).

3.2.4. Partial population coefficient for determination based on kernel smoothing

Similar to Section 1.9.4 we can define a partial population coefficient for determination based on kernel smoothing. We consider the expression

$$R_{X_1}^{ks^2}(Y|X_2, ..., X_k) = \frac{R_{\mathbf{X}}^{ks^2}(Y) - R_{\tilde{\mathbf{X}}}^{ks^2}(Y)}{1 - R_{\tilde{\mathbf{X}}}^{ks^2}(Y)}, \tag{3.3}$$

where $R_{\mathbf{X}}^{ks^2}(Y) = \iota_{\mathbf{X}}^{ks^2}(Y)/Var_{\mathbf{P}}(Y)$ and $R_{\tilde{\mathbf{X}}}^{ks^2}(Y) = \iota_{\tilde{\mathbf{X}}}^{ks^2}(Y)/Var_{\mathbf{P}}(Y)$. Thus, we can rewrite (3.3) as

$$R_{X_1}^{ks^2}(Y|X_2, ..., X_k) = \frac{\iota_{\mathbf{X}}^{ks^2}(Y) - \iota_{\tilde{\mathbf{X}}}^{ks^2}(Y)}{Var_{\mathbf{P}}(Y) - \iota_{\tilde{\mathbf{X}}}^{ks^2}(Y)}.$$

This can be estimated by

$$\hat{R}_{X_1}^{ks^2}(Y|X_2, ..., X_k) = \frac{\hat{\iota}_{\mathbf{X}}^{ks^2}(Y) - \hat{\iota}_{\tilde{\mathbf{X}}}^{ks^2}(Y)}{\widehat{Var}_{\mathbf{P}}(Y) - \hat{\iota}_{\tilde{\mathbf{X}}}^{ks^2}(Y)}, \tag{3.4}$$

where $\widehat{Var}_{\mathbf{P}}(Y) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j)^2$. As a next step we will derive an asymptotic normality result for (3.4). To this end we use the notation of Section 3.2.2, and define $g_9(Z_i, Z_j, Z_l) = (Y_i - Y_j)^2$ and $\tilde{\iota}_9 = n^{-3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n g_9(Z_i, Z_j, Z_l)$. Hence, we obtain

$$\tilde{\iota} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n w(Z_i, Z_j, Z_l),$$

where $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_9)^T$ and still $w(Z_i, Z_j, Z_l) = \frac{1}{6}\{g(Z_i, Z_j, Z_l)+g(Z_i, Z_l, Z_j)+g(Z_j, Z_i, Z_l)+ g(Z_j, Z_l, Z_i) + g(Z_l, Z_i, Z_j) + g(Z_l, Z_j, Z_i)\}$. Assume, that Assumption 3.2 still holds for this w. With this assumption we can follow the argumentation of Section 2.2.1 to obtain

$$\sqrt{n}(\tilde{\iota} - \vartheta) \xrightarrow{\mathcal{L}} N(0, V)$$

with $\vartheta = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l))$ and

$$V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

where $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l)|Z_i) - \vartheta$. Application of the delta method with

$$F((a_1, ..., a_9)^T) = \frac{\left(\frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}\right)^2 - \left(\frac{a_5 - a_6}{\sqrt{a_7 - a_8^2}}\right)^2}{a_9 - \left(\frac{a_5 - a_6}{\sqrt{a_7 - a_8^2}}\right)^2}$$

yields

$$\sqrt{n}\left(\hat{R}^{ks^2}_{X_1}(Y|X_2, ..., X_k) - R^{ks^2}_{X_1}(Y|X_2, ..., X_k)\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$. The considerations of Section 2.2.1 regarding an estimate for the variance σ^2 do also apply in this section. Hence, we can compute a consistent estimator $\hat{\sigma}^2$ for σ^2. Using this estimator we can compute an asymptotic level α confidence interval for $R^{ks^2}_{X_1}(Y|X_2, ..., X_k)$ by

$$CI = \left[\hat{R}^{ks^2}_{X_1}(Y|X_2, ..., X_k) - \hat{\sigma}/\sqrt{n}z_{1-\alpha}, \infty\right).$$

Application of Lemma 2.5 together with the results of Bickel and Freedman (1981) gives consistency of the bootstrap in this case as well.

Doksum and Samarov (1995) define a "measure of relative importance" which is very similar to our partial population coefficient for determination based on kernel smoothers. However, they use a different estimator than we do. Translating their estimator into our framework would lead to the estimator

$$\widehat{Corr}^2\left(\hat{\delta}_i(\mathbf{X}) - \hat{\delta}_i(\tilde{\mathbf{X}}), Y - \hat{\delta}_i(\tilde{\mathbf{X}})\right)$$

where $\widehat{Corr}$ is the empirical correlation coefficient and $\hat{\delta}_i$ is the leave-one-out version of $\hat{\delta}$ already introduced in Section 2.2.1. A comparison between this estimator and ours could be the subject of future research and is not done in this thesis.

4. Simulations - Comparison of methods

In this section we want to investigate the performance of the different methods derived in the previous sections in a simulation study. We aim to compare the different methods with respect to the coverage probability of the confidence intervals as well as the power in cases where the respective impact is greater than zero. We will also compare the newly derived methods to the linear mean impact of Scharpenberg (2012) and Brannath and Scharpenberg (2014). For the results in this section we computed 1,000 repetitions with n=100 observations. The comparably small number of repetitions and observations is due to the processor-intensive nature of the bootstrap methods applied. A more thorough investigation with different sample sizes and more repetitions should be the subject of future research. All confidence intervals are computed at a significance level of $\alpha = 0.05$. Bootstrap intervals are based on 1,000 bootstrap repetitions. All simulations where run using the statistical software `R`.

4.1. Single Covariate Case

In the single covariate case we investigate the performance of the methods derived in the previous sections on six models. The first model is a linear one, where $Y = 0.3X + \epsilon$, where $\epsilon \sim N(0,1)$ is independent from $X \sim N(0,1)$. In the second model we have $Y = \sin((X+1)3\pi/2) + \epsilon$. Here we have again that $\epsilon \sim N(0,1)$ is independent from X, but $X \sim U(-1,1)$. Model 3 is given by $Y = \sin(5X) + \epsilon$ where X and ϵ are independent and both follow a standard normal distribution. In the fourth model we have $X \sim U(0,1)$ independent from ϵ and $Y = \sin(12(X+0.2))/(X+0.2) + \epsilon$. The fifth model states that Y and X are independent and both follow a standard normal distribution. The sixth model is a heteroscedastic one where we chose the impact to equal zero. We assume that $X \sim N(0,1)$ and $Y \sim N(0, x^2/2)$. Figure 2 gives a graphical overview of the different models, and how the different regression techniques fit data, that arise from this model. Furthermore, the value of the mean impact in each scenario is given in Table 9.

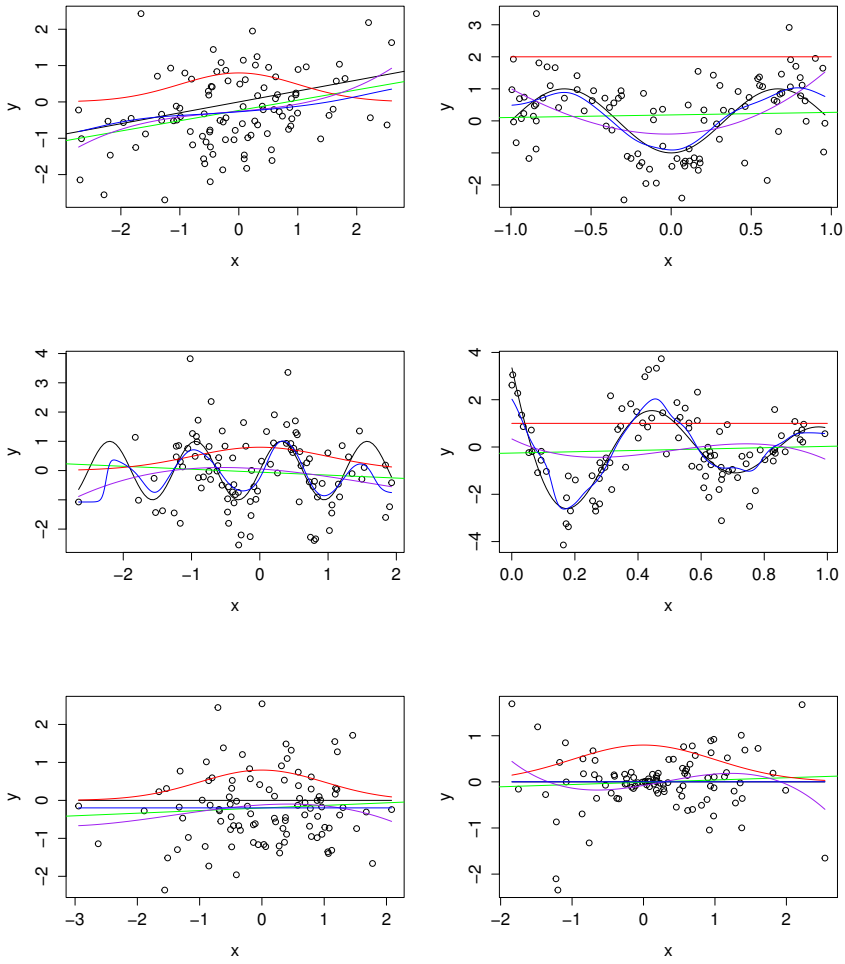

Figure 2: Display of the models of the simulation study. Upper left: Model 1, upper right: Model 2, mid left: Model 3, mid right: Model 4, lower left panel: Model 5, lower right panel: Model 6. In each case a randomly generated data set according is plotted. The black curves show the underlying true relationship between X and Y. The red curves give the density of X. Three different regression fits are also given: Linear model (green), Polynomial of degree 3 (purple) and Kernel Smoother (blue).

Model	$\iota_X(Y)$
$Y = 0.3X + \epsilon$	0.3
$Y = \sin((X + 1)3\pi/2) + \epsilon$	0.675
$Y = \sin(5X) + \epsilon$	0.707
$Y = \sin(12(X + 0.2))/(X + 0.2) + \epsilon$	1.270
$Y = \epsilon$	0
Heteroscedasticity	0

Table 9: Value of mean impact in the different simulation models

In Tables 10, 11 and 12 we can see an overview of the estimated mean impacts, their bias as well as their variance and MSE for the methods based on linear regression, polynomial regression and kernel smoothing. We can see that the kernel smoother based impact tends to have smaller bias than the other two methods in the scenarios where the mean impact is not equal to zero. However, in the scenarios where it equals zero the bias is clearly greater than with the other two methods. We chose a data dependent bandwidth for the kernel smoother based mean impact because it will turn out in Section 4.1.3 that doing so yields better results than using a fixed bandwidth. Whenever we will speak of bootstrap intervals using "transformations" in the course of this section we mean that the bootstrap confidence bounds are computed based on an estimate for the squared mean impact (allowing us to make use of the smooth function model which yields second order accurate bootstrap intervals) and then transform these bounds (just by taking the square root whenever they are positve) to obtain confidence bounds on the (unsquared) impact scale as described in the derivation of the theory.

Model	$\iota_X(Y)$	$\hat{\iota}_X^{lin}(Y)$	Bias	Variance	MSE
$Y = 0.3X + \epsilon$	0.3	0.297	-0.003	0.010	0.010
$Y = \sin((X + 1)3\pi/2) + \epsilon$	0.675	0.089	-0.586	0.005	0.347
$Y = \sin(5X) + \epsilon$	0.707	0.097	-0.610	0.006	0.378
$Y = \sin(12(X + 0.2))/(X + 0.2) + \epsilon$	1.270	0.277	-0.993	0.024	1.010
$Y = \epsilon$	0	0.080	0.080	0.004	0.010
Heteroscedasticity	0	0.093	0.093	0.005	0.014

Table 10: Overview of the estimated linear mean impact. Given are the mean estimate as well as the empirical bias (for the unrestricted mean impact), variance and mse based on 1,000 simulation runs.

Model	$\iota_X(Y)$	$\hat{\iota}_X^{pol}(Y)$	Bias	Variance	MSE
$Y = 0.3X + \epsilon$	0.3	0.331	0.031	0.009	0.010
$Y = \sin((X+1)3\pi/2) + \epsilon$	0.675	0.447	-0.227	0.012	0.063
$Y = \sin(5X) + \epsilon$	0.707	0.197	-0.510	0.007	0.267
$Y = \sin(12(X+0.2))/(X+0.2) + \epsilon$	1.270	0.401	-0.869	0.022	0.777
$Y = \epsilon$	0	0.160	0.160	0.005	0.030
Heteroscedasticity	0	0.216	0.216	0.012	0.058

Table 11: Overview of the estimated polynomial mean impact. Used were cubic polynomials. Given are the mean estimate as well as the empirical bias (for the unrestricted mean impact), variance and mse based on 1,000 simulation runs.

Model	$\iota_X(Y)$	$\hat{\iota}_X^{ks}(Y)$	Bias	Variance	MSE
$Y = 0.3X + \epsilon$	0.3	0.319	0.019	0.008	0.008
$Y = \sin((X+1)3\pi/2) + \epsilon$	0.675	0.695	0.020	0.010	0.011
$Y = \sin(5X) + \epsilon$	0.707	0.481	-0.226	0.015	0.066
$Y = \sin(12(X+0.2))/(X+0.2) + \epsilon$	1.270	1.184	-0.086	0.015	0.022
$Y = \epsilon$	0	0.200	0.200	0.005	0.044
Heteroscedasticity	0	0.207	0.207	0.006	0.049

Table 12: Overview of the estimated kernel smoother mean impact with data dependent bandwidth. Given are the mean estimate as well as the empirical bias (for the unrestricted mean impact), variance and mse based on 1,000 simulation runs.

We will compare many different methods in the sequel. Table 13 gives an overview over all performed simulations and displays which table contains which results.

Type of Impact	Type of Bootstrap	Transformation used	Bandwidth	Pre-performed tests	Table	Page
Linear	Studentized	Yes	NA	No	Table 14	Page 111
Polynomial	Shrinkage*	Yes	NA	Yes	Table 15	Page 112
	Studentized	Yes	NA	No	Table 16	Page 112
	Studentized	No	NA	Yes	Table 17	Page 113
	Basic	Yes	NA	No	Table 18	Page 114
	Basic	No	NA	Yes	Table 19	Page 115
Kernel-Smoother	Basic	No	Fixed	No	Table 20	Page 116
	Basic	No	Optimal	Yes	Table 23	Page 118
	Studentized	No	Optimal	Yes	Table 24	Page 119
	Studentized**	No	Optimal	Yes	Table 25	Page 119
	Basic***	No	Fixed	No	Table 26	Page 120
	Basic***	No	Optimal	Yes	Table 27	Page 120
Local Linear	Basic	No	Fixed	No	Table 21	Page 116
Local Squared	Basic	No	Fixed	No	Table 22	Page 117

Table 13: Overview of simulation scenarios in the single covariate case. *: The shrinkage approach for the calculation of confidence intervals from Section 1.7.2, no bootstrap performed. **: In this scenario the regression functions were normalized such that the true mean impact equals one. ***: Kernel-Smoother based impact estimate without deletion of denominator.

4.1.1. Linear mean impact

First we compute the simulations for the linear mean impact analysis. This allows us to compare the performance of the intervals based on the normal approximation (see (1.15)) to the studentized bootstrap interval (where we consider the "common" impact of X alone). Furthermore, these results will be used for the comparison to the non-linear methods to assess their benefit. One can see from Table 14 that both types of confidence intervals maintain the pre specified level in all cases but the case of heteroscedasticity. Here the confidence interval based on the normal approximation shows slight undercoverage. With both methods one has very low power in all scenarios but the linear one. The normal approximation based confidence interval outperforms the bootstrap interval in terms of power. In the scenario of a linear underlying regression function both methods maintain the confidence level. However the power of the normal approximation interval (0.853) is greater than the one of the bootstrap interval (0.625).

Model	$\iota_X(Y)$	$Cover_{norm}$	$Power_{norm}$	$Cover_{boot}$	$Power_{boot}$
$Y = 0.3X + \epsilon$	0.3	0.974	0.853	0.957	0.625
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.060	1.000	0.008
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.051	1.000	0.009
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.380	1.000	0.150
$Y = \epsilon$	0	0.941	0.059	0.989	0.011
Heteroscedasticity	0	0.939	0.061	0.983	0.017

Table 14: Simulation results for the linear mean impact. Compared are the confidence interval based on the asymptotic normality (norm) result and the bootstrap confidence interval (boot). For each interval the coverage probability for the (unrestricted) mean impact and the probability of exclusion of zero (power) are given.

4.1.2. Polynomial based impact

As a next step we investigate the confidence intervals based on the shrinkage like approach of Section 1.7.2 for polynomial regression. We chose to use polynomials of degree 3. Table 15 shows that the confidence intervals resulting from the shrinkage like approach perform very good in all scenarios where the mean impact is greater than zero. However, the coverage probability of 0.837 in the case where the mean impact is zero is very low and far from tolerable. Hence, other methods for the computation of confidence

Model	$\iota_X(Y)$	*Cover*	*Power*
$Y = 0.3X + \epsilon$	0.3	0.988	0.731
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.820
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.169
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.541
$Y = \epsilon$	0	0.837	0.163
Heteroscedasticity	0	0.693	0.307

Table 15: Simulation results for the shrinkage approach confidence intervals for the cubic polynomial based mean impact. Given are the coverage probability of the interval for the (unrestricted) mean impact and the probability of exclusion of zero (power).

intervals are needed.

Since the confidence intervals based on the shrinkage like approach to not perform very good, we want to investigate the performance of the intervals based on the polynomial mean impact. We now compute studentized bootstrap intervals. We used the functional delta method variance estimate which is further introduced in Section A.3.2. The results

Model	$\iota_X(Y)$	*Coverage*	*Power*
$Y = 0.3X + \epsilon$	0.3	0.949	0.610
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.732
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.066
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.180
$Y = \epsilon$	0	0.936	0.064
Heteroscedasticity	0	0.855	0.145

Table 16: Simulation results for studentized bootstrap intervals for the cubic polynomial based mean impact. Given are the coverage probability of the interval for the (unrestricted) mean impact and the probability of exclusion of zero (power).

of Table 16 indicate, that the coverage probability of the studentized intervals is maintained in scenarios 2-4. In the case of a linear model and when there is no relationship between X and Y the coverage probability lies slightly below 0.95. However, this small deviations from the nominal confidence level is explainable by simulation error.

In Section 1.7.5 it was also mentioned that it is possible to compute bootstrap intervals based directly on the estimate $\hat{\iota}^{lin}_{\mathbf{X}}(Y)$ instead of using its square (as we did in

the calculations for Table 16), when we can preclude that the mean impact is zero. To this end we computed studentized bootstrap intervals, again using the functional delta method estimate of the variance, and pre performed different test for the hypothesis that the polynomial based impact is zero. The first test we used is the test from Section 1.7.1. Since this test is based on the results of White (1980b) we denote the coverage probability and power of the confidence intervals arising from pre performing this test by $Cover_{white}$ and $Power_{white}$. We also investigated whether or not the use of the global F-test ($H_0 : \xi_1 = \xi_2 = \xi_3 = 0$ where $\xi_0, ..., \xi_3$ are the coefficients from the projection of Y onto $\text{span}(1, X, X^2, X^3)$) from the linear regression (without robust variance estimate) as pre-performed test delivers better results (coverage probability and power of this procedure are denoted by $Cover_F$ and $Power_F$ in Table 17). The performance of the procedure where both, the F-test and the test from Section 1.7.1 are performed prior to the calculation of confidence intervals was also investigated. The results of Table 17

Model	$\iota_X(Y)$	$Cover_{no\ test}$	$Power_{no\ test}$	$Cover_{white}$	$Power_{white}$
$Y = 0.3X + \epsilon$	0.3	0.931	0.927	0.931	0.689
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.950	1.000	0.855
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.315	1.000	0.045
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.592	1.000	0.334
$Y = \epsilon$	0	0.688	0.312	0.960	0.040
Heteroscedasticity	0	0.543	0.457	0.592	0.408

Model	$\iota_X(Y)$	$Cover_F$	$Power_F$	$Cover_{both}$	$Power_{both}$
$Y = 0.3X + \epsilon$	0.3	0.931	0.836	0.931	0.685
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.862	1.000	0.841
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.167	1.000	0.043
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.391	1.000	0.320
$Y = \epsilon$	0	0.821	0.179	0.960	0.040
Heteroscedasticity	0	0.671	0.329	0.679	0.321

Table 17: Simulation results for studentized bootstrap intervals for the cubic polynomial based mean impact (not using transformations). Given are the coverage probability of the interval for the (unrestricted) mean impact and the probability of exclusion of zero (power), when pre-performing different test for the impact being zero.

show that the studentized bootstrap intervals based on the polynomial mean impact, not using the transformation performs very bad in the case, where the mean impact is

zero. This is due to the fact, that in this case the smooth function model does not hold, and we do not have any theoretical justification for the use of bootstrap methods in this case. This issue is resolved by the use of the test from Section 1.7.1. However, when using this test, we still have slight undercoverage in the case of a linear relationship between X and Y. The use of the F-test does not give any benefit. The improvement of the coverage probability to 0.821 in the zero impact case is not sufficient. Therefore, the best choice seems to be to use the test from Section 1.7.1 prior to the calculation of the confidence intervals.

When comparing the results of the procedure where we compute the studentized bootstrap intervals via transformation of the estimated polynomial mean impact (Table 16) with the results of the procedure where we did not use the transformation but pre performed a test for the mean impact being zero (Table 17), we can see that the latter procedure yields a higher coverage probability as well as a higher power in most scenarios. In the presence of heteroscedasticity both procedures do not perform very good, although we should mention that the transformation procedure beats the non-transformation procedure in this case. However, this scenario is not covered by the theory derived in this thesis.

The computation of studentized bootstrap confidence intervals can be very time consuming, especially when using the functional delta method variance estimate. This is why it might be preferable to calculate the less cpu-intensive basic bootstrap intervals. We performed the same simulations as for Tables 16 and 17 but with basic bootstrap intervals instead of studentized intervals. We can see from Table 18 that the basic bootstrap

Model	$\iota_X(Y)$	*Coverage*	*Power*
$Y = 0.3X + \epsilon$	0.3	0.988	0.010
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.358
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.001
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.010
$Y = \epsilon$	0	0.997	0.003
Heteroscedasticity	0	0.994	0.006

Table 18: Simulation results for basic bootstrap intervals for the cubic polynomial based mean impact. Given are the coverage probability of the interval for the (unrestricted) mean impact and the probability of exclusion of zero (power).

intervals using the transformation of the estimated mean impact are very conservative

and have close to no power.

Model	$\iota_X(Y)$	$Cover_{no\ test}$	$Power_{no\ test}$	$Cover_{white}$	$Power_{white}$
$Y = 0.3X + \epsilon$	0.3	0.931	0.940	0.931	0.700
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.973	1.000	0.857
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.329	1.000	0.045
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.614	1.000	0.361
$Y = \epsilon$	0	0.554	0.446	0.573	0.427
Heteroscedasticity	0	0.457	0.180	0.200	0.673

Model	$\iota_X(Y)$	$Cover_f$	$Power_f$	$Cover_{both}$	$Power_{both}$
$Y = 0.3X + \epsilon$	0.3	0.931	0.835	0.931	0.685
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.864	1.000	0.842
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.172	1.000	0.043
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.411	1.000	0.338
$Y = \epsilon$	0	0.818	0.182	0.959	0.041
Heteroscedasticity	0	0.656	0.344	0.662	0.338

Table 19: Simulation results for basic bootstrap intervals for the cubic polynomial based mean impact (not using transformations). Given are the coverage probability of the interval for the (unrestricted) mean impact and the probability of exclusion of zero (power), when pre-performing different test for the impact being zero.

Moving to the intervals where we do not use the transformation of the estimated impact (Table 19) increases the power at the cost of resulting severe undercoverage in the case where the mean impact is zero. In this case pre-performing the test from Section 1.7.1 does not resolve the issue. However, using this test and the F-test leads to confidence intervals that maintain the coverage probability and have a power which is comparable to the power of the studentized intervals using no transformation but the test from Section 1.7.1.

4.1.3. Kernel-smoother based impact analysis

In this section we compare the performance of the mean impact analysis based on kernel methods. Since we assumed fixed bandwidths in the derivation of the asymptotic results of the kernel-method based impact analysis we consider the three cases of $h \in \{0.05,\ 0.1,\ 0.5\}$. The computation of the variance estimates derived in Section 2.2 is very time consuming, which is why we compare the methods using basic bootstrap

intervals. We use a normal kernel for the kernel-smoother and higher order regression based mean impact. We can see from Table 20 that the performance of the kernel-

Model	$\iota_X(Y)$	$Cov_{0.05}$	$Pow_{0.05}$	$Cov_{0.1}$	$Pow_{0.1}$	$Cov_{0.5}$	$Pow_{0.5}$
$Y = 0.3X + \epsilon$	0.3	0.227	1.000	0.770	0.998	0.975	0.895
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.893	1.000	0.936	1.000	0.969	1.000
$Y = \sin(5X) + \epsilon$	0.707	0.872	1.000	0.952	1.000	0.998	0.989
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.969	1.000	0.995	1.000	1.000	0.592
$Y = \epsilon$	0	0.528	0.472	0.838	0.162	0.923	0.077
Heteroscedasticity	0	0.000	1.000	0.028	0.972	0.648	0.352

Table 20: Simulation results for basic bootstrap intervals of the kernel smoother based impact for different bandwidths. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power).

smoother based impact analysis is highly dependent on the choice of the bandwidth h. The fact which of the three bandwidths performs best is also depends on the underlying model. Small bandwidths allow for more flexible modeling which is advantageous when the true regression function is very curvy (e.g. in Model 2) while large bandwidths are preferable when the true structure is less curvy (e.g. Model 5). Later in this section we will investigate whether a data dependent choice of the bandwidth (which we have no theoretical justification for) gives good results in simulations. The results for the local

Model	$\iota_X(Y)$	$Cov_{0.05}$	$Pow_{0.05}$	$Cov_{0.1}$	$Pow_{0.1}$	$Cov_{0.5}$	$Pow_{0.5}$
$Y = 0.3X + \epsilon$	0.3	1.000	0.108	1.000	0.209	0.987	0.714
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.097	1.000	0.064	1.000	0.022
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.110	1.000	0.147	1.000	0.118
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.374	1.000	0.406	1.000	0.513
$Y = \epsilon$	0	0.907	0.093	0.922	0.078	0.926	0.074
Heteroscedasticity	0	0.927	0.073	0.929	0.071	0.926	0.074

Table 21: Simulation results for basic bootstrap intervals of the local linear regression based impact for different bandwidths. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power).

linear regression based impact analysis in Table 21 show that in this case too the performance is highly dependent on the choice of the kernel bandwidth. However, we can also see that this method has low power compared to the kernel smoother based impact.

Hence, we will not further investigate the local linear regression based mean impact. In

Model	$\iota_X(Y)$	$Cov_{0.05}$	$Pow_{0.05}$	$Cov_{0.1}$	$Pow_{0.1}$	$Cov_{0.5}$	$Pow_{0.5}$
$Y = 0.3X + \epsilon$	0.3	0.989	0.505	0.989	0.553	0.995	0.598
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	1.000	0.106	1.000	0.104	0.995	0.105
$Y = \sin(5X) + \epsilon$	0.707	1.000	0.242	1.000	0.222	1.000	0.160
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	1.000	0.286	1.000	0.125	1.000	0.311
$Y = \epsilon$	0	0.877	0.123	0.914	0.086	0.925	0.075
Heteroscedasticity	0	0.772	0.228	0.793	0.207	0.873	0.127

Table 22: Simulation results for basic bootstrap intervals of the local quadratic regression based impact for different bandwidths. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power).

Table 22 the simulation results for the local quadratic regression based impact analysis are given. Similar to the kernel-smoother based and the local linear regression based mean impact analysis the results depend on the chosen bandwidth h. This methods does not show a good power and is therefore not followed up any further.

As a next step we want to examine how the kernel-smoother based impact analysis performs when we choose the bandwidth data dependent. For the simulations the bandwidth was chosen with the `R` function `h.select` which selects a bandwidth associated with approximate degrees of freedom equal to 6 (this is the default setup for non-parametric regression, for further details of degrees of freedom of kernel smoothers see Hastie et al. (2001)). Since it turned out that the data-based choice of h leads to undercoverage in scenarios where the mean impact equals zero we also computed some intervals where tests for the the hypothesis that the mean impact is zero were pre-performed. The first test is a permutation test. Another test is the wild bootstrap test explained in Section A.3.5. The third test is a residual bootstrap test, which is similar to the wild bootstrap test with the difference that the residuals for each observation are drawn from the full set of residuals instead of from a two point distribution. The results of these simulations can be found in Table 23. Note, that the results of Doksum and Samarov (1995) suggest that using a leave-one-out type estimator for the mean impact may lead to more conservative results, even without pre-performing any tests. However, since the simulations of this thesis were very computer intensive, such methods were not investigated. This is an interesting subject of possible future research.

We can see in Table 23 that the coverage probability of the confidence intervals where we do not perform any test prior to their calculation is very poor in the cases where

Model	$\iota_X(Y)$	$Cover_{no\ test}$	$Power_{no\ test}$	$Cover_{perm}$	$Power_{perm}$
$Y = 0.3X + \epsilon$	0.3	0.970	0.873	0.970	0.482
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.949	1.000	0.949	1.000
$Y = \sin(5X) + \epsilon$	0.707	0.996	0.919	0.996	0.850
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.992	1.000	0.992	1.000
$Y = \epsilon$	0	0.601	0.399	0.943	0.057
Heteroscedasticity	0	0.673	0.327	0.726	0.274

Model	$\iota_X(Y)$	$Cover_{wild}$	$Power_{wild}$	$Cover_{resid}$	$Power_{resid}$
$Y = 0.3X + \epsilon$	0.3	0.971	0.429	0.971	0.494
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.949	1.000	0.948	1.000
$Y = \sin(5X) + \epsilon$	0.707	0.996	0.830	0.997	0.853
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.992	1.000	0.992	1.000
$Y = \epsilon$	0	0.951	0.049	0.943	0.057
Heteroscedasticity	0	0.963	0.037	0.726	0.274

Table 23: Simulation results for basic bootstrap intervals of the kernel-smoother based impact for data dependent bandwidth. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power) when performing no test (no test), when pre-performing a permutation test (perm), when pre-performing a wild bootstrap test (wild) and when pre-performing a residual bootstrap test (resid).

the mean impact equals zero. All three tests improve the coverage probability of the tests, where it should be mentioned that only the wild bootstrap test leads to confidence intervals that maintain the desired coverage probability. Furthermore, the use of the wild bootstrap procedure gives good results in terms of power and coverage even in the presence of heteroscedasticity. The confidence intervals with pre performed wild bootstrap test and data dependent bandwidth are therefore preferable.

As it is well known, studentized bootstrap intervals are sometimes second order accurate. In order to investigate whether or not the calculation of studentized bootstrap intervals has a benefit in the case of kernel-smoother based mean impact analysis (where there is no proof of second order accuracy), we computed studentized bootstrap intervals using the functional delta method variance estimate. In the simulations we preperformed a wild bootstrap test to rule out that the mean impact equals zero. One can see (Table 24) that the studentized intervals are more conservative than the basic bootstrap intervals in most cases which results in a loss of power of up to 0.4 in the case of a linear model. Therefore, there is no benefit from calculating studentized intervals. For

Model	$\iota_X(Y)$	Coverage	Power
$Y = 0.3X + \epsilon$	0.3	1	0.037
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.994	0.881
$Y = \sin(5X) + \epsilon$	0.707	0.997	0.853
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.9644	0.971
$Y = \epsilon$	0	0.996	0.004
Heteroscedasticity	0	0.999	0.001

Table 24: Simulation results for studentized bootstrap intervals for the kernel-smoother based impact with data dependent bandwidth and pre-performed wild bootstrap test.

explorational purposes we also computed the same simulations as for Table 24 but with standardized regression functions where the mean impact equals one in each scenario. The results can be found in Table 25 and show that the power for the standardized scenarios is much higher than for the unstandardized scenarios.

Model	$\hat{\iota}_X^{ks}(Y)$	Coverage	Power
$Y = X + \epsilon$	0.8719	1	0.992
$Y = \sin((X+1)\frac{3}{2}\pi)/0.6745 + \epsilon$	1.018	0.985	1
$Y = \sin(5X)/\sqrt{0.5} + \epsilon$	0.9185	1	0.917
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)1.2698} + \epsilon$	1.0140	0.980	0.998

Table 25: Simulation results for studentized bootstrap intervals for the kernel-smoother based impact with data dependent bandwidth and pre-performed wild bootstrap test. The regression functions have been modified so that the resulting mean impact equals 1 in each scenario.

Additionally to the fact that the bandwidth of the kernel smoother has to be fixed, we also dropped the denominator of the kernel smoother when deriving the theory of the kernel smoother based impact analysis. In the following we will show some exemplary simulation results where we did not drop the denominator of the kernel smoother. Table 26 shows that the performance of the intervals based on kernel smoothers with fixed bandwidth but inclusion of the denominator also depends on the choice of the bandwidth h. In the zero impact scenario these intervals have much poorer coverage probability than the intervals based on the kernel smoothers without denominator. The

Model	$\iota_X(Y)$	$Cov_{0.05}$	$Pow_{0.05}$	$Cov_{0.1}$	$Pow_{0.1}$	$Cov_{0.5}$	$Pow_{0.5}$
$Y = 0.3X + \epsilon$	0.3	0.043	1.000	0.414	1.000	0.917	0.972
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.839	1.000	0.921	1.000	0.997	1.000
$Y = \sin(5X) + \epsilon$	0.707	0.511	1.000	0.800	1.000	0.999	0.999
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.950	1.000	0.997	1.000	1.000	0.570
$Y = \epsilon$	0	0.001	0.999	0.015	0.985	0.443	0.557
Heteroscedasticity	0	0.000	1.000	0.001	0.999	0.331	0.669

Table 26: Simulation results for basic bootstrap intervals of the kernel-smoother based impact without deletion of the denominator for different bandwidths. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power).

results in the remaining setups are comparable. Table 27 gives simulation results for the kernel smoother based impact with denominator but data dependent bandwidth. We also performed a wild bootstrap test prior to the calculation of the intervals. Table 27

Model	$\iota_X(Y)$	$Cover_{no\ test}$	$Power_{no\ test}$	$Cover_{wild}$	$Power_{wild}$
$Y = 0.3X + \epsilon$	0.3	0.889	0.978	0.889	0.735
$Y = \sin((X+1)\frac{3}{2}\pi) + \epsilon$	0.6745	0.942	1.000	0.942	1.000
$Y = \sin(5X) + \epsilon$	0.707	0.999	0.983	0.999	0.937
$Y = \frac{\sin(12(X+0.2))}{(X+0.2)} + \epsilon$	1.2698	0.956	1.000	0.956	1.000
$Y = \epsilon$	0	0.318	0.682	0.839	0.161
Heteroscedasticity	0	0.245	0.755	0.605	0.395

Table 27: Simulation results for basic bootstrap intervals of the kernel-smoother based impact without deletion of the denominator for data dependent bandwidth. Given are the coverage probability for the (unrestricted) mean impact and the probability of excluding zero (power).

shows that when regarding the kernel smoother based mean impact with data dependent bandwidth where we do not leave out the denominator of the kernel smoother we observe severe undercoverage in the scenarios where the mean impact is zero and in the linear scenario of the first model. In contrast to the case where we dropped the denominator, pre performing of the wild bootstrap test does not improve the coverage probability to an acceptable level.

4.2. Partial impact analysis

To assess the performance of the different methods for a non-linear partial mean impact analysis we use four different models. In all models we have one target variable Y and two independent variables X_1 and X_2. In each case we are interested in the partial mean impact of X_1 on Y. The first two scenarios were also investigated in Scharpenberg (2012) and are given a follows. Model I is given by a linear relationship between Y and X_2, whereas X_1 has no direct influence on the target variable. The model equation is given by $Y = 2 + 0.2X_2 + \epsilon$, where $\epsilon \sim N(0,1)$ is independent from (X_1, X_2), $X_1 \sim N(0, 1.5625)$, $X_2 \sim N(0,1)$ and the correlation between the independent variables is given by $Corr(X_1, X_2) = 0.6$. The second model is similar to the fist one, but now X_1 has a linear influence on Y. The model is given by $Y = 0.3X_1 + 0.2X_2 + \epsilon$, where $\epsilon \sim N(0,1)$ is independent from (X_1, X_2), $X_1 \sim N(0, 1.5625)$, $X_2 \sim N(0,1)$ and the correlation between the independent variables is given by $Corr(X_1, X_2) = 0.6$. In the third model, we assume the independent variables to be stochastically independent. We write the model as $Y = X_1^2 + X_2 + \epsilon$, where $\epsilon \sim N(0,1)$ is independent from (X_1, X_2) and $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$ are stochastically independent. In this case, X_1 has a quadratic influence on the target variable. However this scenario is constructed in a way that the linear partial mean impact equals zero, which is why we expect the non-linear mean impact analysis to outperform the linear partial mean impact when trying to infer about the unrestricted partial mean impact, which is strictly positive in this case. In the fourth scenario we also have that X_1 and X_2 are independent. The model is given as $Y = \sin(12(X_1 + 0.2))/(X_1 + 0.2) + X_2 + \epsilon$, where $\epsilon \sim N(0,1)$ independent from $X_1 \sim U[0,1]$ and $X_2 \sim N(0,1)$. We use a normal kernel for the kernel-smoother based partial mean impact.

4.2.1. Partial linear mean impact analysis

As a first step we investigate the performance of the partial linear mean impact analysis. To this end, we perform simulation runs in the three models specified above. Similar to the singe covariate case we compute the confidence intervals based on the normal approximation and studentized confidence intervals (using the functional delta method variance estimate). The results of Table 28 show that the confidence interval based on the normal approximation for the partial linear mean impact outperforms the studentized bootstrap interval in scenario II. The power loss when moving from the normal approximation to the bootstrap interval amounts to 0.2. In the first model, where the partial mean impact $\iota_{X_1}(Y|X_2)$ is zero, both methods hold the coverage probability.

Model	$\iota_{X_1}(Y\|X_2)$	$Cover_{norm}$	$Power_{norm}$	$Cover_{boot}$	$Power_{boot}$
I	0	0.950	0.005	0.989	0.011
II	0.3	0.980	0.838	0.964	0.613
III	1.414	1.000	0.071	1.000	0.023
IV	1.2698	1.000	0.383	1.000	0.145

Table 28: Simulation results for the partial linear mean impact. Compared are the confidence intervals based on the asymptotic normality (norm) result and the bootstrap confidence interval (boot). For each interval the coverage probability for the (unrestricted) partial mean impact and the probability of exclusion of zero (power) are given.

In models III an IV, where the true linear partial mean impact $\iota^{lin}_{X_1}(Y|X_2)$ equals zero but the unrestricted partial mean impact is strictly positive, we can observe that both methods have very low power, which shows the limitations of the linear partial mean impact.

4.2.2. Partial polynomial impact analysis

Now we want to compare the two procedures for a partial mean impact based on polynomials (again of degree 3). Calculated are the studentized bootstrap confidence intervals for $\iota_{X_1}(Y|X_2)$ and for $\iota^2_{X_1,X_2}(Y) - \iota^2_{X_1}(Y)$. We learn from Table 29 that the polynomial

Model	$\iota_{X_1}(Y\|X_2)$	$\iota^2_{X_1,X_2}(Y) - \iota^2_{X_1}(Y)$	$Cover_{part}$	$Power_{part}$	$Cover_{alt}$	$Power_{alt}$
I	0	0	0.934	0.066	0.934	0.066
II	0.3	0.266	0.924	0.619	1.000	0.619
III	1.414	2	0.973	1.000	0.973	1.000
IV	1.2698	1.6212	0.979	0.184	0.979	0.184

Table 29: Simulation results for the partial polynomial mean impact. Compared are the confidence intervals based on the direct approach via density changes, i.e. the partial polynomial impact (part), and the intervals for the alternative approach (alt). For each interval the coverage probability for the (unrestricted) partial mean impact (respectively for the difference of the squared impacts) and the probability of exclusion of zero (power) are given.

based partial mean impact leads to anti-conservative confidence intervals in the first two scenarios. The alternative approach leads to under coverage in scenario I ($\iota_{X_1}(Y|X_2)$=0)

and a coverage probability of 1 in scenario II. In scenario III both methods perform equally well with a coverage of 0.975 and a power equal to 1. When regarding the results of scenario IV, we see that the power of the partial mean impact analysis based on polynomials drops to 0.184. We observe that both methods lead to the same power, which is due to the fact that

$$\iota^{lin^2}_{X_1}(Y|X_2) = \iota^{lin^2}_{X_1,X_2}(Y) - \iota^{lin^2}_{X_1}(Y),$$

which implies that the confidence bounds for $\iota^{lin^2}_{X_1}(Y|X_2)$ and $\iota^{lin^2}_{X_1,X_2}(Y) - \iota^{lin^2}_{X_1}(Y)$ are the same. Since we obtain from this confidence bound the bound for $\iota^{lin^2}_{X_1}(Y|X_2)$ by a simple transformation, which does not change the fact if zero is included or not in the confidence interval, we obtain the same power in both approaches.

4.2.3. Kernel-smoother based partial impact analysis

In this section we investigate the performance of the kernel smoother based partial mean impact. Compared are the performances of the intervals based on the "direct" approach via density changes and of the approach where we consider the difference $\iota_{X_1,X_2}(Y) - \iota_{X_1}(Y)$. We computed basic bootstrap intervals in all scenarios. The results

Model	$\iota_{X_1}(Y\|X_2)$	$\iota_{X_1,X_2}(Y) - \iota_{X_1}(Y)$	$Cover_{part}$	$Power_{part}$	$Cover_{alt}$	$Power_{alt}$
I	0	0	0.941	0.059	0.971	0.029
II	0.3	0.320	0.999	0.159	1.000	0.019
III	1.414	0.732	1.000	0.700	0.984	0.550
IV	1.2698	0.6163	0.997	1.000	0.996	0.993

Table 30: Simulation results for the partial kernel smoother based mean impact. Compared are the confidence intervals based on the direct approach via density changes, i.e. the partial polynomial impact (part), and the intervals for the alternative approach (alt). For each interval the coverage probability for the (unrestricted) partial mean impact and the probability of exclusion of zero (power) are given.

of Table 30 show that the kernel-smoother based partial mean impact tends to slight undercoverage (0.941) when the mean impact is zero. In each scenario the partial mean impact approach is more powerful than the alternative approach, the loss of power, when applying the alternative approach amounts to 0.15 in scenario III. However, when regarding scenario IV, we can see that the kernel smoother based partial mean impact

analysis outperforms both the linear and the polynomial based analysis in that setup with a gain in power of about 0.61 respectively 0.81. This indicates that the kernel smoother based partial mean impact analysis is the best method when the underlying relationship is highly non-linear.

4.3. Summary of simulation results

In this section we give a summary of the simulation results above. In the single covariate case we observed that the linear mean impact analysis works good when the underlying model is indeed linear, but has very low power in highly non-linear setups.

The shrinkage like approach to the construction of confidence intervals for the polynomial based mean impact led to undercoverage in the case where the mean impact equals zero. This implies that the test from Section 1.7.1 which is essentially an application of the result of White (1980b) does not maintain its level in small sample sizes (e.g. $n = 100$ like it is the case in this thesis). However, increasing the sample size to 200 and 500 in additional simulations not shown here improved the results substantially and led to coverage probabilities close to the nominal level.

The bootstrap confidence intervals for the polynomial based mean impact resolved the issue of undercoverage for small sample sizes. Using studentized bootstrap confidence intervals as described in Section 1.7.5 led to higher coverage probabilities. However, in the case where the mean impact equals zero we still observed a slight undercoverage (0.936, see Table 16). The power of this method could be improved by using studentized bootstrap intervals which do not make use of the transformation to the $\iota^2_{\mathbf{X}}(Y)$-scale and back. In that case, when pre performing the test of Section 1.7.1, we obtain confidence intervals that hold the level of significance with the exception of the linear case, where the coverage probability drops to 0.931. Using basic bootstrap intervals leads to very similar results. Nevertheless, the intervals for the polynomial based mean impact have much larger probability to exclude zero, when the mean impact is greater than zero. Hence, using these intervals instead of the intervals from the linear mean impact analysis leads to a more powerful procedure. The gain in power amounts to approximately 0.79 in scenario 2. However, when the underlying model gets too curvy the gain from fitting polynomials instead of straight lines vanishes, which is due to the limited flexibility of polynomials.

The performance of the kernel-smoother based impact analysis was shown to depend highly on the chosen kernel-bandwidth h. Furthermore, it could be seen that moving from kernel-smoothers to higher order local regression has no benefit. On the contrary

doing so leads to confidence intervals which perform worse in terms of power. In the simulations it could be shown that kernel-smoothing with data dependent bandwidth leads to severe undercoverage when the mean impact is small. However, pre performing a wild bootstrap test for the hypothesis that the mean impact is zero resolved this issue and lead to confidence intervals which maintained the coverage probability and had high power in the non-linear scenarios. However, these intervals experience a huge loss of power when the underlying relationship is linear compared to the intervals from the linear mean impact analysis (0.429 vs. 0.853).

To summarize the simulation results for the single covariate case, we can conclude that the linear mean impact analysis performs good when the underlying relationship between X and Y is indeed linear. When we assume non-linear relationships, the kernel-smoother based impact analysis (with pre-performed wild bootstrap test) was shown to give the most satisfying results, since we had powers of 1, 0.83 and 1 in the scenarios 2-4, where the linear mean impact analysis was not able to identify an effect. When the non-linearities are of moderate order (e.g. in scenario 2), the polynomial based mean impact also gave good results.

The simulations for the linear partial mean impact indicate, that the confidence intervals based on the normal approximation outperform the studentized bootstrap intervals. Both types of confidence intervals maintained the coverage probability in all scenarios under investigation. However, as was expected, the linear partial mean impact performed well in the scenarios where the relationship between Y and X_1 was linear (Model I and II) but had very low power (0.071) when moving the a quadratic relationship. In model IV (non-linear in X_1 and linear in X_2; $X_1,\ X_2$ independent) however the power increased to 0.383, which is still not very high.

For the polynomial partial mean impact we observed undercoverage for the confidence intervals originating from the approach via density changes. In the setup where the influence of X_1 on Y is linear (scenario II) we observe a loss of power of about 0.2 compared to the linear mean impact. A similar power loss is observed in scenario IV. However, in scenario III (quadratic influence of X_1) we observe a power of 1 which is considerably higher than that of the linear mean impact.

When regarding the partial mean impact based on kernel-smoothers we observed slight undercoverage (0.941) in the scenario where the mean impact is zero. The alternative approach to the quantification of the influence of X_1 showed a better coverage (0.971). However, in the other two scenarios this approach led to a loss of power of about 0.2 compared to the direct approach. It was noticeable that the kernel-smoother based

partial mean impact only has a power of 0.159 in the linear scenario resulting in a loss of about 0.68 compared to the linear partial mean impact. In the non-linear scenario III the kernel-smoother based partial mean impact showed a power of 0.7 which is considerably higher than the power of the linear partial mean impact but still 0.3 less than the power of the polynomial based approach. In scenario IV the partial mean impact based on kernel smoothing showed the best performance with a coverage probability and a power both close to 1, clearly outperforming the other methods at hand.

The simulation results for the partial mean impact analysis can be summarized as follows. The linear partial mean impact analysis only performed well, when the underlying relationship between Y and X_1 was indeed linear. In some non-linear setups we observed close to no power for this methods. In the setup of a quadratic relationship the polynomial based mean impact outperformed its opponents. When we face higher order non-linear relationships the kernel smoother based mean impact is the best method at hand. Hence, when we expect the true relationship to be non-linear but not of a high order, we should use the polynomial based mean impact analysis. In cases of higher order non linearities the method of choice should be the kernel smoother based partial mean impact analysis.

5. Conclusion and outlook

In this thesis we filled a gap in the work of Scharpenberg (2012) and derived an asymptotic normality result for the linear signed (partial) mean impact. From this followed a test and consequently a method to the construction of confidence intervals for the linear (partial) mean impact. Furthermore we extended the idea of the mean impact which was originally derived in Scharpenberg (2012) and Brannath and Scharpenberg (2014) to non-linear associations. We defined a common mean impact of several variables which could be estimated by restriction to linear functions. This common mean impact allows for the analysis of associations based on polynomial regression, spline regression with fixed knots or (nearly) any other additive model in one ore more covariates. Another application of the scenario of a common linear mean impact is given by the scenario where we have a zero inflated covariate, i.e. a covariate which has a high probability of becoming zero. In this case the common linear mean impact allows us to model the part where the covariate is zero independent from the part where it differs from zero. This means that we can, in a sense, combine an ANVOA and a linear model and obtain a single measure of association. Using the smooth function model of Hall (1988) and Hall (1992), we have shown that bootstrap BC_a and studentized bootstrap intervals are second order accurate in the setup of the linear common mean impact.

To obtain higher flexibility we also regarded a mean impact based on kernel-smoothing. In this case we chose the distributional disturbance for the estimation of the mean impact as the standardized prediction of a kernel smoother fit where the denominator of the kernel smoother is left out. Using this estimate we were able to show that the resulting estimate is a function of a U-statistics and thereby asymptotically normally distributed. Furthermore we justified the use of bootstrap methods in this case. Higher order local regression was also looked at but did not perform good in simulations. Finally a modification of the kernel-smoother based mean impact gave a consistent and asymptotically normally distributed estimate for the unrestricted mean impact. However, the fact that we have do not use all data in this approach makes it hard to apply in praxis.

In all single-covariate non-linear mean impact analyses we also derived a mean slope and a measure for determination, as generalizations of the measured derived in Scharpenberg (2012).

In extension to the single covariate case we also defined non-linear partial mean impacts which quantify the association between the target variable Y and an independent variable X_1 which goes beyond the possible associations driven by other covariates $X_2, ..., X_k$.

In this setup as well we defined and investigated a common linear partial mean impact. Applications of this common linear mean impact are again polynomial impacts which account for possible polynomial influences or more general we can fit (almost) any additive model in X_1 and account for (almost) any influences of $X_2, ..., X_k$ which can be expressed by additive models.

For the kernel-smoother based mean impact we also derived a partial mean impact. The partial mean impact from Scharpenberg (2012) uses orthogonal projections. In this thesis we also derived an approach that does not need such projections. It quantifies the influence of X_1 on Y which goes beyond the possible influence of other covariates by the difference of the common mean impact of all variables and the common mean impact of all variables except X_1. In all partial non-linear impact analyses we also derived partial non-linear mean slopes and partial non-linear measures for determination.

Simulations indicated that in the single covariate case the kernel-smoother based mean impact is the most powerful approach except when the true underlying regression relationship is linear. In that case, obviously, the linear mean impact performed best. The results from the simulation for the non-linear partial mean impact analysis showed that the performance of the different methods are more dependent on the underlying scenario than in the single covariate case. The linear partial mean impact analysis did by far outperform the other methods in a linear scenario. In moderately non-linear setups the polynomial partial mean impact performed best, while the kernel smoother based partial mean impact analysis was the only method that still had reasonable power in highly non-linear scenarios.

The framework of the mean impact analysis still offers many opportunities for further research. First of all it is desirable to justify the use of data dependent bandwidth in the case of kernel-smoothing theoretically. We are also interested in a theoretically justified method allowing for the use of a kernel-smoother where we do not need to drop the denominator, which uses the full data set available and maintains the coverage probability. Furthermore, it might be valuable to allow for splines with a data dependent knot sequence. Another interesting topic for further research is the application of the mean impact analysis to high dimensional setups. It might be possible to apply data reducing methods like the principal component analysis and use the mean impact analysis to obtain an interpretable and sensible measure of association.

References

Bickel, P. J. and D. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics 9*.

Brannath, W. and M. Scharpenberg (2014). Interpretation of linear regression coefficients under mean model miss-specification. *arXiv:1409.8544v4 [stat.ME]*.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*.

Davison, A. and D. Hinkley (2009). *Bootstrap Methods and their Application (11th printing)*. Cambridge University Press.

Doksum, K. and A. Samarov (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics 23*.

Efron, B. (1979). Bootstrap methods: Another look at the jacknife. *Annals of Statistics 7*.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications 14*.

Fischer, G. (2005). *Lineare Algebra (15.Auflage)*. Wiesbaden: Vieweg.

Gasser, T. and H.-G. Müller (1979). *Kernel estimation of regression functions*. Heidelberg: Springer-Verlag.

Gasser, T., H.-G. Müller, W. Kohler, L. Molinari, and A. Prader (1984). Nonparametric regression analysis of growth curves. *The Annals of Statistics 12*.

Gasser, T., H.-G. Müller, and V. Mammschitz (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Methodological) 47*.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics 16*.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York, Berlin, Heidelberg, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest: Springer Verlag.

Härdle, W. and J. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics 19*.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning (2nd edition)*. New York, Berlin, Heidelberg: Springer-Verlag.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics 19*.

Huber, P. J. (1981). *Robust Statistics*. New York, Chichester, Brisbane, Toronto: John Wiley & Sons.

Karunamuni, R. and T. Alberts (2005). On boundary correction in kernel density estimation. *Statistical Methodology 2*.

Klenke, A. (2008). *Wahrscheinlichkeitstheorie (2. Auflage)*. Berlin, Heidelberg: Springer-Verlag.

Kowalski, J. and X. Tu (2008). *Modern Applied U-Statistics*. Wiley.

Mack, Y. P. and B. W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift fuer Wahrscheinlichkeitstheorie und verwandte Gebiete 61*.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications 9*.

Paulson, D. S. (2007). *Handbook of Regression And Modeling - Applications for the Clinical and Pharmaceutical Industries*. Chapman & Hall /CRC.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York, Berlin, Heidelberg, Tokyo: Springer-Verlag.

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

Scharpenberg, M. (2012). *A population-based approach to analyze the influence of covariates*. University of Bremen: Diploma thesis.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics 18*.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics 26*.

White, H. (1980a). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*.

White, H. (1980b). Least squares to approximate unknown regression functions. *International Economic Review 21*.

Wu, C. F. J. (1968). Jacknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics 14*.

A. Methodology

All literature used in this appendix can be found in the main literature list.

A.1. Nonparametric regression

Since we make extensive use of non-linear regression techniques in the course of this thesis we will give an introduction to them in the following sections. Besides polynomial regression, which is expected to be known to the reader we will make use of kernel- and spline-methods.

A.1.1. Kernel methods

This section is based on Chapter 6 of Hastie et al. (2001). As in the case of linear regression we are interested in estimating the regression function $f(x) = E_{\mathbf{P}}(Y|X = x)$ in order to characterize the dependence of Y on X. Since linear regression delivers a rather rough idea of the regression function we are searching for more flexible estimates which give a closer fit to the data.

Kernel regression methods are regression techniques which achieve such flexibility. Let us assume that we have independent observations $(x_i, y_i)_{i=1,\dots,n}$ of the covariate X and the response Y. As we will see, kernel regression estimates are mainly weighted averages of the observations of Y, where the weight depends on the distance of the observations of X to the point where we try to estimate the regression function. They result from fitting different models at each evaluation point. In the sequel we will explain how this is done. One way to estimate the regression function is using the k-nearest neighbor estimator which is simply the average of the responses of the k observations of X which are nearest to the evaluation point x. This means the k-nearest neighbor estimator is given by

$$\hat{f}(x) = \frac{1}{k} \sum_{i=1}^{n} y_i \mathbb{1}_{\{x_i \in N_k(x)\}},$$

where $N_k(x)$ is the set of k points nearest to x, where "closeness" is defined by the Euclidean distance. This method leads to a very bumpy fit (see right panel of Figure 3) since $\hat{f}$ is discontinuous. This is because the fit remains constant as we move on the X-axis until one point to the right becomes closer to the evaluation point than the farthest point to the left in $N_k(x)$. In that moment the point to the right replaces the one to the left in the fit and leads to a different value of $\hat{f}$. Hence $\hat{f}$ changes in a discrete way and is therefore discontinuous.

Since this discontinuity seems inappropriate we use methods which produce a smoother fit.

Kernel Smoother

One method leading to smoother fits is the Nadaraya-Watson kernel-weighted average which goes back to Nadaraya (1964) and Watson (1964) and is given by

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} K_h(x_i - x_0) y_i}{\sum_{i=1}^{n} K_h(x_i - x_0)},$$

where $K_h(x_i - x_0)$ is a kernel weight function with window width h. Typical kernel weight functions are

- The Epanechnikov quadratic kernel which is given by

  $$K_h(x_i - x_0) = D\left(\frac{|x_i - x_0|}{h}\right),$$

  where

  $$D(t) = \begin{cases} \frac{3}{4}(1 - t^2) & \text{if } |t| \leq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

  Epanechnikov (1969) suggested

  $$D(t) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{t^2}{5}) & \text{if } |t| \leq \sqrt{5}; \\ 0 & \text{otherwise} \end{cases}$$

  instead of (A.1) which is also commonly referred to as "Epanechnikov quadratic kernel",

- The tri-cube function where

  $$K_h(x_i - x_0) = D\left(\frac{|x_i - x_0|}{h}\right),$$

  with

  $$D(t) = \begin{cases} (1 - |t|^3)^3 & \text{if } |t| \leq 1; \\ 0 & \text{otherwise,} \end{cases}$$

- The Gaussian Kernel, where $D(t) = \varphi(t)$ is the density function of the standard normal distribution.

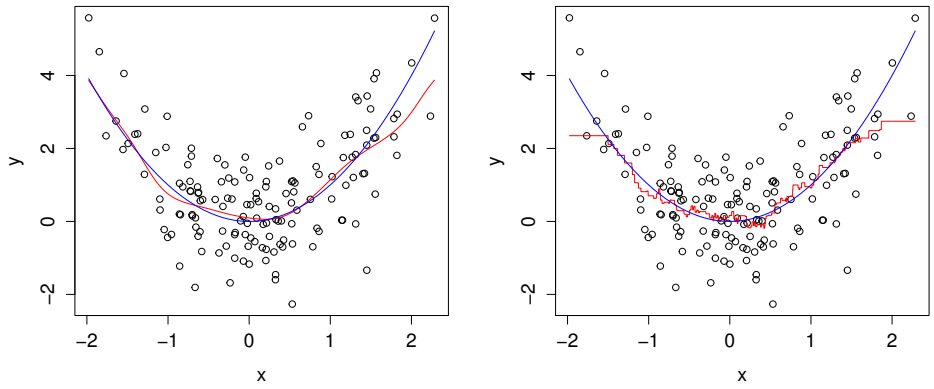

Figure 3: Comparison of a kernel-smoother fit (red curve in the left panel) and a 15-nearest neighbor fit (red curve in the right panel) to a data set of 150 pairs $x_i,\ y_i$ generated at random from $Y = X^2 + \epsilon,\ X \sim N(0,1),\ \epsilon \sim N(0,1)$. The blue curve displays the underlying relationship $f(X) = X^2$. For the kernel smoother a Gaussian kernel with automatically chosen window width $h = 0.247$ was used.

Figure 3 shows a kernel smoother fit and the k-nearest neighbor fit for a given dataset. One can see that the kernel smoother fit is much smoother than the bumpy fit of the nearest neighbor estimator. In practice one has to choose either the bandwidth h when using the kernel smoother or the number k of neighbors involved in the nearest neighbor fit. When choosing this parameter one has to do a trade off between bias and variance. Large bandwidths (respectively large number of neighbors) will decrease the variance of the estimator, since one averages over more observations, while increasing its bias. Vice versa a small bandwidths leads to higher variance and lower bias. There are asymptotic results which state that the kernel regression smoother is consistent for the regression function under certain conditions on the kernel, its bandwidth and the common density of X and Y.

Mack and Silverman (1982) show such a convergence result. Let $(X,Y),(X_i,Y_i), i = 1,2,...$ be i.i.d. bivariate random variables with common joint density $r(x,y)$. Furthermore, let $g(x)$ be the marginal density of X and $f(x) = E_{\mathbf{P}}(Y|X = x)$ the regression function of Y on X and h_n the bandwidth of the Kernel $K_{h_n}(u) = K(u/h_n)$. The assumptions used in their consistency proof are already given in Assumption 2.27 and Assumption 2.28 but are repeated here for better readability.

Assumption A.1.
- *K is uniformly continuous with modulus of continuity w_K, i.e. $|K(x) - K(y)| \leq w_K(|x - y|)$ for all $x, y \in \text{supp}(K)$ and $w_K : [0, \infty] \to [0, \infty]$ is continuous at zero with $w_K(0) = 0$. Furthermore K is of bounded variation $V(K)$;*
- *K is absolutely integrable with respect to the Lebesgue measure on the line;*
- *$K(x) \to 0$ as $|x| \to \infty$;*
- *$\int |x \log |x||^{\frac{1}{2}} |dK(x)| < \infty$,*

and

Assumption A.2.
- *$E_{\mathbf{P}}|Y|^s < \infty$ and $\sup_x \int |y|^s r(x, y)dy < \infty$, $s \geq 2$;*
- *r, g and l are continuous on an open interval containing the bounded interval J, where $l(x) = \int yr(x, y)dy$.*

Theorem A.3. *Suppose K satisfies Assumption A.1 and Assumption A.2 holds. Suppose J is a bounded interval on which g is bounded away from zero. Suppose that $\sum_n h_n^\lambda < \infty$ for some $\lambda > 0$ and that $n^\eta h_n \to \infty$ for some $\eta < 1 - s^{-1}$. Then*

$$\sup_J |\hat{f}(x) - f(x)| = o(1)$$

with probability one.

Hence, under suitable conditions the Nadaraya-Watson kernel regression estimator is consistent for the regression function.

Local Linear Regression

As one can see in Figure 3 the kernel regression estimator can be biased at the boundary because of the asymmetry of the kernel in that region. There are several methods that address the boundary issues of kernel smoothing. For example Gasser and Müller (1979), Gasser et al. (1984) and Gasser et al. (1985) recommend using "*boundary kernels*", which are kernels with asymmetric support. This approach is not followed further in this thesis. Karunamuni and Alberts (2005) give an overview of other methods which could be applied. One of these methods, which reduces the bias to first order is the local linear regression. Note that the kernel regression estimator $\hat{\alpha}(x_0)$ at the point x_0 is the solution to the weighted least squares problem

$$\min_{\alpha \in \mathbb{R}} \sum_{i=1}^{n} K_h(x_i - x_0)[y_i - \alpha]^2.$$

Hence, by using the Nadaraya-Watson kernel regression estimator we essentially fit local constants to the data. Local linear regression (also called loess) was initially proposed by Cleveland (1979) and goes one step further. It does local linear fits at each evaluation point. Thus, we consider at each point x_0 the weighted least squares problem

$$\min_{\alpha,\beta} \sum_{i=1}^{n} K_h(x_i - x_0)[y_i - \alpha - \beta x_i]^2. \tag{A.2}$$

The estimator for the regression function is then given by $\hat{f}(x_0) = \hat{\alpha}(x_0) + x_0\hat{\beta}(x_0)$, where $\hat{\alpha}(x_0)$ and $\hat{\beta}(x_0)$ are solutions to (A.2). With $b(x)^T = (1, x)$, $\mathbf{B}$ the $n \times 2$ matrix with ith row $b(x_i)^T$ and $\mathbf{W}(x_0) = \text{diag}\left(K_h(x_1 - x_0), ..., K_h(x_n - x_0)\right)$ we have

$$\hat{f}(x_0) = b(x_0)^T \left(\mathbf{B}^T\mathbf{W}(x_0)\mathbf{B}\right)^{-1} \mathbf{B}^T\mathbf{W}(x_0)\mathbf{y}.$$

For the analyses of Section 2.2.4 we rearrange this expression in the following way

$$\begin{aligned}
&\hat{f}(x_0) \\
=&(1, x_0) \begin{pmatrix} \sum_{j=1}^n K_h(x_j - x_0) & \sum_{j=1}^n x_j K_h(x_j - x_0) \\ \sum_{j=1}^n x_j K_h(x_j - x_0) & \sum_{j=1}^n x_j^2 K_h(x_j - x_0) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n y_j K_h(x_j - x_0) \\ \sum_{j=1}^n x_j y_j K_h(x_j - x_0) \end{pmatrix} \\
=&\frac{1}{\det(\mathbf{B}^T\mathbf{W}(x_0)\mathbf{B})}(1, x_i) \begin{pmatrix} \sum_{j=1}^n \sum_{l=1}^n (x_j^2 - x_j x_l) K_h(x_j - x_0) K_h(x_l - x_0) y_l \\ \sum_{j=1}^n \sum_{l=1}^n (x_l - x_j) K_h(x_j - x_0) K_h(x_l - x_0) y_l \end{pmatrix} \\
=&\frac{\sum_{j=1}^n \sum_{l=1}^n (x_j - x_0)(x_j - x_l) K_h(x_j - x_0) K_h(x_l - x_0) y_l}{\sum_{j=1}^n \sum_{l=1}^n (x_j^2 - x_j x_l) K_h(x_j - x_0) K_h(x_l - x_0)}.
\end{aligned} \tag{A.3}$$

It can be shown that local linear regression reduces bias to first order. This means that $E_{\mathbf{P}}(\hat{f}(x_0)) - f(x_0)$ only depends on quadratic and higher-order terms in $(x_0 - x_i)$, $i = 1, .., n$ (cf. Hastie et al. (2001)).

Local Polynomial Regression

As a generalization to the Nadaraya-Watson kernel regression estimator and the local linear regression we now introduce local polynomial fitting. Hence, we locally fit polynomials of arbitrary degree $k \geq 0$ and therefore regard the weighted least squares problem

$$\min_{\alpha,\beta_j,j=1,...,k} \sum_{i=1}^{n} K_h(x_i - x_0)[y_i - \alpha - \sum_{j=1}^{k} \beta_j x_i^j]^2$$

at x_0, whose solution we denote by $(\hat{\alpha}(x_0), \hat{\beta}_1(x_0), ..., \hat{\beta}_k(x_0))$ and estimate the regression function via $\hat{f}(x_0) = \hat{\alpha}(x_0) + \sum_{j=1}^{d} \hat{\beta}_j(x_0) x_0^j$. We can obtain $\hat{f}$ via

$$\hat{f}(x_0) = b(x_0)^T \left(\mathbf{B}^T \mathbf{W}(x_0) \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{W}(x_0) \mathbf{y},$$

where $b(x_0)^T = (1, x_0, ..., x_0^k)$, $\mathbf{B}$ is the $n \times (k+1)$ matrix with ith row $b(x_i)$ and $\mathbf{W}(x_0) = \text{diag}\left(K_h(x_1 - x_0), ..., K_h(x_n - x_0)\right)$. Local polynomial regression reduces bias in regions of high curvature of the regression function compared to local linear regression. The price to be paid for this is an increase of the variance. Hastie et al. (2001) summarize the behavior of local fits as follows:

- "Local linear fits can help bias dramatically at the boundaries at a modest cost in variance. Local quadratic fits do little at the boundaries for bias, but increase the variance a lot.
- Local quadratic fits tend to be most helpful in reducing bias due to curvature in the interior of the domain.
- Asymptotic analysis suggest that local polynomials of odd degree dominate those of even degree. This is largely due to the fact that asymptotically the MSE is dominated by boundary effects."

As mentioned before we have to choose the bandwidth h when applying kernel methods. For the theory of non-linear impact analysis derived in this thesis this choice is not allowed to depend on the data. In practice, when one is only interested in estimating the regression function (and not necessarily in impact analysis) the choice of the bandwidth can be done by cross-validation.

Local Regression in $\mathbb{R}^k$

Up to this point we only considered one-dimensional kernel methods. We can easily generalize this concept to the multidimensional case where we observe a set of variables $X_1, ..., X_k$ and want to fit local polynomials in this variables with maximum degree d to the data in order to describe the regression function $f(x_1, ..., x_k) = E_{\mathbf{P}}(Y | X_1 = x_1, ..., X_k = x_k)$. Hence, letting $b(x)$ consist of all polynomial terms with maximum degree d (e.g. we have $b(X) = (1, X_1, X_1^2, X_1^3, X_2, X_2^2, X_2^3, X_1X_2, X_1^2X_2, X_1X_2^2)$ for $d = 3$ and $k = 2$), one solves at x_0 the following equation for $\beta \in \mathbb{R}^m$, where m is the dimension of $b(X)$

$$\min_{\beta \in \mathbb{R}^m} \sum_{i=1}^{n} K_h(x_i - x_0)[y_i - b(x_i)^T \beta]^2. \tag{A.4}$$

Usually we have for the kernel function

$$K_h(x_i - x_0) = D\left(\frac{\|x_i - x_0\|}{h}\right),$$

with $\|\cdot\|$ being the Euclidean norm and D a one-dimensional kernel. The least squares fit at a specified point x_0 is then given by $\hat{f}(x_0) = b(x_0)^T\hat{\beta}(x_0)$, where $\hat{\beta}(x_0)$ is a solution to (A.4). We can rewrite this as

$$\hat{f}(x_0) = b(x_0)^T\left(\mathbf{B}^T\mathbf{W}(x_0)\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{W}(x_0)\mathbf{y},$$

where $\mathbf{B}$ is the matrix with ith row $b(x_i)$ and $\mathbf{W}(x_0) = \text{diag}\left(K_h(x_1 - x_0), ..., K_h(x_n - x_0)\right)$. Hastie et al. (2001, p 174) recommend the standardization of each predictor prior to smoothing "since the Euclidean norm depends on the units in each coordinate".

Additionally to this, the rise in dimensionality comes along with undesired side effects. The boundary effects of kernel smoothing in one dimension "are a much bigger problem in two or higher dimensions, since the fraction of points on the boundary is larger" (Hastie et al., 2001, p. 147). Furthermore, it is claimed that local regression loses it usefulness in dimension much higher than two or three, due to the impossibility of simultaneously maintaining low bias and low variance without a sample size which increases exponentially fast in k.

A.1.2. Spline methods

In this section we will present spline methods including cubic splines and natural cubic splines. This section is based on Chapter 5 of Hastie et al. (2001).

Piecewise Polynomials and Splines

We are still interested in fitting functions to data which are obtained from i.i.d. observations $(X_i, Y_i)_{i=1,...,n}$. In this section we present methods that divide the domain of X into several areas and fit functions in each area. The function fitted in a certain area is used to derive predictions for all points in this area. To this end we will choose a so called knot sequence $\xi_1 < ... < \xi_k$. Between each two knots a function will be fitted to the data. Then all functions are combined to obtain a global fit. When we are for example interested in fitting constant functions between each two knots we can do so by performing a linear regression with target variable Y and dependent variables

$$h_1(X) = \mathbb{1}_{\{X<\xi_1\}}, h_2(X) = \mathbb{1}_{\{\xi_1\leq X<\xi_2\}}, ..., h_k(X) = \mathbb{1}_{\{\xi_{k-1}\leq X<\xi_k\}}, h_{k+1}(X) = \mathbb{1}_{\{\xi_k\leq X\}}.$$

The fitted function is then given by $\hat{f}(x) = \sum_{j=1}^{k+1} \hat{\theta}_j h_j(x)$, where $\hat{\theta}_j$ are the least squares estimators from the linear model. The functions $h_1, ..., h_{k+1}$ are called basis functions in the sequel. If we want a piecewise linear fit to the data we need the additional basis functions

$$h_{m+k+1}(X) = h_m(X)X, \quad m = 1, ..., k+1$$

in the model. As a next step we can make the piecewise linear fit continuous by imposing adequate conditions on the coefficients of the linear model. Alternatively one could use the following set of basis functions, which already incorporates the constrains:

$$h_1(X) = 1, \ h_2(X) = X, \ h_{j+2}(X) = (X - \xi_j)_+, \ j = 1, ..., k,$$

where $(\cdot)_+$ denotes the positive part. The fact that the constraint of continuous piecewise linearity is already incorporated in these basis functions is due to the fact that we use the positive part in h_{j+2}. Thereby the function h_{j+2} changes the fit $\hat{f}(x)$ only if $x > \xi_j$ and only by changing the slope (by its coefficient $\hat{\theta}_{j+2}$) of the fit after ξ_j. Hence the resulting fit is continuous and piecewise linear. Figure 4 shows a piecewise constant, a piecewise linear and a continuous piecewise linear fit to the same data. The continuous piecewise linear fit seems to provide the best fit of the three. Since it is very angular we proceed by fitting local polynomials. Additionally to the continuity we can demand continuous derivatives up to a certain order. We define an order-M spline with knots $\xi_1, ..., \xi_k$ as a piecewise polynomial of order $M-1$ with continuous derivatives up to order $M-2$. We call an order 4 spline *cubic spline*. One can see that the local constant fit is an order 1 spline, while the continuous piecewise linear fit is an order 2 spline. We can compute an order M spline via linear regression with target variable Y and covariates

$$\begin{aligned} h_j(X) &= X^{j-1}, \quad j = 1, ..., M, \\ h_{M+l} &= (X - \xi_l)_+^{M-1}, \quad l = 1, ..., k. \end{aligned}$$

Hastie et al. (2001) state that "it is claimed that cubic splines are the lowest-order splines for which the knot-discontinuity is not visible to the human eye." Hence, unless one is interested in smooth derivatives there is no reason to go beyond cubic splines. When using splines, the goodness of the fit depends on the placement of the knots. In the applications of splines to impact analysis, we will assume that the knot sequence is chosen independent from the data in order to prove consistency of the bootstrap. Simulations indicate that the impact analysis also works with a data dependent choice of the knots. Usually, a convenient procedure to choose the knots, which is also applied

in practice, is to place them at empirical quantiles of X.

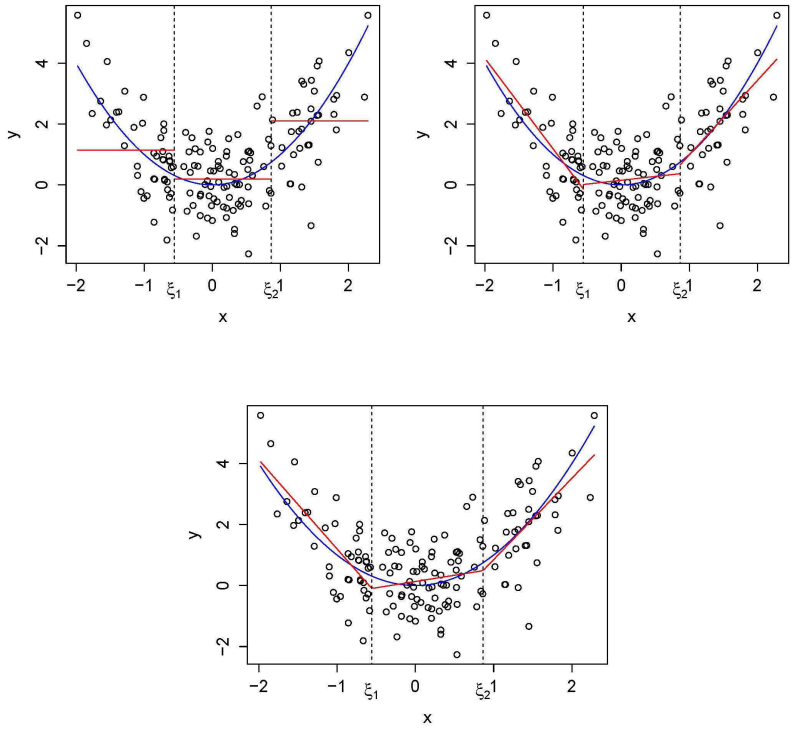

Figure 4: Piecewise constant (upper left panel), piecewise linear (upper right panel) and continuous piecewise linear (lower panel) fit with two knots to a data set of 150 pairs $x_i,\ y_i$ generated at random from $Y = X^2 + \epsilon,\ X \sim N(0,1),\ \epsilon \sim N(0,1)$. The blue curve displays the underlying relationship $f(X) = X^2$.

Natural Cubic Splines

It is well known that polynomial fits tend to be very variable at the boundaries and extrapolation may be dangerous. Hastie et al. (2001) claim that these problems get worse when using splines. Therefore, we want to impose additional constrains to the fitted functions to reduce the variability of the fit outside the range of the data. One common constrain is that the function is linear beyond the boundary knots. A cubic spline which fulfills this condition is called *natural cubic spline*. Of course, reducing variance by forcing the fitted function to be linear at the boundaries leads to bias in that regions. Nevertheless, according to Hastie et al. (2001) the assumption of linearity

near the boundaries is often considered reasonable. We can represent a natural cubic spline with k knots by k basis functions. One such set is

$$N_1(X) = 1,\ N_2(X) = X,\ N_{l+2}(X) = d_l(X) - d_{k-1}(X),$$

where

$$d_l(X) = \frac{(X - \xi_l)_+^3 - (X - \xi_k)_+^3}{\xi_k - \xi_l}.$$

Each of these basis functions has zero second and third derivative outside the boundary knots.

There are many sets of basis functions that represent the same cubic spline. The set of basis functions given here is very simple to understand but lacks numerically attractiveness. Therefore, for numerically reasons, other basis functions such as the so called *B-spline* basis (which will not illustrated here) are used in the practical computation of splines. In R the function `ns` of the package *splines* can be used to compute the basis functions for natural splines for a given data set.

Multidimensional splines

One can also fit smooth functions of several variables $X_1, ..., X_k$ to a given data set. One can do so by choosing appropriate multivariable basis functions. For example in the case of two variables with basis functions $h_{1j}(X_1)$, $j = 1, .., M_1$ for representing functions of X_1 and basis functions $h_{2j}(X_2)$, $j = 1, .., M_2$ for representing functions of X_2 we can define the so called tensor product basis by

$$g_{jl} = h_{1j}(X_1)h_{2l}(X_2), \quad j = 1, ..., M_1,\ l = 1, ..., M_2$$

to represent the two-dimensional function

$$g(X) = \sum_{j=1}^{M_1} \sum_{l=1}^{M_2} \theta_{jl} g_{jl}(X_1, X_2).$$

It is important to note that the dimension of the basis grows exponentially fast in the number of covariates included.

A.2. U-Statistics

In this section, which is based on Kowalski and Tu (2008) we give an introduction to the theory of U-statistics, which is a powerful tool to investigate the asymptotic behavior of sums of correlated random variables. The methods explained here will be used extensively in sections 2.2.1, 2.2.4 and 2.2.5. U-Statistics were originally introduced by Hoeffding (1948) and are closely related to the von Mises statistics which were presented by von Mises (1947).

Often it is the case that we have a statistic which is not the sum of i.i.d. variables. In these cases the classical theorems concerning the asymptotic behavior of the statistic do not apply. In some of these scenarios the theory of U-statistics might be used to derive the asymptotic properties. Let $Z_1, Z_2, ..., Z_n$ be i.i.d. random variables.

Definition A.4. *A statistic U_n is called a k-dimensional order-m U-statistic if it can be written as*

$$U_n = \binom{n}{m}^{-1} \sum_{(j_1,...,j_m)\in C_m^n} w(Z_{j_1}, ..., Z_{j_m}), \tag{A.5}$$

where w is a k-dimensional function which is symmetric in its arguments (i.e. all permutations of its arguments lead to the same value of w) and $C_m^n = \{(j_1, ..., j_m)|1 \leq j_1 < ... < j_m \leq n\}$ is the set of all distinct combinations of m indices from the integer set $\{1, ..., n\}$.

One example of an one-dimensional order-2 U-statistic is the estimator of the variance

$$\hat{\sigma} = \frac{1}{n-1} \sum_{i=1}^{n} \left(Z_i - \frac{1}{n} \sum_{j=1}^{n} Z_j \right)^2 .$$

We have

$$\begin{aligned}
\hat{\sigma}^2 =& \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} \left(Z_i - \frac{1}{n} \sum_{j=1}^{n} Z_j \right)^2 \right) \\
=& \frac{1}{n-1} \left(\sum_{i=1}^{n} Z_i^2 - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_i Z_j \right) \\
=& \frac{1}{n(n-1)} \left((n-1) \sum_{i=1}^{n} Z_i^2 - \sum\sum_{i\neq j} Z_i Z_j \right)
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n(n-1)}\left(\sum_{i\neq j}\sum\frac{1}{2}(Z_i^2+Z_j^2)-\sum_{i\neq j}\sum Z_iZ_j\right)\\
=&\frac{1}{n(n-1)}\sum_{i\neq j}\sum\frac{1}{2}(Z_i-Z_j)^2=\frac{2}{n(n-1)}\sum_{(i,j)\in C_2^n}\frac{1}{2}(Z_i-Z_j)^2.
\end{aligned}
$$

Since $\frac{1}{2}(Z_i-Z_j)^2$ is obviously symmetric, $\hat{\sigma}^2$ is an one-dimensional order-2 U-statistic.

It is obvious, since we have i.i.d. random variables, that U_n defined in (A.5) is an unbiased estimator for $E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_m}))$, provided this exists. Under mild conditions we can say much more about U_n than just that it is unbiased. To this end we need the following assumption.

Assumption A.5.

$$\theta = E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_m})) \quad \textit{and} \quad E_{\mathbf{P}}(w^2(Z_{j_1},...,Z_{j_m}))$$

exist.

We define the "*projection*" of U_n by

$$\hat{U}_n = \sum_{i=1}^{n} E(U_n|Z_i) - \theta(n-1). \tag{A.6}$$

We have

$$E_{\mathbf{P}}(U_n|Z_i) = \binom{n}{m}^{-1} \sum_{(j_1,...,j_m)\in C_m^n} E_{\mathbf{P}}(w(Z_{j_1},...,Z_{j_m})|Z_i)$$

with the notation $E_{\mathbf{P}}[w(Z_{j_1},...,Z_{j_m})|Z_i] = g(Z_i)$ if $i \in \{j_1,...,j_m\}$ we obtain (note that $E_{\mathbf{P}}[w(Z_{j_1},...,Z_{j_m})|Z_i] = \theta$ if $i \notin \{j_1,...,j_m\}$)

$$
\begin{aligned}
&= \binom{n}{m}^{-1}\left[\binom{n-1}{m-1}g(Z_i)+\binom{n-1}{m}\theta\right]\\
&= \frac{m}{n}g(Z_i)+\frac{n-m}{n}\theta.
\end{aligned}
$$

Hence, we can rewrite the projection (A.6) as

$$\hat{U}_n = \frac{m}{n}\sum_{i=1}^{n}(g(Z_i)-\theta)+\theta.$$

It follows for the centered projection

$$\hat{U}_n - \theta = \frac{m}{n} \sum_{i=1}^{n} \tilde{g}(Z_i),$$

where $\tilde{g}(Z_i) = g(Z_i) - \theta$. Since $\hat{U}_n - \theta$ is a sum of i.i.d. random vectors with mean zero, the law of large numbers and the central limit theorem imply

$$\hat{U}_n \xrightarrow{p} \theta, \quad \sqrt{n}\left(\hat{U}_n - \theta\right) \xrightarrow{\mathcal{L}} N(0, m^2\Sigma_g), \tag{A.7}$$

where Σ_g is the covariance matrix of $\tilde{g}(Z_1)$. Since $\tilde{g}(Z_1)$ has mean zero Σ_g is given by

$$\Sigma_g = E_{\mathbf{P}}\left[g(Z_1)g^T(Z_1)\right].$$

One can show (cf Kowalski and Tu (2008)) that under Assumption A.5 the following lemma holds.

Lemma A.6.

$$Var_{\mathbf{P}}(U_n) = \frac{m^2}{n} Var_{\mathbf{P}}(g(Z_1)) + O(n^{-2}),$$

$$Var_{\mathbf{P}}(\hat{U}_n) = \frac{m^2}{n} Var(g(Z_1)),\ Cov_{\mathbf{P}}(U_n, \hat{U}_n) = \frac{m^2}{n}\Sigma_g.$$

We are now able to proof the following theorem which states the consistency and asymptotic normality of U-statistics.

Theorem A.7. *(cf. Kowalski and Tu (2008)). Given Assumption A.5 we have*

$$U_n \xrightarrow{p} \theta, \quad \sqrt{n}\left(U_n - \theta\right) \xrightarrow{\mathcal{L}} N(0, m^2\Sigma_g). \tag{A.8}$$

Proof. We write

$$\sqrt{n}(U_n - \theta) = \sqrt{n}(\hat{U}_n - \theta) + \sqrt{n}(U_n - \hat{U}_n) = \sqrt{n}(\hat{U}_n - \theta) + e_n,$$

where $e_n = \sqrt{n}(U_n - \hat{U}_n) = \sqrt{n}((U_n - \theta) - (\hat{U}_n - \theta))$. The statement of the theorem follows from (A.7) and Slutzky´s lemma when we show that $e_n \xrightarrow{p} 0$. We show this by showing $E_{\mathbf{P}}(e_n e_n^T) \xrightarrow{p} 0$. We have

$$E_{\mathbf{P}}(e_n e_n^T) = nVar_{\mathbf{P}}(U_n) - 2nCov_{\mathbf{P}}(U_n, \hat{U}_n) + nVar_{\mathbf{P}}(U_n)$$

which equals according to Lemma A.6

$$\begin{aligned}
&= m^2 Var_{\mathbf{P}}(g(Z_1)) - 2m^2\Sigma_g + m^2 Var_{\mathbf{P}}(g(Z_1)) + O(n^{-2}) \\
&= m^2\Sigma_g - 2m^2\Sigma_g + m^2\Sigma_g + O(n^{-2}) \xrightarrow{p} 0.
\end{aligned}$$

□

Since θ is usually unknown the covariance matrix of the limiting distribution in (A.8) is unknown too. Hence, in many applications (like those in the impact analysis derived in Section 2.2) we have to estimate it. To this end we consider

$$\begin{aligned}
\Sigma_g =& E_{\mathbf{P}}\left[(g(Z_1)-\theta)\,(g(Z_1)-\theta)^T\right] \\
=& E_{\mathbf{P}}\left[E_{\mathbf{P}}\left\{w(Z_1,...,Z_m)|Z_1\right\} E_{\mathbf{P}}\left\{w^T(Z_1,...,Z_m)|Z_1\right\}\right] - \theta\theta^T \\
=& E_{\mathbf{P}}\left[E_{\mathbf{P}}\left\{w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})|Z_1\right\}\right] - \theta\theta^T \\
=& E_{\mathbf{P}}\left[w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})\right] - \theta\theta^T.
\end{aligned}$$

This means that we have to estimate $E_{\mathbf{P}}\left[w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})\right]$ and θ in order to estimate Σ_g. θ can simply be consistently estimated by the U-statistic defined in (A.5). To estimate $E_{\mathbf{P}}\left[w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})\right]$ we construct an other multivariate U-statistic. To this end, let

$$f(Z_1,...,Z_{2m-1}) = w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})$$

and $\tilde{f}(Z_1,...,Z_{2m-1})$ a symmetric version of $f(Z_1,...,Z_{2m-1})$ for example

$$\tilde{f}(Z_1,...,Z_{2m-1}) = \frac{1}{(2m-1)!}\sum_{\pi\in S(\{1,...,2m-1\})} f(Z_{\pi(1)},...,Z_{\pi(2m-1)}).$$

Then according to Theorem A.7 the U-statistic

$$\binom{n}{2m-1}^{-1}\sum_{(j_1,...,j_{2m-1})\in C^n_{2m-1}} \tilde{f}(Z_1,...,Z_{2m-1})$$

is a consistent estimator for $E_{\mathbf{P}}\left[w(Z_1,...,Z_m)w^T(Z_1,Z_{m+1},...,Z_{2m-1})\right]$. Hence we can estimate Σ_g consistently by

$$\hat{\Sigma}_g = \binom{n}{2m-1}^{-1}\sum_{(j_1,...,j_{2m-1})\in C^n_{2m-1}} \tilde{f}(Z_1,...,Z_{2m-1}) - U_n U_n^T.$$

In the course of Section 2.2 we will derive estimators for the impact (which is introduced in Section 1) that are functions of multivariate U-statistics. Then by application of the delta method the asymptotic normality of the derived estimators follows from the asymptotic normality of the U-statistics.

Note that the function w in (A.5) is not allowed to depend on n. Powell et al. (1989) generalize the results above for the case $m = 2$ to functions w_n that do depend on n. They define for an i.i.d. random sample $Z_1, Z_2, ..., Z_n$ a *general second order U-statistic* by

$$U_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_n(Z_i, Z_j),$$

where w_n is a k-dimensional symmetric function. With the additional definitions

$$r_n(Z_i) = E_{\mathbf{P}}[w_n(Z_i, Z_j)|Z_i], \theta_n = E_{\mathbf{P}}[r_n(Z_i)] = E_{\mathbf{P}}[w_n(Z_i, Z_j)]$$

and

$$\hat{U}_n = \theta_n + \frac{2}{n} \sum_{i=1}^{n} [r_n(Z_i) - \theta_n],$$

where it is assumed that θ_n exists they show the following lemma.

Lemma A.8. *If* $E_{\mathbf{P}}[\|w_n(Z_i, Z_j)\|^2] = o(n)$, *then* $\sqrt{n}(U_n - \hat{U}_n) = o_p(1)$.

Note that this does not necessarily imply the normality of U_n, since further assumptions must be fulfilled for the central limit theorem to be applicable to $\hat{U}_n$.

A.3. The Bootstrap

This section, which is based on Davison and Hinkley (2009) gives a brief introduction to the bootstrap and explains the methods used in the course of this thesis. The name bootstrap goes back to Efron (1979). A lot of research was performed on this field and bootstrap methods were applied to various statistical problems.

A.3.1. The idea of the bootstrap

The setup for this section is the following one: Assume we have data $z_1, ..., z_n$ which are realizations of i.i.d. random variables $Z_1, ..., Z_n$. Let F be the distribution function of Z_i, while f denotes its density. We are interested in a (scalar) parameter θ which is estimated by the statistic T which has the realization t. In order to be able to construct confidence intervals for θ we have to know the distribution of T. Usually one distinguishes between two cases:

- Parametric, i.e. we have a model with parameters ψ that uniquely determine $f = f_\psi$ and $F = F_\psi$. θ is then a component or function of ψ;
- Nonparametric, i.e. we do not have a model.

In the applications of bootstrap methods in this thesis we only use nonparametric bootstrap methods. The parametric bootstrap is used once, in order to introduce the adjusted percentile method for constructing confidence intervals in the following section. The idea of the bootstrap is basically to replace the unknown distribution function F in the quantities of interest by an estimate. When using the parametric bootstrap one estimates ψ by $\hat{\psi}$ (for example by maximum likelihood estimation) and replaces the true distribution function F_ψ by the estimated distribution function $F_{\hat{\psi}}$. In the non-parametric case the empirical distribution function $\hat{F}$ is used instead.

We explain the idea of the nonparametric bootstrap by a simple example. Assume $\theta = t(F)$ and we are interested in the bias β and the variance v of T:

$$\beta = b(F) = E_{\mathbf{P}}(T|F) - t(F), \quad v = v(F) = Var_{\mathbf{P}}(T|F). \tag{A.9}$$

Since F is unknown we are not able to make any statements about β or v. The idea of the bootstrap is to replace the unknown F by its estimator, the empirical distribution function $\hat{F}$. Note that we have

$$\hat{F}(u) = \frac{\#\{z_i \leq u\}}{n} = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{z_i \leq u\}}.$$

Hence when replacing F by $\hat{F}$ in (A.9) we obtain

$$B = b(\hat{F}) = E_{\mathbf{P}}(T|\hat{F}) - t(\hat{F}), \quad V = v(\hat{F}) = Var_{\mathbf{P}}(T|\hat{F}).$$

B and V are called bootstrap estimates of β and v. Often it is very hard or even impossible to determine these bootstrap estimators. In these cases it helps to use Monte Carlo simulations. These are performed following these steps:

1. Draw $Z_1^*, ..., Z_n^*$ independently from $\hat{F}$;
2. Compute from this data the statistic T and name it T^*;
3. Repeat 1. and 2. R-times to obtain $T_1^*, ..., T_R^*$;
4. Regard
$$B = b(\hat{F}) = E_{\mathbf{P}}(T|F) - t = E^*(T^*) - t$$
and estimate it by
$$B_R = R^{-1} \sum_{r=1}^{R} T_r^* - t = \bar{T}^* - t; \tag{A.10}$$
5. Analogously obtain
$$V_R = \frac{1}{R-1} \sum_{r=1}^{R} (T_R^* - \bar{T}^*)^2. \tag{A.11}$$

Here and in the following the superscript "*" denotes the distribution respectively expectation according to $\hat{F}$. From this procedure it gets clear why the bootstrap is also called a resampling method. We estimate the bias and variance by resampling independently from the data. For large R we expect according to the laws of large numbers B_R to be near the true bootstrap estimator B.

A.3.2. Bootstrap confidence intervals

All confidence intervals which are presented here are meant to be computed by non-parametric bootstrap. Only in the introduction of the bias corrected percentile method we assume a parametric model. For the computation of confidence intervals for θ we will need quantiles of the distribution of $T - \theta$. Since this distribution is unknown we approximate it by the distribution of $T^* - t$. Hence we estimate the distribution function G of $T - \theta$ by

$$\hat{G}_R(u) = \frac{\#\{t_r^* - t \leq u\}}{R} = R^{-1} \sum_{r=1}^{R} \mathbb{1}_{\{t_r^* - t \leq u\}}.$$

Note that there are two sources of error: One comes from variability of the data (we only have n observations) and the other arises from finite simulation. In order to keep the error due to finite simulation small we have to choose R "large". Usually one sets $R \geq 1000$. Nevertheless for the choice of R it should be considered that large R may require large amounts of computational time. This issue has become smaller in the past years due to the improvement of computers.

Since we approximate G by $\hat{G}_R$ we approximate the p-Quantile $q(p)$ of G by the p-quantile $\hat{q}(p)$ of $\hat{G}_R$, which is given by $\hat{q}(p) = t^*_{(p)} - t$, where $t^*_{(p)}$ is the p-quantile of the bootstrapped values $t^*_1, ..., t^*_R$. There are several methods for computing bootstrap confidence intervals for θ. We will give five of them here.

Basic bootstrap confidence intervals

We have for given $\alpha \in (0, 1)$ that

$$\begin{aligned} &\mathbb{P}(q(\alpha) \leq T - \theta \leq q(1-\alpha)) = 1 - 2\alpha \\ \Rightarrow &\mathbb{P}(T - q(1-\alpha) \leq \theta \leq T - q(\alpha)) = 1 - 2\alpha. \end{aligned} \tag{A.12}$$

Hence, by replacing the quantiles by their estimates and T by its realization t we obtain the following approximate $1 - 2\alpha$ confidence interval for θ:

$$\begin{aligned} CI_{basic} &= (t - \hat{q}(1-\alpha), t - \hat{q}(\alpha)) \\ &= \left(2t - t^*_{(1-\alpha)}, 2t - t^*_{(\alpha)}\right). \end{aligned}$$

This confidence interval is called *basic bootstrap confidence interval*. As explained before the accuracy of this interval depends on R, so we would choose R large to make the interval more accurate.

Studentized intervals

We now consider a "studentized" version of $T - \theta$ namely

$$Y = \frac{T - \theta}{V^{1/2}}, \tag{A.13}$$

where V is an estimator of $Var(T|F)$. One way to find such an estimator for $Var(T|F)$ is the so called nonparametric delta method or functional delta method. We introduce the nonparametric delta method for functionals $t(\cdot)$ which are Fréchet-differentiable. Functional delta method theorems for more general cases can be found in Serfling (1980)

and van der Vaart (2000, ch. 20). Let $t(\cdot)$ be Fréchet-differentiable at F, this means that for all distribution functions G exists a linear functional $L(F-G)$ such that

$$t(G) = t(F) + L(G-F) + o(\|G-F\|), \quad \text{as } \|G-F\| \to 0, \tag{A.14}$$

where $\|\cdot\|$ is a norm on the linear space generated by differences of distribution functions. An example for such a norm is $\|G-F\| = \sup_{x\in\mathbb{R}} |G(x)-F(x)|$. $L(G-F)$ is called *Fréchet-derivative* of t at F in the direction $G-F$. Huber (1981, p. 37) shows that if $t(\cdot)$ is weakly continuous in a neighborhood of F the function $L(G-F)$ can be represented as

$$L(G-F) = \int L_t(z,F)dG(z),$$

where

$$L_t(z;F) = \lim_{\epsilon\to 0} \frac{t\{(1-\epsilon)F + \epsilon H_z\} - t(F)}{\epsilon} = \left.\frac{\partial t\{(1-\epsilon)F+\epsilon H_z\}}{\partial \epsilon}\right|_{\epsilon=0},$$

with $H_y(u) = \mathbb{1}_{\{u\geq z\}}$ is called the *influence function* of T. According to van der Vaart (2000, p.292) the name "influence function" originated in developing robust statistics. "The function measures the change in the value $t(F)$ if an infinitesimally small part of F is replaced by a pointmass at x". One can see by setting $G=F$ in the approximation (A.14), that

$$\int L_t(x,F)dF(x) = 0.$$

Choosing $\hat{F}$ for G in (A.14) gives

$$t(\hat{F}) = t(F) + \int L_t(z;F)d\hat{F}(z) + o(\|\hat{F}-F\|) \approx t(F) + \frac{1}{n}\sum_{j=1}^{n} L_t(z_j;F). \tag{A.15}$$

Application of the central limit theorem to the sum in (A.15) gives

$$T - \theta \dot{\sim} N(0, v_L(F)),$$

where since we have $\int L_t(z;F)dF(z) = 0$

$$v_L(F) = n^{-1}Var(L_t(Z)) = n^{-1}\int L_t^2(z)dF(z).$$

Since F is unknown we replace F by $\hat{F}$ and obtain the *nonparametric delta method variance estimate*

$$\hat{v}_l = n^{-2} \sum_{j=1}^{n} l_j^2,$$

where

$$l_j = L_t(z_j; \hat{F}) \tag{A.16}$$

are called the *empirical influence values*. The R function `empinf` of the library `boot` delivers several methods for the computation of the empirical influence values.

The idea behind using the studentized statistic (A.13) is to mimic the Student-t statistic which has this form and eliminates the unknown standard deviation when making inference about a normal distribution mean. Recall the Student-t $(1 - 2\alpha)$ confidence interval for a normal distribution mean which is given by

$$\left(\bar{z} - v^{1/2} t_{n-1}(1-\alpha), \bar{z} - v^{1/2} t_{n-1}(\alpha) \right),$$

where v is an estimator for the variance and $t_{n-1}(p)$ is the p-quantile of a central t-distribution with $n - 1$ degrees of freedom. Confidence intervals for θ based on (A.13) have the analogue form

$$\left(t - v^{1/2} y_{(1-\alpha)}, t - v^{1/2} y_{(\alpha)} \right),$$

where $y_{(p)}$ is the p-quantile of the distribution of Y. We estimate the quantiles of Y by the empirical quantiles of repetitions of the studentized bootstrap statistic

$$Y^* = \frac{T^* - t}{V^{*1/2}},$$

where T^* and V^* are based on a simulated random sample $Z_1^*, ..., Z_n^*$. Let $\hat{y}_{(p)}$ denote the p-quantile of $(Y_1^*, ..., Y_R^*)$ then an approximate level $(1 - 2\alpha)$ confidence interval for θ is given by

$$CI_{stud} = \left(t - v^{1/2} \hat{y}_{(1-\alpha)}, t + v^{1/2} \hat{y}_{(\alpha)} \right).$$

This confidence interval is called *studentized bootstrap confidence interval*.

Bootstrap normal confidence interval

Another way to construct a confidence interval for θ is to assume the normal approximation

$$T - \theta \overset{approx}{\sim} N(\beta, v),$$

where β and v are the bias and the variance of T. Since we then have

$$(T - \beta - \theta)/v^{1/2} \overset{approx}{\sim} N(0,1)$$

an approximate level $(1-2\alpha)$ confidence interval is given by

$$CI_{norm} = (t - B_R \mp V_R^{1/2} z_{1-\alpha}),$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution and B_R and V_R are the bootstrap estimates for bias and variance defined in (A.10) and (A.11). This confidence interval is referred to as *bootstrap normal confidence interval*.

Percentile interval

For the percentile method we assume that there exists a transformation $U = h(T)$ which has a symmetric distribution. We want to construct a confidence interval for $\phi = h(\theta)$ by applying the basic bootstrap confidence interval method. The equation (A.12) becomes

$$\mathbb{P}(U - q(1-\alpha) \leq \theta \leq U - q(\alpha)) = 1 - 2\alpha,$$

where now $q(\alpha)$ is the α quantile of the distribution of $U - \phi$. Because of the assumed symmetry we can replace $q(\alpha)$ by $-q(1-\alpha)$ and $q(1-\alpha)$ by $-q(\alpha)$ and obtain

$$\mathbb{P}(U + q(\alpha) \leq \theta \leq U + q(1-\alpha)) = 1 - 2\alpha.$$

Replacing the quantiles $q(\alpha)$ and $q(1-\alpha)$ by their estimates $\hat{q}(\alpha) = u^*(\alpha) - u$ and $\hat{q}(1-\alpha) = u^*(1-\alpha) - u$ (where $u^*(p)$ is the p-quantile of the bootstrapped values $u_1^*, ..., u_R^*$) gives the interval

$$CI_\phi = \left(u^*_{(\alpha)}, u^*_{(1-\alpha)}\right)$$

for ϕ. Transformation back to the θ scale gives us the *bootstrap percentile intervall*

$$CI_{perc} = \left(t^*_{(\alpha)}, t^*_{(1-\alpha)}\right).$$

According to Davison and Hinkley (2009, p.203) this "method turns out to not work very well with the nonparametric bootstrap even when a suitable transformation h does exist."

Adjusted percentile interval

Since the percentile method does not work very well we need improvements of this method. One such improvement is the so called *adjusted percentile method*. This method can be explained simplest by transformation theory in the parametric framework with no nuisance parameters and then be extended to the non-parametric case. Hence, for the beginning we assume that the data are described by a parametric model with the single unknown parameter θ. We estimate θ by its maximum likelihood estimate $t = \hat{\theta}$ and make the assumption that there exists a monotone increasing transformation h as well as an unknown bias correction factor w and an unknown skewness correction factor a such that we have for $h(T) = U$ and $h(\theta) = \phi$

$$U \sim N(\phi - w\sigma(\phi), \sigma^2(\phi)), \quad \text{where } \sigma(\phi) = 1 + a\phi. \tag{A.17}$$

We will derive confidence limits for ϕ and transform them to the θ scale with the help of the bootstrap distribution of T (note that we use parametric bootstrapping here). Assuming that a and w are known we obtain

$$U = \phi + (1 + a\phi)(Z - w),$$

where $Z \sim N(0, 1)$. Efron (1987, ch. 3) discusses that via suitable transformation one can obtain the level α confidence limit for ϕ as

$$\hat{\phi}_\alpha = u + \sigma(u)\frac{w + z_\alpha}{1 - a(w + z_\alpha)}.$$

The confidence limit for θ is then given as $\hat{\theta}_\alpha = h^{-1}(\hat{\phi}_\alpha)$. Since we do not know h we cannot compute this bound. We can overcome this lack of knowledge by using the distribution function of the (parametric) bootstrap replications T^* which we denote by $\hat{G}$. We then have

$$\begin{aligned}\hat{G}(\hat{\theta}_\alpha) &= \mathbb{P}^*(T^* < \hat{\theta}_\alpha | t) = \mathbb{P}^*(U^* < \hat{\phi}_\alpha | u) \\ &= \Phi\left(\frac{\hat{\phi}_\alpha - u}{\sigma(u)} + w\right) = \Phi\left(w + \frac{w + z_\alpha}{1 - a(w + z_\alpha)}\right),\end{aligned}$$

hence,

$$\hat{\theta}_\alpha = \hat{G}^{-1}\left(\Phi\left(w + \frac{w + z_\alpha}{1 - a(w + z_\alpha)}\right)\right).$$

Thus, a level α confidence limit for θ is given by

$$\hat{\theta}_\alpha = t^*_{(\tilde{\alpha})} \quad \text{with} \quad \tilde{\alpha} = \Phi\left(w + \frac{w + z_\alpha}{1 - a(w + z_\alpha)}\right).$$

We see that we do not need to know the underlying transformation h for the computation of the confidence bound $\hat{\theta}_\alpha$. Nevertheless, the constants w and a remain unknown and need to be estimated. To this end we make use of the normal distribution of U and write

$$\mathbb{P}^*(T^* < t|t) = \mathbb{P}^*(U^* < u|u) = \mathbb{P}(U < \phi|\phi) = \Phi(w),$$

which implies

$$w = \Phi^{-1}(\hat{G}(t)).$$

Hence we estimate w by

$$\hat{w} = \Phi^{-1}\left(\frac{\#\{t^*_r \leq t\}}{R}\right). \tag{A.18}$$

For the estimation of a we denote the log-likelihood which is given by the transformation (A.17) by $l(\phi)$. One can show that a good transformation for a (ignoring terms of order n^{-1}) is

$$a = \frac{1}{6}\frac{E_{\mathbf{P}}\left\{l'(\phi)^3\right\}}{Var_{\mathbf{P}}\left\{l'(\phi)\right\}^{3/2}},$$

which, transformed back to the θ scale (again ignoring terms of order n^{-1}), gives

$$a = \frac{1}{6}\frac{E_{\mathbf{P}}\left\{l'(\theta)^3\right\}}{Var_{\mathbf{P}}\left\{l'(\theta)\right\}^{3/2}}.$$

Hence we can estimate a by

$$\hat{a} = \frac{1}{6}\frac{E^*\left\{l^{*\prime}(\hat{\theta})^3\right\}}{Var\left\{l^{*\prime}(\hat{\theta})\right\}^{3/2}}, \tag{A.19}$$

where l^* is the log likelihood of a set of data simulated form the fitted model. This implies that an approximate level α confidence limit for θ is given by

$$\hat{\theta}_\alpha = t^*_{(\tilde{\alpha})} \quad \text{with} \quad \tilde{\alpha} = \Phi\left(\hat{w} + \frac{\hat{w} + z_\alpha}{1 - \hat{a}(\hat{w} + z_\alpha)}\right), \tag{A.20}$$

with $\hat{w}$ from (A.18) and $\hat{a}$ from (A.19). The limit $\hat{\theta}_\alpha$ is commonly referred to as *bootstrap BC_a limit*.

(A.20) gives the BC_a limit only for the parametric case without nuisance parameters. This method can be extended to the nonparametric case as well which is done by applying the method for the parametric case to a specially constructed exponential tilted distribution. This approach does not affect the form of $\hat{\theta}_\alpha$ but only the way we estimate a. The estimator for a in the nonparametric case is given by

$$\hat{a} = \frac{1}{6} \frac{\sum_{j=1}^n l_j^3}{\{\sum_{j=1}^n l_j^2\}^{3/2}},$$

where l_j is the empirical influence value of t at z_j derived in (A.16). More details are given in Davison and Hinkley (2009).

A.3.3. Second order accuracy and the smooth function model

A desirable property of some bootstrap confidence intervals is that they are second order accurate in many cases. We call a level α lower bound u_n for θ *first order accurate*, if

$$\mathbb{P}(\theta \leq u_n) = \alpha + O(n^{-1/2}).$$

u_n is said to be *second order accurate* if

$$\mathbb{P}(\theta \leq u_n) = \alpha + O(n^{-1}).$$

Confidence intervals derived via limiting normal distributions are often only first order accurate. Hall (cf. Hall (1988), Hall (1992)) gives a setup, the so called *smooth function model*, in which bootstrap BC_a intervals and studentized bootstrap intervals are second order accurate. The smooth function model is given as follows: Assume we have i.i.d. d-vectors $X_1, ..., X_n$ with $E(X_i) = \mu$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Assume further, that our parameter of interest is $\theta = f(\mu)$ for a smooth real-valued function f. Let $\hat{\theta} = f(\bar{X})$ be the estimator of θ with asymptotic variance $n^{-1}\sigma^2$, where $\sigma^2 = g(\mu)$ for a real-valued smooth function g. (Hall shows without using the smoothness of g that σ^2 is a smooth function of μ, which means that it is not necessary to demand the smoothness of g.) Hall then shows that the BC_a critical points and the studentized bootstrap critical points are second order accurate.

A.3.4. Bootstrapping U-statistics

In the course of this thesis we will apply bootstrap methods to U-statistics. A theoretical justification for this is given by Bickel and Freedman (1981). Let $X_1, ..., X_n$ be an i.i.d. sample of d-vectors with distribution function F and empirical distribution function F_n. Define

$$g(F) = \int \int w(x,y) dF(x) dF(y)$$

respectively

$$g(F_n) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} w(X_i, X_j),$$

where w is a symmetric function. g is called a von Mises statistic (cf. von Mises (1947)) and is closely related to U-statistics (see (A.5)). It follows similar to the case of U-statistics that if

$$\int w^2(x,y) dF(x) dF(y) < \infty \tag{A.21}$$

and

$$\int w^2(x,x) dF(x) < \infty$$

we obtain

$$\sqrt{n}\{g(F_n) - g(F)\} \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

where σ^2 is given by

$$\sigma^2 = 4 \left[\int \left\{ \int w(x,y) dF(y) \right\}^2 dF(x) - g^2(F) \right].$$

Bickel and Freedman (1981) show that under the same conditions, for almost all $X_1, X_2, ...$, given $(X_1, ..., X_n)$,

$$\sqrt{n}\{g(G_n) - g(F_n)\} \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where G_n is the empirical distribution function of $X_1^*, ..., X_n^*$. Note, that the condition (A.21) is necessary for the bootstrap to work. Bickel and Freedman (1981) give a counterexample to show the inconsistency of the bootstrap, when (A.21) does not hold. Nevertheless, the considerations above show that under suitable conditions the bootstrap is valid for second order von Mises statistics. Bickel and Freedman (1981) argue that under the respective conditions on w analogous results also hold for von Mises statistics and U-statistics of arbitrary order. Hence the use of the bootstrap when handling U-statistics is justified. However, these results do not offer second order accuracy.

A.3.5. Wild-bootstrap

In this section we introduce the method of the wild-bootstrap, which is used for the computation of confidence intervals for the kernel smoother based impact of Section 2.2.1. Let $\hat{m}(x)$ be an estimator for $E_{\mathbf{P}}(Y|X = x)$, e.g. a kernel-smoothing fit. The wild-bootstrap is a method where we do not sample form the data pairs (X_i, Y_i) but from the residuals $\epsilon_i = Y_i - \hat{m}(X_i)$. The method was introduced by Wu (1968). In this thesis we use the modification of Härdle and Marron (1991). They draw the bootstrap residual ϵ_i^* from the distribution

$$\epsilon_i^* = \begin{cases} \hat{\epsilon}_i(1-\sqrt{5})/2 & \text{with probability } (5+\sqrt{5})/10, \\ \hat{\epsilon}_i(1+\sqrt{5})/2 & \text{with probability } 1-(5+\sqrt{5})/10 \end{cases}. \tag{A.22}$$

Using this distribution they obtain that $E_{\mathbf{P}}(\epsilon_i^*) = 0$, $E_{\mathbf{P}}(\epsilon_i^{*^2}) = \hat{\epsilon}_i^2$ and $E_{\mathbf{P}}(\epsilon_i^{*^3}) = \hat{\epsilon}_i^3$, which means that the first three moments of ϵ_i^* coincide with those of $\hat{\epsilon}_i$. Thus, Härdle and Marron (1991) note that "In a certain sense the resampling distribution [...] can be thought of as attempting to reconstruct the distribution of each residual through the use of one single observation." Using this resampling distribution Härdle and Marron (1991) construct R bootstrap repetitions $Y_i^* = \hat{m}(X_i) + \epsilon_i^*$ for $i = 1, ..., n$ and calculate bootstrap confidence bands for kernel smoothers on their basis.
We will use the wild bootstrap approach to perform a test for the null hypothesis H_0 : $\iota_X^{ks}(Y) = 0$. The procedure is as follows:

1. Perform a kernel smoother fit to the data, obtaining residuals $\epsilon_i = Y_i - \hat{m}(X_i)$,
2. Generate R sets of bootstrap residuals according to (A.22),
3. Compute $\hat{\iota}_X^{ks}(Y)$ in each of the R data sets $(\hat{\epsilon}_i^*, X_i)_{i=1,...,n}$ obtaining $\hat{\iota}_X^{ks}(Y)_r$ for $r = 1, ..., R$,
4. Calculate the wild bootstrap p-value for H_0 as:

$$p = \#\{\hat{\iota}_X^{ks,*}(Y)_r \geq \hat{\iota}_X^{ks}(Y)\}/R,$$

   where $\hat{\iota}_X^{ks}(Y)$ is calculated using the original data,
5. Reject H_0 if p falls below the level of significance.

B. Theorems and Proofs

Lemma 1.13.
Let $(X_1, ..., X_k) \in L^2_{\mathbf{P}}$ and $(X_{i1}, ..., X_{ik})_{i=1,...,n}$ i.i.d. observations of $(X_1, ..., X_k)$. We have that

$$\hat{\xi} \xrightarrow{p} \xi,$$

where $\hat{\xi}$ is the vector of coefficients from the projection of $\mathbf{X}_1 = (X_{11}, ..., X_{n1})$ onto $\text{span}(\mathbf{1}, \mathbf{X}_2, ..., \mathbf{X}_k) \subseteq \mathbb{R}^n$ with $\mathbf{1} = (1, ..., 1) \in \mathbb{R}^n$ and ξ are the coefficients form the corresponding projection of X_1 onto $\text{span}(1, X_2, ..., X_k)$ in $L^2_{\mathbf{P}}$.

Proof. Let $D_n = (\mathbf{1}, \mathbf{X}_2, ..., \mathbf{X}_k)$. Since we have i.i.d. observations

$$\begin{aligned} \frac{1}{n} D_n^T D_n &= \frac{1}{n} \begin{pmatrix} n & ... & \sum_{i=1}^n X_{ik} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} & ... & \sum_{i=1}^n X_{ik}^2 \end{pmatrix} \\ &\xrightarrow{p} \begin{pmatrix} 1 & ... & E_{\mathbf{P}}(X_k) \\ \vdots & \ddots & \vdots \\ E_{\mathbf{P}}(X_k) & ... & E_{\mathbf{P}}(X_k^2) \end{pmatrix} =: C \end{aligned} \tag{B.1}$$

where C is obviously symmetric and positive definite.
In order to simplify the following exposition let $W_1 = 1$ in $L^2_{\mathbf{P}}$ and $W_j = X_j$ in $L^2_{\mathbf{P}}$, $j = 2, ..., k$. We know that ξ minimizes the expression

$$\begin{aligned} E_{\mathbf{P}}\{(X_1 - \sum_{j=1}^k \xi_j W_j)^2\} &= E_{\mathbf{P}}(X_1^2) - 2E_{\mathbf{P}}(\sum_{j=1}^k \xi_j W_j X_1) + \sum_{j=1}^k \sum_{l=1}^k \xi_j \xi_l E_{\mathbf{P}}(W_j W_l) \\ &= E_{\mathbf{P}}(Y^2) - 2\xi^T A + \xi C \xi^T \end{aligned} \tag{B.2}$$

where $A = (E_{\mathbf{P}}(W_1 X_1), ..., E_{\mathbf{P}}(W_k X_1))^T$ and $C = (E_{\mathbf{P}}(W_i W_j))_{ij}$ is the same symmetric matrix as in (B.1). Since C is positive definite, the minimum of the quadratic form (B.2) is the root of its derivative.

$$\frac{\partial}{\partial \xi}[E_{\mathbf{P}}(Y^2) - 2\xi^T A + \xi C \xi^T] = 2(-A + C\xi).$$

Setting the derivative to zero leads to $\xi = C^{-1}A$. As a next step we consider

$$\hat{\xi} = (D_n^T D_n)^{-1} D_n^T \mathbf{X}_1 = n(D_n^T D_n)^{-1} \frac{1}{n} D_n^T \mathbf{X}_1.$$

Since $n(D_n^T D_n)^{-1} \xrightarrow{p} C^{-1}$ by (B.1) and $\frac{1}{n} D_n^T \mathbf{X}_1 \xrightarrow{p} A$ by the law of large numbers we obtain

$$\hat{\xi} \xrightarrow{p} C^{-1} A = \xi.$$

□

Theorem 2.13.
Under Assumption 2.12 we have that

$$\sqrt{n}(\iota_X^{loess}(Y) - \iota_X^{loess}(Y)) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$,

$$F\left((a_1, ..., a_4)^T\right) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}} \quad V = 9E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right),$$

as well as $\iota_X^{loess}(Y) = F(\vartheta)$ and $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l, Z_k, Z_m)|Z_i) - \vartheta$.

Proof. We have

$$\begin{aligned}\tilde{\iota}_1 =& \frac{1}{n}\sum_{i=1}^{n} Y_i\hat{\delta}(X_i) = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_lY_i \\ =& \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} g_1(Z_i, Z_j, Z_l, Z_k, Z_m).\end{aligned}$$

Furthermore,

$$\begin{aligned}\tilde{\iota}_2 =& \frac{1}{n}\sum_{i=1}^{n} Y_i\bar{\delta} = \frac{1}{n^4}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j - X_k)(X_j - X_l)K_h(X_k - X_j)K_h(X_k - X_l)Y_lY_i \\ =& \frac{1}{n^4}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} g_2(Z_i, Z_j, Z_l, Z_k, Z_m).\end{aligned}$$

Analogous to this we obtain

$$\begin{aligned}\tilde{\iota}_3 =& \frac{1}{n}\sum_{i=1}^{n}\delta(X_i)^2 \\ =& \frac{1}{n^5}\sum_{i,j,l,k,m}\{(X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l \\ & (X_k - X_i)(X_k - X_m)K_h(X_i - X_k)K_h(X_i - X_m)Y_m\} \\ =& \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} g_3(Z_i, Z_j, Z_l, Z_k, Z_m)\end{aligned}$$

and

$$\begin{aligned}\tilde{\iota}_4 =& \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}(X_j - X_i)(X_j - X_l)K_h(X_i - X_j)K_h(X_i - X_l)Y_l \\ =& \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} g_4(Z_i, Z_j, Z_l, Z_k, Z_m).\end{aligned}$$

Hence, we obtain that

$$\tilde{\iota} = \frac{1}{n^5}\sum_{i,j,l,k,m} g(Z_i, Z_j, Z_l, Z_k, Z_m) = \frac{1}{n^5}\sum_{i,j,l,k,m} w(Z_i, Z_j, Z_l, Z_k, Z_m).$$

Analogous to the case of $\hat{\iota}_X^{ks}(Y)$ Lemma 2.1 gives that under Assumption 2.12, we have

$$\sqrt{n}\tilde{\iota} = \frac{\sqrt{n}}{n^5}\sum_{C(\{i,j,l,k,m\})} w(Z_i, Z_j, Z_l, Z_k, Z_m) + o_p(1),$$

where $C(\{i,j,l,k,m\})$ is the set of all combinations which can be drawn without replacement from $\{1,...,n\}$ in five draws. It follows that

$$\begin{aligned}\sqrt{n}\tilde{\iota} =& \frac{\sqrt{n}}{n^5}\sum_{C(\{i,j,l,k,m\})} w(Z_i, Z_j, Z_l, Z_k, Z_m) + o_p(1) \\ =& \frac{\sqrt{n}}{n^4}5!\sum_{i<j<l<k<m} w(Z_i, Z_j, Z_l, Z_k, Z_m) + o_p(1) \\ =& \frac{n(n-1)(n-2)(n-3)(n-4)}{n^5}\sqrt{n}\underbrace{\binom{n}{5}^{-1}\sum_{i<j<l<k<m} w(Z_i, Z_j, Z_l, Z_k, Z_m)}_{=:U_n} + o_p(1)\end{aligned}$$

where U_n is a fifth-order U-statistics. We now have

$$\begin{aligned}\sqrt{n}(\tilde{\iota} - \vartheta) =& \frac{n(n-1)(n-2)(n-3)(n-4)}{n^5}\sqrt{n}U_n + o_p(1) - \sqrt{n}\vartheta \\ =& \underbrace{c_n}_{\to 1}\underbrace{\sqrt{n}(U_n - \vartheta)}_{\xrightarrow{\mathcal{L}} N(0,V)} + o_p(1) + \underbrace{\left(c_n\sqrt{n} - \sqrt{n}\right)\vartheta}_{\to 0} \\ \xrightarrow{\mathcal{L}}& N(0, V),\end{aligned}$$

where $c_n \frac{n(n-1)(n-2)(n-3)(n-4)}{n^5}$ and

$$V = 25E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$$

with $\tilde{w}(Z_i) = E_{\mathbf{P}}(w(Z_i, Z_j, Z_l, Z_k, Z_m)|Z_i) - \vartheta$. Since the mapping

$$F\left((a_1, ..., a_4)^T\right) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}$$

is continuously differentiable, application of the delta-method with $\iota_X^{loess}(Y) = F(\vartheta)$ yields

$$\sqrt{n}(\hat{\iota}_X^{loess}(Y) - \iota_X^{loess}(Y)) = \sqrt{n}\left(F(\tilde{\iota}) - F(\vartheta)\right) \xrightarrow{\mathcal{L}} DF(\vartheta)^T N(0, V) = N(0, \sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T V DF(\vartheta)$.

□

Lemma 2.14. A consistent estimator for σ^2 is given by

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{V} DF(\tilde{\iota}),$$

where

$$\hat{V} = 25 \left(\binom{n}{9}^{-1} \sum_{i<\ldots<d} \frac{1}{9!} \sum_{\pi \in S(\{i,\ldots,d\})} \tilde{g}(Z_{\pi(i)}, ..., Z_{\pi(d)}) - \tilde{\iota}\tilde{\iota}^T \right)$$

and

$$g(Z_i, Z_j, Z_l, Z_k, Z_m, Z_a, Z_b, Z_c, Z_d) = w(Z_i, Z_j, Z_l, Z_k, Z_m) w^T(Z_i, Z_a, Z_b, Z_c, Z_d).$$

Proof. Since $\tilde{\iota}$ is consistent for ϑ, $DF(\vartheta)$ can be consistently estimated by $DF(\tilde{\iota})$. To find a consistent estimator for V we make the following considerations (which are similar to those in (Kowalski and Tu, 2008, p. 259) and to those in the kernel smoother case in Section 2.2.1).

$$\begin{aligned}
V/25 =& E_{\mathbf{P}} \left(\tilde{w}(Z_i) \tilde{w}^T(Z_i) \right) \\
=& E_{\mathbf{P}} \left(E_{\mathbf{P}} \left\{ w(Z_i, Z_j, Z_l, Z_k, Z_m) | Z_i \right\} E_{\mathbf{P}} \left\{ w^T(Z_i, Z_j, Z_l, Z_k, Z_m) | Z_i \right\} \right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}} \left(E_{\mathbf{P}} \left\{ w(Z_i, Z_j, Z_l, Z_k, Z_m) w^T(Z_i, Z_a, Z_b, Z_c, Z_d) | Z_i \right\} \right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}} \underbrace{\left(w(Z_i, Z_j, Z_l, Z_k, Z_m) w^T(Z_i, Z_a, Z_b, Z_c, Z_d) \right)}_{=: \tilde{g}(Z_i, Z_j, Z_l, Z_k, Z_m, Z_a, Z_b, Z_c, Z_d)} - \vartheta\vartheta^T.
\end{aligned}$$

$E_{\mathbf{P}} \left(\tilde{g}(Z_i, Z_j, Z_l, Z_k, Z_m, Z_a, Z_b, Z_c, Z_d) \right)$ can be consistently estimated by

$$\binom{n}{9}^{-1} \sum_{i<j<l<k<m<a<b<c<d} \tilde{\tilde{g}}(Z_i, Z_j, Z_l, Z_k, Z_m, Z_a, Z_b, Z_c, Z_d),$$

where $\tilde{\tilde{g}}(Z_i ..., Z_d)$ is a symmetric version of $\tilde{g}(Z_i, ..., Z_d)$ say

$$\tilde{\tilde{g}}(Z_i, ..., Z_d) = \frac{1}{9!} \sum_{\pi \in S(\{i,\ldots,d\})} \tilde{g}(Z_{\pi(i)}, ..., Z_{\pi(d)})$$

with $S\left(\{i, ..., d\}\right)$ being the set off all permutations of $\{i, ..., d\}$. Hence a consistent estimator for V is given by

$$\hat{V} = 25 \left(\binom{n}{9}^{-1} \sum_{i<\ldots<d} \frac{1}{9!} \sum_{\pi \in S(\{i,\ldots,d\})} \tilde{g}(Z_{\pi(i)}, ..., Z_{\pi(d)}) - \tilde{\iota}\tilde{\iota}^T \right)$$

which leads to

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{V} DF(\tilde{\iota})$$

as consistent estimator for σ. □

Theorem 2.16. Under Assumption 2.15 we have that

$$\sqrt{n}\left\{\hat{\iota}_X^{locpol}(Y) - \iota_X^{locpol}(Y)\right\} \xrightarrow{\mathcal{L}} N(0,\sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T \Sigma DF(\vartheta)$, $F\left((a_1,...,a_4)^T\right) = \frac{a_1-a_2}{\sqrt{a_3-a_4^2}}$, $\iota_X^{locpol}(Y) = F(\vartheta)$, and

$$\Sigma = (2k+3)^2 E_{\mathbf{P}}\left\{\left(E_{\mathbf{P}}\left[w(Z_{j_1},...,Z_{j_{2k+3}})|Z_{j_1}\right] - \vartheta\right)\left(E_{\mathbf{P}}\left[w^T(Z_{j_1},...,Z_{j_{2k+3}})|Z_{j_1}\right] - \vartheta^T\right)\right\}.$$

Proof. We have that

$$\begin{aligned}
&\operatorname{cof}\left\{n^{-1}\mathbf{B}^T\mathbf{W}(X_i)\mathbf{B}\right\}\frac{1}{n}\mathbf{B}^T\mathbf{W}(X_i)\mathbf{Y} \\
&= \begin{pmatrix} \sum_{m=1}^{k+1}\frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n} h_{1m}(Z_{j_1},...,Z_{j_k})K_h(X_{j_{k+1}}-X_i)Y_{j_{k+1}} \\ \vdots \\ \sum_{m=1}^{k+1}\frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n} h_{(k+1)m}(Z_{j_1},...,Z_{j_k})K_h(X_{j_{k+1}}-X_i)Y_{j_{k+1}}X_{j_{k+1}}^k \end{pmatrix}
\end{aligned}$$

and thereby

$$\begin{aligned}
\hat{\delta}(X_i) =& \sum_{l=1}^{k+1}\sum_{m=1}^{k+1}\frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n} h_{lm}(Z_{j_1},...,Z_{j_k})K_h(X_{j_{k+1}}-X_i)Y_{j_{k+1}}X_{j_{k+1}}^{m-1}X_i^{m-1} \\
=& \frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n}\underbrace{\sum_{l=1}^{k+1}\sum_{m=1}^{k+1} h_{lm}(Z_{j_1},...,Z_{j_k})K_h(X_{j_{k+1}}-X_i)Y_{j_{k+1}}X_{j_{k+1}}^{m-1}X_i^{m-1}}_{=\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)}.
\end{aligned}$$

With this we obtain

$$\begin{aligned}
\tilde{\iota}_1 =& \frac{1}{n}\sum_{i=1}^{n} Y_i\hat{\delta}(X_i) = \frac{1}{n^{k+2}}\sum_{i=1}^{n}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n} Y_i\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i) \\
=& \frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{2k+3}=1}^{n} g_1(Z_{j_1},...,Z_{j_{2k+3}}).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\tilde{\iota}_2 =& \frac{1}{n}\sum_{i=1}^{n} Y_i\frac{1}{n}\sum_{l=1}^{n}\hat{\delta}(X_l) \\
=& \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n} Y_i\frac{1}{n^{k+2}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n}\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_l)
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n^{k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+3}=1}^{n}Y_{k+3}\tilde{w}(Z_{j_1},...,Z_{k+2})\\
=&\frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{2k+3}=1}^{n}g_2(Z_{j_1},...,Z_{j_{2k+3}}).
\end{aligned}
$$

For $\tilde{\iota}_3$ and $\tilde{\iota}_4$ we obtain

$$
\begin{aligned}
\tilde{\iota}_3 =&\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)^2\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n}\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)\right.\\
&\left.\frac{1}{n^{k+1}}\sum_{m_1=1}^{n}\cdots\sum_{m_{k+1}=1}^{n}\tilde{w}(Z_{m_1},...,Z_{m_{k+1}},Z_i)\right\}\\
=&\frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{2k+3}=1}^{n}\underbrace{\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_{2k+3})\tilde{w}(Z_{j_{k+2}},...,Z_{j_{2k+2}},Z_{2k+3})}_{=g_3(Z_{j_1},...,Z_{j_{2k+3}})}
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\iota}_4 =&\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(X_i)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n^{k+1}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+1}=1}^{n}\tilde{w}(Z_{j_1},...,Z_{j_{k+1}},Z_i)\\
=&\frac{1}{n^{k+2}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+2}=1}^{n}\tilde{w}(Z_{j_1},...,Z_{j_{k+2}})\\
=&\frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{k+2}=1}^{n}g_4(Z_{j_1},...,Z_{j_{2k+3}}).
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
\tilde{\iota} =&\frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{2k+3}=1}^{n}g(Z_{j_1},...,Z_{j_{2k+3}})\\
=&\frac{1}{n^{2k+3}}\sum_{j_1=1}^{n}\cdots\sum_{j_{2k+3}=1}^{n}w(Z_{j_1},...,Z_{j_{2k+3}}).
\end{aligned}
$$

We now consider

$$\sqrt{n}\tilde{\iota} = \frac{\sqrt{n}}{n^{2k+3}} \sum_{j_1=1}^{n} \cdots \sum_{j_{2k+3}=1}^{n} w(Z_{j_1}, ..., Z_{j_{2k+3}})$$

which we can decompose under Assumption 2.15 according to Lemma 2.1 to

$$\begin{aligned}
&= \frac{\sqrt{n}}{n^{2k+3}} \sum_{C(\{j_1,...,j_{2k+3}\})}^{n} w(Z_{j_1}, ..., Z_{j_{2k+3}}) + o_p(1) \\
&= \frac{\sqrt{n}}{n^{2k+3}} (2k+3)! \sum_{j_1<...<j_{2k+3}}^{n} w(Z_{j_1}, ..., Z_{j_{2k+3}}) + o_p(1) \\
&= \underbrace{\frac{n!/(n-[2k+3])!}{n^{2k+3}}}_{=:c_n} \sqrt{n} \underbrace{\binom{n}{2k+3}^{-1} \sum_{j_1<...<j_{2k+3}}^{n} w(Z_{j_1}, ..., Z_{j_{2k+3}})}_{=:U_n} + o_p(1).
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
\sqrt{n}(\tilde{\iota} - \vartheta) &= c_n \sqrt{n} U_n - \sqrt{n}\vartheta \\
&= \underbrace{c_n}_{\to 1} \underbrace{\sqrt{n}(U_n - \vartheta_n)}_{\xrightarrow{\mathcal{L}} N(0,\Sigma)} + \underbrace{(c_n - 1)\sqrt{n}\vartheta}_{\to 0} \xrightarrow{\mathcal{L}} N(0, \Sigma),
\end{aligned}$$

where

$$\Sigma = (2k+3)^2 E_{\mathbf{P}} \left\{ \left(E_{\mathbf{P}} \left[w(Z_{j_1}, ..., Z_{j_{2k+3}}) | Z_{j_1} \right] - \vartheta \right) \left(E_{\mathbf{P}} \left[w^T(Z_{j_1}, ..., Z_{j_{2k+3}}) | Z_{j_1} \right] - \vartheta^T \right) \right\}.$$

With $F\left((a_1, ..., a_4)^T\right) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}$ and $\iota_X^{locpol}(Y) = F(\vartheta)$ application of the delta-method yields

$$\sqrt{n} \left\{ \hat{\iota}_X^{locpol}(Y) - \iota_X^{locpol}(Y) \right\} = \sqrt{n} \left\{ F(\tilde{\iota}) - F(\vartheta) \right\} \xrightarrow{\mathcal{L}} N(0, \underbrace{DF(\vartheta)^T \Sigma DF(\vartheta)}_{=:\sigma^2}).$$

□

Lemma 2.17. The variance σ^2 can be consistently estimated by

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{\Sigma} DF(\tilde{\iota}),$$

where

$$\hat{\Sigma} = (2k+3)^2 \left[\binom{n}{4k+5}^{-1} \sum_{j_1<\ldots<j_{4k+5}} \frac{1}{(4k+5)!} \sum_{\pi \in S(\{1,\ldots,4k+5\})} \tilde{g}(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(4k+5)}}) - \tilde{\iota}\tilde{\iota}^T \right]$$

and

$$\tilde{g}(Z_{j_1}, ..., Z_{j_{4k+5}}) = w(Z_{j_1}, ..., Z_{j_{2k+3}}) w^T(Z_{j_1}, Z_{j_{2k+4}}, ..., Z_{j_{4k+5}}).$$

Proof. We note that as a direct consequence of the theory of U-statistics $DF(\vartheta)$ can be consistently estimated by $DF(\tilde{\iota})$. Hence, if we find a consistent estimator for Σ we can estimate σ consistently. Again we will estimate Σ based on U-statistics, following the idea in Kowalski and Tu (2008) and the previous sections. To this end we consider

$$\begin{aligned}
\Sigma/(2k+3)^2 =& E_{\mathbf{P}} \left\{ E[w(Z_{j_1}, ..., Z_{j_{2k+3}})|Z_{j_1}] E_{\mathbf{P}}[w^T(Z_{j_1}, ..., Z_{j_{2k+3}})|Z_{j_1}] \right\} - \vartheta\vartheta^T \\
=& E_{\mathbf{P}} \left\{ E_{\mathbf{P}}[w(Z_{j_1}, ..., Z_{j_{2k+3}}) w^T(Z_{j_1}, Z_{j_{2k+4}}, ..., Z_{j_{4k+5}})|Z_{j_1}] \right\} - \vartheta\vartheta^T \\
=& \underbrace{E_{\mathbf{P}} \left\{ w(Z_{j_1}, ..., Z_{j_{2k+3}}) w^T(Z_{j_1}, Z_{j_{2k+4}}, ..., Z_{j_{4k+5}}) \right\}}_{\Sigma_h} - \vartheta\vartheta^T.
\end{aligned}$$

We can estimate Σ_h by a $(4k+5)$-th oder U-statistics. To this end let

$$\tilde{g}(Z_{j_1}, ..., Z_{j_{4k+5}}) = w(Z_{j_1}, ..., Z_{j_{2k+3}}) w^T(Z_{j_1}, Z_{j_{2k+4}}, ..., Z_{j_{4k+5}})$$

and $\tilde{\tilde{g}}$ a symmetric version of $\tilde{g}$, for example

$$\tilde{\tilde{g}}(Z_{j_1}, ..., Z_{j_{4k+5}}) = \frac{1}{(4k+5)!} \sum_{\pi \in S(\{1,\ldots,4k+5\})} \tilde{g}(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(4k+5)}}).$$

It follows that the U-statistics

$$\hat{\Sigma}_h = \binom{n}{4k+5}^{-1} \sum_{j_1<\ldots<j_{4k+5}} \tilde{\tilde{g}}(Z_{j_1}, ..., Z_{j_{4k+5}})$$

is a consistent estimator for Σ_h. Hence, a consistent estimator for Σ is given by

$$\hat{\Sigma} = (2k+3)^2 \left[\binom{n}{4k+5}^{-1} \sum_{j_1<\ldots<j_{4k+5}} \frac{1}{(4k+5)!} \sum_{\pi \in S(\{1,\ldots,4k+5\})} \tilde{g}(Z_{j_{\pi(1)}}, ..., Z_{j_{\pi(4k+5)}}) + \tilde{\iota}\tilde{\iota}^T \right].$$

This implies that σ^2 can be consistently estimated by

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{\Sigma} DF(\tilde{\iota}).$$

□

Lemma 2.26. The variance σ^2 can be consistently estimated by $DF(\tilde{\iota})^T 9(\hat{g}^* - \tilde{\iota}\tilde{\iota}^T) DF(\tilde{\iota})$, where

$$\hat{g}^* = \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \hat{w}(Z_i, Z_j, Z_l) \hat{w}^T(Z_i, Z_a, Z_b)$$

and $\hat{w}(Z_i, Z_j, Z_l)$ is obtained from $w(Z_i, Z_j, Z_l)$ by replacing all $\tilde{f}s$ by $\hat{f}s$.

Proof. Since $\tilde{\iota}$ is consistent for ϑ we can estimate $DF(\vartheta)$ consistently by $DF(\tilde{\iota})$. Hence, a consistent estimator for V remains to be found. Applying the same calculus as in Section 2.2.1 we obtain that

$$V/9 = E_{\mathbf{P}}(\underbrace{w(Z_i, Z_j, Z_l) w^T(Z_i, Z_a, Z_b)}_{\tilde{g}(Z_i, \ldots Z_b)}) - \vartheta\vartheta^T.$$

In this case as well, for estimation, we can replace ϑ by $\tilde{\iota}$. The mean of $\tilde{g}$ can be consistently estimated by the U-statistic

$$\hat{g} = \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} \tilde{g}(Z_{\pi(i)}, ..., Z_{\pi(b)}).$$

Since $\tilde{g}$ contains the unknown $\tilde{f}$ we can not compute $\tilde{g}$. Nevertheless, we can replace $\tilde{f}$ by its estimator $\hat{f}$ without changing the asymptotic behavior of the estimator. This can be seen as follows:

$$\begin{aligned}
(\hat{g})_{u,v} =& \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \sum_{\pi \in S(\{i,j,l,a,b\})} w_u(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}) w_v(Z_{\pi(i)}, Z_{\pi(a)}, Z_{\pi(b)}) \\
=& \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!} \frac{1}{6!^2} \sum_{\pi \in S(\{i,j,l,a,b\})} \sum_{\substack{\psi \in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho \in S(\{\pi(i),\pi(j),\pi(l)\})}} \big\{ g_u(Z_{\rho(\pi(i))}, Z_{\rho(\pi(j))}, Z_{\rho(\pi(l))}) \\
& g_v(Z_{\psi(\pi(i))}, Z_{\psi(\pi(a))}, Z_{\psi(\pi(b))}) \big\}.
\end{aligned}$$

The replacement of $\tilde{f}$ by $\hat{f}$ can be justified similar to the reverse replacement before. Let $u \neq 3$ and define

$$L_{\pi,\rho,\psi} = \big(g_u(Z_{\rho(i)}, Z_{\rho(j)}, Z_{\rho(l)}) g_v(Z_{\psi(i)}, Z_{\psi(a)}, Z_{\psi(b)}) \big) \frac{\tilde{f}(X_{\rho(\pi(i))})}{\hat{f}(X_{\rho(\pi(i))})}$$

and

$$A_{\pi,\rho,\psi} = \left(g_u(Z_{\rho(i)}, Z_{\rho(j)}, Z_{\rho(l)}) g_v(Z_{\psi(i)}, Z_{\psi(a)}, Z_{\psi(b)})\right) \frac{\hat{f}(X_{\rho(\pi(i))}) - \tilde{f}(X_{\rho(\pi(i))})}{\hat{f}(X_{\rho(\pi(i))})}.$$

With this definitions we can rewrite the estimator $\hat{g}$ as

$$\begin{aligned}
&(\hat{g})_{u,v} \\
=&\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} (L_{\pi,\rho,\psi} + A_{\pi,\rho,\psi}) \\
=&\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} L_{\pi,\rho,\psi} \\
&+ \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} L_{\pi,\rho,\psi} \frac{\hat{f}(X_{\rho(\pi(i))}) - \tilde{f}(X_{\rho(\pi(i))})}{\tilde{f}(X_{\rho(\pi(i))})}.
\end{aligned}$$

The last term can be shown to be $o_p(1)$. To this end regard

$$\left| \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} L_{\pi,\rho,\psi} \frac{\hat{f}(X_{\rho(\pi(i))}) - \tilde{f}(X_{\rho(\pi(i))})}{\tilde{f}(X_{\rho(\pi(i))})} \right|$$

$$\leq \sup_{x\in\mathbb{R}} \left| \frac{\hat{f}(x) - \tilde{f}(x)}{\tilde{f}(x)} \right| \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} |L_{\pi,\rho,\psi}|.$$

The first term converges to zero in probability by Lemma 2.23 and the second term converges to $E(|L_{\pi,\rho,\psi}|)$, provided this exists, by Theorem A.7. Therefore,

$$\binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi\in S(\{i,j,l,a,b\})} \sum_{\substack{\psi\in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho\in S(\{\pi(i),\pi(j),\pi(l)\})}} L_{\pi,\rho,\psi} \frac{\hat{f}(X_{\rho(\pi(i))}) - \tilde{f}(X_{\rho(\pi(i))})}{\tilde{f}(X_{\rho(\pi(i))})}$$

is $o_p(1)$ and we have

$$(\hat{g})_{u,v} = \binom{n}{5}^{-1} \sum_{i<j<l<a<b} \frac{1}{5!}\frac{1}{6!^2} \sum_{\pi \in S(\{i,j,l,a,b\})} \sum_{\substack{\psi \in S(\{\pi(i),\pi(a),\pi(b)\}) \\ \rho \in S(\{\pi(i),\pi(j),\pi(l)\})}} \big\{ \nu_u(Z_{\rho(\pi(i))}, Z_{\rho(\pi(j))}, Z_{\rho(\pi(l))})$$

$$g_v(Z_{\psi(\pi(i))}, Z_{\psi(\pi(a))}, Z_{\psi(\pi(b))}) \big\} + o_p(1),$$

where ν_u is g_u except for $\tilde{f}$ is replaced by $\hat{f}$. By symmetry this argumentation also holds for $v \neq 3$. If either u or v (or possibly both) equals 3 we can also replace $\tilde{f}^2$ by $\hat{f}^2$ by twofold application of the argumentation above. Hence, when replacing all $\tilde{f}$s by $\hat{f}$s in the definition of $\hat{g}$ we obtain a consistent estimator $\hat{g}^*$ for the mean of $\tilde{g}$. Thus a consistent estimator for σ^2 is given by

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T 9(\hat{g}^* - \tilde{\iota}\tilde{\iota}^T) DF(\tilde{\iota}).$$

□

Theorem 2.34. Under the assumptions of Section 2.2.8 we have that

$$\sqrt{m_n}\left(\hat{\iota}_X^{ks,mod2}(Y) - \iota_X(Y)\right) \xrightarrow{\mathcal{L}} N(0,\sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T \Sigma DF(\vartheta)$, $\Sigma = 2E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$, where $\tilde{w}(Z_i) = E_{\mathbf{P}}\left(w(Z_i,Z_j)|Z_i\right) - \vartheta$ and $F(a_1,...,a_4) = \frac{a_1-a_2}{\sqrt{a_3-a_4^2}}$.

Proof. We choose the same procedure as in the preceding sections to show the assertion. This means that we show that the vector $\tilde{\iota} = (\tilde{\iota}_1,...,\tilde{\iota}_4)^T$, where $\tilde{\iota}_1 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i\hat{\delta}(X_i)$, $\tilde{\iota}_2 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i\overline{\hat{\delta}(X)}$, $\tilde{\iota}_3 = \frac{1}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)^2$ and $\tilde{\iota}_4 = \frac{1}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)$ is essentially a U-statistics. After that application of the delta method will give the desired asymptotic distribution result. For the first element of $\tilde{\iota}$ we obtain

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_1 =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i\hat{\delta}(X_i)\\ =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i E_{\mathbf{P}}(Y|X_i) + \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i\left(\hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i)\right).\end{aligned}$$

For the last term we have

$$\left|\frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i\left(\hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i)\right)\right| \leq \underbrace{\sqrt{m_n}\sup_{x\in J}\left|\hat{\delta}(x) - E_{\mathbf{P}}(Y|x)\right|}_{=o_p(1)}\underbrace{\frac{1}{m_n}\sum_{i=1}^{m_n}|Y_i|}_{\to E_{\mathbf{P}}(|Y|)<\infty} = o_p(1). \tag{B.3}$$

The fact that the first term here is $o_p(1)$ follows from Corollary 2.32. As a consequence to (B.3) we have that

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_1 =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i E_{\mathbf{P}}(Y|X_i) + o_p(1)\\ =& \frac{\sqrt{m_n}}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n} Y_i E_{\mathbf{P}}(Y|X_i) + o_p(1).\end{aligned} \tag{B.4}$$

Furthermore, for $\tilde{\iota}_2$ we can show that

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_2 =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} Y_i\overline{\hat{\delta}(X)} = \frac{\sqrt{m_n}}{m_n}\bar{\mathbf{Y}}\sum_{i=1}^{m_n}\hat{\delta}(X_i)\\ =& \frac{\sqrt{m_n}}{m_n}\bar{\mathbf{Y}}\sum_{i=1}^{m_n} E_{\mathbf{P}}(Y|X_i) + \frac{\sqrt{m_n}}{m_n}\bar{\mathbf{Y}}\sum_{i=1}^{m_n}\left(\hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i)\right)\end{aligned}$$

where it follows again from Corollary 2.32 that the second summand is $o_p(1)$, hence

$$= \frac{\sqrt{m_n}}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Y_i E_{\mathbf{P}}(Y|X_j) + o_p(1). \tag{B.5}$$

For the third element of $\tilde{\iota}$ we need the twofold application of Corollary 2.32 to get our desired result. As a first step note that similar to the considerations before, we have that

$$\begin{aligned}
\sqrt{m_n}\tilde{\iota}_3 =& \frac{\sqrt{m_n}}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \hat{\delta}(X_i)\hat{\delta}(X_j) \\
=& \frac{\sqrt{m_n}}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \hat{\delta}(X_i) E_{\mathbf{P}}(Y|X_j) + \underbrace{\frac{\sqrt{m_n}}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \hat{\delta}(X_i) \left(\hat{\delta}(X_j) - E_{\mathbf{P}}(Y|X_j) \right)}_{(*)}.
\end{aligned}$$

As a next step we show that the term $(*)$ is $o_p(1)$. We have that

$$|(*)| \leq \sqrt{m_n} \sup_{x \in J} \left| \hat{\delta}(x) - E_{\mathbf{P}}(Y|x) \right| \frac{1}{m_n} \sum_{i=1}^{m_n} \left| \hat{\delta}(X_i) \right|.$$

The first summand is $o_p(1)$ by Corollary 2.32. Hence, it remains to show that $\frac{1}{m_n} \sum_{i=1}^{m_n} \left| \hat{\delta}(X_i) \right|$ is bounded as $n \to \infty$. To this end, we regard

$$\begin{aligned}
\frac{1}{m_n} \sum_{i=1}^{m_n} \left| \hat{\delta}(X_i) \right| =& \frac{1}{m_n} \sum_{i=1}^{m_n} \left| \hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i) + E_{\mathbf{P}}(Y|X_i) \right| \\
\leq& \frac{1}{m_n} \sum_{i=1}^{m_n} |E_{\mathbf{P}}(Y|X_i)| + \frac{1}{m_n} \sum_{i=1}^{m_n} \left| \hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i) \right| \\
\leq& \underbrace{\frac{1}{m_n} \sum_{i=1}^{m_n} |E_{\mathbf{P}}(Y|X_i)|}_{\to E_{\mathbf{P}}\{|E_{\mathbf{P}}(Y|X)|\}} + \underbrace{\sup_{x \in J} \left| \hat{\delta}(x) - E_{\mathbf{P}}(Y|x) \right|}_{=o_p(1)} \\
\xrightarrow{p}& E_{\mathbf{P}} \left\{ |E_{\mathbf{P}}(Y|X)| \right\} < \infty.
\end{aligned}$$

Thus, we have that

$$\sqrt{m_n}\tilde{\iota}_3 = \frac{\sqrt{m_n}}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \hat{\delta}(X_i) E_{\mathbf{P}}(Y|X_j) + o_p(1).$$

Applying Corollary 2.32 again leads to

$$\sqrt{m_n}\tilde{\iota}_3 = \frac{\sqrt{m_n}}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n} E_{\mathbf{P}}(Y|X_i)E_{\mathbf{P}}(Y|X_j) + o_p(1). \tag{B.6}$$

The considerations for $\tilde{\iota}_4$ are

$$\begin{aligned}\sqrt{m_n}\tilde{\iota}_4 =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}\hat{\delta}(X_i)\\ =& \frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n} E_{\mathbf{P}}(Y|X_i) + \underbrace{\frac{\sqrt{m_n}}{m_n}\sum_{i=1}^{m_n}\left(\hat{\delta}(X_i) - E_{\mathbf{P}}(Y|X_i)\right)}_{=o_p(1)}\\ =& \frac{\sqrt{m_n}}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n} E_{\mathbf{P}}(Y|X_i) + o_p(1).\end{aligned} \tag{B.7}$$

We have according to Assumption 2.33

$$g_1(Z_i, Z_j) = Y_i E_{\mathbf{P}}(Y|X_i) \quad g_2(Z_i, Z_j) = Y_i E_{\mathbf{P}}(Y|X_j)$$

$$g_3(Z_i, Z_j) = E_{\mathbf{P}}(Y|X_i)E_{\mathbf{P}}(Y|X_j) \quad g_4(Z_i, Z_j) = E_{\mathbf{P}}(Y|X_i)$$

as well as $g = (g_1, ..., g_4)^T$ and $w(Z_i, Z_j) = \frac{1}{2}\left(g(Z_i, Z_j) + g(Z_j, Z_i)\right)$. With these definitions and equations (B.4), (B.5), (B.6) and (B.7) we obtain

$$\sqrt{m_n}\tilde{\iota} = \sqrt{m_n}\frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{i=1}^{m_n} w(Z_i, Z_j) + o_p(1).$$

With Assumption 2.33 it follows from Lemma 2.1 that

$$\sqrt{m_n}\tilde{\iota} = \sqrt{m_n}\frac{2}{m_n^2}\sum\sum_{i<j} w(Z_i, Z_j) + o_p(1)$$

which implies

$$\sqrt{m_n}\left(\tilde{\iota} - \vartheta\right) = c_n\sqrt{m_n}\left(U_n - \vartheta\right) + o_p(1),$$

where $c_n = \frac{m_n - 1}{m_n} \to 1$ and

$$U_n = \binom{m_n}{2}^{-1}\sum\sum_{i<j} w(Z_i, Z_j)$$

is a second order U-statistics. Consequently, we have that

$$\sqrt{m_n}\,(\tilde{\iota} - \vartheta) \xrightarrow{\mathcal{L}} N_4(\mathbf{0}, \Sigma),$$

with $\Sigma = 2E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right)$, where $\tilde{w}(Z_i) = E_{\mathbf{P}}\left(w(Z_i, Z_j)|Z_i\right) - \vartheta$. As a next step, we apply the delta-method to this result and obtain,

$$\sqrt{m_n}\left(\hat{\iota}_X^{ks,mod2}(Y) - \iota_X(Y)\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = DF(\vartheta)^T \Sigma DF(\vartheta)$, and $F(a_1, ..., a_4) = \frac{a_1 - a_2}{\sqrt{a_3 - a_4^2}}$. □

Lemma 2.35. Under the setup of this section a consistent estimate for σ^2 is given by

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{\Sigma} DF(\tilde{\iota}),$$

where $\hat{\Sigma} = \hat{\tilde{g}} - \tilde{\iota}\tilde{\iota}^T$ with $\tilde{\iota} = (\tilde{\iota}_1, ..., \tilde{\iota}_4)^T$, where $\tilde{\iota}_1 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i \hat{\delta}(X_i)$, $\tilde{\iota}_2 = \frac{1}{m_n}\sum_{i=1}^{m_n} Y_i \overline{\hat{\delta}(X)}$, $\tilde{\iota}_3 = \frac{1}{m_n}\sum_{i=1}^{m_n} \hat{\delta}(X_i)^2$ and $\tilde{\iota}_4 = \frac{1}{m_n}\sum_{i=1}^{m_n} \hat{\delta}(X_i)$ and

$$\hat{\tilde{g}}_{u,v} = \binom{m_n}{3}^{-1} \sum_{i<j<l} \frac{1}{24} \sum_{\pi \in S(\{i,j,l\})} \sum_{\substack{\psi \in S(\{\pi(i),\pi(j)\}) \\ \rho \in S(\{\pi(i),\pi(l)\})}} \check{g}_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) \check{g}_v(Z_{\rho(\pi(i))}, Z_{\rho(\pi(l))}).$$

As in the proof of the previous theorem $\check{g}$ is obtained from g by replacing $E_{\mathbf{P}}(Y|x)$ with $\hat{\delta}(x)$.

Proof. Since we can estimate ϑ consistently by $\tilde{\iota}$ it suffices to find a consistent estimate for Σ. To this end we note that

$$\begin{aligned}
\frac{1}{2}\Sigma =& E_{\mathbf{P}}\left(\tilde{w}(Z_i)\tilde{w}^T(Z_i)\right) \\
=& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j)|Z_i\right\} E_{\mathbf{P}}\left\{w^T(Z_i, Z_j)|Z_i\right\}\right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}}\left(E_{\mathbf{P}}\left\{w(Z_i, Z_j)|Z_i\right\} E_{\mathbf{P}}\left\{w^T(Z_i, Z_l)|Z_i\right\}\right) - \vartheta\vartheta^T \\
=& E_{\mathbf{P}} \underbrace{\left(w(Z_i, Z_j) w^T(Z_i, Z_l)\right)}_{=\tilde{g}(Z_i, Z_j, Z_l)} - \vartheta\vartheta^T.
\end{aligned}$$

Again we can estimate ϑ consistently by $\tilde{\iota}$. $E_{\mathbf{P}}\left(w(Z_i, Z_j)w^T(Z_i, Z_l)\right)$ can be, according to the theory of U-statistics, consistently estimated by

$$\hat{g} = \binom{m_n}{3}^{-1} \sum_{i<j<l} \tilde{\tilde{g}}(Z_i, Z_j, Z_l),$$

where

$$\tilde{\tilde{g}}(Z_i, Z_j, Z_l) = \frac{1}{3!} \sum_{\pi \in S(\{i,j,l\})} \tilde{g}(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}),$$

where $S(\{i, j, l\})$ is the symmetric group of the set $\{i, j, l\}$. The estimate $\hat{g}$ can then be written as

$$\hat{g} = \binom{m_n}{3}^{-1} \sum_{i<j<l} \frac{1}{3!} \sum_{\pi \in S(\{i,j,l\})} \tilde{g}(Z_{\pi(i)}, Z_{\pi(j)}, Z_{\pi(l)}).$$

However, $\tilde{g}$ contains the unknown $E_{\mathbf{P}}(Y|X_i)$ which means that we cannot compute this

estimate. The solution to this problem is to replace this conditional expectation by the kernel smoother $\hat{\delta}(X_i)$. Note, that from the considerations that lead to (B.4), (B.5), (B.6) and (B.7) it follows that

$$\frac{\sqrt{m_n}}{m_n^2}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}|g_u(Z_i,Z_j)-\check{g}_u(Z_i,Z_j)|=o_p(1)\quad\forall u\in\{1,...,4\},\tag{B.8}$$

where $\check{g}_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})$ is the same function as $g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})$ but with $\hat{\delta}(x)$ instead of $E_{\mathbf{P}}(Y|x)$. In the following, we will show that the replacement of the conditional expectation by the kernel smoother does not affect the consistency of $\hat{g}$. We do so by showing that replacement does not affect the consistency in each element of $\hat{g}$. Therefore, have a look at

$$\begin{aligned}&\hat{g}_{u,v}\\ =&\binom{m_n}{3}^{-1}\sum_{i<j<l}\frac{1}{3!}\sum_{\pi\in S(\{i,j,l\})}w_u(Z_{\pi(i)},Z_{\pi(j)})w_v(Z_{\pi(i)},Z_{\pi(l)})\\ =&\binom{m_n}{3}^{-1}\sum_{i<j<l}\frac{1}{3!}\sum_{\pi\in S(\{i,j,l\})}\frac{1}{2!^2}\sum_{\substack{\psi\in S(\{\pi(i),\pi(j)\})\\ \rho\in S(\{\pi(i),\pi(l)\})}}g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})g_v(Z_{\rho(\pi(i))},Z_{\rho(\pi(l))})\end{aligned}$$

We show that we can replace $g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})$ in this expression by $\check{g}_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})$. By symmetry it then follows that we can also replace $g_v(Z_{\rho(\pi(i))},Z_{\rho(\pi(l))})$ by $\check{g}_v(Z_{\rho(\pi(i))},Z_{\rho(\pi(l))})$. Obviously, it suffices to show that

$$\binom{m_n}{3}^{-1}\sum_{i<j<l}\frac{1}{3!2}\sum_{\pi\in S(\{i,j,l\})}\sum_{\psi\in S(\{\pi(i),\pi(j)\})}\left\{g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})-\check{g}_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})\right\}\tag{B.9}$$

converges to zero as $m_n\to\infty$. To this end, we make the following considerations.

$$\left|\binom{m_n}{3}^{-1}\sum_{i<j<l}\frac{1}{3!2}\sum_{\pi\in S(\{i,j,l\})}\sum_{\psi\in S(\{\pi(i),\pi(j)\})}\left\{g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})-\check{g}_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})\right\}\right|$$

$$\leq\binom{m_n}{3}^{-1}\sum_{i<j<l}\frac{1}{3!2}\sum_{\pi\in S(\{i,j,l\})}\sum_{\psi\in S(\{\pi(i),\pi(j)\})}\left|\left\{g_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})-\check{g}_u(Z_{\psi(\pi(i))},Z_{\psi(\pi(j))})\right\}\right|.$$

To show that this converges to zero it suffices to regard each of the 12 summands for

which π and ψ are fixed. Hence, we regard for fixed π and ψ

$$\begin{aligned}
&\binom{m_n}{3}^{-1} \sum_{i<j<l} \frac{1}{3!2} \left| \{ g_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) - \check{g}_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) \} \right| \\
\leq &\frac{1}{12} \binom{m_n}{3}^{-1} \sum_{i,j,l} \left| \{ g_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) - \check{g}_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) \} \right| \\
= &\frac{m_n}{12} \binom{m_n}{3}^{-1} \sum_{i,j} |\{ g_u(Z_i, Z_j) - \check{g}_u(Z_i, Z_j \}| = o_p(1)
\end{aligned}$$

by (B.8). From this it follows that (B.9) is also $o_p(1)$, which implies that the estimate $\hat{\hat{g}}$ with

$$\hat{\hat{g}}_{u,v} = \binom{m_n}{3}^{-1} \sum_{i<j<l} \frac{1}{24} \sum_{\pi \in S(\{i,j,l\})} \sum_{\substack{\psi \in S(\{\pi(i),\pi(j)\}) \\ \rho \in S(\{\pi(i),\pi(l)\})}} \check{g}_u(Z_{\psi(\pi(i))}, Z_{\psi(\pi(j))}) \check{g}_v(Z_{\rho(\pi(i))}, Z_{\rho(\pi(l))})$$

is also consistent for $E_{\mathbf{P}}\left(w(Z_i, Z_j) w^T(Z_i, Z_l)\right)$. Consequently, Σ can be consistently estimated by

$$\hat{\Sigma} = \hat{\hat{g}} - \tilde{\iota}\tilde{\iota}^T.$$

This leads to the estimate

$$\hat{\sigma}^2 = DF(\tilde{\iota})^T \hat{\Sigma} DF(\tilde{\iota})$$

which is consistent for σ^2. □