Robust nonparametric estimation with missing data

Graciela Boente
Universidad de Buenos Aires and CONICET, Argentina
Wenceslao González–Manteiga
Universidad de Santiago de Compostela, Spain
and
Ana Pérez–González
Universidad de Vigo, Spain

Abstract

In this paper, under a nonparametric regression model, we introduce a family of robust procedures to estimate the regression funtion when missing data occur in the response. Our proposal is based on a local $M-$functional applied to the conditional distribution function estimate adapted to the presence of missing data. We show that the robust procedure is consistent and asymptotically normally distributed.

Corresponding Author
Graciela Boente
Moldes 1855, 3^o A
Buenos Aires, C1428CRA, Argentina
email: gboente@mate.dm.uba.ar

AMS Subject Classification 1990: Primary 62F35, Secondary 62H25.
Key words and phrases: Robust Estimation, Missing Data, Nonparametric Regression, Kernel Weights, Asymptotic Properties.

1 Introduction

Nonparametric regression models assume that we have a response $y_i \in I\!R$ and covariates or design points $\mathbf{x}_i \in I\!R^p$ satisfying

$$y_i = m(\mathbf{x}_i) + \sigma(\mathbf{x}_i)\,\varepsilon_i \qquad 1 \le i \le n\ , \tag{1}$$

with the errors ε_i independent and independent of $\mathbf{x}_i$, with symmetric distribution $F_0(\cdot)$. The nonparametric nature of model (1) offers more flexibility than the standard linear model, when modelling a complicated relationship between the response variable with the covariates.

Two of the most common methods in nonparametric regression are kernel and k-nearest neighbor kernel methods, introduced by Nadaraya-Watson (1964) and Collomb (1981) respectively. But these estimators are not robust. Robust estimators can be obtained via M-estimates. See Tsybakov (1982) and Härdle (1984), who studied pointwise asymptotic properties of a robust version of the Nadaraya- Watson method and later they extended their results to M-type scale equivariant kernel estimates (Härdle and Tsybakov (1988)); Boente and Fraiman (1989), who considered robust equivariant nonparametric estimates using nearest neighbor weights and weights based on kernel methods.

Most of the statistical methods in nonparametric regression are designed for complete data sets and problems arise when missing obervations are present which is a common situation in biomedical or socioeconomic studies, for example. Classic examples are found in the field of social sciences with the problem of non-response in sample surveys, in Physics, in Genetics (Meng, 2000), among others.

The objective of this paper is to introduce a robust nonparametric estimator of the regression function when the response variable has missing observations but the covariate $\mathbf{x}$ is totally observed. This pattern is common, for example, in the scheme of double sampling proposed by Neyman (1938), where first a complete sample is obtained and then some additional covariate values are computed since perhaps this is less expensive than to obtain more response values.

In the regression setting, a common method is to impute the incomplete observations and then proceed to carry out the estimation of the conditional or unconditional mean of the response variable with the completed sample. The methods considered include linear regression (Yates, 1933), kernel smoothing (Cheng, 1994; Chu and Cheng, 1995) nearest neighbor imputation (Chen and Shao, 2000), semiparametric estimation (Wang et al., 2004), nonparametric multiple imputation (Aerts et al., 2002), empirical likelihood over the imputed values (Wang and Rao, 2002), among others. González–Manteiga and Pérez–Gonzalez (2004) considered an approach based on local polynomials to estimate the regression function when the response variable y is missing but the covariate $\mathbf{x}$ is totally observed. All these proposals are very sensitive to anomalous observations since they are based on a local least squares approach.

As is well known, most nonparametric regression estimates with complete data suffer from the same lack of robustness that their linear counterparts in parametric models. In this setting, outlying observations can be even more dangerous since the shape of the estimated curve is highly sensitive to outlying observations. The treatment of outliers is an important step in highlighting features of a data set. Extreme points affect the scale and the shape of any estimate of the regression function based on local averaging, leading to possible wrong conclusions. This has motivated the interest in combining the ideas of robustness with those of smoothed regression, to develop procedures which will be resistant to deviations from the central model in nonparametric regression models. The first proposal of robust estimates for nonparametric regression was given by Cleveland (1979) who adapted a local polynomial fit by introducing weights to deal with large residuals. A review of several methods leading to robust nonparametric regression estimators for complete data sets can be seen in Härdle (1990).

In this paper, we propose two robust nonparametric regression estimators when we are dealing with missing observations in the response based on the robust nonparametric estimators for complete data studied by Boente and Fraiman (1989). The first one is the *simplified multivariate local* $M-$*smoother* which uses only the complete observations for the estimation and discards the incomplete vectors. The second one is the *imputed multivariate local* $M-$*smoother* which uses the simplified local $M-$smoother in order to impute the missing observations of the response y and then estimates the regression function with the completed sample.

The paper is organized as follows. Section 2 introduces the Robust Nonparametric estimators. The asymptotic properties of the Simplified estimator and for the Imputed estimator are studied in the sections 3 and 4 respectively. In Section 5 we present results of a simulation study. The concluding Remarks are in Section 6. And finally, technical proofs are in the Appendix.

2 Robust Proposals

We will consider robust inference with an incomplete data set $(y_i, \mathbf{x}_i, \delta_i)$, $1 \leq i \leq n$ where $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ if y_i is missing. Let $(Y, \mathbf{X}, \delta)$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i, \delta_i)$. Our aim is to estimate the nonparametric regression function in a robust way with the data set at hand. An ignorable missing mechanism will be imposed by assuming that δ and Y are conditionally independent given $\mathbf{X}$, i.e.,

$$P(\delta = 1 | Y, \mathbf{X}) = P(\delta = 1 | \mathbf{X}) = p(\mathbf{X}) \tag{2}$$

We will consider two type of smoothers. The first one is based on kernel weights which are given by

$$w_i(\mathbf{x}) = \frac{K\left(\dfrac{\mathbf{x}_i - \mathbf{x}}{h_n}\right)\delta_i}{\displaystyle\sum_{j=1}^{n} K\left(\dfrac{\mathbf{x}_j - \mathbf{x}}{h_n}\right)\delta_j}, \tag{3}$$

with K a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h the bandwidth parameter, while the nearest neighbor with kernel approach considers as weight function

$$w_i(\mathbf{x}) = \frac{K\left(\dfrac{\mathbf{x}_i - \mathbf{x}}{H_n(\mathbf{x})}\right)\delta_i}{\displaystyle\sum_{j=1}^{n} K\left(\dfrac{\mathbf{x}_i - \mathbf{x}}{H_n(\mathbf{x})}\right)\delta_i}, \tag{4}$$

with $H_n(\mathbf{x})$ the distance between $\mathbf{x}$ and its k_n-nearest neighbor among $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

2.1 Simplified Local $M-$Smoother

The simplified local $M-$smoother (*SLMS*) uses the information at hand and defines the estimator with the complete observations only. Denote by $\widehat{F}(y|\mathbf{X}=\mathbf{x})$ the empirical conditional distribution function which is defined as

$$\widehat{F}(y|\mathbf{X}=\mathbf{x}) \quad = \quad \sum_{i=1}^{n} w_i(\mathbf{x}) I_{(-\infty,y]}(y_i)\,, \tag{5}$$

with $w_i(\mathbf{x})$ the kernel weights defined in (3) or the nearest neighbor with kernel weights given in (4). $\widehat{F}(y|\mathbf{X}=\mathbf{x})$ provides an estimate of the distribution of $Y|\mathbf{X}=\mathbf{x}$ which will be denoted $F(y|\mathbf{X}=\mathbf{x})$ and which has been studied by Cheng and Chu (1996). Note also that the kernel weights are modified multiplying by the indicator of the missing variables in order to adapt to the complete sample and avoid bias.

The simplest one is the local median, $\widehat{m}_{\mathrm{MED}}(\mathbf{x})$, computed as the median of $\widehat{F}(y|\mathbf{X}=\mathbf{x})$. An interesting feature of this estimate is that it does not need any consistent scale estimate, when scale is unknown.

On the other hand, the local $M-$type estimate, $\widehat{m}_{\mathrm{M}}(\mathbf{x})$ is defined as the location $M-$estimates related to $\widehat{F}(y|\mathbf{X}=\mathbf{x})$. Thus, it is the solution of

$$\sum_{i=1}^{n} w_i(\mathbf{x})\psi\left(\frac{y_i-\widehat{m}_{\mathrm{M}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)=0\,, \tag{6}$$

where $w_i(\mathbf{x})$ are given in (3) or (4), ψ is an odd, increasing, bounded and continuous function and $\widehat{s}(\mathbf{x})$ is a local robust scale estimate. Possible choices for the score function ψ are the Huber or the bisquare $\psi-$function, while the scale $\widehat{s}(\mathbf{x})$ can be taken as the local median of the absolute deviations from the local median (local MAD), i.e., the MAD (Huber (1981)) with respect to the distribution $\widehat{F}(y|\mathbf{X}=\mathbf{x})$ defined in (5). Note that $\widehat{m}_{\mathrm{MED}}(\mathbf{x})$ corresponds to the choice $\psi(t)=\mathrm{sg}(t)$.

2.2 Imputed Local $M-$Smoother

As in the classical setting, see González–Manteiga and Pérez–González (2004), an imputation method can be developed. The imputed local $M-$smoother is constructed in two stages. In the first step, the *SLMS* is used to predict the missing observations so as to complete the sample. In this way, a complete sample of the form $(\mathbf{x}_i,\widehat{y}_i)$, $1\le i\le n$, where $\widehat{y}_i=\delta_i\,y_i+(1-\delta_i)\widehat{m}(\mathbf{x}_i)$, is obtained. The predictor $\widehat{m}(\mathbf{x}_i)$ can be taken as the local median, $\widehat{m}_{\mathrm{MED}}(\mathbf{x}_i)$, defined as the median of the empirical conditional distribution function given in (5), or as the local $M-$estimator, $\widehat{m}_{\mathrm{M}}(\mathbf{x}_i)$ defined through (6). Also, a local one–step, $\widehat{m}_{\mathrm{OS}}(\mathbf{x})$, or a reweighted estimator, $\widehat{m}_{\mathrm{RW}}(\mathbf{x})$, can be consider to improve the efficiency of the local median and to reduce computations. These estimators are defined through

$$\widehat{m}_{\mathrm{OS}}(\mathbf{x}) \quad = \quad \widehat{m}_{\mathrm{MED}}(\mathbf{x})+\widehat{s}(\mathbf{x})\frac{\displaystyle\sum_{i=1}^{n} w_i(\mathbf{x})\psi\left(\frac{y_i-\widehat{m}_{\mathrm{MED}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)}{\displaystyle\sum_{i=1}^{n} w_i(\mathbf{x})\psi\prime\left(\frac{y_i-\widehat{m}_{\mathrm{MED}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)}$$

$$\widehat{m}_{\mathrm{RW}}(\mathbf{x}) \quad = \quad \frac{\sum_{i=1}^{n} w_i(\mathbf{x}) w\left(\frac{y_i - \widehat{m}_{\mathrm{MED}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) y_i}{\sum_{i=1}^{n} w_i(\mathbf{x}) w\left(\frac{y_i - \widehat{m}_{\mathrm{MED}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)}$$

respectively, with $w(t) = \frac{\psi(t)}{t}$. To make explicit the dependence on the smoothing parameter, we will denote the preliminary robust simplified smoother as $\widehat{m}_{\mathrm{S}}(\mathbf{x}_i, h_n)$, when dealing with kernel weights and $\widehat{m}_{\mathrm{S}}(\mathbf{x}_i, k_n)$ when using nearest neighbor with kernel weights.

The kernel–based *ILMS*, $\widehat{m}_{\mathrm{M,I}}(\mathbf{x})$, is then defined as the solution of

$$\sum_{i=1}^{n} L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \psi\left(\frac{\widehat{y}_i - \widehat{m}_{\mathrm{M,I}}(\mathbf{x})}{\widehat{s}(\mathbf{x}, h_n)}\right) = 0 \,, \tag{7}$$

where $\widehat{s}(\mathbf{x}, h_n)$ is the simplified estimator for the scale function, to avoid extra computations and $\widehat{y}_i = \delta_i \, y_i + (1 - \delta_i)\widehat{m}_{\mathrm{S}}(\mathbf{x}_i, h_n)$. Note that a different smoothing parameter and a different kernel L can be used in this step. The nearest neighbor with kernel estimate is defined similarly.

3 Asymptotic Properties of the SLMS

3.1 Consistency

We will derive consistency for both kernel or nearest neighbor with kernel estimates. For this reason, assumptions are split according to the weights used. Denote $f_{\mathbf{X}}$ the density of $\mathbf{X}$. When dealing with kernel weights, there will be no need to require a density to the distribution μ of $\mathbf{X}$. In that sense, as in the complete sample setting, the results will be robust and distribution free. We will consider the following set of assumptions.

H1. $\psi : I\!R \to I\!R$ is an odd function, strictly increasing, bounded and continuous function.

H2. $F(y|\mathbf{X} = \mathbf{x})$ is symmetric around $m(\mathbf{x})$ and a continuous function of y for each fixed $\mathbf{x}$.

H3. $0 < p(\mathbf{x})$

H4. The kernel $K : I\!R^p \to I\!R$ is a bounded nonnegative function such that

$$\begin{aligned} a\, I_{\|\mathbf{x}\| \le r}(\mathbf{x}) \quad &\le \quad K(\mathbf{x}) \quad \text{for some } a > 0, r > 0 \\ a_1\, H(\|\mathbf{x}\|) \quad &\le \quad K(\mathbf{x}) \le a_2\, H(\|\mathbf{x}\|) \end{aligned}$$

where a_1, a_2 are positive numbers and $H : I\!R^+ \to I\!R^+$ is bounded decreasing and $u^p H(u) \to 0$ as $u \to \infty$.

H5. The sequence $h = h_n$ is such that $h_n \to 0$, $nh_n^p \to \infty$ and $\frac{nh_n^p}{\log n} \to \infty$.

Note that **H3** implies that, locally, some response variables are observed, which is a common assumption in the literature. The following result ensures consistency of the regression function estimator of m, when the smoothing is based either on local medians or local $M-$smoothers and kernel weights.

Proposition 3.1.1. *Assume that **H1** to **H5** hold. Then, we have that* $\widehat{m}_{\mathrm{M}}(\mathbf{x}) \xrightarrow{a.s.} m(\mathbf{x})$, *for almost all* $\mathbf{x}(\mu)$.

When dealing with nearest neighbor with kernel weights we will need the following additional assumptions

H6. The vector $\mathbf{X}$ has a density $f_{\mathbf{X}}$ positive at $\mathbf{x}$.

H7. The kernel $K : I\!R^p \to I\!R$ is a bounded nonnegative function such that $\int K(\mathbf{x})d\mathbf{x} < \infty$ and either of the following hold

i) $K(\mathbf{x}) \le c\, I_{\|\mathbf{x}\|\le r}(\mathbf{x})$ for some $c > 0$ and $r > 0$

ii) $f_{\mathbf{X}}$ is bounded and $\int K^2(\mathbf{x})d\mathbf{x} < \infty$, $\lim_{\|\mathbf{x}\|\to\infty} \|\mathbf{x}\|^p K(\mathbf{x}) = 0$.

H8. $K(u\,\mathbf{x}) \ge K(\mathbf{x})$ for $u \in (0,1)$.

H9. The sequence $k = k_n$ is such that $\dfrac{k_n}{n} \to 0$, $k_n \to \infty$ and $\dfrac{k_n}{\log n} \to \infty$.

Proposition 3.1.2. *Assume that **H1** to **H3** and that **H6** to **H8** hold. Then, we have that the local $M-$smoothers based on nearest neighbor with kernel weights satisfy* $\widehat{m}_{\mathrm{M}}(\mathbf{x}) \xrightarrow{a.s.} m(\mathbf{x})$, *for almost all* $\mathbf{x}(\mu)$.

Remark 3.1.1. Using the continuity of the median, similar arguments allow to show the consistency of local medians if $F(y|\mathbf{X} = \mathbf{x})$ has a unique median at $m(\mathbf{x})$, for almost all $\mathbf{x}$.

3.2 Strong Convergence Rates

In order to obtain strong consistency rates we will need some additional regularity conditions.

H10. $F(y|\mathbf{X} = \mathbf{x})$ is Lipschitz in $\mathbf{x}$ uniformly in y, i.e., there exists $\eta > 0$ and $c > 0$ such that $\|\mathbf{u} - \mathbf{x}\| < \eta$ entail $|F(y|\mathbf{X} = \mathbf{u}) - F(y|\mathbf{X} = \mathbf{x})| \le c\|\mathbf{u} - \mathbf{x}\|$ for all y.

H11. $p(\mathbf{x})$ and $f_{\mathbf{X}}(\mathbf{x})$ satisfy a Lipschitz condition of order one.

H12. The function H defined in **H4** satisfies $u^{p+2}H(u)$ is bounded.

H13. $\theta_n^{-1} h_n \le A < \infty$ for all n with $\theta_n = \left(\dfrac{\log n}{n h_n^p}\right)^{\frac{1}{2}}$.

Proposition 3.2.1. *Assume that **H1** to **H5** and that **H10** to **H13** hold. If in addition ψ is continuously differentiable with derivative $\psi\prime$ positive and bounded, we have that the local $M-$smoothers based on kernel weights satisfy* $\theta_n^{-1}|\widehat{m}_{\mathrm{M}}(\mathbf{x}) - m(\mathbf{x})| = O(1)$, *almost surely.*

3.3 Uniform Consistency

We will now derive uniform consistency on a compact set $\mathbf{C} \subset \mathbb{R}^p$, for both kernel or nearest neighbor with kernel estimates. We will consider the following set of assumptions.

A1. $\psi : \mathbb{R} \to \mathbb{R}$ is an odd function, strictly increasing, bounded and continuous differentiable, with bounded derivative $\psi\prime$ such that $\eta(u) = u\,\psi\prime(u) \leq \psi(u)$.

A2. The functions $f_{\mathbf{X}}(\mathbf{x})$ and $p\,(\mathbf{x})$ are bounded functions on $\mathbf{C}$ such that $A_p = \inf_{\mathbf{x}\in\mathbf{C}} p(\mathbf{x}) > 0$ and $A_f = \inf_{\mathbf{x}\in\mathbf{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$. Moreover, $p(\mathbf{x})$ is a continuous function in a neighborhood of $\mathbf{C}$.

A3. $F(y|\mathbf{X}=\mathbf{x})$ is a continuous function of $\mathbf{x}$ in a neighborhood of $\mathbf{C}$. Furthermore, it satisfies the following equicontinuity condition:

$$\forall\epsilon > 0 \quad \exists\delta > 0 : \quad |u - v| < \delta \Rightarrow \sup_{\mathbf{x}\in\mathbf{C}} \left(|F(u|\mathbf{X}=\mathbf{x}) - F(v|\mathbf{X}=\mathbf{x})|\right) < \epsilon\ .$$

A4. The kernel $K : \mathbb{R}^p \to \mathbb{R}$ is a bounded nonnegative function such that $0 < \int K(\mathbf{u})\,d\mathbf{u} < \infty$, $\int |u|K(u)\,du < \infty$, $\|\mathbf{u}\|^p K(\mathbf{u}) \to 0$ as $\|\mathbf{u}\| \to \infty$ and satisfies a Lipschitz condition of order one.

A5. The function $f_{\mathbf{X}}$ is a continuous function in a neighborhood of $\mathbf{C}$.

Remark 3.3.1. This set of assumptions can be divided in three groups. The first one establishes standard conditions on the score function ψ. The second one states regularity conditions on the marginal density of $\mathbf{X}$ and on the conditional distribution function which imply that, for any compact set $\mathbf{C}$, $0 < \inf_{\mathbf{x}\in\mathbf{C}} s(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathbf{C}} s(\mathbf{x}) < \infty$ and that $m(\mathbf{x})$ is a continuous function of $\mathbf{x}$. The third group restricts the class of kernel functions to be chosen and establishes conditions on the rate of convergence of the smoothing parameters, which are standard in nonparametric regression.

The following result ensures uniform consistency of the regression function m, when the smoothing is based either on local medians or local $M-$smoothers.

Proposition 3.3.1. *Assume that **A2** to **A4** hold. Moreover, assume that **H5** holds, for kernel weights and that **A5**, **H8** and **H9** hold for nearest neighbor with kernel weights. Then, for any compact set* $\mathbf{C}$,

a) under **A1** *and* **H2**, *we have that* $\sup_{\mathbf{x}\in\mathbf{C}} |\widehat{m}_{\mathrm{M}}(\mathbf{x}) - m(\mathbf{x})| \xrightarrow{a.s.} 0$,

b) if, in addition, $F(y|\mathbf{X}=\mathbf{x})$ *have a unique median at* $m(\mathbf{x})$, *we have that*

$$\sup_{\mathbf{x}\in\mathbf{C}} |\widehat{m}_{\mathrm{MED}}(\mathbf{x}) - m(\mathbf{x})| \xrightarrow{a.s.} 0. \tag{8}$$

Uniform strong convergence rates can also be derived similarly to the complete sample setting.

3.4 Asymptotic Distribution

We will state the result giving the asymptotic normality of the kernel–based estimates. The result for the nearest neighbor with kernel weights can be derived similarly.

We will derive the asymptotic normality under the following set of assumptions

N1. The kernel $K : \mathbb{R}^d \to \mathbb{R}$ is bounded, nonnegative, $0 < \int K(\mathbf{u})d\mathbf{u} < \infty$, $0 < \int \|\mathbf{u}\|^2 K(\mathbf{u})d\mathbf{u} < \infty$ and $\|\mathbf{u}\|^p K(\mathbf{u}) \to 0$ as $\|\mathbf{u}\| \to \infty$.

N2. There exists $0 \le \beta < \infty$ such that $h_n n^{\frac{1}{p+2}} \to \beta$

N3. There exists a continuous symmetric function F_0 such that the conditional distribution $F(y|\mathbf{X} = \mathbf{u}) = F_0\left(\dfrac{y - m(\mathbf{u})}{\sigma(\mathbf{u})}\right)$ with m and σ such that σ is continuous in a neighborhood of $\mathbf{x}$ and m satisfies a Lipschitz condition of order one and there exists $\lim_{\epsilon \to 0} \dfrac{m(\mathbf{x} + \epsilon\mathbf{u}) - m(\mathbf{x})}{\epsilon} = m\prime(\mathbf{x}, \mathbf{u})$.

N4. The function ψ is twice continuously differentiable with bounded derivatives and with second derivative $\psi\prime\prime$ verifying that there exists positive constants c, M and ϵ such that $\psi\prime\prime(t) \le c|t|^{-2+\epsilon}$ for $|t| > M$.

N5. $A_0(\psi) = \int \psi\prime(u)dF_0(u) \neq 0$

N6. $g(\mathbf{u}) = p(\mathbf{u})\, f_{\mathbf{X}}(\mathbf{u})$ is positive and continuous at $\mathbf{x}$. Moreover, $f_{\mathbf{X}}(\mathbf{u})$ is a bounded function.

Proposition 3.4.1. *Under **H1**, **H2**, **N1** to **N6**, if in addition $\widehat{m}_{\mathrm{M}}(\mathbf{x}) \xrightarrow{p} m(\mathbf{x})$ and $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$, we have that*

$$(nh_n^p)^{\frac{1}{2}} \left(\widehat{m}_{\mathrm{M}}(\mathbf{x}) - m(\mathbf{x})\right) \xrightarrow{\mathcal{D}} N\left(b_1, \frac{\int \psi^2(u)dF_0(u)}{\left[\int \psi\prime(u)dF_0(u)\right]^2} V(\mathbf{x})\right)$$

with

$$\begin{aligned} b_1 &= \beta^{1+\frac{p}{2}} \frac{\int m\prime(\mathbf{x}, \mathbf{u})K(\mathbf{u})d\mathbf{u}}{\int K(\mathbf{u})d\mathbf{u}} \\ V(\mathbf{x}) &= \frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})} \frac{\int K^2(\mathbf{u})d\mathbf{u}}{\left[\int K(\mathbf{u})d\mathbf{u}\right]^2} . \end{aligned}$$

Remark 3.4.1. It is worthwhile noticing that the asymptotic distribution of the simplified local $M-$estimator is analogous to that of the kernel $M-$smoother based on the complete sample except for the factor $p(\mathbf{x})$ appearing in the asymptotic variance that corrects the effect of having missing responses.

4 Asymptotic Properties of the ILMS

4.1 Consistency

We will derive consistency for the kernel estimates defined through (7), under mild conditions on the smoother used to predict the missing observations, the results for nearest neighbor with kernel weights follow similarly.

Proposition 4.1.1. *Assume that **H1** to **H3**, and that the kernel L satisfy **H4** and the bandwidth γ_n verify **H5**. Let $\widehat{m}_{\mathrm{S}}(\mathbf{x})$ be the robust simplified estimator used to predict the missing responses. Assume that for any compact set $\mathbf{C}$, $\sup_{\mathbf{u}\in\mathbf{C}} |\widehat{m}_{\mathrm{S}}(\mathbf{u}) - m(\mathbf{u})| \xrightarrow{a.s.} 0$. Then, if in addition $\widehat{s}(\mathbf{x}) \xrightarrow{a.s.} \sigma(\mathbf{x})$, we have that $\widehat{m}_{\mathrm{M,I}}(\mathbf{x}) \xrightarrow{a.s.} m(\mathbf{x})$.*

Note that **H2**, **H5** and **A2** to **A4** entail that the simplified local $M-$smoother $\widehat{m}_{\mathrm{M,S}}(\mathbf{x}, h)$ can be used as predictor.

4.2 Asymptotic Distribution

We will derive the asymptotic distribution under two different conditions on the robust estimators used to predict the missing responses.

Proposition 4.2.1. *Assume that **H1**, **H2**, **N3** to **N6** hold, that L satisfies **N1** and that there exists $0 \le \beta < \infty$ such that $\gamma_n n^{\frac{1}{p+2}} \to \beta$. Moreover, assume that L has compact support and that $\widehat{m}_{\mathrm{M,I}}(\mathbf{x}) \xrightarrow{p} m(\mathbf{x})$ and $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$.*

i) If for any compact neighborhood $\mathbf{C}$ of $\mathbf{x}$, $\widehat{\upsilon}(\mathbf{C}) = (n\gamma_n^p)^{\frac{1}{2}} \sup_{\mathbf{u}\in\mathbf{C}} |\widehat{m}_{\mathrm{S}}(\mathbf{u}) - m(\mathbf{u})| = o_p(1)$, we have that

$$(n\gamma_n^p)^{\frac{1}{2}} \left(\widehat{m}_{\mathrm{M,I}}(\mathbf{x}) - m(\mathbf{x})\right) \xrightarrow{\mathcal{D}} N\left(b_1, \frac{\int \psi^2(u) dF_0(u)}{\left[p(\mathbf{x})A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)\right]^2} V(\mathbf{x})\right)$$

with

$$\begin{aligned} b_1 &= \beta^{1+\frac{p}{2}} \frac{\int m\prime(\mathbf{x},\mathbf{u}) L(\mathbf{u}) d\mathbf{u}}{\int L(\mathbf{u}) d\mathbf{u}} \\ V(\mathbf{x}) &= \frac{\sigma^2(\mathbf{x})\, p(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \frac{\int L^2(\mathbf{u}) d\mathbf{u}}{\left[\int L(\mathbf{u}) d\mathbf{u}\right]^2} \end{aligned}$$

*ii) Assume now that $\widehat{m}_{\mathrm{M,S}}(\mathbf{u})$ is the simplified local $M-$smoother defined in (6) with score function ψ_1 satisfying **H1**, **N4** and **N5**, bandwidth h_n and kernel K with compact support satisfying **N1**. Moreover, assume that $\dfrac{\gamma_n}{h_n} \to \kappa \neq 0$, so that, for any compact neighborhood $\mathbf{C}$ of $\mathbf{x}$, $\widehat{\upsilon}(\mathbf{C}) = (n\gamma_n^p)^{\frac{1}{2}} \sup_{\mathbf{u}\in\mathbf{C}} |\widehat{m}_{\mathrm{M,S}}(\mathbf{u}) - m(\mathbf{u})| = O_p(1)$. Denote $\Delta(\mathbf{x},\mathbf{u},\mathbf{v}) = m\prime(\mathbf{x}, \mathbf{u} + \kappa\mathbf{v}) -$*

$\kappa m\prime(\mathbf{x},\mathbf{v})$ *and* $\Gamma(\mathbf{v},a) = \int L(\mathbf{u})\, K((\mathbf{v}-\mathbf{u})\, a)\, d\mathbf{u}$. *If, in addition,* $\sup_{\mathbf{u}\in\mathbf{C}} |\widehat{s}(\mathbf{u}) - \sigma(\mathbf{u})| \xrightarrow{p} 0$, *we have that*

$$(n\gamma_n^p)^{\frac{1}{2}}\left(\widehat{m}_{\mathrm{M,I}}(\mathbf{x}) - m(\mathbf{x})\right) \xrightarrow{\mathcal{D}} N\left(b_1, \frac{\int \psi^2(u)dF_0(u)}{\left[p(\mathbf{x})A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)\right]^2} V(\mathbf{x})\right)$$

with

$$b_1 = \beta^{\frac{p}{2}+1}\left\{\frac{\int L(\mathbf{v})\, m\prime(\mathbf{x},\mathbf{v})d\mathbf{v}}{\int L(\mathbf{v})d\mathbf{v}} + \frac{\kappa^{-1}(1-p(\mathbf{x}))\psi\prime(0)\int L(\mathbf{v})\, K(\mathbf{u})\, \Delta(\mathbf{x},\mathbf{u},\mathbf{v})d\mathbf{v}d\mathbf{u}}{\left[p(\mathbf{x})A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)\right]\int K(\mathbf{u})d\mathbf{u}\int L(\mathbf{v})d\mathbf{v}}\right\}$$

$$V(\mathbf{x}) = \frac{\sigma^2(\mathbf{x})\, p(\mathbf{x})\int\left[L(\mathbf{v})\,\psi(\epsilon) + \kappa^p \frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\, p(\mathbf{x})}\psi\prime(0)\,\psi_1(\epsilon)\Gamma(\mathbf{v},\kappa)\right]^2 dF_0(\epsilon)\, d\mathbf{v}}{f_{\mathbf{X}}(\mathbf{x})\int \psi^2(u)dF_0(u)\left[\int L(\mathbf{u})d\mathbf{u}\right]^2}$$

Remark 4.2.1. Note that the asymptotic behavior of the imputed $M-$estimator depends on the rate of convergence of the initial simplified estimator. If the initial estimate has a higher rate of convergence, then the asymptotic bias does not depend on the score function, only the asymptotic variance depends on the score function used and the efficiency involves now the value $\psi\prime(0)$ weighted with the probability of having missing observations. On the other hand, if the simplified $M-$estimator has rate of convergence $(n\gamma_n^p)^{\frac{1}{2}}$ the bias depend on th score function used to compute the imputed estimate while the asymptotic variance depend on the score functions used in both steps. In particular, if the same score function is used, i.e., $\psi = \psi_1$, the expression for the asymptotic variance reduces to $\int \psi^2(u)dF_0(u)[p(\mathbf{x})A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)]^{-2}V(\mathbf{x})$ with

$$V(\mathbf{x}) = \frac{\sigma^2(\mathbf{x})\, p(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}\, \frac{\int\left[L(\mathbf{v}) + \kappa^p \frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\, p(\mathbf{x})}\psi\prime(0)\,\Gamma(\mathbf{v},\kappa)\right]^2 d\mathbf{v}}{\left[\int L(\mathbf{u})d\mathbf{u}\right]^2}.$$

5 Monte Carlo Study

This section contains the results of a simulation study, in dimension $p = 1$, designed to evaluate the performance of the robust procedure defined in Section 2 when there are missing observations in the response variable. The S–code is available upon request to the authors. The aims of this study are

- to compare the behavior of the classical and robust estimators under contamination and under normal samples.
- to study the behavior of the two robust proposals, simplified and imputed, among them and compared to that of the robust nonparametric regression estimator with complete data.

5.1 General Description

Once the smoothing parameter was selected, we performed 1000 replications generating independent samples of size $n = 100$ following the model

$$z_i = 0.25\,\pi\ \sin(\pi x_i) + \varepsilon_i, \qquad 1 \leq i \leq n$$

where $x_i \sim \mathcal{U}(0,1)$, ε_i are i.i.d. and independent of x_i, $\varepsilon_i \sim (1-\alpha)N\left(0,\sigma^2\right) + \alpha N\left(0, 25\sigma^2\right)$ with $\sigma = 0.5$. We considered three contamination proportions $\alpha = 0$, 0.1 and 0.2, the first one corresponding to the central normal model.

We then define $y_i = z_i$ if $\delta_i = 1$ and missing otherwise to obtain the missing responses, where the model for the missing probability considered, (2), is $p\,(x) = 0.3 + 0.5\,(\sin(5\,(x+0.2)))^2$ which gives a proportion of missing data in each sample near the 40%.

The robust smoothing procedure uses local $M-$estimates with bisquare score function, with tuning constant 4.685, and local medians as initial estimate. In both the classical and the robust estimators, we have used the gaussian kernel with standard deviation $\frac{0.25}{0.675} = 0.37$ such that the interquartile range is 0.5.

The estimators considered were:

- The Nadaraya–Watson and the local $M-$estimates with the complete data set, denoted respectively, $\widehat{m}_{\text{LS,C}}$ and $\widehat{m}_{\text{M,C}}$, in Tables and Figures
- The simplified version of the Nadaraya–Watson and $M-$estimates, denoted respectively, $\widehat{m}_{\text{LS,S}}$ and $\widehat{m}_{\text{M,S}}$,
- The imputed Nadaraya–Watson and $M-$estimates with predictors the simplified ones, denoted respectively, $\widehat{m}_{\text{LS,I}}$ and $\widehat{m}_{\text{M,I}}$.

5.2 Selection of the smoothing parameter

The smoothing parameters was selected for each of these estimators and for each contamination using as goodness of fit criterium the mean integrated square error, MISE,

$$\text{MISE}\,(h) = \text{E}\int \left(m\,(x) - \widehat{m}_h\,(x)\right)^2 dx,$$

where $\widehat{m}_h$ denotes the estimator to be considered (classical or robust, with the complete data set, simplified or imputed).

We performed 100 replications generating independent samples of size $n = 100$ following the model described above. For each value of the smoothing parameter , the value of the MISE was approximated by Monte Carlo as $\dfrac{1}{100}\sum_{k=1}^{100} M(h,k)$, where for each replication k,

$$M(h,k) = \frac{1}{\ell}\sum_{j=1}^{\ell}\left(m\,(v_j) - \widehat{m}_h\,(v_j)\right)^2$$

with $v_j = j/\ell$, $1 \leq j \leq \ell$, $\ell = 50$. The smoothing parameter h was selected on a grid of 20 points in $[0.2, 0.4]$ for the complete and the simplified estimators while for the imputed one, the minimization for each parameter h and γ, was carried out over a two dimensional grid on $[0.2, 0.4] \times [0.2, 0.4]$. When the minimization process, leads to a value on the boundary the search was carried on over the limits of the interval. Table 1 reports the values obtained in each situation.

5.3 Results

Once the smoothing parameters were obtained, we compared the different estimators for sample sizes $n = 100$, under the scenario described in Section 5.1. The performance of an estimate $\widehat{m}$ of m is measured using two measures computed over the replications

$$\text{ISE}(\widehat{m}) = \frac{1}{\ell} \sum_{j=1}^{\ell} \left(m\left(v_j\right) - \widehat{m}\left(v_j\right)\right)^2$$

and $\text{MSE}(\widehat{m}, x) = \dfrac{1}{1000} \displaystyle\sum_{k=1}^{1000} \text{M}(\widehat{m}, x, k)$ where for each replication k, $\text{M}(\widehat{m}, x, k) = \left(m\left(x\right) - \widehat{m}\left(x\right)\right)^2$. As above $v_j = j/\ell$, $1 \leq j \leq \ell$, $\ell = 50$. An approximation to the MISE was obtained as the mean over the 1000 replications of ISE. Note that $\text{ISE}(\widehat{m})$ is simply the value of $M(h_0, k)$ obtained for the bandwidth h_0 at the k/th replication, where h_0 denotes the smoothing parameter used in the estimation procedure and reported in Table 1.

Tables 2 to 4 summarize the results of the simulations. Table 2 gives the values of the MISE for the linear and robust nonparametric estimators for complete data and when considering the simplified and imputed estimators.

The reported efficiency of the robust estimators with respect to their linear relatives was computed as

$$\text{EF}_{\text{LS,M}} = \frac{\text{MISE}_{\text{LS}} - \text{MISE}_{\text{M}}}{\text{MISE}_{\text{LS}}} \times 100\ .$$

where MISE_{LS} denotes the MISE of the local linear estimator and MISE_{M} that of the local $M-$smoother for each method, complete data, simplified or imputed one that will be indicated, respectively, by C, S and I after the comma.

Table 3 shows the percentage of times that the ISE for robust estimators is less than the ISE of the classic estimators.

Table 4 gives the percentage of times that the imputed robust estimator is less than the ISE of the simplified together with the efficiency of the first with respect to the latter

$$\text{EF}_{\text{M; S,I}} = \frac{\text{MISE}_{\text{M,S}} - \text{MISE}_{\text{M,I}}}{\text{MISE}_{\text{M,S}}} \times 100\ ,$$

with $\text{MISE}_{\text{M,S}}$ the MISE of the local simplified $M-$estimator and $\text{MISE}_{\text{M,I}}$ that of the imputed local $M-$smoother.

Figures 1 to 3, present the density estimation of ratio between ISE for robust estimator and that of the classical estimators when $\alpha = 0$, 0.1 and 0.2 for the complete data and the simplified and imputed estimators, respectively. In order to compare the simplified and imputed robust estimators, Figure 4 shows the density estimator ratio between the ISE for the imputed robust estimator and that of the simplified robust one while Figure 6 represents the ratio betwen the MSE of both estimators across the values of x. The density estimates were evaluated using the normal kernel with bandwidth 0.6 in all cases.

Finally, Figure 7 shows the boxplots of the ISE for the simplified and imputed robust estimators.

The results reported in Tables 2 to 4 show that when there are no contamination, the linear estimator performs better than the robust ones that show a loss of efficiency related to that of the $M-$smoother used. On the other hand, the performance of the classical Nadaraya–Watson estimator is highly sensitive to the presence of outliers in the sample. The MISE increases with the contamination level. For instance, when $\alpha = 0.2$, the MISE of the linear estimator computed with complete data is almost four times that observed under no contamination. The robust estimators are much more estable under contamination increasing at most a 50% their MISE. This explains the better efficiency observed in Table 2 for the robust estimators as contamination increases. On the other hand, this is also reflected in Table 3 that shows that under contamination the robust procedure reaches more than 70% of the times lower ISE than the linear estimates. Note also, that when there is no contamination only 60% of the times the classical estimator is better than the robust one been this fact related to the efficiency of the $M-$smoother.

From Figures 1 to 3, it should be noticed that when there is no contamination ($\alpha = 0$) the behavior of all estimators is quite similar, improving, in some cases, the robust procedure the performance of the classical ones as it was observed in Table 3. However, as the contamination increases, Figures 1 to 3 clarify the phenomena observed in Table 3 with respect to the better performance of the robust estimators, since the density functions move towards the left of the point 1. Notice that the behavior of ratio is less than 1 as long as α increases.

Figures 4 to 8 show that the imputed estimator has a better performance than the simplified, both at each point and globally. Moreover, its behavior improves as the contamination α increases.

5.4 Data–driven selection of the smoothing parameter

An important issue in any smoothing procedure is the choice of the smoothing parameter. Under a nonparametric regression model, two commonly used approaches are cross–validation and plug–in. However, these procedures may not be robust and their sensitivity to anomalous data was discussed by several authors, including Leung, Marrot and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997), Cantoni and Ronchetti (2001) and and Leung (2005). Wang and Scott (1994) note that, in the presence of outliers, the least squares cross–validation function is nearly constant on its whole domain and thus, essentially worthless for the purpose of choosing a bandwidth. The robustness issue remains for the estimators considered in this paper, specially since we are dealing with missing responses. With a small bandwidth, a small number of outliers with similar values of $\mathbf{x}_i$ could easily drive the estimate of m to dangerous levels. Therefore, we may consider a robust cross-validation approach analogous to that described in Leung (2005) that takes into account the missing observations. To avoid a complete search in both parameters (h_n, γ_n), in the first step, we choose the smoothing parameter for the simplified local $M-$estimator as follows

- For each given h, compute $\widehat{m}_{\mathrm{M,S}}^{-i}(\mathbf{x}, h)$ as the solution of

$$\sum_{j\neq i}^{n} w_j(\mathbf{x}, h)\psi\left(\frac{y_i - a}{\widehat{s}_{-i}(\mathbf{x})}\right) = 0 \ ,$$

i.e., the estimator computed without the $i-$th observation where

$$w_j(\mathbf{x}, h) = \frac{K\left(\dfrac{\mathbf{x}_j - \mathbf{x}}{h}\right)\delta_j}{\sum_{\ell=1}^{n} K\left(\dfrac{\mathbf{x}_\ell - \mathbf{x}}{h}\right)\delta_\ell}, .$$

- Calculate

$$RCV_{1,\mathrm{S}}(h) = \sum_{i=1}^{n} \delta_i \psi_H^2\left(\frac{\widehat{u}_i(h)}{\widehat{\sigma}_n}\right) w\left(\mathbf{x}_i\right)$$

where ψ_H is the Huber's function with tunning constant 1.345, $\widehat{u}_i(h) = y_i - \widehat{m}_{\mathrm{M,S}}^{-i}(\mathbf{x}_i, h)$, $\widehat{\sigma}_n$ is an estimator of the error's scale that does relatively little smoothing (see Cantoni and Ronchetti (2001) for the complete case), $w\left(\mathbf{x}_i\right)$ is a function to control boundary effects. As estimator of the scale and since we are dealing with an homocedastic model we can take $\widehat{\sigma}_n = \mathrm{median}_i\, |y_{i+1} - y_i|/(0.6745\sqrt{2})$ for the complete case if $\mathbf{x} \in I\!R$, $\mathbf{x}_1 \leq \ldots \leq \mathbf{x}_n$ or a robust scale estimator using the observations at hand and computed using a preliminary regression estimator. To be more precise, let $\widehat{m}_{\mathrm{MED,S}}(\mathbf{x}, h_0)$ the simplified local median computed with a pilot bandwidth h_0 and $\widehat{u}_i(h_0) = y_i - \widehat{m}_{\mathrm{MED,S}}(\mathbf{x}_i, h)$, then $\widehat{\sigma}_n$ can be taken as the median of the absolute deviation with respect to the median of the residuals $\widehat{u}_i(h_0)$, i.e., $\widehat{\sigma}_n = \mathrm{MAD}_i\left(\widehat{u}_i(h_0)\right)$. The function w is a function that can be taken as $w \equiv 1$ or as

$$w(x) = \begin{cases} 1 & \text{if } \left|\dfrac{x - m_x}{s_x}\right| < 3 \\ 0 & \text{otherwise} \ , \end{cases}$$

with $m_x = \mathrm{median}_i(x_i)$ and $s_x = \mathrm{MAD}(x_i)$ to avoid boundary points to be influential in the selection of the smoothing parameter.

- Choose $\widehat{h}_n = \underset{h}{\mathrm{argmin}}\, RCV_{1,\mathrm{S}}(h)$.

With this data–driven bandwidth, to complete the sample the missing observations are imputed as $\widehat{y}_i = \delta_i\, y_i + (1 - \delta_i)\widehat{m}_{\mathrm{S}}\left(\mathbf{x}_i, \widehat{h}_n\right)$. In the second step, to select γ_n, we apply the robust cross–validation procedure described in Leung (2005) to the completed sample, i.e, we select $\widehat{\gamma}_n = \underset{h}{\mathrm{argmin}}\, RCV_{1,\mathrm{I}}(\gamma)$ with

$$RCV_{1,\mathrm{I}}(\gamma) = \sum_{i=1}^{n} \psi_H^2\left(\frac{\widehat{\widehat{u}}_i(\gamma)}{\widehat{\sigma}_n}\right) w\left(\mathbf{x}_i\right)$$

where $\widehat{\widehat{u}}_i(\gamma) = \widehat{y}_i - \widehat{m}_{\mathrm{M,I}}^{-i}(\mathbf{x}_i, \gamma)$ and $\widehat{m}_{\mathrm{M,I}}^{-i}(\mathbf{x}, \gamma)$ is the solution of

$$\sum_{j\neq i}^{n} L\left(\frac{\mathbf{x}_j - \mathbf{x}}{\gamma}\right)\psi\left(\frac{\widehat{y}_j - a}{\widehat{s}(\mathbf{x})}\right) = 0 \ .$$

It is worth noticing that this method is not a direct application of the cross–validation criteria but it reduces considerably the computations and our results show that it worked favorably.

We have also considered a cross–validation criterium analogous to that defined in Bianco and Boente (2006). As mentioned by these authors, the cross–validation criterium tries to provide a measure both of bias and variance, and so it would make sense to introduce a new measure that establishes a trade–off between bias and variance. The robust cross–validation criterium when there are no missing observations is thus based on a robust estimator of the bias, defined through a location estimator μ_n, and on a robust scale estimator σ_n, as follows,

$$RCV_2(h) = \mu_n^2 \left(\widehat{u}_i(h) w\left(\mathbf{x}_i\right)\right) + \sigma_n^2 \left(\widehat{u}_i(h) w\left(\mathbf{x}_i\right)\right) \ .$$

For the situation we are dealing with, it is enough, in the first step, to compute RCV_2 with the observations at hand, i.e, using only the complete observations for the estimation and discarding the incomplete vectors. In the second step we proceed as described above. We can consider as μ_n the median and as σ_n the MAD, the bisquare a–scale estimator or the Huber $\tau-$scale estimator, in our simulation study we choose the , the $\tau-$scale estimator.

For this preliminary study, the search for the bandwidth parameter was performed searching over a grid of step 0.02 on the interval $[0.05, 0.99]$, for the complete estimators and on the interval $[0.1, 0.99]$, for the simplified and imputed one. So, too small or too large bandwidths are not allowed in this procedure. Due to the expensive computing time, we have only performed 200 replications. We have also evaluate the performance of the classical estimators using least squares cross–validation.

Table 7 shows the percentage of times that the ISE for the robust imputed estimator is less than the ISE for the robust simplified estimator when using robust cross–validation together with the efficiency of the first with respect to the latter while Table 5 shows the number of times that the ISE for the robust estimators is less than that of their linear counterparts. The results given in Table 5 show that RCV_2 is slightly better than RCV_1. Figure 5 plots the density estimator of the ratio between the ISE of the imputed robust estimator and that of the simplified robust one when using the two robust cross–validation criterium while Figure 9 shows the boxplots of the ISE for the simplified and the imputed robust estimators. The results are quite similar to those obtained with the fixed bandwdiths and the comments given in Section 5.3 still hold.

On the other hand, to study the sensitivity of the data selector to the contamination considered, Figure 10 shows the boxplots of $\log(\widehat{h}_{n,C_0}/\widehat{h}_{n,C_1})$ and $\log(\widehat{h}_{n,C_0}/\widehat{h}_{n,C_2})$ for each scenario. These boxplots show the sensitivity of least squares cross–validation and the stability of the robust bandwidth selector when contaminating 10% or 20% of the data. The results being quite similar with both robust cross–validation measures.

6 Concluding Remarks

We have introduced two robust procedures to estimate the regression function when there are missing obervations in the response variable and it can be suspected that anomalous observations are present in the sample. Both procedures are strongly consistent and asymptotically normally distributed.

Under the contaminations considered, they show their advantage over the Nadaraya-Watson estimators. Moreover, the imputed local $M-$estimator, even if it is computationally more expensive, should be used, since it performs better as contamination increases.

Acknowledgment

This work began while Graciela Boente was visiting the Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela. This research was partially supported by Grant PID 5505 from CONICET, Grant X-094 from the Universidad de Buenos Aires and Grant 13900-6 from the Fundación Antorchas at Buenos Aires, Argentina and also by the DGICYT Spanish Grant MTM2005-00820 (European FEDER support included).

A Appendix: Proofs.

PROOF OF PROPOSITION 3.1.1. Using Theorem 2.2 in Boente and Fraiman (1989), it will be enough to show that

$$\sup_{y\in \mathbb{R}} \left|\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})\right| \xrightarrow{a.s.} 0 \quad \text{for almost all } \mathbf{x}\ ,$$

which will follow easily, if we show that for any measurable $A \subset \mathbb{R}$

$$\widehat{\phi}_A(\mathbf{x}) \xrightarrow{a.s.} \phi_A(\mathbf{x}) \quad \text{for almost all } \mathbf{x}\ ,$$

where

$$\begin{aligned} \widehat{\phi}_A(\mathbf{x}) &= \sum_{i=1}^n w_i(\mathbf{x}) I_A(y_i)\,, \\ \phi_A(\mathbf{x}) &= P(Y\in A|\mathbf{X}=\mathbf{x})\ . \end{aligned}$$

Note that $\widehat{\phi}_A(\mathbf{x}) = \dfrac{\widehat{r}_A(\mathbf{x})}{\widehat{p}(\mathbf{x})}$ where

$$\widehat{r}_A(\mathbf{x}) = \sum_{i=1}^n W_{i,n}(\mathbf{x})\delta_i I_A(y_i)\,, \tag{A.1}$$

$$\widehat{p}(\mathbf{x}) = \sum_{i=1}^n W_{i,n}(\mathbf{x})\delta_i\,, \tag{A.2}$$

$$W_{i,n}(\mathbf{x}) = \frac{K\left(\dfrac{\mathbf{x}_i-\mathbf{x}}{h_n}\right)}{\displaystyle\sum_{j=1}^n K\left(\dfrac{\mathbf{x}_j-\mathbf{x}}{h_n}\right)}. \tag{A.3}$$

Theorem 1 of Greblicki, Krzyzak and Pawlak (1984) entails that

$$\begin{aligned}\widehat{r}_A(\mathbf{x}) &\xrightarrow{a.s.} p(\mathbf{x})\phi_A(\mathbf{x}) \quad \text{for almost all } \mathbf{x}\,,\\ \widehat{p}(\mathbf{x}) &\xrightarrow{a.s.} p(\mathbf{x}) \quad \text{for almost all } \mathbf{x}\,,\end{aligned}$$

which concludes the proof. □

Proof of Proposition 3.1.2. The proof follows as that of Proposition 3.1.1 using Proposition 2 in Collomb (1980). □

Proof of Proposition 3.2.1. The proof follows from Theorem 2.3 in Boente and Fraiman (1988) if we show that

$$\theta_n^{-1} \sup_{y\in I\!R} \left|\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})\right| = O(1)$$

which is a consequence of Lemma 2.1 in Boente and Fraiman (1990b) using that

$$\sup_{y\in I\!R} |\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})| \le \frac{\sup_{y\in I\!R} |\widehat{r}(y,\mathbf{x}) - r(y,\mathbf{x})| + |\widehat{p}(\mathbf{x}) - p(\mathbf{x})|}{p(\mathbf{x})\widehat{p}(\mathbf{x})}$$

with $\widehat{r}(y,\mathbf{x}) = \widehat{\phi}_{(-\infty,y]}(\mathbf{x})$, where $\widehat{\phi}_{(-\infty,y]}(\mathbf{x})$ and $\widehat{p}(\mathbf{x})$ defined in (A.1) and (A.2). □

Proof of Proposition 3.3.1. a) Arguing as in Theorem 3.3 in Boente and Fraiman (1991) we will only need to show that

$$\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})| \xrightarrow{a.s.} 0,. \tag{A.4}$$

Theorems 3.1 or 3.2 from Boente and Fraiman (1991), entail that

$$\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{r}(y,\mathbf{x}) - r(y,\mathbf{x})| \xrightarrow{a.s.} 0, \tag{A.5}$$

$$\sup_{\mathbf{x}\in\mathbf{C}} |\widehat{p}(\mathbf{x}) - p(\mathbf{x})| \xrightarrow{a.s.} 0, \tag{A.6}$$

where $r(y,\mathbf{x}) = \phi_{(-\infty,y]}(\mathbf{x}) = p(\mathbf{x})\,F(y|\mathbf{X}=\mathbf{x})$, $\widehat{r}(y,\mathbf{x}) = \widehat{\phi}_{(-\infty,y]}(\mathbf{x})$, with $\widehat{\phi}_{(-\infty,y]}(\mathbf{x})$ and $\widehat{p}(\mathbf{x})$ defined in (A.1) and (A.2), respectively. The weights $W_{i,n}$ are the kernel weights given by (A.3) or the nearest with kernel weights

$$W_{i,n}(\mathbf{x}) = \frac{K\left(\dfrac{\mathbf{x}_i - \mathbf{x}}{H_n(\mathbf{x})}\right)}{\sum_{j=1}^{n} K\left(\dfrac{\mathbf{x}_j - \mathbf{x}}{H_n(\mathbf{x})}\right)}\,.$$

Note that (A.6) can be derived for kernel weights using Proposition 2 in Collomb (1979). Now, (A.4) follows using **A2** and the inequality

$$\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})| \le \frac{\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{r}(y,\mathbf{x}) - r(y,\mathbf{x})| + \sup_{\mathbf{x}\in\mathbf{C}} |\widehat{p}(\mathbf{x}) - p(\mathbf{x})|}{A_p \widehat{A}_p}$$

where $A_p = \inf_{\mathbf{x}\in\mathbf{C}} p(\mathbf{x})$ and $\widehat{A}_p = \inf_{\mathbf{x}\in\mathbf{C}} \widehat{p}(\mathbf{x})$.

b) The equicontinuity condition required in **A3** and the uniqueness of the conditional median imply that $m(\mathbf{x})$ is a continuous function of $\mathbf{x}$ and thus, for any fixed $a \in I\!R$ the function $h_a(\mathbf{x}) = F\left(a + m(\mathbf{x})|\mathbf{X} = \mathbf{x}\right)$ will also be continuous as a function of $\mathbf{x}$.

Given $\epsilon > 0$, let $0 < \delta < \epsilon$ be such that

$$|u - v| < \delta \Rightarrow \sup_{\mathbf{x}\in\mathbf{C}} \left(|F(u|\mathbf{X} = \mathbf{x}) - F(v|\mathbf{X} = \mathbf{x})|\right) < \frac{\epsilon}{2}\,. \tag{A.7}$$

Then, from the uniqueness of the conditional median and (A.7) we get that,

$$\frac{1}{2} < \quad F\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right) \quad < \frac{1}{2} + \frac{\epsilon}{2} \tag{A.8}$$

$$\frac{1}{2} - \frac{\epsilon}{2} < \quad F\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right) \quad < \frac{1}{2}\,. \tag{A.9}$$

Write $\iota(\delta) = \inf_{\mathbf{x}\in\mathbf{C}} F\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right)$ and $\nu(\delta) = \sup_{\mathbf{x}\in\mathbf{C}} F\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right)$.

The continuity of $h_\delta(\mathbf{x})$ and $h_{-\delta}(\mathbf{x})$ together with (A.8) and (A.9) entail that, $\nu(\delta) < \dfrac{1}{2} < \iota(\delta)$ and thus $\eta = \min\left(\iota(\delta) - \dfrac{1}{2}, \dfrac{1}{2} - \nu(\delta)\right) > 0$.

Since (A.4) holds, let $\mathcal{N}$ be such that $P(\mathcal{N}) = 0$ and for any $\omega \notin \mathcal{N}$, $\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{F}(y|\mathbf{X} = \mathbf{x}) - F(y|\mathbf{X} = \mathbf{x})| \to 0$. Thus, for n large enough we have that $\sup_{\mathbf{x}\in\mathbf{C}} \sup_{y\in I\!R} |\widehat{F}(y|\mathbf{X} = \mathbf{x}) - F(y|\mathbf{X} = \mathbf{x})| < \min\left(\dfrac{\eta}{2}, \dfrac{\epsilon}{2}\right) = \epsilon_1$. Therefore, for $\mathbf{x} \in \mathbf{C}$, we have that

$$\begin{aligned} F\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right) - \epsilon_1 < \quad & \widehat{F}\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right) \quad < F\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right) + \epsilon_1 \\ F\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right) - \epsilon_1 < \quad & \widehat{F}\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right) \quad < F\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right) + \epsilon_1\,, \end{aligned}$$

which entail that $\frac{1}{2} < \widehat{F}\left(m(\mathbf{x}) + \delta|\mathbf{X} = \mathbf{x}\right) < \frac{1}{2} + \epsilon$ and $\frac{1}{2} - \epsilon < \widehat{F}\left(m(\mathbf{x}) - \delta|\mathbf{X} = \mathbf{x}\right) < \frac{1}{2}$ and so, $\sup_{\mathbf{x}\in\mathbf{C}} |\widehat{m}_{\mathrm{MED}}(\mathbf{x}) - m(\mathbf{x})| \leq \delta < \epsilon$ which concludes the proof. $\square$

PROOF OF PROPOSITION 3.4.1. Using a Taylor's expansion of order one, we get that

$$\widehat{m}_{\mathrm{M}}(\mathbf{x}) - m(\mathbf{x}) = \widehat{s}(\mathbf{x}) A_{0,n}^{-1}(\mathbf{x}) A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x}))$$

with

$$\begin{aligned} A_{0,n}(\mathbf{x}) &= \frac{1}{nh_n^p} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi\prime\left(\frac{y_i - \xi(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \\ A_{1,n}(\mathbf{x}, \sigma) &= \frac{1}{nh_n^p} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi\left(\frac{y_i - m(\mathbf{x})}{\sigma}\right) \end{aligned}$$

where $\xi(\mathbf{x})$ is an intermediate point. It is enough to show that

a) $A_{0,n}(\mathbf{x}) \xrightarrow{p} p(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) A_0(\psi) \int K(\mathbf{u}) d\mathbf{u}$

b) $(nh_n^p)^{\frac{1}{2}} \left(\widetilde{A}_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) - \widetilde{A}_{1,n}(\mathbf{x}, \sigma(\mathbf{x})) \right) \xrightarrow{p} 0$

c) $(nh_n^p)^{\frac{1}{2}} \left(A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) - \widetilde{A}_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) \right) \xrightarrow{p} a_1$

d) $(nh_n^p)^{\frac{1}{2}} \widetilde{A}_{1,n}(\mathbf{x}, \sigma(\mathbf{x})) \xrightarrow{\mathcal{D}} N(0, \sigma_1)$,

where

$$\begin{aligned}
\widetilde{A}_{1,n}(\mathbf{x}, \sigma) &= \frac{1}{nh_n^p} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi\left(\frac{y_i - m(\mathbf{x}_i)}{\sigma}\right) \\
a_1 &= b_1 \frac{p(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) A_0(\psi) \int K(\mathbf{u}) d\mathbf{u}}{\sigma(\mathbf{x})} = \beta^{1+\frac{p}{2}} \frac{p(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) A_0(\psi) \int m\prime(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) d\mathbf{u}}{\sigma(\mathbf{x})} \\
\sigma_1 &= p(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \int K^2(\mathbf{u}) d\mathbf{u} \int \psi^2(t) dF_0(t) .
\end{aligned}$$

a)and b) follows as in Boente and Fraiman (1990a). c) follows using a Taylor's expansion of order two. Effectively, denote $Z_i(\sigma) = \delta_i \psi' \left(\frac{\epsilon_i \sigma(\mathbf{x}_i)}{\sigma}\right) [m(\mathbf{x}_i) - m(\mathbf{x})]$, $Z_i = \delta_i \psi'(\epsilon_i) [m(\mathbf{x}_i) - m(\mathbf{x})]$. We have the following expansion

$$\begin{aligned}
(nh_n^p)^{\frac{1}{2}} \left(A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) - \widetilde{A}_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) \right) &= \frac{1}{\widehat{s}(\mathbf{x})} \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i Z_i(\widehat{s}(\mathbf{x})) + \\
&+ \frac{1}{2} \frac{1}{\widehat{s}(\mathbf{x})^2} \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi''\left(\frac{\xi_i}{\widehat{s}(\mathbf{x})}\right) [m(\mathbf{x}_i) - m(\mathbf{x})]^2 \\
&= \frac{1}{\widehat{s}(\mathbf{x})} \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i Z_i + \\
&+ \frac{1}{2} \frac{1}{\widehat{s}(\mathbf{x})^2} \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i [\sigma(\mathbf{x}_i) - \sigma(\mathbf{x})] \psi''\left(\frac{\theta_i}{\widehat{s}(\mathbf{x})}\right) [m(\mathbf{x}_i) - m(\mathbf{x})] \\
&+ \frac{1}{2} \frac{1}{\widehat{s}(\mathbf{x})^2} [\sigma(\mathbf{x}) - \widehat{s}(\mathbf{x})] \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi''\left(\frac{\theta_i}{\widehat{s}(\mathbf{x})}\right) [m(\mathbf{x}_i) - m(\mathbf{x})] \\
&+ \frac{1}{2} \frac{1}{\widehat{s}(\mathbf{x})^2} \frac{1}{\sqrt{nh_n^p}} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h_n}\right) \delta_i \psi''\left(\frac{\xi_i}{\widehat{s}(\mathbf{x})}\right) [m(\mathbf{x}_i) - m(\mathbf{x})]^2 \\
&= S_{1n} + S_{2n} + S_{3n} + S_{4n}
\end{aligned}$$

with θ_i and ξ_i intermediate points. Using the boundness of ψ'', the Lipschitz continuity of m and the continuity of $\sigma()$ together with the consistency of the scale estimator, **N1** and **N2** we get that $S_{jn} \xrightarrow{p} 0$ for $2 \leq j \leq 4$. On the other hand, using that $\lambda(\mathbf{u}) = E(Z_1 | \mathbf{X}_1 = \mathbf{u}) = p(\mathbf{u}) A_0(\psi) [m(\mathbf{u}) - m(\mathbf{x})]$, we get that $S_{1n} \xrightarrow{p} a_1$.

In order to prove d) denote by $Z_i = \delta_i \psi\left(\dfrac{y_i - m(\mathbf{x}_i)}{\sigma(\mathbf{x})}\right)$, $E(Z_1|\mathbf{X}_1) = 0$. Therefore, the results follows using the asymptotic distribution for the classical Nadaraya–Watson estimates for bounded variables applied to $(\mathbf{x}_i, Z_i)$ (see, for instance, Theorem 2 in Schuster (1972)). □

Proof of Proposition 4.1.1. Let $W_{i,n,\mathrm{L}}(\mathbf{x})$ be the kernel weights defined as

$$W_{i,n,\mathrm{L}}(\mathbf{x}) = \frac{L\left(\dfrac{\mathbf{x}_i - \mathbf{x}}{h_n}\right)}{\sum_{j=1}^{n} L\left(\dfrac{\mathbf{x}_j - \mathbf{x}}{h_n}\right)}. \tag{A.10}$$

Then, the imputed estimate are the solution of

$$\int \psi\left(\frac{y - \widehat{m}_{\mathrm{M,I}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) d\widehat{\widehat{F}}_n(y|\mathbf{X} = \mathbf{x}) = 0$$

where

$$\widehat{\widehat{F}}_n(y|\mathbf{X} = \mathbf{x}) = \sum_{i=1}^{n} W_{i,n,\mathrm{L}}(\mathbf{x}) I_{(-\infty,y]}(\widehat{y}_i)\ .$$

Denote by $\widetilde{F}(y|\mathbf{X} = \mathbf{x}) = p(\mathbf{x})F(y|\mathbf{X} = \mathbf{x}) + (1 - p(\mathbf{x}))\Delta_{m(\mathbf{x})}$ with $\Delta_{m(\mathbf{x})}$ the point mass at $m(\mathbf{x})$ and by $\widetilde{F}_n(y|\mathbf{X} = \mathbf{x}) = \sum_{i=1}^{n} W_{i,n,\mathrm{L}}(\mathbf{x}) I_{(-\infty,y]}(\widetilde{y}_i)$ where $\widetilde{y}_i = \delta_i y_i + (1 - \delta_i)m(\mathbf{x}_i)$. Note that $m(\mathbf{x})$ is the unique solution of $\lambda(\mathbf{x}, a, \sigma) = 0$ for all $\sigma > 0$ with

$$\lambda(\mathbf{x}, a, \sigma) = p(\mathbf{x})E\left(\psi\left(\frac{y - a}{\sigma}\right)|\mathbf{X} = \mathbf{x}\right) + (1 - p(\mathbf{x}))\psi\left(\frac{m(\mathbf{x}) - a}{\sigma}\right)\ .$$

It is easy to see that $\sup_{y\in I\!R} |\widetilde{F}_n(y|\mathbf{X} = \mathbf{x}) - \widetilde{F}(y|\mathbf{X} = \mathbf{x})| \xrightarrow{a.s.} 0$. Thus, in order to obtain the strong consistency and using Theorem 2.2 in Boente and Fraiman (1989), it will be enough to show that $\sup_{y\in I\!R} |\widetilde{F}_n(y|\mathbf{X} = \mathbf{x}) - \widehat{\widehat{F}}_n(y|\mathbf{X} = \mathbf{x})| \xrightarrow{a.s.} 0$, which will follow if we show that, there exists a measurable set $\mathcal{N}$ with probability 0, such that, for any $\omega \notin \mathcal{N}$ it holds that for any bounded and Lipschitz f, we have that

$$\int f(y) d\widehat{\widehat{F}}_n(y|\mathbf{X} = \mathbf{x}) - \int f(y) d\widetilde{F}_n(y|\mathbf{X} = \mathbf{x}) \to 0 \tag{A.11}$$

Note that

$$\begin{aligned}\Delta_{n,f}(\mathbf{x}) &= \int f(y) d\widehat{\widehat{F}}_n(y|\mathbf{X} = \mathbf{x}) - \int f(y) d\widetilde{F}_n(y|\mathbf{X} = \mathbf{x}) \\ &= \sum_{i=1}^{n} W_{i,n,\mathrm{L}}(\mathbf{x})(1 - \delta_i)\left[f\left(\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)\right) - f\left(m(\mathbf{x}_i)\right)\right]\end{aligned}$$

Let $\mathbf{C}$ be a compact set such that $\mathbf{x} \in \text{int}(\mathbf{C})$. Denote

$$\begin{aligned} B_{1,n} &= \sum_{i=1}^n W_{i,n,\mathrm{L}}(\mathbf{x})(1-\delta_i) I_{\mathbf{C}^c}(\mathbf{x}_i) \left[f(\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)) - f(m(\mathbf{x}_i))\right] \\ B_{2,n} &= \sum_{i=1}^n W_{i,n,\mathrm{L}}(\mathbf{x})(1-\delta_i) I_{\mathbf{C}}(\mathbf{x}_i) \left[f(\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)) - f(m(\mathbf{x}_i))\right] \end{aligned}$$

Then, $\Delta_{n,f}(\mathbf{x}) = B_{1,n} + B_{2,n}$. Using that the kernel K is a positive function, we get that $|B_{1,n}| \leq \|f\|_\infty \sum_{i=1}^n W_{i,n,\mathrm{L}}(\mathbf{x}) I_{\mathbf{C}^c}(\mathbf{x}_i)$ and $|B_{2,n}| \leq C_f \sup_{\mathbf{u}\in\mathbf{C}} |\widehat{m}_{\mathrm{S}}(\mathbf{u}) - m(\mathbf{u})|$, where C_f denotes the Lipschitz constant of f. Note that $\sum_{i=1}^n W_{i,n,\mathrm{L}}(\mathbf{x}) I_{\mathbf{C}^c}(\mathbf{x}_i) \xrightarrow{a.s.} 0$ since $\mathbf{x} \notin \mathbf{C}$. Let $\mathcal{N}$ the probability 0 set such that, for any $\omega \notin \mathcal{N}$, we have

$$\begin{aligned} \sum_{i=1}^n W_{i,n,\mathrm{L}}(\mathbf{x}) I_{\mathbf{C}^c}(\mathbf{x}_i) &\to 0 \\ \sup_{\mathbf{u}\in\mathbf{C}} |\widehat{m}_{\mathrm{S}}(\mathbf{u}) - m(\mathbf{u})| &\to 0\,. \end{aligned}$$

Then, the bounds given entail that, for $\omega \notin \mathcal{N}$, $\Delta_{n,f}(\mathbf{x}) \to 0$, concluding the proof. □

PROOF OF PROPOSITION 4.2.1. Using a Taylor's expansion of order one, we get that

$$\widehat{m}_{\mathrm{M,I}}(\mathbf{x}) - m(\mathbf{x}) = \widehat{s}(\mathbf{x}) A_{0,n}^{-1}(\mathbf{x}) A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x}))$$

with

$$\begin{aligned} A_{0,n}(\mathbf{x}) &= \frac{1}{n\gamma_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \psi\prime\left(\frac{\widehat{y}_i - m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) + \\ &+ \frac{1}{2}\left[m(\mathbf{x}) - \widehat{m}_{\mathrm{M,I}}(\mathbf{x})\right] \frac{1}{n\gamma_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \psi\prime\prime\left(\frac{\widehat{y}_i - \xi(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \\ A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) &= \frac{1}{n\gamma_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \psi\left(\frac{\widehat{y}_i - m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \end{aligned}$$

where $\xi(\mathbf{x})$ is an intermediate point between $\widehat{m}_{\mathrm{M,I}}(\mathbf{x})$ and $m(\mathbf{x})$. It is enough to show that

a) $A_{0,n}(\mathbf{x}) \xrightarrow{p} f_{\mathbf{X}}(\mathbf{x}) \left[p(\mathbf{x}) A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)\right] \int L(\mathbf{u}) d\mathbf{u}$

b) $(n\gamma_n^p)^{\frac{1}{2}} A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x})) \xrightarrow{\mathcal{D}} N(a_1, \sigma_1)$ with

$$\begin{aligned} a_1 &= b_1 \frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})} \left[p(\mathbf{x}) A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)\right] \int L(\mathbf{u}) d\mathbf{u} \\ \sigma_1 &= \int \psi^2(u) dF_0(u)\, V(\mathbf{x}) \left(\frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\right)^2 \left[\int L(\mathbf{u}) d\mathbf{u}\right]^2 \\ &= p(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \begin{cases} \int \psi^2(u) dF_0(u) \int L^2(\mathbf{u}) d\mathbf{u} & \text{under i)} \\ \int \left[L(\mathbf{v})\, \psi(\epsilon) + \dfrac{\kappa^p\, (1-p(\mathbf{x}))\, \psi\prime(0)\, \psi_1(\epsilon)}{A_0(\psi_1)\, \int K(\mathbf{u}) d\mathbf{u}\, p(\mathbf{x})} \Gamma(\mathbf{v}, \kappa)\right]^2 d\mathbf{v}\, dF_0(\epsilon) & \text{under ii)} \end{cases} \end{aligned}$$

a) follows using similar arguments to those considered in Proposition 4.1.1. To obtain b) note that

$$\begin{aligned}
n\gamma_n^p\ A_{1,n}(\mathbf{x},\sigma) &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widehat{y}_i-m(\mathbf{x})}{\sigma}\right) \\
&= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left[\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\sigma}\right)+(1-\delta_i)\frac{\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)-m(\mathbf{x}_i)}{\sigma}\psi\prime\left(\frac{\xi(\mathbf{x}_i)-m(\mathbf{x})}{\sigma}\right)\right] \quad (A.12) \\
&= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left[\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\sigma}\right)+(1-\delta_i)\frac{\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)-m(\mathbf{x}_i)}{\sigma}\psi\prime\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\sigma}\right)\right] \\
&+ \frac{1}{2}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\left[\widehat{m}_{\mathrm{S}}(\mathbf{x}_i)-m(\mathbf{x}_i)\right]^2}{\sigma}\psi\prime\prime\left(\frac{\xi(\mathbf{x}_i)-m(\mathbf{x})}{\sigma}\right) \quad (A.13)
\end{aligned}$$

where $\widetilde{y}_i=\delta_i y_i+(1-\delta_i)m(\mathbf{x}_i)$ and $\xi(\mathbf{x}_i)$ denotes an intermediate point.

We begin by proving i).

For n large enough, we have that $L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)=0$ for $\mathbf{x}_i\notin\mathbf{C}$ which entails that $(n\gamma_n^p)^{\frac{1}{2}}\ A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x}))$ has the same asymptotic behavior as

$$(n\gamma_n^p)^{-\frac{1}{2}}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)$$

since from (A.12)

$$(n\gamma_n^p)^{\frac{1}{2}}\left|A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x}))-\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)\right|\le\widehat{\upsilon}(\mathbf{C})\|\psi\prime\|_\infty\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)$$

Now the proof follows as in Proposition 3.4.1 by showing that

a1) $(n\gamma_n^p)^{-\frac{1}{2}}\left|\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x}_i)}{\widehat{s}(\mathbf{x})}\right)-\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x}_i)}{\sigma(\mathbf{x})}\right)\right|\xrightarrow{p}0$

b1) $(n\gamma_n^p)^{-\frac{1}{2}}\left|\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)-\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x}_i)}{\widehat{s}(\mathbf{x})}\right)\right|\xrightarrow{p}a_1$

c1) $(n\gamma_n^p)^{-\frac{1}{2}}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x}_i)}{\sigma(\mathbf{x})}\right)\xrightarrow{\mathcal{D}}N(0,\sigma_1)$

$$\begin{aligned}
a_1 &= b_1\frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\ [p(\mathbf{x})A_0(\psi)+(1-p(\mathbf{x}))\psi\prime(0)]\int L(\mathbf{u})d\mathbf{u} \\
\sigma_1 &= \int\psi^2(u)dF_0(u)p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\int L^2(\mathbf{u})d\mathbf{u}\ .
\end{aligned}$$

a1) and b1) are derived as in Proposition 3.4.1. On the other hand, to prove c1) let $Z_i=\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x}_i)}{\sigma(\mathbf{x})}\right)$, $E\left(Z_1|\mathbf{X}_1=\mathbf{u}\right)=0$. The result follows now from the asymptotic distribution of the classical Nadaraya–Watson estimator.

Let us derive ii).

Using (A.13) we obtain that $(n\gamma_n^p)^{\frac{1}{2}} A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x}))$ has the same asymptotic behavior as

$$(n\gamma_n^p)^{-\frac{1}{2}} \left[\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)+\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\widehat{m}_{\mathrm{M,S}}(\mathbf{x}_i)-m(\mathbf{x}_i)}{\widehat{s}(\mathbf{x})}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)\right]$$

Using the consistency of $\widehat{s}(\mathbf{x})$ straightforward calculations allow to show that $(n\gamma_n^p)^{\frac{1}{2}} A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x}))$ has the same asymptotic behavior as

$$(n\gamma_n^p)^{-\frac{1}{2}} \left[\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)+\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\widehat{m}_{\mathrm{M,S}}(\mathbf{x}_i)-m(\mathbf{x}_i)}{\sigma(\mathbf{x})}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]$$

As in Proposition 3.4.1, denote

$$A_{0,n,\mathrm{S}}(\mathbf{x},\psi_1)=\frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{\mathbf{x}_i-\mathbf{x}}{h_n}\right)\delta_i\psi_1\prime\left(\frac{y_i-\xi(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)\ .$$

Using that $\sup_{\mathbf{u}\in\mathbf{C}}|\widehat{s}(\mathbf{u})-\sigma(\mathbf{u})| \xrightarrow{p} 0$, and that $\sup_{\mathbf{x}\in\mathbf{C}}\left|A_{0,n,\mathrm{S}}(\mathbf{x},\psi_1)-A_0(\psi_1)p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\int K(\mathbf{u})d\mathbf{u}\right| \xrightarrow{p} 0$ and expanding as in Proposition 3.4.1, we get

$$\widehat{m}_{\mathrm{M,S}}(\mathbf{x}_i)=m(\mathbf{x}_i)+\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\left(\int K(\mathbf{u})d\mathbf{u}A_0(\psi_1)\right)^{-1}\frac{1}{nh_n^p}\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\frac{y_j-m(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)+R_n(\mathbf{x}_i)$$

where $(nh_n^p)^{\frac{1}{2}}\sup_{\mathbf{u}\in\mathbf{C}}|R_n(\mathbf{u})| \xrightarrow{p} 0$. Since $\dfrac{\gamma_n}{h_n}\to\kappa>0$, using the boundness of $\psi\prime$, we obtain that $(n\gamma_n^p)^{\frac{1}{2}} A_{1,n}(\mathbf{x}, \widehat{s}(\mathbf{x}))$ has the same asymptotic behavior as

$$\begin{aligned}
U_n &= (n\gamma_n^p)^{-\frac{1}{2}}\left[B_{1,n}+\left(\sigma(\mathbf{x})A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\right)^{-1}B_{2,n}\right]\\
B_{1,n} &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\frac{\widetilde{y}_i-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
B_{2,n} &= \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\frac{y_j-m(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)
\end{aligned}$$

We begin by splitting $B_{1,n}$ and $B_{2,n}$ in the terms leading to the bias and those leading to the asymptotic disribution. Using that $\widetilde{y}_i=\delta_i y_i+(1-\delta_i)m(\mathbf{x}_i)$, we get

$$\begin{aligned}
B_{1,n} &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\left(\frac{y_i-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)+\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\psi\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\left(\frac{\sigma(\mathbf{x}_i)\epsilon_i}{\sigma(\mathbf{x})}\right)+\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\prime\left(\frac{\sigma(\mathbf{x}_i)\epsilon_i}{\sigma(\mathbf{x})}\right)\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi''\left(\frac{\xi_{i,1}}{\sigma(\mathbf{x})}\right)\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)^2 + \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\psi\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\left(\frac{\sigma(\mathbf{x}_i)\epsilon_i}{\sigma(\mathbf{x})}\right) + \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi'(\epsilon_i)\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&+ \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\left[\psi''\left(\frac{\xi_{i,1}}{\sigma(\mathbf{x})}\right)\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)^2 + \psi''\left(\frac{\xi_{i,2}}{\sigma(\mathbf{x})}\right)\left(\frac{\sigma(\mathbf{x}_i)-\sigma(\mathbf{x})}{\sigma(\mathbf{x})}\right)\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]\\
&+ \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\psi\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&= D_{11,n}+D_{12,n}+D_{13,n}+D_{14,n}
\end{aligned}
$$

where $\xi_{i,1}$ and $\xi_{i,2}$ denote intermediate points. Using that ψ'' is bounded and that m satisfies a Lipschitz condition of order one (with Lipschitz constant C_m) we get that

$$
\begin{aligned}
(n\gamma_n^p)^{-\frac12}|D_{13,n}| &\le \frac{C_m^2\|\psi''\|_\infty}{\sigma(\mathbf{x})^2}\left(n\gamma_n^{p+2}\right)^{\frac12}\left[\gamma_n\,\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}\|}{\gamma_n}\right)^2\right.\\
&+\left.\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}\|}{\gamma_n}\right)(\sigma(\mathbf{x}_i)-\sigma(\mathbf{x}))\right],
\end{aligned}
$$

which together with $n\gamma_n^{p+2}\to\beta^{p+2}$, the continuity of σ and

$$
\begin{aligned}
\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}\|}{\gamma_n}\right)^2 &\xrightarrow{p} f_{\mathbf{X}}(\mathbf{x})\int\|\mathbf{u}\|^2L(\mathbf{u})d\mathbf{u}\\
\frac{1}{n\gamma_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}\|}{\gamma_n}\right)(\sigma(\mathbf{x}_i)-\sigma(\mathbf{x})) &\xrightarrow{p} 0
\end{aligned}
$$

entail that $(n\gamma_n^p)^{-\frac12}D_{13,n}\xrightarrow{p}0$. On the other hand, we have that

$$
\begin{aligned}
(n\gamma_n^p)^{-\frac12}E\left(D_{12,n}+D_{14,n}\right) &= \left(\frac{n}{\gamma_n^p}\right)^{\frac12}E\left(L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\left[p(\mathbf{x}_1)A_0(\psi)\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)+\right.\right.\\
&+\left.\left.(1-p(\mathbf{x}_1))\psi\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]\right)\\
&= \left(n\gamma_n^{p+2}\right)^{\frac12}\int L(\mathbf{u})\left[p(\mathbf{x}+\gamma_n\,\mathbf{u})A_0(\psi)\frac{1}{\sigma(\mathbf{x})}\left(\frac{m(\mathbf{x}+\gamma_n\,\mathbf{u})-m(\mathbf{x})}{\gamma_n}\right)+\right.\\
&+\left.(1-p(\mathbf{x}+\gamma_n\,\mathbf{u}))\frac{1}{\gamma_n}\psi\left(\frac{m(\mathbf{x}+\gamma_n\,\mathbf{u})-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]f_{\mathbf{X}}(\mathbf{x}+\gamma_n\,\mathbf{u})d\mathbf{u}
\end{aligned}
$$

Using the continuity of $p(\mathbf{u})$, the boundness pf $f_{\mathbf{X}}$, the Lipschitz continuity of m and ψ, **N3** and the fact that $\psi(0)=0$ and $n\gamma_n^{p+2}\to\beta^{p+2}$, from the dominated convergence Theorem, we get that

$$
(n\gamma_n^p)^{-\frac12}E\left(D_{12,n}+D_{14,n}\right)\to\beta^{\frac{p+2}{2}}\left[p(\mathbf{x})A_0(\psi)+(1-p(\mathbf{x}))\psi'(0)\right]\frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\int L(\mathbf{u})\,m'(\mathbf{x},\mathbf{u})d\mathbf{u}\,. \tag{A.14}
$$

Finally, since

$$\begin{aligned}
\frac{Var\left(D_{12,n}+D_{14,n}\right)}{n\gamma_n^p} &= \frac{1}{\gamma_n^p}Var\left(L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\left[\delta_1\psi\prime\left(\epsilon_1\right)\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)+(1-\delta_1)\psi\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]\right)\\
&\leq \frac{1}{\gamma_n^p}E\left(L^2\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\left[\delta_1\psi\prime\left(\epsilon_1\right)\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)+(1-\delta_1)\psi\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]^2\right)\\
&\leq \|\psi\prime\|_\infty\frac{1}{\gamma_n^p}E\left(L^2\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\left[\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right]^2\right)\\
&\leq \|\psi\prime\|_\infty C_m^2\,\frac{1}{\sigma(\mathbf{x})^2}\gamma_n^pE\left(L^2\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\left(\frac{\|\mathbf{x}_1-\mathbf{x}\|}{\gamma_n^p}\right)^2\right)\\
&\leq \|\psi\prime\|_\infty C_m^2\,\frac{1}{\sigma(\mathbf{x})^2}\,\gamma_n^p\int L^2\left(\mathbf{u}\right)\|\mathbf{u}\|^2 f_{\mathbf{X}}(\mathbf{x}+\gamma_n\,\mathbf{u})d\mathbf{u}
\end{aligned}$$

we obtain that $(n\gamma_n^p)^{-1}Var\left(D_{12,n}+D_{14,n}\right)\to 0$ and so using (A.14), we have

$$(n\gamma_n^p)^{-\frac{1}{2}}\left(D_{12,n}+D_{14,n}\right)\xrightarrow{p}\beta^{\frac{p+2}{2}}\,\left[p(\mathbf{x})A_0(\psi)+(1-p(\mathbf{x}))\psi\prime(0)\right]\,\frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\,\int L\left(\mathbf{u}\right)m\prime(\mathbf{x},\mathbf{u})d\mathbf{u}\,. \tag{A.15}$$

Moreover, note that

$$\begin{aligned}
D_{11,n} &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\left(\epsilon_i\right)+\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\left[\psi\left(\frac{\sigma(\mathbf{x}_i)\epsilon_i}{\sigma(\mathbf{x})}\right)-\psi\left(\epsilon_i\right)\right]\\
&= \Delta_{11,n}+\Delta_{12,n}
\end{aligned}$$

with $E(\Delta_{11,n})=0$ and $E(\Delta_{12,n})=0$. We have the following bound for the variance of $(n\gamma_n^p)^{-\frac{1}{2}}\Delta_{12,n}$

$$\begin{aligned}
Var\left((n\gamma_n^p)^{-\frac{1}{2}}\Delta_{12,n}\right) &= \frac{1}{\gamma_n^p}Var\left(L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\delta_1\left[\psi\left(\frac{\sigma(\mathbf{x}_1)\epsilon_1}{\sigma(\mathbf{x})}\right)-\psi\left(\epsilon_1\right)\right]\right)\\
&= \frac{1}{\gamma_n^p}E\left(L^2\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\delta_1\left[\psi\left(\frac{\sigma(\mathbf{x}_1)\epsilon_1}{\sigma(\mathbf{x})}\right)-\psi\left(\epsilon_1\right)\right]^2\right)\\
&\leq \int L^2\left(\mathbf{u}\right)\left[\psi\left(\frac{\sigma(\mathbf{u}\gamma_n+\mathbf{x})\epsilon}{\sigma(\mathbf{x})}\right)-\psi\left(\epsilon\right)\right]^2 f_{\mathbf{X}}(\mathbf{u}\gamma_n+\mathbf{x})dF_0(\epsilon)\,.
\end{aligned}$$

Using the dominated convergence Theorem, we get that $Var\left((n\gamma_n^p)^{-\frac{1}{2}}\Delta_{12,n}\right)\to 0$, which together with (A.15) entail that

$$(n\gamma_n^p)^{-\frac{1}{2}}B_{1,n}=(n\gamma_n^p)^{-\frac{1}{2}}\Delta_{11,n}+\frac{\beta^{\frac{p+2}{2}}\left[p(\mathbf{x})A_0(\psi)+(1-p(\mathbf{x}))\psi\prime(0)\right]f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\int L\left(\mathbf{u}\right)m\prime(\mathbf{x},\mathbf{u})d\mathbf{u}+o_p(1) \tag{A.16}$$

where $\Delta_{11,n}=\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi\left(\epsilon_i\right)$ and $E(\Delta_{11,n})=0$.

We will made an expansion to the term $(n\gamma_n^p)^{-\frac{1}{2}}\left(\sigma(\mathbf{x})A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\right)^{-1}B_{2,n}$ analogous to that made for $B_{1,n}$. Using a Taylor's expansion of order 2, we obtain

$$\begin{aligned}
B_{2,n} &= \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\frac{y_j-m(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)\\
&= \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times \sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\frac{\sigma(\mathbf{x}_j)\epsilon_j+m(\mathbf{x}_j)-m(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)\\
&= \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\epsilon_j\frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right)\\
&+ \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\frac{(1-\delta_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times \sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\prime\left(\epsilon_j\frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right)(m(\mathbf{x}_j)-m(\mathbf{x}_i))\\
&+ \frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times \sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\prime\prime\left(\frac{\xi_{ij,1}}{\sigma(\mathbf{x}_i)}\right)\left(\frac{m(\mathbf{x}_j)-m(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)^2\\
&= D_{21,n}+D_{22,n}+D_{23,n}
\end{aligned}$$

where $\xi_{ij,1}$ denote intermediate points. Using that $\psi_1\prime\prime$ is bounded and that m satisfies a Lipschitz condition of order one (with Lipschitz constant C_m) we get that

$$\begin{aligned}
(n\gamma_n^p)^{-\frac{1}{2}}|D_{23,n}| &\le C_m^2\|\psi\prime\|_\infty\|\psi_1\prime\prime\|_\infty(n\gamma_n^p)^{-\frac{1}{2}}\frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}_j\|}{\sigma(\mathbf{x}_i)}\right)^2\\
&\le C_m^2\|\psi\prime\|_\infty\|\psi_1\prime\prime\|_\infty\left(n\gamma_n^{p+2}\right)^{\frac{1}{2}}\gamma_n\Delta_{23,n}\\
\Delta_{23,n} &= \frac{1}{n\gamma_n^p}\frac{1}{nh_n^p}\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\frac{1}{\sigma(\mathbf{x}_i)p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\left(\frac{\|\mathbf{x}_i-\mathbf{x}_j\|}{\gamma_n}\right)^2 .
\end{aligned}$$

Thus, it will be enough to show that $\lim E\Delta_{23,n}<\infty$, since $\gamma_n\to 0$ and $n\gamma_n^{p+2}\to\beta^{p+2}$.

$$\begin{aligned}
E\Delta_{23,n} &= \frac{1}{\gamma_n^p}\frac{1}{nh_n^p}EL\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{1}{\sigma(\mathbf{x}_1)p(\mathbf{x}_1)f_{\mathbf{X}}(\mathbf{x}_1)}\sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_1}{h_n}\right)\left(\frac{\|\mathbf{x}_1-\mathbf{x}_j\|}{\gamma_n}\right)^2\\
&= \frac{1}{\gamma_n^p}\frac{1}{h_n^p}\frac{n-1}{n}E\left[L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{1}{\sigma(\mathbf{x}_1)p(\mathbf{x}_1)f_{\mathbf{X}}(\mathbf{x}_1)}K\left(\frac{\mathbf{x}_2-\mathbf{x}_1}{h_n}\right)\left(\frac{\|\mathbf{x}_1-\mathbf{x}_2\|}{\gamma_n}\right)^2\right]\\
&= \frac{1}{\gamma_n^p}\frac{1}{h_n^p}\frac{n-1}{n}\int L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{1}{\sigma(\mathbf{x}_1)p(\mathbf{x}_1)}f_{\mathbf{X}}(\mathbf{x}_2)K\left(\frac{\mathbf{x}_2-\mathbf{x}_1}{h_n}\right)\left(\frac{\|\mathbf{x}_1-\mathbf{x}_2\|}{\gamma_n}\right)^2 d\mathbf{x}_1 d\mathbf{x}_2
\end{aligned}$$

$$= \quad \frac{n-1}{n}\left(\frac{h_n}{\gamma_n}\right)^2 \int L\left(\mathbf{u}\right) \frac{1}{\sigma(\mathbf{x}+\gamma_n\mathbf{u})p(\mathbf{x}+\gamma_n\mathbf{u})} f_{\mathbf{X}}(\mathbf{x}+\gamma_n\mathbf{u}+h_n\mathbf{v})K\left(\mathbf{v}\right)\|\mathbf{v}\|^2 d\mathbf{u}d\mathbf{v}\ .$$

Therefore, using L has compact support, $\dfrac{\gamma_n}{h_n} \to \kappa \neq 0$ and the dominated convergence Theorem we get that $E\Delta_{23,n} \to \kappa^{-2}\frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})p(\mathbf{x})}\int L\left(\mathbf{u}\right)d\mathbf{u}\int K\left(\mathbf{v}\right)\|\mathbf{v}\|^2 d\mathbf{v}$ and so, $(n\gamma_n^p)^{-\frac{1}{2}}\, D_{23,n} \stackrel{p}{\longrightarrow} 0$.

Let us show that $(n\gamma_n^p)^{-\frac{1}{2}}\, D_{22,n}$ converges in probability to the second component of the bias. Denote

$$H_{n,i,j} = L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\frac{(1-\delta_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\prime\left(\epsilon_j\frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right)\left(m(\mathbf{x}_j)-m(\mathbf{x}_i)\right)$$

$$\begin{aligned}
\frac{ED_{22,n}}{(n\gamma_n^p)^{\frac{1}{2}}} &= (n\gamma_n^p)^{-\frac{1}{2}}\frac{1}{nh_n^p}E\Bigg[\sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\frac{(1-\delta_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times \sum_{j=1}^n K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\prime\left(\epsilon_j\frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right)\left(m(\mathbf{x}_j)-m(\mathbf{x}_i)\right)\Bigg]\\
&= (n\gamma_n^p)^{-\frac{1}{2}}\frac{1}{nh_n^p}n(n-1)E\left(H_{n,1,2}\right)\\
&= (n\gamma_n^p)^{\frac{1}{2}}\frac{n-1}{n}\frac{1}{\gamma_n^p\, h_n^p}E\Bigg[L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{(1-p(\mathbf{x}_1))}{p(\mathbf{x}_1)f_{\mathbf{X}}(\mathbf{x}_1)}\psi\prime\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times K\left(\frac{\mathbf{x}_2-\mathbf{x}_1}{h_n}\right)p(\mathbf{x}_2)\psi_1\prime\left(\epsilon_2\frac{\sigma(\mathbf{x}_2)}{\sigma(\mathbf{x}_1)}\right)\left(m(\mathbf{x}_2)-m(\mathbf{x}_1)\right)\Bigg]\\
&= (n\gamma_n^p)^{\frac{1}{2}}\frac{n-1}{n}\frac{1}{\gamma_n^p\, h_n^p}\int L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{(1-p(\mathbf{x}_1))}{p(\mathbf{x}_1)}\psi\prime\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times K\left(\frac{\mathbf{x}_2-\mathbf{x}_1}{h_n}\right)p(\mathbf{x}_2)\psi_1\prime\left(\epsilon\frac{\sigma(\mathbf{x}_2)}{\sigma(\mathbf{x}_1)}\right)\left(m(\mathbf{x}_2)-m(\mathbf{x}_1)\right)f_{\mathbf{X}}(\mathbf{x}_2)d\mathbf{x}_1 d\mathbf{x}_2 dF_0(\epsilon)\\
&= (n\gamma_n^p)^{\frac{1}{2}}\frac{n-1}{n}\frac{1}{\gamma_n^p}\int L\left(\frac{\mathbf{x}_1-\mathbf{x}}{\gamma_n}\right)\frac{(1-p(\mathbf{x}_1))}{p(\mathbf{x}_1)}\psi\prime\left(\frac{m(\mathbf{x}_1)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times K\left(\mathbf{u}\right)p(\mathbf{x}_1+h_n\mathbf{u})\psi_1\prime\left(\epsilon\frac{\sigma(\mathbf{x}_1+h_n\mathbf{u})}{\sigma(\mathbf{x}_1)}\right)\left(m(\mathbf{x}_1+h_n\mathbf{u})-m(\mathbf{x}_1)\right)f_{\mathbf{X}}(\mathbf{x}_1+h_n\mathbf{u})d\mathbf{x}_1 d\mathbf{u}dF_0(\epsilon)\\
&= (n\gamma_n^p)^{\frac{1}{2}}\frac{n-1}{n}\int L\left(\mathbf{v}\right)\frac{(1-p(\mathbf{x}+\gamma_n\mathbf{v}))p(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})}{p(\mathbf{x}+\gamma_n\mathbf{v})}\psi\prime\left(\frac{m(\mathbf{x}+\gamma_n\mathbf{v})-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times K\left(\mathbf{u}\right)\psi_1\prime\left(\epsilon\frac{\sigma(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})}{\sigma(\mathbf{x}+\gamma_n\mathbf{v})}\right)\left(m(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})-m(\mathbf{x}+\gamma_n\mathbf{v})\right)f_{\mathbf{X}}(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})d\mathbf{v}d\mathbf{u}dF_0(\epsilon)\\
&= \left(n\gamma_n^{p+2}\right)^{\frac{1}{2}}\frac{n-1}{n}\frac{h_n}{\gamma_n}\int L\left(\mathbf{v}\right)\frac{(1-p(\mathbf{x}+\gamma_n\mathbf{v}))p(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})}{p(\mathbf{x}+\gamma_n\mathbf{v})}\psi\prime\left(\frac{m(\mathbf{x}+\gamma_n\mathbf{v})-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\times K\left(\mathbf{u}\right)\psi_1\prime\left(\epsilon\frac{\sigma(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})}{\sigma(\mathbf{x}+\gamma_n\mathbf{v})}\right)\frac{m(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})-m(\mathbf{x}+\gamma_n\mathbf{v})}{h_n}f_{\mathbf{X}}(\mathbf{x}+\gamma_n\mathbf{v}+h_n\mathbf{u})d\mathbf{v}d\mathbf{u}dF_0(\epsilon)
\end{aligned}$$

Using that $n\gamma_n^{p+2}\to\beta^{p+2}$, $\frac{h_n}{\gamma_n}\to\kappa^{-1}$, **N3** and the dominated convergence Theorem, we get that

$$(n\gamma_n^p)^{-\frac{1}{2}}\,E\left(D_{22,n}\right)\to A_0(\psi_1)f_{\mathbf{X}}(\mathbf{x})(1-p(\mathbf{x}))\psi\prime\left(0\right)\beta^{\frac{p+2}{2}}\kappa^{-1}\int L\left(\mathbf{v}\right)K\left(\mathbf{u}\right)\Delta(\mathbf{x},\mathbf{u},\mathbf{v})d\mathbf{v}d\mathbf{u}\ . \quad \text{(A.17)}$$

It remains to show that $(n\gamma_n^p)^{-1} Var(D_{22,n}) \to 0$.

$$
\begin{aligned}
\frac{Var(D_{22,n})}{n\gamma_n^p} &= \frac{1}{n\gamma_n^p}\frac{1}{(nh_n^p)^2} Var\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} H_{n,i,j}\right) \\
&= \frac{1}{n\gamma_n^p}\frac{1}{(nh_n^p)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{s=1}^{n}\sum_{\ell\neq s}^{n} Cov(H_{n,i,j}, H_{n,s,\ell}) \\
&= \frac{1}{n\gamma_n^p}\frac{1}{(nh_n^p)^2}\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{\ell\neq i}^{n} Cov(H_{n,i,j}, H_{n,i,\ell}) + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{s\neq i}^{n} Cov(H_{n,i,j}, H_{n,s,i})\right. \\
&+ \left.\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{\ell\neq i,j}^{n} Cov(H_{n,i,j}, H_{n,j,\ell}) + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{s\neq i,j}^{n} Cov(H_{n,i,j}, H_{n,s,j})\right] \\
&= \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}\left[(n-2)Cov(H_{n,1,2}, H_{n,1,3}) + Var(H_{n,1,2})\right. \\
&+ \left[(n-2)Cov(H_{n,1,2}, H_{n,3,1}) + Cov(H_{n,1,2}, H_{n,2,1})\right] \\
&+ \left.(n-2)\left[Cov(H_{n,1,2}, H_{n,2,3}) + Cov(H_{n,1,2}, H_{n,3,2})\right]\right] \\
&= \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}(n-2)\left[Cov(H_{n,1,2}, H_{n,1,3}) + Cov(H_{n,1,2}, H_{n,3,1})\right. \\
&+ \left.Cov(H_{n,1,2}, H_{n,2,3}) + Cov(H_{n,1,2}, H_{n,3,2})\right] \\
&+ \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}\left[Var(H_{n,1,2}) + Cov(H_{n,1,2}, H_{n,2,1})\right]
\end{aligned}
$$

From the equality

$$
\alpha_n = \frac{ED_{22,n}}{(n\gamma_n^p)^{\frac{1}{2}}} = (n\gamma_n^p)^{-\frac{1}{2}}\frac{1}{nh_n^p} n(n-1)E(H_{n,1,2})
$$

and the fact that

$$
\begin{aligned}
\Lambda_n &= \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}(n-2)\left[E(H_{n,1,2})\right]^2 + \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}\left[E(H_{n,1,2})\right]^2 \\
&= \alpha_n^2\frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}\left\{(n-2)\left[\frac{(n\gamma_n^p)^{\frac{1}{2}}\, nh_n^p}{n(n-1)}\right]^2 + \left[\frac{(n\gamma_n^p)^{\frac{1}{2}}\, nh_n^p}{n(n-1)}\right]^2\right\} \\
&= \alpha_n^2\frac{1}{n}
\end{aligned}
$$

using (A.17), we have that it will be enough to show that

$$
\begin{aligned}
&\frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}(n-2)\left[E(H_{n,1,2}H_{n,1,3}) + E(H_{n,1,2}H_{n,3,1}) + E(H_{n,1,2}H_{n,2,3}) + E(H_{n,1,2}H_{n,3,2})\right] \\
&\qquad + \frac{1}{n\gamma_n^p}\frac{n(n-1)}{(nh_n^p)^2}\left[E\left(H_{n,1,2}^2\right) + E(H_{n,1,2}H_{n,2,1})\right] \to 0
\end{aligned}
$$

that follows easily using that $\frac{\gamma_n}{h_n} \to \kappa \neq 0$, $E\left(H_{n,1,2} H_{n,2,3}\right) = (h_n^p)^3\, o(1)$, $E\left(H_{n,1,2}^2\right) = (h_n^p)^2\, o(1)$ and the fact that $n\gamma_n^p \to \infty$.

Therefore, using (A.17), we have that

$$
\begin{aligned}
(n\gamma_n^p)^{-\frac{1}{2}} \; \frac{B_{2,n}}{\sigma(\mathbf{x}) A_0(\psi_1) \int K(\mathbf{u}) d\mathbf{u}} &= (n\gamma_n^p)^{-\frac{1}{2}} \left(\sigma(\mathbf{x}) A_0(\psi_1) \int K(\mathbf{u}) d\mathbf{u}\right)^{-1} D_{21,n} \\
&+ \frac{(1-p(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x}) \int K(\mathbf{u}) d\mathbf{u}} \psi\prime(0)\, \beta^{\frac{p+2}{2}} \kappa^{-1} \int L(\mathbf{v})\, K(\mathbf{u})\, \Delta(\mathbf{x},\mathbf{u},\mathbf{v}) d\mathbf{v} d\mathbf{u} + o_p(1) \quad \text{(A.18)}
\end{aligned}
$$

with

$$
D_{21,n} = \frac{1}{n h_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) (1-\delta_i) \frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i) f_{\mathbf{X}}(\mathbf{x}_i)} \psi\prime \left(\frac{m(\mathbf{x}_i) - m(\mathbf{x})}{\sigma(\mathbf{x})}\right) \sum_{j=1}^n K\left(\frac{\mathbf{x}_j - \mathbf{x}_i}{h_n}\right) \delta_j \psi_1\left(\epsilon_j \frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right) .
$$

Denote

$$
\begin{aligned}
\Delta_{21,n} &= \frac{1}{n h_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \frac{(1-\delta_i)\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i) f_{\mathbf{X}}(\mathbf{x}_i)} \psi\prime \left(\frac{m(\mathbf{x}_i) - m(\mathbf{x})}{\sigma(\mathbf{x})}\right) \sum_{j=1}^n K\left(\frac{\mathbf{x}_j - \mathbf{x}_i}{h_n}\right) \delta_j \psi_1(\epsilon_j) \quad \text{(A.19)} \\
\Delta_{22,n} &= D_{21,n} - \Delta_{21,n}
\end{aligned}
$$

Using that $E\left(\Delta_{22,n}\right) = 0$ and bounding the variance of $(n\gamma_n^p)^{-\frac{1}{2}} \Delta_{22,n}$ as we have done with that of $(n\gamma_n^p)^{-\frac{1}{2}} D_{22,n}$, replacing $\psi_1\prime\left(\epsilon_j \frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right)(m(\mathbf{x}_j) - m(\mathbf{x}_i))$ by $\psi_1\left(\epsilon_j \frac{\sigma(\mathbf{x}_j)}{\sigma(\mathbf{x}_i)}\right) - \psi_1(\epsilon_j)$, the continuity of σ and straightforward calculations lead to $(n\gamma_n^p)^{-\frac{1}{2}} \Delta_{22,n} \xrightarrow{p} 0$.

Thus, putting together (A.16), (A.18) and (A.19), we have the following expansion for U_n

$$
\begin{aligned}
U_n &= (n\gamma_n^p)^{-\frac{1}{2}} \left[B_{1,n} + \left(\sigma(\mathbf{x}) A_0(\psi_1) \int K(\mathbf{u}) d\mathbf{u}\right)^{-1} B_{2,n}\right] \\
&= (n\gamma_n^p)^{-\frac{1}{2}} \left[\Delta_{11,n} + \left(\sigma(\mathbf{x}) A_0(\psi_1) \int K(\mathbf{u}) d\mathbf{u}\right)^{-1} \Delta_{21,n}\right] + a_1 + o_p(1) \\
a_1 &= \frac{\beta^{\frac{p+2}{2}} f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})} \left\{ [p(\mathbf{x}) A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)] \int L(\mathbf{v})\, m\prime(\mathbf{x},\mathbf{v}) d\mathbf{v} \right. \\
&+ \left. \frac{(1-p(\mathbf{x}))}{\int K(\mathbf{u}) d\mathbf{u}} \psi\prime(0)\, \kappa^{-1} \int L(\mathbf{v})\, K(\mathbf{u})\, \Delta(\mathbf{x},\mathbf{u},\mathbf{v}) d\mathbf{v} d\mathbf{u} \right\} \\
&= b_1 \frac{f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})} [p(\mathbf{x}) A_0(\psi) + (1-p(\mathbf{x}))\psi\prime(0)] \int L(\mathbf{u}) d\mathbf{u} \\
\Delta_{11,n} &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) \delta_i \psi(\epsilon_i) \\
\Delta_{21,n} &= \frac{1}{n h_n^p} \sum_{i=1}^n L\left(\frac{\mathbf{x}_i - \mathbf{x}}{\gamma_n}\right) (1-\delta_i) \frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i) f_{\mathbf{X}}(\mathbf{x}_i)} \psi\prime \left(\frac{m(\mathbf{x}_i) - m(\mathbf{x})}{\sigma(\mathbf{x})}\right) \sum_{j=1}^n K\left(\frac{\mathbf{x}_j - \mathbf{x}_i}{h_n}\right) \delta_j \psi_1(\epsilon_j)
\end{aligned}
$$

It only remains to show that $(n\gamma_n^p)^{-\frac{1}{2}} \left[\Delta_{11,n} + (\sigma(\mathbf{x}) A_0(\psi_1) \int K(\mathbf{u}) d\mathbf{u})^{-1} \Delta_{21,n}\right] \xrightarrow{\mathcal{D}} N(0, \sigma_1)$.

Let us begin by writting $(n\gamma_n^p)^{-1}\,\Delta_{21,n}$ as a $U-$statistic with symmetric kernel. Denote $L_{\gamma_n}(\mathbf{v}) = \gamma_n^{-p}L(\mathbf{v}/\gamma_n)$, $K_{h_n}(\mathbf{v}) = h_n^{-p}K(\mathbf{v}/h_n)$ and $\mathbf{z}_i = \left(\mathbf{x}_i^{\mathrm{T}}, \delta_i, \epsilon_i\right)^{\mathrm{T}}$. Then,

$$
\begin{aligned}
\frac{1}{n\gamma_n^p}\Delta_{21,n} &= \frac{1}{n\gamma_n^p}\frac{1}{nh_n^p}\sum_{i=1}^{n} L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)(1-\delta_i)\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\sum_{j=1}^{n} K\left(\frac{\mathbf{x}_j-\mathbf{x}_i}{h_n}\right)\delta_j\psi_1\left(\epsilon_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} L_{\gamma_n}\left(\mathbf{x}_i-\mathbf{x}\right)K_{h_n}\left(\mathbf{x}_j-\mathbf{x}_i\right)(1-\delta_i)\delta_j\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\psi_1\left(\epsilon_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} H_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) \\
H_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) &= L_{\gamma_n}\left(\mathbf{x}_i-\mathbf{x}\right)K_{h_n}\left(\mathbf{x}_j-\mathbf{x}_i\right)(1-\delta_i)\delta_j\frac{\sigma(\mathbf{x}_i)}{p(\mathbf{x}_i)f_{\mathbf{X}}(\mathbf{x}_i)}\psi\prime\left(\frac{m(\mathbf{x}_i)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\psi_1\left(\epsilon_j\right)
\end{aligned}
$$

Note that $H_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_i\right) = 0$, thus if we denote

$$
\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) = \frac{1}{2}\left(H_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) + H_{\gamma_n,h_n}\left(\mathbf{z}_j,\mathbf{z}_i\right)\right)
$$

we obtain that

$$
\begin{aligned}
\frac{1}{n\gamma_n^p}\Delta_{21,n} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) = \frac{n-1}{n}U_{\lambda,n} \\
U_{\lambda,n} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right)
\end{aligned}
$$

We have that $E\left(\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right)\right) = 0$ and for any $j \neq i$

$$
\begin{aligned}
E\left(U_{\lambda,n}|\mathbf{z}_i\right) &= \frac{2}{n(n-1)}\sum_{j=1}^{n} E\left(\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right)|\mathbf{z}_i\right) \\
E\left(U_{\lambda,n}|\mathbf{z}_i\right) &= \frac{2}{n}E\left(\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right)|\mathbf{z}_i\right) \\
E\left(\lambda_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right)|\mathbf{z}_i\right) &= \frac{1}{2}E\left(\left(H_{\gamma_n,h_n}\left(\mathbf{z}_i,\mathbf{z}_j\right) + H_{\gamma_n,h_n}\left(\mathbf{z}_j,\mathbf{z}_i\right)\right)|\mathbf{z}_i\right) \\
&= \frac{1}{2}E\left(H_{\gamma_n,h_n}\left(\mathbf{z}_j,\mathbf{z}_i\right)|\mathbf{z}_i\right)\ .
\end{aligned}
$$

The last equality holds since $E\psi_1(\epsilon_j) = 0$. Therefore, for any $j \neq i$, we have

$$
\begin{aligned}
E\left(U_{\lambda,n}|\mathbf{z}_i\right) &= \frac{1}{n}E\left(H_{\gamma_n,h_n}\left(\mathbf{z}_j,\mathbf{z}_i\right)|\mathbf{z}_i\right) \\
&= \frac{1}{n}\,\delta_i\psi_1\left(\epsilon_i\right)E\left(L_{\gamma_n}\left(\mathbf{x}_j-\mathbf{x}\right)\frac{\sigma(\mathbf{x}_j)(1-\delta_j)}{p(\mathbf{x}_j)f_{\mathbf{X}}(\mathbf{x}_j)}\psi\prime\left(\frac{m(\mathbf{x}_j)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)K_{h_n}\left(\mathbf{x}_i-\mathbf{x}_j\right)|\mathbf{z}_i\right) \\
&= \frac{1}{n}\,\delta_i\psi_1\left(\epsilon_i\right)E\left(L_{\gamma_n}\left(\mathbf{x}_j-\mathbf{x}\right)\frac{\sigma(\mathbf{x}_j)(1-p(\mathbf{x}_j))}{p(\mathbf{x}_j)f_{\mathbf{X}}(\mathbf{x}_j)}\psi\prime\left(\frac{m(\mathbf{x}_j)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)K_{h_n}\left(\mathbf{x}_i-\mathbf{x}_j\right)|\mathbf{z}_i\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\,\delta_i\psi_1(\epsilon_i)\int L_{\gamma_n}(\mathbf{x}_j-\mathbf{x})\,\frac{\sigma(\mathbf{x}_j)(1-p(\mathbf{x}_j))}{p(\mathbf{x}_j)}\psi\prime\left(\frac{m(\mathbf{x}_j)-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)K_{h_n}(\mathbf{x}_i-\mathbf{x}_j)\,d\mathbf{x}_j\\
&= \frac{1}{n}\,\delta_i\psi_1(\epsilon_i)\int L(\mathbf{u})\,\frac{\sigma(\mathbf{x}+\gamma_n\mathbf{u})(1-p(\mathbf{x}+\gamma_n\mathbf{u}))}{p(\mathbf{x}+\gamma_n\mathbf{u})}\psi\prime\left(\frac{m(\mathbf{x}+\gamma_n\mathbf{u})-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\quad\times K_{h_n}(\mathbf{x}_i-\mathbf{x}-\gamma_n\mathbf{u})\,d\mathbf{u}\\
&= \frac{\gamma_n^p}{n\,h_n^p}\frac{1}{\gamma_n^p}\,\delta_i\psi_1(\epsilon_i)\int L(\mathbf{u})\,\frac{\sigma(\mathbf{x}+\gamma_n\mathbf{u})(1-p(\mathbf{x}+\gamma_n\mathbf{u}))}{p(\mathbf{x}+\gamma_n\mathbf{u})}\psi\prime\left(\frac{m(\mathbf{x}+\gamma_n\mathbf{u})-m(\mathbf{x})}{\sigma(\mathbf{x})}\right)\\
&\quad\times K\left(\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}-\mathbf{u}\right)\frac{\gamma_n}{h_n}\right)d\mathbf{u}\\
&= \frac{\gamma_n^p}{n\,h_n^p}\,\delta_i\psi_1(\epsilon_i)\left[\frac{\sigma(\mathbf{x})(1-p(\mathbf{x}))}{p(\mathbf{x})}\psi\prime(0)\int L(\mathbf{u})\,\frac{1}{\gamma_n^p}K\left(\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}-\mathbf{u}\right)\frac{\gamma_n}{h_n}\right)d\mathbf{u}+R_{1,n}(\mathbf{x}_i)\right]\\
&= \frac{\kappa^p}{n}\,\delta_i\psi_1(\epsilon_i)\left[\frac{\sigma(\mathbf{x})(1-p(\mathbf{x}))}{p(\mathbf{x})}\psi\prime(0)\int L(\mathbf{u})\,\frac{1}{\gamma_n^p}K\left(\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}-\mathbf{u}\right)\frac{\gamma_n}{h_n}\right)d\mathbf{u}+R_{2,n}(\mathbf{x}_i)\right]\ ,
\end{aligned}
$$

where $(n\gamma_n^p)^{\frac{1}{2}}\,\frac{1}{n}\sum_{i=1}^n|R_{2,n}(\mathbf{x}_i)|\to 0$. Let $\Gamma(\mathbf{v},a)=\int L(\mathbf{u})\,K((\mathbf{v}-\mathbf{u})\,a)\,d\mathbf{u}$ and $\Gamma_{\gamma_n}(\mathbf{v},a)=\gamma_n^{-p}\Gamma(\mathbf{v}/\gamma_n,a)$

$$
\begin{aligned}
E(U_{\lambda,n}|\mathbf{z}_i) &= \frac{\kappa^p}{n}\,\delta_i\psi_1(\epsilon_i)\left[\frac{\sigma(\mathbf{x})(1-p(\mathbf{x}))}{p(\mathbf{x})}\psi\prime(0)\,\Gamma_{\gamma_n}\left(\mathbf{x}_i-\mathbf{x},\frac{\gamma_n}{h_n}\right)+R_{2,n}(\mathbf{x}_i)\right]\\
&= \frac{1}{n}\Lambda_{\gamma_n,h_n}(\mathbf{z}_i)+\frac{\kappa^p}{n}\,\delta_i\psi_1(\epsilon_i)\,R_{2,n}(\mathbf{x}_i)\ ,
\end{aligned}
$$

with $\Lambda_{\gamma_n,h_n}(\mathbf{z}_i)=\kappa^p\sigma(\mathbf{x})(1-p(\mathbf{x}))p^{-1}(\mathbf{x})\psi\prime(0)\,\delta_i\psi_1(\epsilon_i)\,\Gamma_{\gamma_n}\left(\mathbf{x}_i-\mathbf{x},\frac{\gamma_n}{h_n}\right)$.

Denote $\widehat{U}_{\lambda,n}=\frac{1}{n}\sum_{i=1}^n\Lambda_{\gamma_n,h_n}(\mathbf{z}_i)$. Then, standard U–statistic arguments allow to show that $n\gamma_n^pE\left(U_{\lambda,n}-\widehat{U}_{\lambda,n}\right)^2\le(1/(n-1))Var(\lambda_{\gamma_n,h_n}(\mathbf{z}_1,bz_2))$. Using that $h_n^p\ \gamma_n^p\ Var(\lambda_{\gamma_n,h_n}(\mathbf{z}_1,\mathbf{z}_2))=(1/2)E\left(H_{\gamma_n,h_n}(\mathbf{z}_1,\mathbf{z}_2)^2\right)$ is bounded, we get that $n\gamma_n^pE\left(U_{\lambda,n}-\widehat{U}_{\lambda,n}\right)^2\le C/(n\,h_n^p)\to 0$, and so $(n\gamma_n^p)\,U_{\lambda,n}$ and $(n\gamma_n^p)\,\widehat{U}_{\lambda,n}$ have the same asymptotic behavior. Therefore, we obtain that $(n\gamma_n^p)^{\frac{1}{2}}\,A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x}))$ has the same asymptotic behavior as

$$
\begin{aligned}
W_n &= (n\gamma_n^p)^{-\frac{1}{2}}\left[\Delta_{11,n}+\left(\sigma(\mathbf{x})A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\right)^{-1}\Delta_{31,n}\right]+a_1\\
\Delta_{11,n} &= \sum_{i=1}^n L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\delta_i\psi(\epsilon_i)\\
\Delta_{31,n} &= \gamma_n^p\sum_{i=1}^n\Lambda_{\gamma_n,h_n}(\mathbf{z}_i)=\sum_{i=1}^n\kappa^p\frac{\sigma(\mathbf{x})(1-p(\mathbf{x}))}{p(\mathbf{x})}\psi\prime(0)\,\delta_i\psi_1(\epsilon_i)\,\Gamma\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n},\frac{\gamma_n}{h_n}\right)
\end{aligned}
$$

Therefore, $W_n=a_1+W_{1,n}$ with

$$
W_{1,n} = (n\gamma_n^p)^{-\frac{1}{2}}\sum_{i=1}^n\delta_i\left[L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi(\epsilon_i)+\kappa^p\frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\,p(\mathbf{x})}\psi\prime(0)\,\Gamma\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n},\frac{\gamma_n}{h_n}\right)\psi_1(\epsilon_i)\right]
$$

Since

$$
\begin{aligned}
Var\left(W_{1,n}\right) &= \gamma_n^{-p} E\left\{\delta_i\left[L\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n}\right)\psi\left(\epsilon_i\right)+\kappa^p\frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\,p(\mathbf{x})}\psi\prime\left(0\right)\Gamma\left(\frac{\mathbf{x}_i-\mathbf{x}}{\gamma_n},\frac{\gamma_n}{h_n}\right)\psi_1\left(\epsilon_i\right)\right]^2\right\} \\
&= \gamma_n^{-p}\int p(\mathbf{v})\left[L\left(\frac{\mathbf{v}-\mathbf{x}}{\gamma_n}\right)\psi\left(\epsilon\right)\right. \\
&\quad \left.+\,\kappa^p\frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\,p(\mathbf{x})}\psi\prime\left(0\right)\Gamma\left(\frac{\mathbf{v}-\mathbf{x}}{\gamma_n},\frac{\gamma_n}{h_n}\right)\psi_1\left(\epsilon\right)\right]^2 f_{\mathbf{X}}(\mathbf{v})d\mathbf{v}\;dF_0(\epsilon) \\
&= \int p(\mathbf{v}\gamma_n+\mathbf{x})\Big[L\left(\mathbf{v}\right)\psi\left(\epsilon\right) \\
&\quad +\frac{\kappa^p(1-p(\mathbf{x}))}{A_0\int K(\mathbf{u})d\mathbf{u}\,p(\mathbf{x})}\psi\prime\left(0\right)\Gamma\left(\mathbf{v},\frac{\gamma_n}{h_n}\right)\psi_1\left(\epsilon\right)\Big]^2 f_{\mathbf{X}}(\mathbf{v}\gamma_n+\mathbf{x})d\mathbf{v}\;dF_0(\epsilon)\longrightarrow\sigma_1 \\
\sigma_1 &= p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\int\left[L\left(\mathbf{v}\right)\psi\left(\epsilon\right)+\kappa^p\frac{(1-p(\mathbf{x}))}{A_0(\psi_1)\int K(\mathbf{u})d\mathbf{u}\,p(\mathbf{x})}\psi\prime\left(0\right)\Gamma\left(\mathbf{v},\kappa\right)\psi_1\left(\epsilon\right)\right]^2 d\mathbf{v}\;dF_0(\epsilon)\;.\;\Box
\end{aligned}
$$

References

[1] Aerts, M.; Claeskens, G.; Hens, N. and Molenberghs, G., 2002. Local multiple imputation. Biometrika, 89, 2, 375–388.

[2] Bianco, A. and Boente, G., 2006. Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behavior and bandwidth selection. In press in *Journal of Time Series Analysis*. Available at `http://www.ic.fcen.uba.ar/preprints/biancoboente2006.pdf` .

[3] Boente, G. and Fraiman, R., 1988. Robust nonparametric regression estimation. J. Multivariate Anal. 29, 180-198.

[4] Boente, G. and Fraiman, R., 1989. Robust nonparametric regression estimation for dependent observations. Ann. Statist. 17, 1242-1256.

[5] Boente, G. and Fraiman, R., 1990a. Asymptotic distribution of robust estimates for nonparametric models from mixing observations. Ann. Statist. 18, 891-906.

[6] Boente, G. and Fraiman, R., 1990b. Asymptotic distribution and strong order of convergence of nonparametric regression estimates. Rev. Un. Mat. Argentina 35, 39-52.

[7] Boente, G. and Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. Internat. Statist. Rev. 59, 355-372.

[8] Chen, J. H. and Shao, J., 2000. Nearest neighbor imputation for survey data. J. Official Statist. 16, 113-131.

[9] Cheng, P. E., 1994. Nonparametric estimation of mean functionals with data missing at random. J. Amer. Statist. Assoc. 89, 81-87.

[10] Cheng, P. E. and Chu, C.K., 1996. Kernel estimation of distribution functions and quantiles with missing data. Statist. Sinica 6, 63-78.

[11] Chu, C. K.and Cheng, P. E., 1995. Nonparametric regression estimation with missing data. J. Statist. Plann. Inference, 48, 85-99.

[12] Cleveland, W., 1989. Robust locally weighted regression and smoothing scatter plots. J. Amer. Statist. Assoc. 74, 829-836.

[13] Collomb, G., 1979. Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. C. R. Acad. Sc. Paris 228, 161-163.

[14] Collomb, G., 1980. Estimation de la regression par la méthode des k points les plus proches avec noyau, pp. 159-175, Lecture Notes in Mathematics # 821, SpringerVerlag, Berlin.

[15] Collomb, G., 1981. Estimation non paramétrique de la regression: revue bibliographique. Internat. Statist. Rev. 49, 75-93.

[16] González–Manteiga, W. and Pérez–González, A., 2004. Nonparametric mean estimation with missing data. Comm. Statist. Theory Methods 33, 277-303.

[17] Greblicki, W., Krzyzak, A. and Pawlak, M., 1984. Distribution free pointwise consistency of kernel regression estimate. Ann. Statist. 12, 1570-1575.

[18] Härdle, W., 1984. Robust regression function estimation. J. Multivariate Anal. 14,169-180.

[19] Härdle, W., 1990. *Applied nonparametric regression*, Cambridge University Press.

[20] Härdle, W. and Tsybakov, A. B., 1988. Robust nonparametric regression with simultaneous scale curve estimation. Ann. Statist. 16, 120-135.

[21] Huber, P., 1981. *Robust statistics*, Wiley, New York.

[22] Leung, D., 2005. Cross-validation in nonparametric regression with outliers. *Ann. Statist.*, **33**, 2291-2310.

[23] Leung, D., Marrot, F. and Wu, E., 1993. Bandwidth selection in robust smoothing. J. Nonparametr. Statist. 4, 333-339.

[24] Little, R.J.A. and Rubin, D.B., 1987. *Statistical Analysis with Missing Data*, Wiley, New York.

[25] Meng, X.-L., 2000. Missing data: Dial M for ???. J. Amer. Statist. Assoc. 95, 452, 1325-1330.

[26] Nadaraya, E. A., 1964. On estimating regression. Theory Probab. Appl. 9, 141-142.

[27] Neyman, J., 1938. Contribution to the theory of sampling human populations. J. Amer. Statist. Assoc. 33 101-116.

[28] Schuster, E., 1972. Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Statist. 43, 84-88.

[29] Tsybakov, A. B., 1982. Robust estimates of a function. Problems Inform. Transmission 18, 190-201.

[30] Wang, F. and Scott, D., 1994. The L_1 method for robust nonparametric regression. J. Amer. Stat. Assoc. 89, 65-76.

[31] Wang, Q.; Linton, O. and Härdle, W., 2004. Semiparametric regression analysis with missing response at random. J. Amer. Statist. Assoc. 99, 466, 334–345.

[32] Wang, W. and Rao, J. N. K., 2002. Empirical likelihood-based inference under imputation for missing response data. Ann. Statist. 30, 896-924.

[33] Watson, G. S., 1964. Smooth regression analysis. Sankhy$\overline{a}$ A 26, 359-372.

[34] Yates, F., 1933. The analysis of replicated experiments when the field results are incomplete. Emporium J. Exp. Agriculture 1, 129-142.

	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$
$\widehat{m}_{\text{LS,C}}$	0.23	0.34	0.41
$\widehat{m}_{\text{M,C}}$	0.23	0.27	0.3
$\widehat{m}_{\text{LS,S}}$	0.25	0.39	0,49
$\widehat{m}_{\text{M,S}}$	0.26	0.29	0.34
		(h, γ)	
$\widehat{m}_{\text{LS,I}}$	$(0.21, 0.24)$	$(0.22, 0.38)$	$(0.25, 0.47)$
$\widehat{m}_{\text{M,I}}$	$(0.19, 0.25)$	$(0.21, 0.28)$	$(0.25, 0.3)$

Table 1: Smoothing parameters for each scenario for the linear and the robust nonparametric estimators.

	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$
$\widehat{m}_{\text{LS,C}}$	0.01275	0.03034	0.04558
$\widehat{m}_{\text{M,C}}$	0.01319	0.01566	0.02051
$\text{EF}_{\text{LS,M;C}}$	**-3.452**	**48.371**	**55.003**
$\widehat{m}_{\text{LS,S}}$	0.02187	0.05190	0.07309
$\widehat{m}_{\text{M,S}}$	0.02264	0.02737	0.03504
$\text{EF}_{\text{LS,M;S}}$	**-3.536**	**47.258**	**52.061**
$\widehat{m}_{\text{LS,I}}$	0.02088	0.05004	0.07133
$\widehat{m}_{\text{M,I}}$	0.02197	0.02592	0.03223
$\text{EF}_{\text{LS,M;I}}$	**-5.207**	**48.197**	**54.809**

Table 2: Mean Integrated Squared Error (MISE) for the linear and the robust nonparametric estimators.

	$\widehat{m}_{\mathrm{M,C}}$	$\widehat{m}_{\mathrm{M,S}}$	$\widehat{m}_{\mathrm{M,I}}$
$\alpha = 0$	39.1	39.7	41.7
$\alpha = 0.1$	79.2	81.5	79.6
$\alpha = 0.2$	83.5	81.9	83.3

Table 3: Percentage of times that the ISE for the robust estimators is less than the ISE for their linear counterparts.

	% $\mathrm{ISE}(\widehat{m}_{\mathrm{M,I}}) < \mathrm{ISE}(\widehat{m}_{\mathrm{M,S}})$	$\mathrm{EF}_{\mathrm{M;\,S,I}}$
$\alpha = 0$	61.4	2.982
$\alpha = 0.1$	63.5	5.294
$\alpha = 0.2$	68.2	8.009

Table 4: Percentage of times that the ISE for the robust imputed estimator is less than the ISE for the robust simplified estimator.

	RCV_1			RCV_2		
	$\widehat{m}_{\mathrm{M,C}}$	$\widehat{m}_{\mathrm{M,S}}$	$\widehat{m}_{\mathrm{M,I}}$	$\widehat{m}_{\mathrm{M,C}}$	$\widehat{m}_{\mathrm{M,S}}$	$\widehat{m}_{\mathrm{M,I}}$
C_0	22.0	36.5	36.0	30.5	39.0	41.0
C_1	65.5	74.0	75.0	71.5	76.0	77.0
C_2	75.5	79.5	78.5	79.5	81.0	81.0

Table 5: Percentage of times that the ISE for the robust estimators is less than the ISE for their linear counterparts when using cross–validation.

	RCV_1		RCV_2	
	% $\mathrm{ISE}(\widehat{m}_{\mathrm{M,I}}) < \mathrm{ISE}(\widehat{m}_{\mathrm{M,S}})$	$\mathrm{EF}_{\mathrm{M;\,S,I}}$	% $\mathrm{ISE}(\widehat{m}_{\mathrm{M,I}}) < \mathrm{ISE}(\widehat{m}_{\mathrm{M,S}})$	$\mathrm{EF}_{\mathrm{M;\,S,I}}$
C_0	51.5	−1.399	51.5	−0.953
C_1	56.5	2.827	54.0	1.480
C_2	61.0	5.601	58.0	3.270

Table 6: Percentage of times that the ISE for the robust imputed estimator is less than the ISE for the robust simplified estimator when using robust cross–validation.

	RCV_1		RCV_2	
	% $\mathrm{ISE}(\widehat{m}_{\mathrm{M,I}}) < \mathrm{ISE}(\widehat{m}_{\mathrm{M,S}})$	$\mathrm{EF}_{\mathrm{M;\,S,I}}$	% $\mathrm{ISE}(\widehat{m}_{\mathrm{M,I}}) < \mathrm{ISE}(\widehat{m}_{\mathrm{M,S}})$	$\mathrm{EF}_{\mathrm{M;\,S,I}}$
C_0	51.5	−1.399	51.5	−0.953
C_1	56.5	2.827	54.0	1.480
C_2	61.0	5.601	58.0	3.270

Table 7: Percentage of times that the ISE for the robust imputed estimator is less than the ISE for the robust simplified estimator when using robust cross–validation.

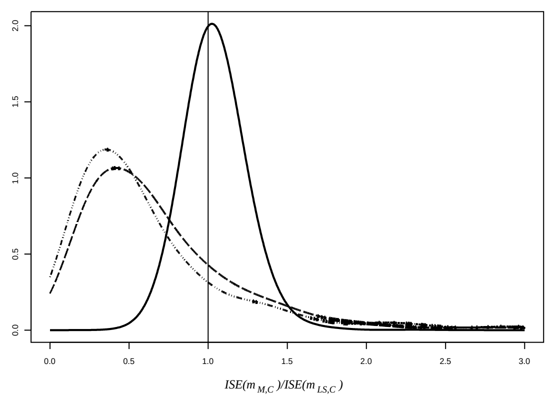

Figure 1: Density estimator of the ratio between the ISE of the robust estimator and that of the linear estimator for the complete data estimator. The solid line corresponds to $\alpha = 0$, while the broken $(- \, -)$ and dashed $(- \cdots -)$ lines correspond to $\alpha = 0.1$ and 0.2, respectively.

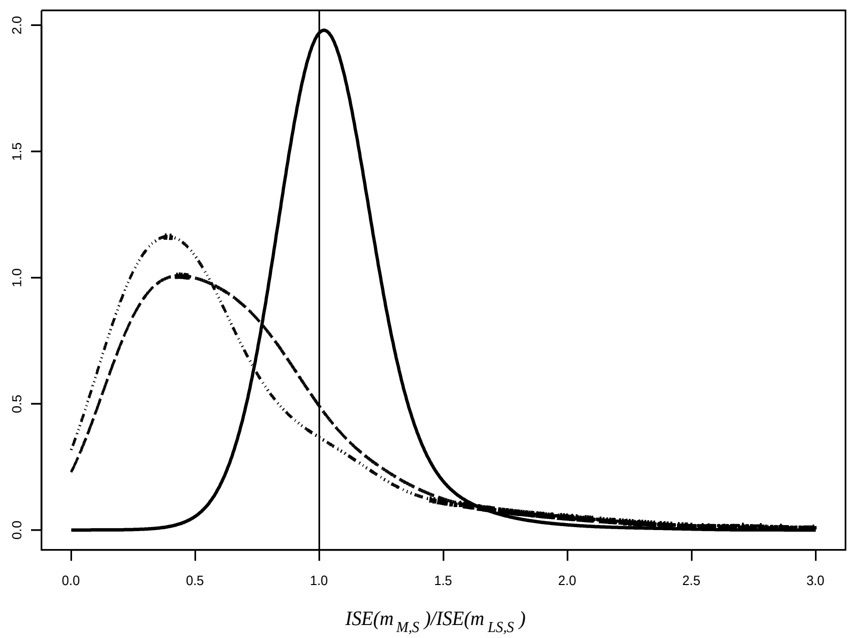

Figure 2: Density estimator of the ratio between the ISE of the robust estimator and that of the linear estimator for the simplified estimator. The solid line corresponds to $\alpha = 0$, while the broken $(- \, -)$ and dashed $(- \cdots -)$lines correspond to $\alpha = 0.1$ and 0.2, respectively.

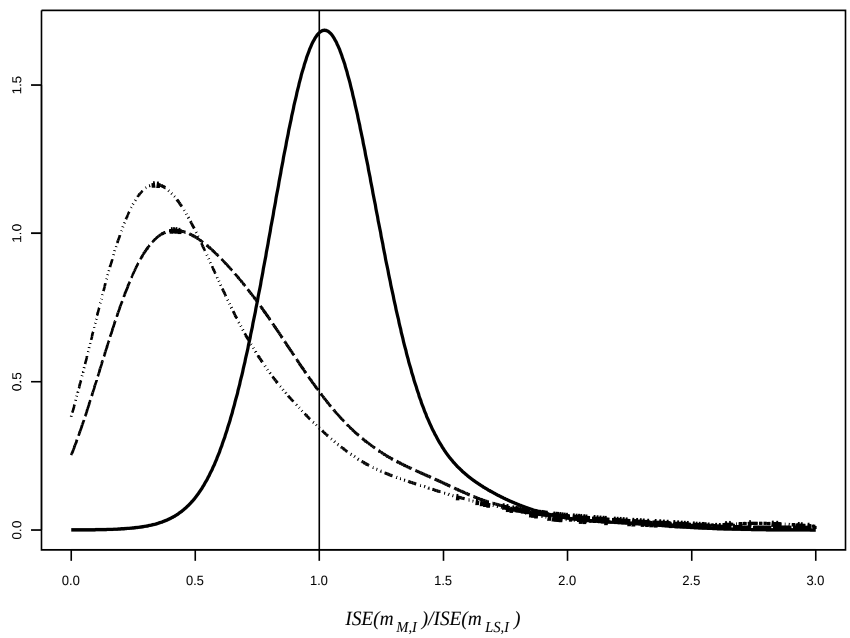

Figure 3: Density estimator of the ratio between the ISE of the robust estimator and that of the linear estimator for the imputed estimator. The solid line corresponds to $\alpha = 0$, while the broken $(- \, -)$ and dashed $(- \cdots -)$lines correspond to $\alpha = 0.1$ and 0.2, respectively.

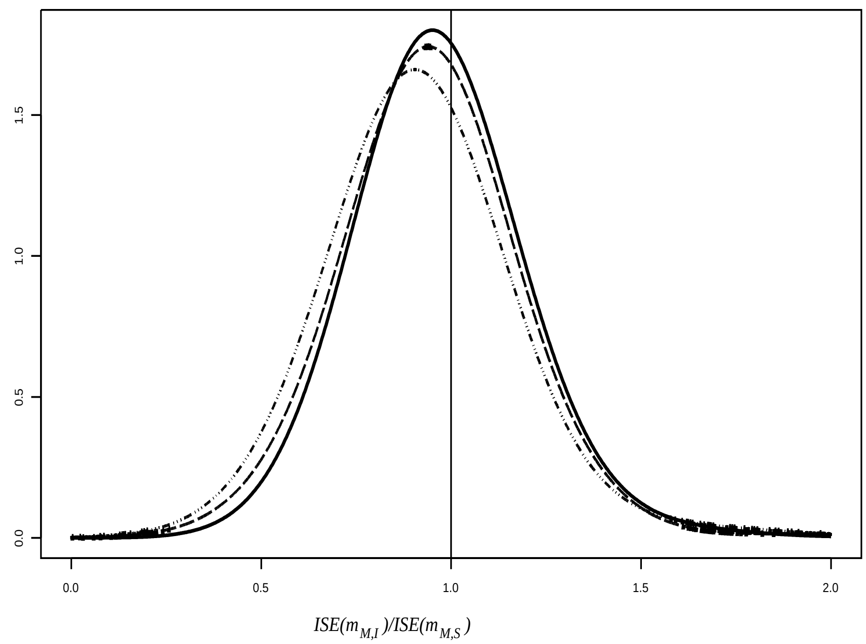

Figure 4: Density estimator of the ratio between the ISE of the imputed robust estimator and that of the simplified robust one. The solid line corresponds to $\alpha = 0$, while the broken $(- \ -)$ and dashed $(- \cdots -)$lines correspond to $\alpha = 0.1$ and 0.2, respectively.

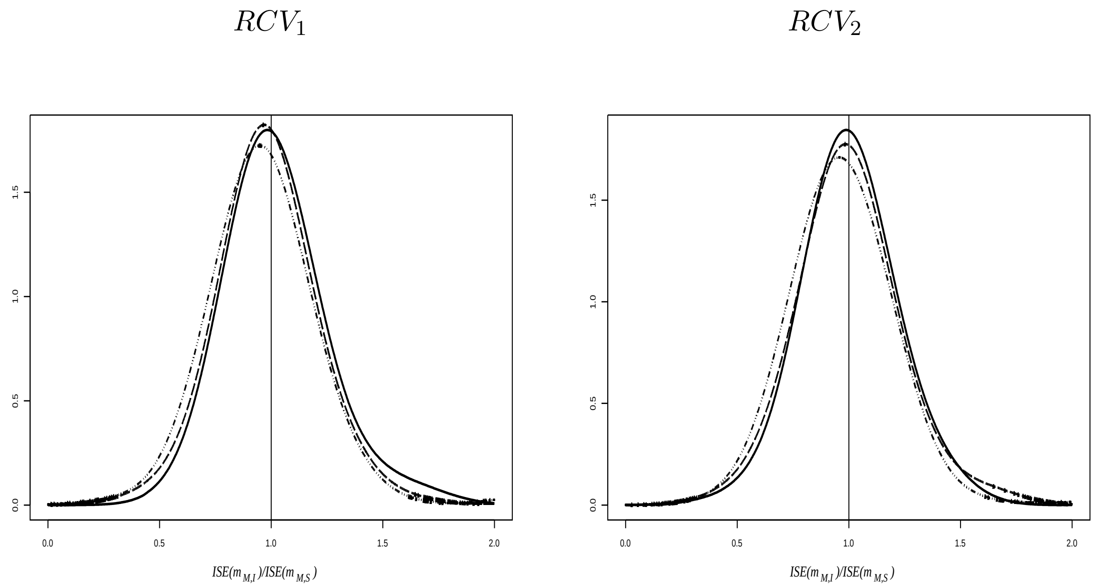

Figure 5: Density estimator of the ratio between the ISE of the imputed robust estimator and that of the simplified robust one when using a robust cross–validation criterium. The solid line corresponds to C_0, while the broken $(- \ -)$ and dashed $(- \cdots -)$ lines correspond to C_1 and C_2, respectively.

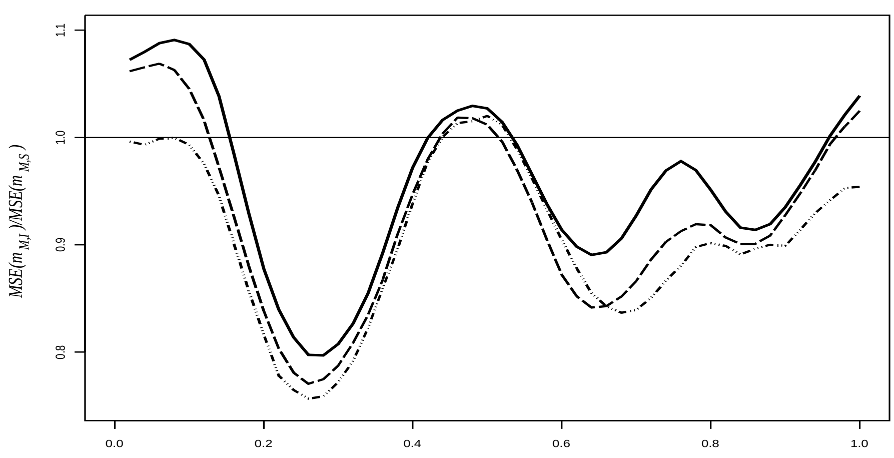

Figure 6: Ratio betwen the MSE of the imputed robust estimator and that of the simplified robust one across the values of x. The solid line corresponds to C_0, while the broken $(- \ -)$ and dashed $(- \cdots -)$lines correspond to C_1 and C_2, respectively.

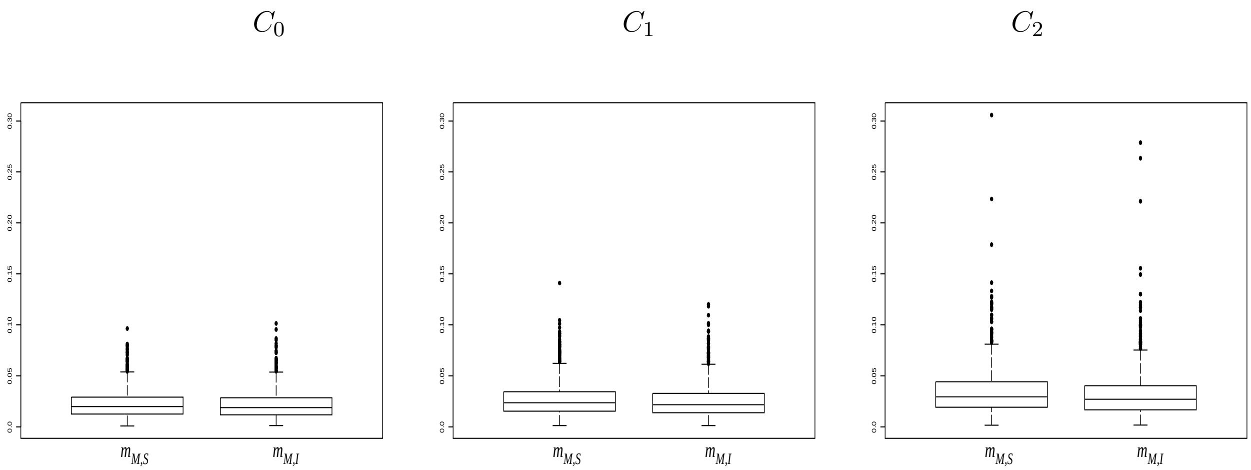

Figure 7: Boxplots of the ISE for the simplified and the imputed robust estimators for $\alpha = 0$, 0.1 and 0.2.

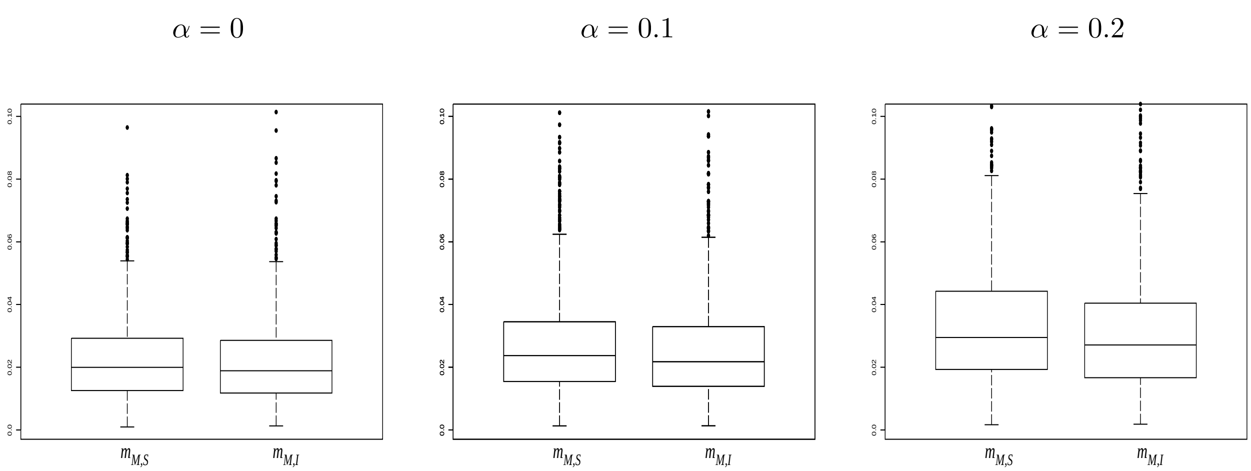

Figure 8: Boxplots of the ISE for the simplified and the imputed robust estimators for $\alpha = 0$, 0.1 and 0.2. Values larger than 0.10 are not plotted.

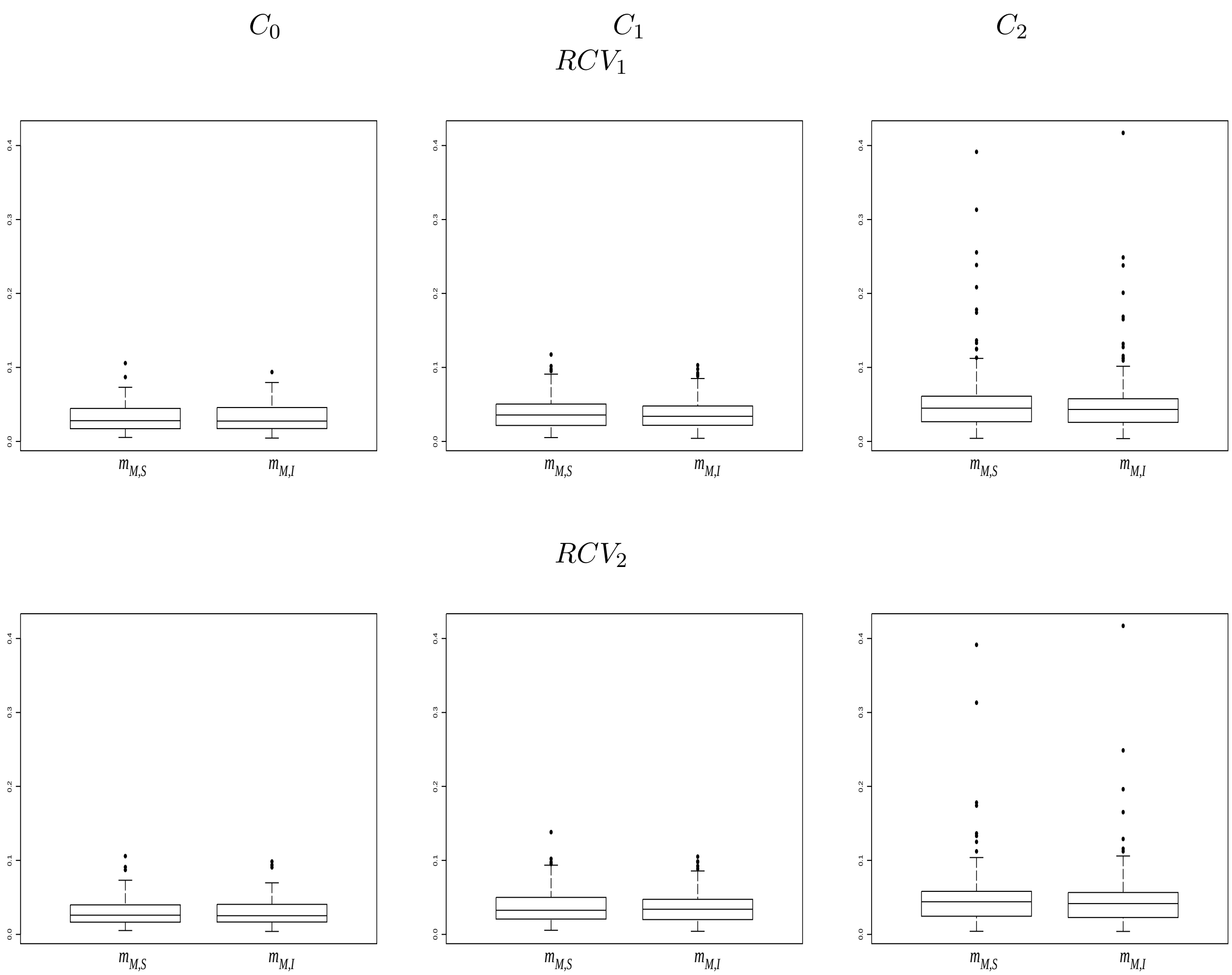

Figure 9: Boxplots of the ISE for the simplified and the imputed robust estimators for $\alpha = 0$, 0.1 and 0.2 using robust cross–validation.

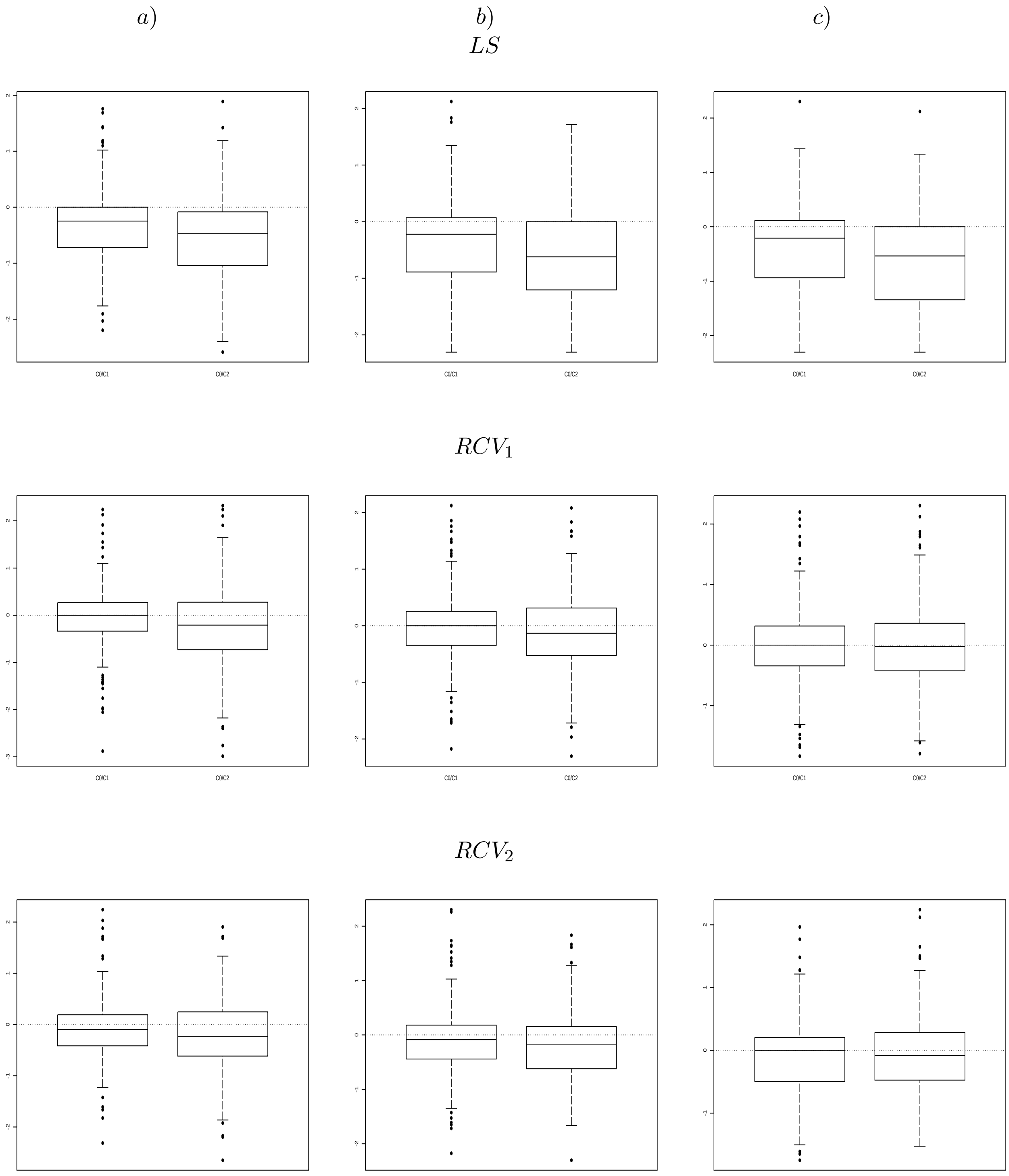

Figure 10: Boxplots of $\log(\widehat{h}_{n,C_0}/\widehat{h}_{n,C_1})$ and $\log(\widehat{h}_{n,C_0}/\widehat{h}_{n,C_2})$ for the complete (a), simplified (b) and imputed (c) robust estimators.