Čížek, Pavel

Working Paper

Robust estimation in nonlinear regression models

SFB 373 Discussion Paper, No. 2001,25

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Čížek, Pavel (2001) : Robust estimation in nonlinear regression models, SFB 373 Discussion Paper, No. 2001,25, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10049502

This Version is available at:
https://hdl.handle.net/10419/62757

Robust Estimation in Nonlinear Regression Models

Pavel Čížek

Initial version: 6th May 2001
Revised version: 21st October 2001

Contents

1 Introduction

Least trimmed squares (LTS) is a statistical technique for estimation of unknown parameters of a linear regression model. It was proposed by Rousseeuw (1985) as a robust alternative to the classical regression method based on minimizing the sum of squared residuals, which, while being frequently used in regression analysis, is quite sensitive to data contamination and model misspecification. Although the asymptotic and robust properties of this estimators were already studied by Rousseeuw and Leroy (1987), at least in the case of regression with one explanatory variable, LTS was not widely used until recently. There are several reason for this, but the main one is computational: it is possible to compute LTS only approximately and even obtaining an approximation was relatively time consuming; moreover, a good approximation algorithm did not previously exist. However, availability of a good and fast approximation algorithm (see, for example, Rousseeuw and Van Driessen (1999)) and faster computers make LTS more attractive recently, of course, hand in hand with the presence of these algorithm in some widely-spread statistical packages. In addition, a general proof of consistency, asymptotic normality of LTS appeared as well as some sensitivity studies of this estimator (Víšek (1999a)).

Still, the LTS estimator has several shortcomings, concerning especially its applicability. There are several classes of regression models in which such a robust estimator cannot be used right now, partly because its properties are not known in such models, partly because it is not adapted to suit such models. This concerns estimation with discrete explanatory variables, estimation of nonlinear model, and more generally, limited-dependent-variable and discrete-choice models and their variants. In this paper, I aim to make a first step concerning the LTS estimation of nonlinear regression models, namely to prove consistency and asymptotic normality of LTS when used in nonlinear regression models. Moreover, the techniques used and developed for these proofs are supposed to suit well a subsequent analysis of LTS in more complicated models such as, for example, limited-dependent-variable models.

Why is it useful to think about nonlinear models at all? Let me exemplify this. It is sometimes not clear, for instance, which functional form describes best the dependence on an explanatory variable. To resolve this point, the Box-Cox transformation can be used (see Box and Cox (1964)), that is a transformation of a random variable Z parameterized by $\lambda \in \mathbb{R}$ of the following form:

$$Z^{(\lambda)} = \begin{cases} \dfrac{Z^\lambda - 1}{\lambda} & \text{for} \quad \lambda \neq 0, \\ \ln Z & \text{for} \quad \lambda = 0. \end{cases}$$

Its advantage is that $Z^{(\lambda)}$ represents various functions of Z for different values of λ: linear ($\lambda = 1$), square root ($\lambda = 1/2$), logarithmic ($\lambda = 0$), inversely proportional ($\lambda = -1$), and so on. Applying the transformation both to the dependent and independent variables provides then parameterized choice between different regression models (linear, log-linear, semi-logarithmic, reciprocal, etc.). Another example of an intrinsically nonlinear model can be a model with an exponential regression function but an additive error term (instead of a multiplicative one). Finally, time series models with state-dependent regression function are becoming more widely used (see Tong (1990) for summary of these models) and they are typically estimated with nonlinear least squares as well. Thus, they represent another class where nonlinear LTS can be applied.

Let me now precise the goal of the work. In this paper, I study the behavior of the LTS estimator applied in the nonlinear regression model

$$y_i = h(x_i, \beta) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $y_i \in \mathbb{R}$ represents the dependent variable and $h(x, \beta)$ is a function of $x_i \in \mathbb{R}^k$, a vector of explanatory variables, and of $\beta \in \mathbb{R}^p$, a vector of unknown parameters. The error terms ε_i are assumed to form a sequence of independent and identically distributed random variables that possess an absolutely continuous distribution function. The LTS estimator used within this framework is further referred to as the *nonlinear least trimmed squares* estimator (NLTS) in order to differentiate it from the LTS estimator used within the linear-regression framework.

The analysis is carried out along the same lines as recent proofs of the LTS asymptotic properties in Víšek (1999a). This means that, starting from a very similar basic assertions, I derive asymptotic linearity of NLTS, and based on this result, consistency and asymptotic normality of this estimator are proved. However, the theoretical tools and results used in these proofs are quite different from the strategy used by Jurečková and Sen (1989), Víšek (1996a) or Víšek (1999a)—I use relatively simple and more direct techniques instead of elegant, but very advanced theoretical results, such as Skorohod embedding in Wiener process, that could eventually be hard to use once we turn our attention to models with more complicated error structure, for example, limited-dependent-variable models. Thus, this simplification of the methodology should render a generalization to these more complicated models. In addition to that, I provide here a most general set of assumption required for the consistency of NLTS estimator.

In the rest of the paper, I first review important facts about LTS that are related to NLTS (Section 2.1), later I discuss necessary assumptions for the asymptotic properties of NLTS (Section 3.2), and finally, I derive asymptotic linearity, consistency, and asymptotic normality of the proposed estimator (Sections 3.4 and 3.5).

2 Definition of nonlinear least trimmed squares

To assure easy understanding, it is beneficial to describe first the least trimmed squares estimator (LTS), introduced by Rousseeuw (1985), and its properties.

2.1 Least trimmed squares

Let us consider a linear regression model for a sample (y_i, x_i) with a dependent variable $y_i \in \mathbb{R}$ and a vector of explanatory variables $x_i \in \mathbb{R}^p$:

$$y_i = \beta^T x_i + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{2}$$

The least trimmed squares estimator $\hat{\beta}_n^{(LTS)}$ is defined as

$$\hat{\beta}_n^{(LTS)} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{h} r_{[i]}^2(\beta), \tag{3}$$

where $r_{[i]}^2(\beta)$ represents the ith order statistics of squared residuals $r_1^2(\beta), \ldots, r_n^2(\beta)$; $r_i(\beta) = y_i - \beta^T x_i$ and $\beta \in \mathbb{R}^p$ (p denotes the number of estimated parameters). The *trimming constant* h have to satisfy $\frac{n}{2} < h \leq n$. This constant determines robustness of the LTS estimator, since definition (3) implies that $n - h$ observations with the largest residuals do not have a direct influence on the estimator. The highest level of robustness is achieved for $h = [n/2] + [(p+1)/2]$ (Theorem 6, Rousseeuw and Leroy (1987)), whereas the LTS robustness is lowest for $h = n$, which corresponds to the least squares estimator. There is, of course, a trade-off: lower values of h lead to a higher degree of robustness, while higher values of h improve efficiency (if the data are not too contaminated) since more (presumably correct) information in the data is utilized. The most robust choice of h is often employed when the LTS is used for diagnostic purposes. It may also be favored when LTS is used for comparison with some less robust estimator, e.g., the least squares, because a comparison of these two estimators can serve as a simple check of data and a model—if the estimates are not similar to each other, special care should be taken throughout the analysis. On the other hand, it may be sensible to evaluate LTS for a wide range of trimming-constant values and to observe how the estimate behaves with increasing h, because this dependence can provide hints about the amount of contamination and possibly about suspicious structures in studied data (for example, that the data contain actually a mixture of two different populations, see Víšek (1999a)).

Before proceeding further, it seems to be useful to discussed several issues, namely, the existence of this estimator and its statistical properties. First, the existence of the optimum in (3) under some reasonable assumptions can be justified in the following way: the minimization of the objective function in (3) can be viewed as a process in which we choose every time a subsample of h observations and find some β minimizing the sum of squared residuals for the selected subsample. Doing this for every subsample, we get $\binom{n}{h}$ candidates for the LTS estimate and the one that commands the smallest value of the objective function is the final estimate. Therefore, the existence of the LTS estimator is basically equivalent to the existence of the least squares estimator for subsamples of size h.

2.2 Definition of nonlinear least trimmed squares

The nonlinear least trimmed squares estimator is defined in this section. Before a definition is given, let me first clarify for which regression models I aim to define NLTS. Let us assume that there is a regression model of the form[1]

$$y_i = h(x_i, \beta) + \varepsilon_i, \tag{4}$$

[1] Although the regression function $h(\cdot, \cdot)$ uses the same symbol as the trimming constant h (both represent traditional notation), no confusion should arise—the regression function is always used with its parameters enclosed in brackets.

where y_i is the dependent variable, $h(x_i, \beta)$ is a known (regression) function of the data x_i and a vector β of p unknown parameters; $(x_i, \varepsilon_i) \in \mathbb{R}^k \times \mathbb{R}, i = 1, \ldots, n$, are is a sequence of independent identically distributed random vectors.[2] Random variables x_i and ε_i are assumed to be mutually independent. Now, having specified the regression model (4), we attempt to define the nonlinear least squares estimator (NLTS), i.e., the estimator based on minimizing the trimmed sum of squared residuals in regression model (4). Given a sample $(y_i, x_i) \in \mathbb{R} \times \mathbb{R}^k, i = 1, \ldots, n$, the NLTS estimate $\hat{\beta}_n^{(NLTS,h)}$ is defined by

$$\hat{\beta}_n^{(NLTS,h)} = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{h} r_{[i]}^2(\beta), \tag{5}$$

where

- $\beta \in B \subset \mathbb{R}^p$ is a p-dimensional vector of unknown parameters and $B \subset \mathbb{R}^p$ is the corresponding parameter space,
- $r_{[i]}^2(\beta)$ represents the ordered sample of squared residuals $r_i^2(\beta) = (y_i - h(x_i, \beta))^2, i = 1, \ldots, n$, for any $\beta \in B$,
- $h \in \left\{ \left[\frac{n+1}{2} \right], \ldots, n \right\}$ is the *trimming constant* (see Section 2.1).

Apparently, this estimator shares its robustness properties with the already reviewed LTS (see Section 2.1) and the same is true for many of its finite sample and asymptotic properties as will become gradually evident in what follows. For instance, the computation of NLTS could be theoretically done by performing the nonlinear least squares method (NLS) for each of $\binom{n}{h}$ subsamples of size h and taking as the final estimate the NLS estimate corresponding to the subsample at which the sum of squared residuals reaches its minimum. Therefore, if the NLS estimator exists for all $\binom{n}{h}$ subsamples of $(y_i, x_i), i = 1, \ldots, n$, NLTS also exists (it is just the minimum of a finite number of values). Consistency of NLS and NLTS estimators are related in a similar way—see Section 3.4 for further information.

3 Consistency and asymptotic linearity of NLTS

In this section I present the main asymptotic results concerning NLTS, namely, its asymptotic linearity, consistency, and asymptotic normality. Before proving these properties, an alternative definition of NLTS and some notational conventions used in the rest of the paper are mentioned as well as the assumptions necessary for the mentioned asymptotic results.

3.1 Alternative definition of NLTS, notation

Given a sample (y_i, x_i), the NLTS estimator of unknown parameter vector β is defined for model (4) by equation (5). The dependent variable is denoted $y_i \in \mathbb{R}$, the vector of explanatory variables $x_i \in \mathbb{R}^k$, and ε_i represents the error term. In addition, Ω_x and Ω_ε refer to probability spaces on which x_i and ε_i are defined, so $\Omega = \Omega_x \times \Omega_\varepsilon$ is the probability space of the random vector (x_i, ε_i). The true underlying value of the vector β in (4) will be referred to by β^0. The nonlinear least squares estimator, which naturally coincides with $\hat{\beta}_n^{(NLTS,n)}$, is denoted

$$\hat{\beta}_n^{(NLS)} = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{n} r_{[i]}^2(\beta) = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta) = \underset{\beta \in B}{\arg\min} \sum_{i=1}^{n} (y_i - h(x_i, \beta))^2.$$

Here and in definition (5), $r_{[i]}^2(\beta)$ stands for the ith order statistics of squared residuals $r_i^2(\beta) = (y_i - h(x_i, \beta))^2$. In other words, it holds that $0 \leq r_{[1]}^2(\beta, \omega) \leq \cdots \leq r_{[n]}^2(\beta, \omega)$ for any $\beta \in B$

[2] In general, the dimensions of x_i and β do not have to be the same, i.e., $k \neq p$.

and $\omega \in \Omega$.[3] Given an $\omega \in \Omega$, we understand by symbol $r_{[i]}(\beta,\omega)$ the value of residual $r_k(\beta,\omega)$ such that $r_k^2(\beta,\omega) = r_{[i]}^2(\beta,\omega)$; hence, $\left|r_{[i]}(\beta)\right| = \sqrt{r_{[i]}^2(\beta)}$. If it is necessary to refer to the order statistics of sample $r_1(\beta),\ldots,r_n(\beta)$, then notation $r_{(i)}(\beta)$ would be used.

Next, an alternative definition of NLTS employed in the theoretical part of this paper instead of (5) is given by[4]

$$\hat{\beta}_n^{(NLTS,h)} = \arg\min_{\beta\in B} \sum_{i=1}^{n} r_i^2(\beta)\cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big) \tag{6}$$

$$= \arg\min_{\beta\in B} \sum_{i=1}^{n} (y_i - h(x_i,\beta))^2\cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big). \tag{7}$$

To obtain this formula, one has to realize that for a given value of $\beta \in B$, the minimization of the h smallest squared residuals means that we include in the objective function only those residuals that satisfy $r_i^2(\beta) \le r_{[h]}^2(\beta)$.[5] One additional note concerns the trimming constant: whenever asymptotic properties of NLTS are studied, that is $n \to +\infty$, we have to work with a sequence of trimming constants h_n (for every sample size n, there has to be a corresponding choice of h). As this constant determines the robustness properties of NLTS, we want to prescribe asymptotically a fixed fraction λ of observations that are considered to be correct, $\frac{1}{2} < \lambda \le 1$, or alternatively, a fraction $1-\lambda$ of observations that are excluded from the objective function of NLTS ($0 \le 1-\lambda < \frac{1}{2}$). The trimming constant for a given $n \in \mathbb{N}$ can be then defined by $h_n = [\lambda n]$, where $[x]$ represents the integer part of x, and hence $h_n/n \to \lambda$. From now on, we assume that there is such a number $\lambda \in (\frac{1}{2}, 1\rangle$ for a sequence h_n of trimming constants defining the NLTS estimator.[6]

To close this section, we discuss some purely mathematical notation. As observations and parameters considered here always belong to an Euclidean space $\mathbb{R}^l$, we shall need to define a neighborhood of a point $x \in \mathbb{R}^l$: an open neighborhood (open ball) $U(x,\delta) = \{z \in \mathbb{R}^l : \|z-x\| < \delta\}$ and a closed neighborhood (closed ball) $\bar{U}(x,\delta) = \{z \in \mathbb{R}^l : \|z-x\| \le \delta\}$. Moreover, let us denote a convex span of $x_1,\ldots,x_m \in \mathbb{R}^l$ by $[x_1,\ldots,x_m]_{\bigstar}$. Finally, several symbols from linear algebra are introduced: 1_n represents n-dimensional vector of ones, $b_1,\ldots,b_n$ are standard basis vectors in $\mathbb{R}^n$, i.e., $b_k = (0,\ldots,0,1,0,\ldots,0)$, and $\mathcal{I}_n$ is the identity matrix of dimension n.

3.2 Assumptions

Now, I specify all the assumptions necessary to prove the asymptotic linearity of NLTS. They form two groups—distributional assumptions D for random variables in (4) and assumptions H concerning properties of function $h(x,\beta)$.

First of all, let me discuss the distributional assumptions D dealing with the random variables used in model (4). All of these conditions that are applicable in linear regression models are analogous to assumptions $\mathcal{A}$ in Víšek (1999a) or to the assumption of the standard linear regression model. Moreover, we argue in a number of remarks that the following assumptions do not restrict us in any way in real applications.

Assumptions D.

[3] Since $y_i = h(x_i,\beta) + \varepsilon_i$ and $r_i = y_i - h(x_i,\beta) = h(x_i,\beta^0) - h(x_i,\beta) + \varepsilon_i$, regression residuals can be written as a function of β and $\omega \in \Omega = \Omega_x \times \Omega_\varepsilon$.

[4] By I(property describing a set A) we denote the indicator of the set A.

[5] In general, this definition is not equivalent to the first one. They are exactly equivalent if and only if all the residuals are different from each other. Under the assumptions given in Section 3.2, this happens with zero probability and definitions (5) and (6) are equivalent almost surely as the cumulative distribution function of $r_i(\beta)$ is assumed to be absolutely continuous. Therefore, we further use definition (6) for convenience.

[6] The case of $\lambda = 1$ will be excluded for the sake of simplicity from some proofs. This case corresponds, indeed, to the usual nonlinear least squares estimator, which is extensively studied in the literature anyway. All the propositions given later are valid for $\lambda = 1$ too, but their proofs are slightly different or trivial in this case.

D1 Let $(x_i, \varepsilon_i) \in \mathbb{R}^k \times \mathbb{R}, i = 1, \ldots, n$, be a sequence of independent identically distributed random vectors with finite fourth moments and let x_i and ε_i be mutually independent. Moreover,

$$n^{-1/4} \max_{i,j} |x_{ij}| = \mathcal{O}_p(1). \tag{8}$$

***Remark* 1** *The necessity to include restriction (8) is caused by the discontinuity of the objective function of NLTS. A nonrandom version of this assumption was used for the first time by Jurečková (1984) and the presented version was introduced by Víšek (1999a). Using Proposition 1, we can say that equation (8) holds even for some distribution functions with polynomial tails, namely for those that have finite second moments. This becomes apparent once we realize that a distribution with tails behaving like one over a polynomial of the third (or lower) order does not have finite second moments. As the existence of finite second moments is almost always utilized, and moreover, it is one of the necessary conditions here, assumption (8) should not pose a considerable restriction on the explanatory variables. You can also notice that random variables with a finite support are not restrained by this assumption in any way.*

Proposition 1 *Let $x_1, x_2, \ldots$ be a sequence of independent identically distributed random variables with a distribution function $F(x)$. Let $b(x)$ be a lower bound for $F(x)$ in a neighborhood U_1 of $+\infty$. If $b(x)$ can be chosen as $1 - \frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \max_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$. Analogously, let $c(x)$ be an upper bound for $F(x)$ in a neighborhood U_2 of $-\infty$. If $c(x)$ can be chosen as $\frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \min_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$.*

D2 We assume $\mathsf{E}\, x_i x_i^T = Q$, where Q is a nonsingular matrix, and

$$\mathsf{E}\left(\varepsilon_i I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \middle| x_i\right) = 0, \qquad \mathsf{E}\left(\varepsilon_i^2 I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \middle| x_i\right) = \sigma^2,$$

where $\sigma^2 \in (0, +\infty)$.

***Remark* 2** *This is nothing but a natural analogy of usual orthogonality condition $\mathsf{E}(\varepsilon|x) = 0$ and spheriality condition $\mathsf{E}(\varepsilon\varepsilon^T|x) = \sigma^2\mathcal{I}$ in the case of the linear regression model. For further discussion, see Víšek (1999a).*

D3 The distribution function F of ε_i is absolutely continuous. Let f denote the probability density of F, which is assumed to be positive, bounded by $M_f > 0$ and differentiable on the whole support of the distribution function F. Let f' denote the first derivative of f.

Furthermore, a piece of notation is added. Let $G(x)$ represents the distribution function of ε_i^2. It follows that $G(x) = F(\sqrt{x}) - F(-\sqrt{x})$ for $x > 0$, $G(x) = 0$ otherwise, and hence, it is also absolutely continuous. Therefore, we can define $g(x)$ to be the corresponding probability density function. It is positive on the support of $G(x)$, zero elsewhere, and $g(x) = \frac{1}{2\sqrt{x}}\left(f(\sqrt{x}) + f(-\sqrt{x})\right)$ for $x > 0$. Consequently, $g(x)$ is bounded on any interval $(K, +\infty), K > 0$ by $M_f/\sqrt{K}$. Moreover, sometimes it is necessary to refer to the distribution function of $r_i(\beta)$ and $r_i^2(\beta)$; in such a case, F_β and G_β are used for the cumulative distribution functions and f_β and g_β for the corresponding probability densities. It follows that $F = F_{\beta^0}$ and $G = G_{\beta^0}$, and similarly, $f = f_{\beta^0}$ and $g = g_{\beta^0}$.

***Remark* 3** *Note that Assumption D3 implies the following property of the distribution function $F(x)$ and its density $f(x)$: for any $0 < \alpha < 1$ we can find $\varepsilon > 0$ such that $\inf_{x \in (F^{-1}(\alpha)-\varepsilon, F^{-1}(\alpha)+\varepsilon)} \min\{F(x), f(x)\} > 0$. The same is true for $G(x)$and $g(x)$.*

***Remark* 4** *In the last assumption, D3, the existence of the probability density function f and its derivative is required. Once we go through the proof of the asymptotic linearity, it will become obvious that if the explanatory variables are uniformly bounded*

$$\sup_{i \in \mathbb{N}, j=1,\ldots,k} |x_{ij}| = \mathcal{O}(1),$$

then it is sufficient for these densities and their derivatives to exist only in a neighborhood of $-\sqrt{G^{-1}(\lambda)}$ *and* $\sqrt{G^{-1}(\lambda)}$. *The same applies for the assumption that the probability densities are bounded.*

As we aim to apply NLTS to nonlinear models, several conditions on the regression function $h(x,\beta)$ have to be specified. Most of them are just regularity conditions that are employed in almost any work concerning nonlinear regression models. For example, the regression function of a nonlinear regression model is almost always assumed to be twice continuously differentiable; see, for example, Amemiya (1983).

As the assumptions stated below rely on the value of β and I do not have to require their validity over the whole parametric space, I restrict β to a neighborhood $U(\beta^0,\delta)$ in these cases and suppose that there exists a positive constant δ such that all the assumptions are valid.

Assumptions H.

H1 Let $h(x_i,\beta)$ be a continuous (uniformly over any compact subset of the support of (x,y)) in $\beta \in B$ and twice differentiable function in β on $U(\beta^0,\delta)$ almost surely:

$$\left(\forall \beta \in U(\beta^0,\delta); \forall x \in A_x \subseteq \mathbb{R}^k, P(A_x)=1; j,k \in \{1,\ldots,p\}\right) \left(\exists \frac{\partial h(x,\beta)}{\partial \beta_j}, \frac{\partial^2 h(x,\beta)}{\partial \beta_j \partial \beta_k}\right).$$

The first derivative is continuous in $\beta \in U(\beta^0,\delta)$.

H2 Furthermore, let us assume that the second derivatives $h''_{\beta_j\beta_k}(x,\beta)$ satisfy locally the Lipschitz property in a neighborhood of β^0, i.e., for any compact subsets of $\operatorname{supp} x$ there exists a constant $L_p > 0$ such that for all $\beta, \beta' \in U(\beta^0,\delta)$, and $j,k = 1,\ldots,p$

$$\left| h''_{\beta_j\beta_k}(x,\beta) - h''_{\beta_j\beta_k}(x,\beta') \right| \leq L_p \cdot \|\beta - \beta'\| .$$

H3 Let

$$n^{-1/4} \max_{1\leq i\leq n} \max_{1\leq j\leq p} \left\| h'_{\beta_j}(x_i,\beta) \right\| = \mathcal{O}_p(1) \tag{9}$$

and

$$n^{-1/2} \max_{1\leq i\leq n} \max_{1\leq j,k\leq p} \left\| h''_{\beta_j\beta_k}(x_i,\beta) \right\| = \mathcal{O}_p(1) \tag{10}$$

as $n \to \infty$ uniformly over $\beta \in B$.

***Remark* 5** *This assumption depicts another regularity condition that is going to be fulfilled in most cases. For example, for a function of the form* $h(x_i^T\beta)$, *where* h *is twice differentiable with bounded derivatives, we can immediately observe that* $h'_{\beta_j}(x,\beta) = h'(x_i^T\beta)x_{ij}$, *and analogously,* $h''_{\beta_j\beta_k}(x,\beta) = h''(x_i^T\beta)x_{ij}x_{ik}$. *Hence, assumptions (9) and (10) are a direct consequence of (8) as long as the first two derivatives of* $h(x,\beta)$ *are bounded on any compact subset of the support of random variable* x.

H4 To proceed further, we have to postulate some assumptions about the following expectations:

- Let $\mathsf{E}\,[h(x_i,\beta)]^m$ and $\mathsf{E}\,[r_i(\beta)]^m = \mathsf{E}\left[\varepsilon_i + h\left(x_i,\beta^0\right) - h(x_i,\beta)\right]^m$ exist and are finite for $m = 1,2$ and any $\beta \in B$.
- Let $\mathsf{E}\left[h'_{\beta_j}\left(x_i,\beta^0\right)\right]^m$, $\mathsf{E}\left[h''_{\beta_j\beta_k}\left(x_i,\beta^0\right)\right]^m$, $\mathsf{E}\left[h'_{\beta_j}\left(x_i,\beta^0\right)\cdot h'_{\beta_k}\left(x_i,\beta^0\right)\right]^m$, and $\mathsf{E}\left[h'_{\beta_l}\left(x_i,\beta^0\right)\cdot h''_{\beta_j\beta_k}\left(x_i,\beta^0\right)\right]$ exist and are finite for $m = 1,2$, and for all $j,k,l = 1,\ldots,p$.
- Moreover, we assume that $\mathsf{E}\,h'_\beta\left(x_i,\beta^0\right)h'_\beta\left(x_i,\beta^0\right)^T = Q_h$, where Q_h is a nonsingular positive definite matrix.

***Remark* 6** *It is important to remember that these assumptions correspond in our nonlinear model to the existence of finite fourth moments (see assumption D1). Moreover, the second part of Assumption H4 is a natural analogy to assumption D2 in the linear regression model.*

3.3 Asymptotic linearity

3.3.1 Normal equations

In order to analyze the behavior of the NLTS estimator, we use normal equations as the starting point, i.e., instead of minimizing the objective function

$$\rho(\beta) = \sum_{i=1}^{n} (y_i - h(x_i, \beta))^2 \cdot I\Big(r_i^2(\beta) \leq r_{[h]}^2(\beta)\Big)$$

over all $\beta \in B$, we consider a solution of $\frac{\partial \rho(\beta)}{\partial \beta} = 0$. The normal equations (for $\beta \in U(\beta^0, \delta)$) can be written as

$$\begin{aligned} 0 = \frac{\partial \rho(\beta)}{\partial \beta} \quad = \quad & \sum_{i=1}^{n} \Bigg[2(y_i - h(x_i, \beta)) \cdot h'_\beta(x_i, \beta) \cdot I\Big(r_i^2(\beta) \leq r_{[h]}^2(\beta)\Big) \\ & + (y_i - h(x_i, \beta))^2 \cdot \frac{\partial}{\partial \beta} I\Big(r_i^2(\beta) \leq r_{[h]}^2(\beta)\Big) \Bigg]. \end{aligned}$$

Now, let us show that the continuity of residuals $r_i^2(\beta)$ and order statistics $r_{[h]}^2(\beta)$ in β (assumptions H) guarantees that the second term is almost everywhere zero. Consider $j = 1, \ldots, p$ and an arbitrary, but fixed $\omega \in \Omega$:

$$\frac{\partial}{\partial \beta_j} I\Big(r_i^2(\beta, \omega) \leq r_{[h]}^2(\beta, \omega)\Big) = \lim_{\Delta \to 0} \frac{1}{\Delta} \left[I\Big(r_i^2(\beta^{(\Delta)}, \omega) \leq r_{[h]}^2(\beta^{(\Delta)}, \omega)\Big) - I\Big(r_i^2(\beta, \omega) \leq r_{[h]}^2(\beta), \omega\Big) \right],$$

where $\beta^{(\Delta)} = (\beta_1, \ldots, \beta_{j-1}, \beta_j + \Delta, \beta_{j+1}, \ldots, \beta_p)$. As the ordering of residuals is constant in a neighborhood of β for all $\omega \in \Omega_1$, where $P(\Omega_1) = 1$ (see Lemma 1), the limit is equal to zero jointly for all $i = 1, \ldots, n$ and $j = 1, \ldots, n$ with probability 1. Consequently, it is enough to study the behavior of

$$\frac{\partial \rho(\beta)}{\partial \beta} = \sum_{i=1}^{n} (y_i - h(x_i, \beta)) \cdot h'_\beta(x_i, \beta) \cdot I(r_i^2(\beta) \leq r_{[h]}^2(\beta)) \quad \text{a.s.}, \tag{11}$$

as the NLTS estimator is a solution of $\frac{\partial \rho(\beta)}{\partial \beta} = 0$.

Lemma 1 *Let $n \in \mathbb{N}$ and $k_h(\beta) : \mathbb{R}^p \to \{1, \ldots, n\}$ be a function that represents an index of an observation such that $r_{k_h(\beta)}^2(\beta) = r_{[h]}^2(\beta)$, $h \in \{1, \ldots, n\}$. Under conditions D and H, there exists a set Ω_1, $P(\Omega_1) = 1$, such that for every $\omega \in \Omega_1$ there is some neighborhood $U(\beta^0, \varepsilon(\omega))$ of β^0 such that the function $k_h(\beta)$ is constant on $U(\beta^0, \varepsilon(\omega))$ for all $h \in \{1, \ldots, n\}$.*

Proof: See Appendix A.1. □

3.3.2 Asymptotic linearity of the derivative of objective function $\rho(\beta)$ at $\beta = \beta^0$

Analogously as for M-estimators or LTS (see Jurečková and Sen (1989), Víšek (1996a) or Víšek (1999a)), we shall investigate the term $S_n(t) = \frac{\partial \rho(\beta^0 - n^{-\frac{1}{2}} t)}{\partial \beta} - \frac{\partial \rho(\beta^0)}{\partial \beta}$, i.e.,

$$\begin{aligned} S_n(t) = \sum_{i=1}^{n} \Bigg[& \Big(y_i - h\Big(x_i, \beta^0 - n^{-\frac{1}{2}} t\Big)\Big) \cdot h'_\beta\Big(x_i, \beta^0 - n^{-\frac{1}{2}} t\Big) \cdot I\Big(r_i^2\left(\beta^0 - n^{-\frac{1}{2}} t\right) \leq r_{[h_n]}^2\left(\beta^0 - n^{-\frac{1}{2}} t\right)\Big) \\ & - \big(y_i - h\big(x_i, \beta^0\big)\big) \cdot h'_\beta\big(x_i, \beta^0\big) \cdot I\Big(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\Big) \Bigg] \end{aligned}$$

for $t \in \mathcal{T}_M = \{t \in \mathbb{R}^p : \|t\| \leq M\}$; $0 < M < \infty$ is an arbitrary, but fixed constant throughout this section. More precisely, we show that $S_n(t)$ behaves asymptotically as a linear function of $n^{\frac{1}{2}}t$ over the whole set $\mathcal{T}_M$.

Theorem 1 *Let assumptions D and H hold. Then for a given $\lambda \in (\frac{1}{2}, 1\rangle$, it holds that*

$$n^{-\frac{1}{4}} \sup_{t \in \mathcal{T}_M} \left\| S_n(t) + n^{\frac{1}{2}} \cdot Q_h t \cdot \left[\lambda - \sqrt{G^{-1}(\lambda)} \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\}\right] \right\| = \mathcal{O}_p(1)$$

as $n \to +\infty$.

Before we actually prove this theorem in Section 3.3.4, it is beneficial to study behavior of indicators $I\left(r_i^2(\beta) \leq r_{[h]}^2(\beta)\right)$ since they determine which observations enter the objective function of NLTS, and thus, determine the main characteristics of $S_n(t)$. In addition, this allows us to make the proof itself more transparent.

3.3.3 Auxiliary propositions and lemmas

In this section, several useful propositions intended to make the proof of Theorem 1 more transparent are shown. First, several lemmas that characterize behavior of order statistics of squared residuals and their distribution functions are introduced.

Lemma 2 $P\left(\left\{\omega = (\omega_1, \ldots, \omega_n) \in \Omega^n : r_i^2(\beta, \omega_i) = r_{[h]}^2(\beta, \omega)\right\}\right) = \frac{1}{n}$ *for any $n \in \mathbb{N}$, $i, h \in \{1, \ldots, n\}$, and $\beta \in B$.*

Proof: See Appendix A.1. □

Lemma 3 *Let $1/2 < \lambda < 1$ and $0 < c < G_\beta^{-1}(\lambda)$ be a real constant, where G_β represents the distribution function of $r_i^2(\beta)$, $\beta \in B$. Then, under Assumptions D, $P\left(r_{[h_n]}^2(\beta) \leq c\right) = \mathcal{O}\left(n^{-k}\right)$ for any $k \in \mathbb{N}$ as $n \to +\infty$.*

Proof: See Appendix A.1. □

Corollary 1 *Analogously, it is possible under Assumptions D to show that for real constants $1/2 < \lambda < 1$ and $G_\beta^{-1}(\lambda) < c < \infty$ it holds that $P\left(r_{[h_n]}^2(\beta) \geq c\right) = \mathcal{O}\left(n^{-k}\right)$ for any $k \in \mathbb{N}$ as $n \to +\infty$.*

Lemma 3 actually says that the probability that the order statistics does not lie in a given neighborhood of the corresponding quantile of the underlying distribution function converges to zero faster than any polynomial. This assertion can be generalized to the set of all $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. This neighborhood actually corresponds to the set of all $\beta = \beta^0 - n^{-\frac{1}{2}}t$, where $t \in \mathcal{T}_M$— $U(\beta^0, n^{-\frac{1}{2}}M) = \left\{\beta = \beta^0 - n^{-\frac{1}{2}}t : t \in M\right\}$.

Corollary 2 *Let $1/2 < \lambda < 1$ and $0 < c < G^{-1}(\lambda) < c' < \infty$ be real constants. Under Assumptions D, it holds that*

$$P\left(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : r_{[h_n]}^2(\beta) \notin (c, c')\right) = \mathcal{O}\left(n^{-k}\right)$$

for any $k \in \mathbb{N}$ as $n \to +\infty$.

Proof: See Appendix A.1. □

To study asymptotic properties of NLTS, we need to know how small the probability that $I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right)$ and $I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)$ differ is, that is, the probability that replacing the order statistics $r_{[h_n]}^2(\beta^0)$ by the corresponding quantile $G^{-1}(\lambda)$ in the indicator $I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right)$ changes its value. A generalized statement follows.

Lemma 4 *Under Assumptions D and H,*

$$P\Big(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big) \neq I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\Big) = \mathcal{O}\Big(n^{-\frac{1}{2}}\Big)$$

as $n \to +\infty$.

Proof: See appendix A.1. □

Corollary 4 *Under Assumptions D and H,*

$$P\Big(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big) \neq I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\Big| x_i\Big) = \mathcal{O}\Big(n^{-\frac{1}{2}}\Big)$$

as $n \to +\infty$.

Proof: See Appendix A.1. □

In order to study $S_n(t)$, we also need to know how small the probability of $I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)$ being different at some $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ and at β^0 is and the expected value of this difference, respectively. We find out that the mentioned probability is proportional to the distance of β and β^0 multiplied by a function of x_i. The following lemma analyzes this probability conditionally on x_i and also unconditionally; moreover, it does so for a certain $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ and all possible $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ as well.

Lemma 5 *Let Assumptions D and H hold and* $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. *Then it holds*

1. *For the conditional probability*

$$\begin{aligned} & P\Big(I\Big(r_i^2(\beta^0) \le r^2_{[h_n]}(\beta^0)\Big) \neq I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)\Big| x_i\Big) \\ = \;& \Big|h'_\beta(x_i, \beta^0)^T(\beta - \beta^0)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) \\ = \;& \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big) \end{aligned}$$

and

$$\begin{aligned} & \mathsf{E}\left(\operatorname{sgn} r_i(\beta^0) \cdot \left(I\Big(r_i^2(\beta^0) \le r^2_{[h_n]}(\beta^0)\Big) - I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)\right)\Big| x_i\right) = \\ & = h'_\beta(x_i, \beta^0)^T(\beta - \beta^0) \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) \end{aligned}$$

as $n \to +\infty$.

2. *For the corresponding unconditional probability*

$$\begin{aligned} & P\Big(I\Big(r_i^2(\beta^0) \le r^2_{[h_n]}(\beta^0)\Big) \neq I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)\Big) \\ = \;& \mathsf{E}_x\Big|h'_\beta(x_i, \beta^0)^T(\beta - \beta^0)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) \\ = \;& \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) \end{aligned}$$

as $n \to +\infty$.

3. *For the conditional probability of supremum over* β

$$\begin{aligned} & P\Big(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\Big(r_i^2(\beta^0) \le r^2_{[h_n]}(\beta^0)\Big) \neq I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)\Big| x_i\Big) \\ = \;& n^{-\frac{1}{2}}M \cdot \sum_{j=1}^{p}\Big|h'_{\beta_j}(x_i, \beta^0)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) \\ = \;& \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big) \end{aligned}$$

as $n \to \infty$.

4. For the corresponding unconditional probability of supremum over β

$$\begin{aligned} & P\Big(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\Big(r_i^2\big(\beta^0\big) \le r^2_{[h_n]}(\beta^0)\Big) \neq I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)\Big) \\ = \; & n^{-\frac{1}{2}}M \cdot \sum_{j=1}^{p} \mathsf{E}_x \left|h'_{\beta_j}\left(x_i, \beta^0\right)\right| \cdot \left\{f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right)\right\} + \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) \\ = \; & \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) \end{aligned}$$

as $n \to \infty$.

Proof: See Appendix A.1. □

Finally, there is a simple, but useful corollary of the previous lemma, which describes behavior of $r_i(\beta^0)$ in all cases when $I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big)$ differs at some $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ and at β^0.

Corollary 6 *Assume that $\lambda \in (\frac{1}{2}, 1\rangle$ and there exists $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ such that Corollary*

$$I\Big(r_i^2\big(\beta^0\big) \le r^2_{[h_n]}\big(\beta^0\big)\Big) \neq I\Big(r_i^2(\beta) \le r^2_{[h_n]}(\beta)\Big).$$

Then for any $\varepsilon > 0$ we can find a constant $K > 0$ such that

$$\big|\big|r_i(\beta^0)\big| - G^{-1}(\lambda)\big| = \big|r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot G^{-1}(\lambda)\big| < n^{-\frac{1}{4}}K$$

for all n high enough with probability greater than $1-\varepsilon$, that is $r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot G^{-1}(\lambda) = \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big)$ as $n \to +\infty$.

Proof: See Appendix A.1. □

3.3.4 Proof of asymptotic linearity

The main focus moves now to the term $S_n(t) = \frac{\partial\rho(\beta^0 - n^{-\frac{1}{2}}t)}{\partial\beta} - \frac{\partial\rho(\beta^0)}{\partial\beta}$, i.e.,

$$\begin{aligned} S_n(t) = \sum_{i=1}^{n} \Big[& \Big(y_i - h\Big(x_i, \beta^0 - n^{-\frac{1}{2}}t\Big)\Big) \cdot h'_\beta\Big(x_i, \beta^0 - n^{-\frac{1}{2}}t\Big) \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r^2_{[h]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) \\ & - \big(y_i - h\big(x_i, \beta^0\big)\big) \cdot h'_\beta\big(x_i, \beta^0\big) \cdot I\Big(r_i^2\big(\beta^0\big) \le r^2_{[h]}\big(\beta^0\big)\Big)\Big] \end{aligned}$$

for $t \in \mathcal{T}_M = \{t \in \mathbb{R}^p : \|t\| \le M\}$. There is apparently a $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$ and $t \in \mathcal{T}_M$ it holds that $\beta^0 - n^{-\frac{1}{2}}t \in U(\beta^0, \delta)$. Therefore, for all $n \ge n_0$ and $t \in \mathcal{T}_M$

$$h\Big(x, \beta^0 - n^{-\frac{1}{2}}t\Big) = h\big(x, \beta^0\big) - h'_\beta(x, \xi)^T n^{-\frac{1}{2}}t$$

and

$$h'_\beta\Big(x, \beta^0 - n^{-\frac{1}{2}}t\Big) = h'_\beta\big(x, \beta^0\big) - h''_{\beta\beta}(x, \xi')n^{-\frac{1}{2}}t,$$

where $\xi, \xi' \in \Big[\beta^0, \beta^0 - n^{-\frac{1}{2}}t\Big]_{\varkappa}$. Consequently, we may write S_n in the following form:

$$\begin{aligned} S_n(t) = & \sum_{i=1}^{n} \Big[\Big\{ \big(y_i - h\big(x_i, \beta^0\big)\big) \cdot h'_\beta\big(x_i, \beta^0\big) \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r^2_{[h]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) \\ & - \big(y_i - h\big(x_i, \beta^0\big)\big) \cdot h'_\beta\big(x_i, \beta^0\big) \cdot I\Big(r_i^2\big(\beta^0\big) \le r^2_{[h]}\big(\beta^0\big)\Big)\Big\} && (12) \\ & - \big(y_i - h\big(x_i, \beta^0\big)\big) \cdot h''_{\beta\beta}(x_i, \xi')n^{-\frac{1}{2}}t \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r^2_{[h]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) && (13) \\ & - h'_\beta(x_i, \xi)^T n^{-\frac{1}{2}}t \cdot h'_\beta\big(x_i, \beta^0\big) \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r^2_{[h]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) && (14) \\ & + h'_\beta(x_i, \xi)^T n^{-\frac{1}{2}}t \cdot h''_{\beta\beta}(x_i, \xi')n^{-\frac{1}{2}}t \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r^2_{[h]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big)\Big] && (15) \end{aligned}$$

$$
\begin{aligned}
= \quad & \sum_{i=1}^{n} \Big[\Big\{ \big(y_i - h(x_i,\beta^0)\big) \cdot h_\beta^{'}(x_i,\beta^0) \times \\
& \qquad \times \Big[I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big] \Big\} && (16) \\
& - \big(y_i - h(x_i,\beta^0)\big) \cdot h_{\beta\beta}^{''}(x_i,\xi') n^{-\frac{1}{2}}t \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) && (17) \\
& - \big(y_i - h(x_i,\beta^0)\big) \cdot h_{\beta\beta}^{''}(x_i,\xi') n^{-\frac{1}{2}}t \times && (18) \\
& \qquad \times \Big[I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big] \\
& - h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_\beta^{'}(x_i,\beta^0) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) && (19) \\
& - h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_\beta^{'}(x_i,\beta^0) \times \\
& \qquad \times \Big[I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big] && (20) \\
& + h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_{\beta\beta}^{''}(x_i,\xi') n^{-\frac{1}{2}}t \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big)\Big] && (21)
\end{aligned}
$$

Let us now analyze parts of the previous expression one by one. We will show that parts (17), (18), (20), and (21) behave like $\mathcal{O}_p\Big(n^{\frac{1}{4}}\Big)$, and therefore, are asymptotically negligible with respect to parts (16) and (19), which behave like $\mathcal{O}_p\Big(n^{\frac{1}{2}}\Big)$. Moreover, we find asymptotic representations of (16) and (19).

First of all, the last part (21) can be bounded from above in the following way (see assumptions H1 and H3):

$$
\begin{aligned}
\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_{\beta\beta}^{''}(x_i,\xi') n^{-\frac{1}{2}}t \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big)\Big\| & \\
\le \mathcal{O}_p\Big(n^{-\frac{3}{4}}\Big) \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_{\beta\beta_k}^{''}(x_i,\xi')\Big\| & \\
\le \mathcal{O}_p\Big(n^{-\frac{3}{4}}\Big) \left(\sum_{i=1}^{n} \Big\| h_{\beta\beta_k}^{''}(x_i,\beta^0)\Big\| + \mathcal{O}_p(1) \right), &
\end{aligned}
$$

where the last result follows from assumption H2 (the Lipschitz property for $h_{\beta\beta}^{''}(x_i,\beta)$). Once we realize that assumption H4 and the law of large numbers guarantees $\sum_{i=1}^{n} \Big\| h_{\beta\beta}^{''}(x_i,\beta^0)\Big\| = \mathcal{O}(n)$ as $n \to +\infty$, we get immediately

$$
\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_{\beta\beta}^{''}(x_i,\xi) n^{-\frac{1}{2}}t \cdot I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big)\Big\| = \mathcal{O}_p\Big(n^{\frac{1}{4}}\Big)
$$

as $n \to +\infty$.

Next, we are going to analyze part (20), i.e.,

$$
\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}}t \cdot h_\beta^{'}(x_i,\beta^0) \cdot \nu_i(n,t)\Big\|,
$$

where $\nu_i(n,t) = I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)$. As

$$
\begin{aligned}
& \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \cdot \nu_i(n,t) \Big\| \\
= & \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| \Big(h_\beta^{'}\big(x_i,\beta^0\big)^T n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) + n^{-\frac{1}{2}} t^T \cdot h_{\beta\beta}^{''}(x_i,\xi) \cdot n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \Big) \cdot \nu_i(n,t) \Big\| \\
\le & \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}\big(x_i,\beta^0\big)^T n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \Big\| \cdot |\nu_i(n,t)| \qquad (22) \\
+ & \mathcal{O}_p(1) \cdot \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \Big\| \cdot |\nu_i(n,t)| \qquad (23)
\end{aligned}
$$

(see condition H3), we need to analyze these two summands. This can be done in the same way for both of them, so we will do it here just for (22). To do this, we employ the Chebyshev inequality for non-negative random variables: for any non-negative random variable X it holds that $P(X > K) < \frac{\mathsf{E}\, X}{K}$. Therefore,

$$
\begin{aligned}
& P\left(\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}\big(x_i,\beta^0\big)^T n^{-\frac{1}{2}} t \cdot h_{\beta_k}^{'}\big(x_i,\beta^0\big) \cdot \nu_i(n,t) \Big\| > K n^{\frac{1}{4}} \right) \\
\le & \frac{1}{K n^{\frac{1}{4}}} \mathsf{E} \left(\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}\big(x_i,\beta^0\big)^T n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \cdot \nu_i(n,t) \Big\| \right) \\
\le & \frac{n^{-\frac{3}{4}}}{K} \sum_{i=1}^{n} \mathsf{E} \left\{ \sup_{t\in\mathcal{T}_M} \Big\| h_\beta^{'}\big(x_i,\beta^0\big)^T t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \Big\| \cdot \sup_{t\in\mathcal{T}_M} |\nu_i(n,t)| \right\}
\end{aligned}
$$

and by the Schwartz inequality and Lemma 5

$$
\begin{aligned}
& \frac{n^{-\frac{3}{4}}}{K} \sum_{i=1}^{n} \mathsf{E} \left\{ \sup_{t\in\mathcal{T}_M} \Big\| h_\beta^{'}\big(x_i,\beta^0\big)^T t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \Big\| \cdot \sup_{t\in\mathcal{T}_M} |\nu_i(n,t)| \right\} \\
\le & \frac{n^{-\frac{3}{4}}}{K} \sum_{i=1}^{n} \sqrt{ \mathsf{E} \left(\sup_{t\in\mathcal{T}_M} \Big\| h_\beta^{'}(x_i,\beta^0)^T t \Big\| \cdot \Big\| h_\beta^{'}(x_i,\beta^0) \Big\| \right)^2 \cdot \mathsf{E} \sup_{t\in\mathcal{T}_M} |\nu_i(n,t)| } \\
\le & \frac{n^{\frac{1}{4}}}{K} \sqrt{ \mathsf{E} \left(M \cdot \sum_{j=1}^{p} \Big\| h_{\beta_j}^{'}(x_i,\beta^0) \Big\| \cdot \Big\| h_\beta^{'}(x_i,\beta^0) \Big\| \right)^2 \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) } \\
\le & \frac{\mathcal{O}(1)}{K} = \frac{\text{const}}{K}. \qquad (24)
\end{aligned}
$$

Apparently, for any $\varepsilon > 0$ there is a $K > 0$ such that the constant term (24), which is proportional to $\frac{1}{K}$, is smaller than ε. Thus, we have shown that

$$
\begin{aligned}
& \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| h_\beta^{'}(x_i,\xi)^T n^{-\frac{1}{2}} t \cdot h_\beta^{'}\big(x_i,\beta^0\big) \times \\
& \quad \times \Big(I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big) \Big\| = \mathcal{O}_p\Big(n^{\frac{1}{4}}\Big)
\end{aligned}
$$

as $n \to +\infty$. Please, note that (18) can be estimated in the same way, so we have also shown how to prove

$$
\begin{aligned}
& \sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \Big\| \big(y_i - h\big(x_i,\beta^0\big)\big) \cdot h_{\beta\beta}^{''}(x_i,\xi') n^{-\frac{1}{2}} t \times \\
& \quad \times \Big(I\Big(r_i^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) \le r_{[h]}^2\Big(\beta^0 - n^{-\frac{1}{2}}t\Big)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big) \Big\| = \mathcal{O}_p\Big(n^{\frac{1}{4}}\Big)
\end{aligned}
$$

as $n \to +\infty$.

The next summand to be analyzed is (17):

$$\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \big(y_i - h(x_i,\beta^0)\big) \cdot h''_{\beta\beta}(x_i,\xi') n^{-\frac{1}{2}} t \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big).$$

This can be rewritten as (assumption H2)

$$\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \big(y_i - h(x_i,\beta^0)\big) \cdot h''_{\beta\beta}(x_i,\beta^0) n^{-\frac{1}{2}} t \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) \tag{25}$$

$$+\mathcal{O}_p(n^{-1}) \cdot \sum_{i=1}^{n} \big(y_i - h(x_i,\beta^0)\big) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big). \tag{26}$$

Assumption D2 implies that the expectation of (25) conditional on x_i is equal to zero, thus the unconditional expectation is zero as well. Moreover, the variance of (25) equals (Assumptions D2 and H4 are used)

$$\begin{aligned}
& \mathsf{var}\Big[\big(y_i - h(x_i,\beta^0)\big) \cdot h''_{\beta_j\beta_k}(x_i,\beta^0) t_l \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big] \\
= \; & \mathsf{var}_x\Big\{h''_{\beta_j\beta_k}(x_i,\beta^0) t_l \cdot \mathsf{E}\Big[\big(y_i - h(x_i,\beta^0)\big) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big| x_i\Big]\Big\} \\
+ \; & \mathsf{E}_x\Big\{\Big(h''_{\beta_j\beta_k}(x_i,\beta^0) t_l\Big)^2 \cdot \mathsf{var}\Big[\big(y_i - h(x_i,\beta^0)\big) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)\Big| x_i\Big]\Big\} \\
= \; & \mathsf{var}_x\Big\{h''_{\beta_j\beta_k}(x_i,\beta^0) t_l \cdot 0\Big\} + \mathsf{E}_x\Big\{\Big(h''_{\beta_j\beta_k}(x_i,\beta^0) t_l\Big)^2 \cdot \sigma^2\Big\} \\
= \; & \sigma^2 t_l^2 \cdot \mathsf{E}\Big(h''_{\beta_j\beta_k}(x_i,\beta^0)\Big)^2,
\end{aligned}$$

so it exists and is finite, and we can employ the central limit theorem for the sum (25). Hence,

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \big(y_i - h(x_i,\beta^0)\big) \cdot h''_{\beta_j\beta_k}(x_i,\beta^0) t_l \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big)$$

converges in distribution to a normally distributed random variable with variance bounded in $t \in \mathcal{T}_M$ for all $j,k,l = 1,\ldots,p$. Because (26) is apparently bounded in probability, it holds that

$$\sup_{t\in\mathcal{T}_M} \sum_{i=1}^{n} \big(y_i - h(x_i,\beta^0)\big) \cdot h''_{\beta\beta}(x_i,\xi') n^{-\frac{1}{2}} t \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) = \mathcal{O}_p(1)$$

as $n \to \infty$.

The last but one term to be estimated is (19), i.e.,

$$\sum_{i=1}^{n} h'_\beta(x_i,\xi)^T n^{-\frac{1}{2}} t \cdot h'_\beta(x_i,\beta^0) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) \tag{27}$$

$$= \sum_{i=1}^{n} h'_\beta(x_i,\beta^0)^T n^{-\frac{1}{2}} t \cdot h'_\beta(x_i,\beta^0) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big) \tag{28}$$

$$+ \sum_{i=1}^{n} n^{-\frac{1}{2}} t^T \cdot h''_{\beta\beta}(x_i,\xi'') \cdot n^{-\frac{1}{2}} t \cdot h'_{\beta_k}(x_i,\beta^0) \cdot I\Big(r_i^2(\beta^0) \le r_{[h]}^2(\beta^0)\Big). \tag{29}$$

The supremum of the second part (29) over $t \in \mathcal{T}_M$ behaves like $\mathcal{O}_p(1)$ by the same kind of argument as the one used before: since by means of assumption H2 (Lipschitz condition for the second derivative) and H3

$$
\begin{aligned}
& \mathsf{E}\left(\sup_{t\in\mathcal{T}_M}\sum_{i=1}^{n}\left\| n^{-\frac{1}{2}}t^T\cdot h''_{\beta\beta}(x_i,\xi'')\cdot n^{-\frac{1}{2}}t\cdot h'_{\beta_k}(x_i,\beta^0)\cdot I\Big(r_i^2(\beta^0)\le r^2_{[h]}(\beta^0)\Big)\right\|\right)\\
\le\ & n^{-1}M^2\cdot\sum_{i=1}^{n}\mathsf{E}\left\| h''_{\beta\beta}(x_i,\xi'')\cdot h'_{\beta}(x_i,\beta^0)\right\|\\
\le\ & n^{-1}M^2\cdot\sum_{i=1}^{n}\mathsf{E}\left(\left\| h''_{\beta\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)\right\|+\left\| h'_{\beta}(x_i,\beta^0)\cdot\mathcal{O}_p(1)\right\|\right)\\
=\ & \mathcal{O}(1),
\end{aligned}
$$

the Chebyshev inequality for nonnegative random variables implies boundedness of (29) in probability.

Let us look now at (28):

$$
\begin{aligned}
& \sum_{i=1}^{n} h'_{\beta}(x_i,\beta^0)^T n^{-\frac{1}{2}}t\cdot h'_{\beta}(x_i,\beta^0)\cdot I\Big(r_i^2(\beta^0)\le r^2_{[h]}(\beta^0)\Big) && (30)\\
=\ & \sum_{i=1}^{n} h'_{\beta}(x_i,\beta^0)^T n^{-\frac{1}{2}}t\cdot h'_{\beta}(x_i,\beta^0)\cdot\Big(I\Big(r_i^2(\beta^0)\le r^2_{[h]}(\beta^0)\Big)-I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big)\Big) && (31)\\
+\ & n^{-\frac{1}{2}}\sum_{i=1}^{n}\Big\{h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\cdot I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big) && (32)\\
& \qquad -\mathsf{E}\left(h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\cdot I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big)\right)\Big\}t\\
+\ & n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathsf{E}\left(h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\cdot I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big)\right)t. && (33)
\end{aligned}
$$

The supremum of the first part, i.e., sum (31), over $t\in\mathcal{T}_M$ behaves again like $\mathcal{O}_p(1)$ for $n\to\infty$. This can be proved in the same manner as we did in the previous paragraph, this time utilizing Lemma 4. Next, using the central limit theorem, each element of (32) converges in distribution to a normally distributed random variable with zero mean and a finite variance (both the expectation and variance of the term exist and are finite due to assumption H3 and $t<M$). Hence, it is bounded in probability as well. Finally, the last element (33) can be rewritten as $n^{\frac{1}{2}}\cdot\lambda\cdot Q_h t$ since

$$
\begin{aligned}
& \mathsf{E}\left(h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\cdot I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big)\right)\\
=\ & \mathsf{E}_x\left(h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\cdot\mathsf{E}\left(I\big(r_i^2(\beta^0)\le G^{-1}(\lambda)\big)\,\middle|\,x_i\right)\right)\\
=\ & \lambda\cdot\mathsf{E}_x\left(h'_{\beta}(x_i,\beta^0)\cdot h'_{\beta}(x_i,\beta^0)^T\right)=\lambda\cdot Q_h.
\end{aligned}
$$

We can conclude that

$$
\sup_{t\in\mathcal{T}_M}\left\|\sum_{i=1}^{n} h'_{\beta}(x_i,\beta^0)n^{-\frac{1}{2}}t\cdot h'_{\beta}(x_i,\beta^0)\cdot I\Big(r_i^2(\beta^0)\le r^2_{[h]}(\beta^0)\Big)-n^{\frac{1}{2}}\cdot\lambda\cdot Q_h t\right\|=\mathcal{O}_p(1)
$$

as $n\to+\infty$.

Finally, let us move our attention to the term (16). Using once again notation

$$
\nu_i(n,t)=\left(I\Big(r_i^2\Big(\beta^0-n^{-\frac{1}{2}}t\Big)\le r^2_{[h]}\Big(\beta^0-n^{-\frac{1}{2}}t\Big)\Big)-I\Big(r_i^2(\beta^0)\le r^2_{[h]}(\beta^0)\Big)\right),
$$

we can rewrite (16) as

$$\sum_{i=1}^{n} \left(y_i - h\left(x_i, \beta^0\right)\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t) \tag{34}$$

$$= \sum_{i=1}^{n} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t) \tag{35}$$

$$= \sum_{i=1}^{n} \left(r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot \sqrt{G^{-1}(\lambda)}\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t) \tag{36}$$

$$+ \sum_{i=1}^{n} \operatorname{sgn} r_i(\beta^0) \cdot \sqrt{G^{-1}(\lambda)} \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t). \tag{37}$$

For the simplicity of notation, let $q_\lambda = \sqrt{G^{-1}(\lambda)}$. The first part (36) multiplied by $n^{-\frac{1}{4}}$ is bounded in probability. This can be shown as follows: Corollary 6 implies

$$P\left(n^{-\frac{1}{4}} \sup_{t \in \mathcal{T}_M} \left\| \sum_{i=1}^{n} \left(r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot q_\lambda\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t) \right\| > K \right) \tag{38}$$

$$\leq \frac{1}{K} \mathsf{E}\left(n^{-\frac{1}{4}} \sup_{t \in \mathcal{T}_M} \left\| \sum_{i=1}^{n} \left(r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot q_\lambda\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t) \right\| \right)$$

$$\leq \frac{n^{\frac{3}{4}}}{K} \mathsf{E}\left(\left| r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot q_\lambda \right| \cdot \left\| h_\beta^{'}\left(x_i, \beta^0\right) \right\| \cdot \sup_{t \in \mathcal{T}_M} |\nu_i(n,t)| \right)$$

$$\leq \frac{n^{\frac{3}{4}}}{K} \mathsf{E}\left(\mathcal{O}_p\left(n^{-\frac{1}{4}}\right) \cdot \left\| h_\beta^{'}\left(x_i, \beta^0\right) \right\| \cdot \sup_{t \in \mathcal{T}_M} |\nu_i(n,t)| \right)$$

and by Lemma 5 ($r_i(\beta^0) \equiv \varepsilon_i$ and x_i are independent random variables)

$$\frac{n^{\frac{3}{4}}}{K} \mathsf{E}_x\left(\left\| h_\beta^{'}\left(x_i, \beta^0\right) \right\| \cdot \mathsf{E}\left[\mathcal{O}_p\left(n^{-\frac{1}{4}}\right) \cdot \sup_{t \in \mathcal{T}_M} |\nu_i(n,t)| \,\middle|\, x_i \right] \right)$$

$$\leq \frac{n^{\frac{1}{2}}}{K} \mathsf{E}\left(\mathcal{O}\left(n^{-\frac{1}{2}}\right) \cdot \left\| h_\beta^{'}\left(x_i, \beta^0\right) \right\| \cdot \left[\sum_{j=1}^{p} \left| h_{\beta_j}^{'}\left(x_i, \beta^0\right) \right| + \mathcal{O}_p(1) \right] \right)$$

$$\leq \frac{\mathcal{O}(1)}{K} = \frac{\text{const.}}{K}.$$

Therefore, the probability (38) can be made smaller than ε by an appropriate choice of K, and hence, (36) multiplied by $n^{-\frac{1}{4}}$ is bounded in probability. In other words, it holds for (36)

$$\sup_{t \in \mathcal{T}_M} \left\| \sum_{i=1}^{n} \left(r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot q_\lambda\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \times \right.$$
$$\left. \times \left[I\left(r_i^2\left(\beta^0 - n^{-\frac{1}{2}}t\right) \leq r_{[h]}^2\left(\beta^0 - n^{-\frac{1}{2}}t\right)\right) - I\left(r_i^2(\beta^0) \leq r_{[h]}^2(\beta^0)\right) \right] \right\| = \mathcal{O}_p\left(n^{\frac{1}{4}}\right)$$

as $n \to \infty$. All we have to do now is to treat

$$\sum_{i=1}^{n} \operatorname{sgn} r_i(\beta^0) \cdot \sqrt{G^{-1}(\lambda)} \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \nu_i(n,t). \tag{39}$$

This is done again in two steps—first, we show that the sum less its expectation is $\mathcal{O}_p(1)$, and second, the expectation of the sum is evaluated. For the first part, we have shown in Lemma 5 that the probability of $\nu_i(n,t) = \left[I\left(r_i^2\left(\beta^0 - n^{-\frac{1}{2}}t\right) \leq r_{[h]}^2\left(\beta^0 - n^{-\frac{1}{2}}t\right)\right) - I\left(r_i^2(\beta^0) \leq r_{[h]}^2(\beta^0)\right)\right]$

being non-zero conditional on x_i (and thus the conditional expectation of this term in absolute value) is equal to $\left|h'_\beta(x_i,\beta^0)^T(\beta-\beta^0)\right|\cdot\left\{f\left(-\sqrt{G^{-1}(\lambda)}\right)+f\left(\sqrt{G^{-1}(\lambda)}\right)\right\}+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ as $n\to+\infty$, and that the expectation of this conditional probability behaves like $\mathcal{O}\left(n^{-\frac{1}{4}}\right)$. Therefore, the random variable $\nu_i(n,t)$ multiplied by $n^{\frac{1}{2}}$ will have its expectation conditional on x_i behaving like $\left\|h'_\beta(x_i,\beta^0)\right\|\cdot\mathcal{O}(1)+\mathcal{O}_p(1)$ in absolute value. Consequently,

$$\begin{aligned}
&\mathsf{E}\left\{n^{1/4}\cdot\operatorname{sgn}r_i(\beta^0)\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)\right\}\\
=\;&\mathsf{E}\left\{n^{-1/4}\cdot\operatorname{sgn}r_i(\beta^0)\cdot h'_\beta(x_i,\beta^0)\cdot n^{\frac{1}{2}}\,\mathsf{E}\left[\nu_i(n,t)|x_i\right]\right\}\\
\le\;&\mathsf{E}\left\{n^{-1/4}\cdot\left\|h'_\beta(x_i,\beta^0)\right\|\cdot\left[\left\|h'_\beta(x_i,\beta^0)\right\|\cdot\mathcal{O}(1)+\mathcal{O}_p(1)\right]\right\}=o(1)
\end{aligned}$$

and, since $\mathsf{var}\,\nu_i(n,t)\le\mathsf{E}\,|\nu_i(n,t)|$,

$$\begin{aligned}
&\mathsf{var}\left\{n^{1/4}\cdot\operatorname{sgn}r_i(\beta^0)\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)\right\}\\
=\;&\mathsf{var}\left\{n^{1/4}\cdot\operatorname{sgn}r_i(\beta^0)\cdot h'_\beta(x_i,\beta^0)\cdot\mathsf{E}\left[\nu_i(n,t)|x_i\right]\right\}\\
+\;&\mathsf{E}\left\{n^{1/2}\cdot h'_\beta(x_i,\beta^0)\cdot\mathsf{var}\left[\nu_i(n,t)|x_i\right]\cdot h'_\beta(x_i,\beta^0)^T\right\}\\
\le\;&\mathsf{var}\left\{n^{-1/4}\cdot\left\|h'_\beta(x_i,\beta^0)\right\|\cdot\left[\left\|h'_\beta(x_i,\beta^0)\right\|\mathcal{O}(1)+\mathcal{O}_p(1)\right]\right\}\\
+\;&\mathsf{E}\left\{\left\|h'_\beta(x_i,\beta^0)\right\|^2\cdot\left[\left\|h'_\beta(x_i,\beta^0)\right\|\mathcal{O}(1)+\mathcal{O}_p(1)\right]\right\}\\
=\;&\mathcal{O}(1).
\end{aligned}$$

Hence, the central limit theorem can be applied to the random variable

$$\mu_i=n^{1/4}\cdot\operatorname{sgn}r_i(\beta^0)\cdot q_\lambda\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)$$

and we obtain that $n^{-\frac{1}{2}}\sum_{i=1}^n n^{1/4}(\mu_i-\mathsf{E}\,\mu_i)$ is asymptotically normally distributed around zero with a finite variance. As a direct consequence, it follows that

$$\begin{aligned}
&\sum_{i=1}^n\operatorname{sgn}r_i(\beta^0)\cdot q_\lambda\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)-\\
-\sum_{i=1}^n\mathsf{E}\left\{\operatorname{sgn}r_i(\beta^0)\cdot q_\lambda\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)\right\}\;&=\;\mathcal{O}_p\left(n^{1/4}\right)
\end{aligned}$$

as $n\to+\infty$.

Finally, the expectation of (39)

$$\mathsf{E}\sum_{i=1}^n\operatorname{sgn}r_i(\beta^0)\cdot\sqrt{G^{-1}(\lambda)}\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)=\tag{40}$$

$$=\mathsf{E}_x\left\{\sum_{i=1}^n q_\lambda\cdot h'_\beta(x_i,\beta^0)\cdot\mathsf{E}\left(\operatorname{sgn}r_i(\beta^0)\cdot\nu_i(n,t)|x_i\right)\right\}\tag{41}$$

can be proved to be a linear function of t by means of Lemma 5. Since

$$\mathsf{E}\left(\operatorname{sgn}r_i(\beta^0)\cdot\nu_i(n,t)|x_i\right)=h'_\beta(x_i,\beta^0)^Tn^{-\frac{1}{2}}t\cdot\{f(-q_\lambda)+f(q_\lambda)\}+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right),$$

(41) can be rewritten as

$$\mathsf{E}_x\left\{\sum_{i=1}^{n} q_\lambda \cdot h'_\beta(x_i,\beta^0)\cdot\left[h'_\beta(x_i,\beta^0)^T n^{-\frac{1}{2}}t\cdot\{f(-q_\lambda)+f(q_\lambda)\}+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)\right]\right\}=$$

$$=q_\lambda\cdot\{f(-q_\lambda)+f(q_\lambda)\}\cdot\left\{\sum_{i=1}^{n}\mathsf{E}_x\left(h'_\beta(x_i,\beta^0)\cdot h'_\beta(x_i,\beta^0)^T\right)+\mathcal{O}\left(n^{-\frac{1}{2}}\right)\right\}\cdot n^{-\frac{1}{2}}t$$

$$=\sqrt{G^{-1}(\lambda)}\cdot\{f(-q_\lambda)+f(q_\lambda)\}\cdot Q_h\cdot n^{\frac{1}{2}}t+\mathcal{O}(1).$$

Therefore, we can conclude that

$$\sup_{t\in\mathcal{T}_M}\left\|\sum_{i=1}^{n}\left(y_i-h(x_i,\beta^0)\right)\cdot h'_\beta(x_i,\beta^0)\cdot\nu_i(n,t)-\right.$$
$$\left.-n^{\frac{1}{2}}\cdot\sqrt{G^{-1}(\lambda)}\cdot\{f(-q_\lambda)+f(q_\lambda)\}\cdot Q_h t\right\| \quad=\quad \mathcal{O}_p\left(n^{\frac{1}{4}}\right)$$

as $n\to+\infty$. This closes the proof. □

3.4 $\sqrt{n}$ consistency

In Sections 2.1 and 2.2, I provided a sort of an argument why consistency and asymptotic normality of (N)LTS and (N)LS are equivalent. In this section, I will prove the consistency of NLTS properly.

First, let me precise the previous wordy argumentation. The main idea of the following reasoning comes from Víšek (1999a). Correctly, we should write $r_i(\beta)=r_i(\beta,\omega_i)$ as the residuals are both functions of parameters $\beta\in B$ and realizations $\omega_i\in\Omega$. As we have already argued, the NLTS estimate for a given sample realization $\omega\in\Omega^n$ can be obtained by means of a simple process: taking all subsamples of size h one after another and applying for each of them the nonlinear least squares estimator, we get $\binom{n}{h}$ candidates for the estimate and the one that governs the smallest value of the sum of squares residuals in its subpopulation is the final estimate. For an arbitrary $\omega\in\Omega^n$, let us take all h-tuples from $\{1,\ldots,n\}$ and apply the NLS estimator on the corresponding subpopulations. To formalize this procedure, let us define a mapping $a:\Omega\to\{1,\ldots,n\}^h$ by expression "$a(\omega)=(i_1,\ldots,i_h)^T$ is the point for which $\min_{\beta\in B}\sum_{i\in a(\omega)}r_i^2(\beta,\omega)$ is attained." More precisely, $a(\omega)$ is an element of $\{1,\ldots,n\}^h$ such that

$$\min_{\beta\in B}\sum_{i\in a(\omega)}r_i^2(\beta,\omega)=\min_{\beta\in B}\sum_{i=1}^{h}r_{[i]}^2(\beta,\omega).$$

Furthermore, let $X(a)=X(a(\omega))$ be the h-dimensional submatrix of X containing all rows of X with indices belonging to $a(\omega)$ (this can be done for any $\omega\in\Omega$). In the same way, let us define $Y(a)=Y(a(\omega))$ as a subvector of Y and $\varepsilon(a)=\varepsilon(a(\omega))$ as a subvector of ε. Then it is obvious that

$$\hat\beta_n^{(NLS,n)}(Y(a(\omega)),X(a(\omega)))=\hat\beta_n^{(NLTS,n)}(\omega),\forall\omega\in\Omega, \tag{42}$$

where $\hat\beta_n^{(NLTS,n)}(\omega)$ represents the value of the NLTS estimate at the realization ω. This confirms that the nonlinear least trimmed squares estimate is consistent under similar conditions as the nonlinear least squares.

Second, to provide as complete picture as possible about the consistency of NLTS, I specify two sets of assumptions. The first group, Assumptions NC, is as general as possible and is sufficient just for proving consistency of NLTS; the second group, Assumptions NN, allows us to derive $\sqrt{n}$-consistency and asymptotic normality of NLTS. In the presented form, Assumptions NC correspond mostly to the assumptions required for the uniform law of large numbers in nonlinear models, which is in a very general form presented in Andrews (1987).

Assumptions NC. Let the following assumptions are satisfied for function $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)$, where $r_i(\beta) = \varepsilon_i + h\big(x_i, \beta^0\big) - h(x_i, \beta)$.[7]

NC1 The parameter space B be a compact metric space (or a compact subset of $\mathbb{R}^p$).

NC2 Let $q(x_i, \varepsilon_i; \beta)$, $q^*(x_i, \varepsilon_i; \beta, \rho) = \sup\{q(x_i, \varepsilon_i; \beta) : \beta' \in U(\beta, \rho)\}$, and $q_*(x_i, \varepsilon_i; \beta, \rho) = \inf\{q(x_i, \varepsilon_i; \beta) : \beta' \in U(\beta, \rho)\}$, where $U(\beta, \rho) = \{\beta' \in B : \|\beta' - \beta\| < \rho\}$, be measurable random variables for all $\beta \in B, i \in \mathbb{N}$, and for all $\rho > 0$ sufficiently small.

NC3 Let $\mathsf{E}\left\{\sup_{\beta \in B} |q(x_i, \varepsilon_i; \beta)|\right\}^{1+\delta} < \infty$ for some $\delta > 0$.

***Remark* 7** *Assumptions NC1–NC3 are necessary (together with the assumption concerning the differentiability of function $h(x, \beta)$ with respect to β) for the uniform law of large number. Assumption NC3 is actually a standard condition used in this context to ensure that functions $\{q^*(x_i, \varepsilon_i; \beta, \rho)\}$ and $\{q_*(x_i, \varepsilon_i; \beta, \rho)\}$ satisfy pointwise the strong law of large numbers for any $\beta \in B$ and all ρ sufficiently small; see Andrews (1987), for instance. Moreover, note that*

$$r_i(\beta) = \varepsilon_i + h\big(x_i, \beta^0\big) - h(x_i, \beta) = \varepsilon_i + h'_\beta(x_i, \xi(\beta)) \cdot (\beta - \beta^0), \tag{43}$$

where $h'_\beta(x_i, \xi)$ is bounded independently of β (under Assumption H3) and $\beta - \beta^0$ is bounded as well (B is compact space)—hence, the existence of an upper bound for $r_i^2(\beta)$ follows from Assumptions H and NC1 and Assumption NC3 just requires existence of a certain expectation of this upper bound.

NC4 For any $\varepsilon > 0$ and $U(\beta^0, \varepsilon)$ such that $B - U(\beta^0, \varepsilon)$ is compact, there exists $\alpha(\varepsilon) > 0$ such that it holds

$$\begin{aligned} \min_{\beta \in B - U(\beta^0, \varepsilon)} \mathsf{E}\, q(x_i, \varepsilon_i; \beta) - \mathsf{E}\, q(x_i, \varepsilon_i; \beta^0) &= \\ \min_{\beta \in B - U(\beta^0, \varepsilon)} \mathsf{E}\left[r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)\right] - \mathsf{E}\left[r_i^2(\beta^0) \cdot I\Big(r_i^2(\beta^0) \leq G_{\beta^0}^{-1}(\lambda)\Big)\right] &> \alpha(\varepsilon). \end{aligned}$$

***Remark* 8** *This is nothing but an analogy of the identification condition for the nonlinear least squares, see for example White (1980).*

NC5 Assume that $I_G = \inf_{\beta \in B} G_\beta^{-1}(\lambda) > 0$ and

$$m_{gg} = \inf_{\beta \in B} \inf_{z \in (-\delta, \delta)} g_\beta\left(G_\beta^{-1}(\lambda) + z\right) > 0,$$

where G_β and g_β are the cumulative distribution function and the probability density function of $r_i^2(\beta)$.

NC6 Let $g_\beta(z)$ is bounded on $(I_G, +\infty)$ uniformly in $\beta \in B$, that is, there is $M_{gg} > 0$ such that $\sup_{\beta \in B} \sup_{z \in (I_G, +\infty)} g_\beta(z) \leq M_{gg}$.

***Remark* 9** *Although Assumption NC5 and NC6 might look unfamiliar at the first sight, they just guarantee that the distribution functions of random variables $r_i^2(\beta)$ do not converge to some extreme cases for some $\beta \in B$. Namely, these conditions exclude cases when the expectation or variance of $r_i^2(\beta)$ converge to infinite values for some $\beta \in B$ or when the distribution function G_β converges to a discrete distribution function for some $\beta \in B$. This does not restrict us in commonly used regression models, because the parametric space B is compact.*

The following theorem confirms that assumptions NC are sufficient for consistency of NLTS.

[7] For the case of nonlinear least squares, $\lambda = 1$ and $G_\beta^{-1}(\lambda) = \infty$. Therefore, this case corresponds to $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta)$.

Theorem 2 *Let Assumptions D, H, and NC hold. Then the nonlinear least trimmed squares estimator defined for model (4) by*

$$\hat{\beta}_n^{(NLTS,h)} = \arg\min_{\beta\in B} \sum_{i=1}^{n} r_i^2(\beta)\cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big) = \arg\min_{\beta\in B} \frac{1}{n}\sum_{i=1}^{n} r_i^2(\beta)\cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big) \quad (44)$$

is consistent, i.e., $\hat{\beta} \to \beta^0$ in probability as $n \to +\infty$.

To prove this theorem, we need one additional lemma showing that we can use the uniform law of large numbers for sum (44) and that weak dependence among $I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big), i \in \mathbb{N}$, does not spoil the result.

Lemma 7 *Let Assumptions D and H hold and assume that $t(x,\varepsilon;\beta)$ is a real-valued function continuous in β uniformly in x and ε over any compact subset of the support of (x,ε). Moreover, assume that Assumptions NC hold for $t(x,\varepsilon;\beta)$. Furthermore, let $r_i(\beta) = \varepsilon_i + h\big(x_i,\beta^0\big) - h(x_i,\beta)$ and G_β denote the distribution function of $r_i^2(\beta)$ (for any $\beta \in B$). Finally, let $h_n/n \to \lambda \in \big(\frac{1}{2},1\big)$. Then*

$$\sup_{\beta\in B}\left|\frac{1}{n}\sum_{i=1}^{n}\Big[t(x_i,\varepsilon_i;\beta)\cdot I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big)\Big] - \mathsf{E}\Big[t(x_i,\varepsilon_i;\beta)\cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\Big]\right| \to 0$$

as $n \to +\infty$ in probability.

Proof: See Appendix A.1. □

Let us continue with the proof of Theorem 2 now.

Proof: The principle of the proof is actually the same as of the proof of the SLS consistency done by Ichimura (1993), and employs the theorem about uniform consistency in nonlinear models that is due to Andrews (1987) by means of Lemma 6. Let us denote ($r_i^2(\beta)$ are independent identically distributed random variables)

$$J(\beta) = \mathsf{E}\left\{r_1^2(\beta)\cdot I\Big(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\right\},$$
$$J_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} r_i^2(\beta)\cdot I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big).$$

By the definition of NLTS, $P\Big(J_n\left(\hat{\beta}_n^{(NLTS,h_n)}\right) < J_n\left(\beta^0\right)\Big) = 1$. For any $\delta > 0$ and an open neighborhood $U(\beta^0,\delta)$ of β^0

$$\begin{aligned}
1 = P\Big(J_n\left(\hat{\beta}_n^{(NLTS,h_n)}\right) < J_n\left(\beta^0\right)\Big) = \\
= \; & P\Big(J_n\left(\hat{\beta}_n^{(NLTS,h_n)}\right) < J_n\left(\beta^0\right) \quad \text{and} \quad \hat{\beta}_n^{(NLTS,h_n)} \in U(\beta^0,\delta)\Big) \\
+ \; & P\Big(J_n\left(\hat{\beta}_n^{(NLTS,h_n)}\right) < J_n\left(\beta^0\right) \quad \text{and} \quad \hat{\beta}_n^{(NLTS,h_n)} \in B - U(\beta^0,\delta)\Big) \\
\le \; & P\Big(\hat{\beta}_n^{(NLTS,h_n)} \in U(\beta^0,\delta)\Big) + P\Big(\inf_{\beta\in B-U(\beta^0,\delta)} J_n(\beta) < J_n\left(\beta^0\right)\Big).
\end{aligned}$$

Therefore, $P\big(\inf_{\beta\in B-U(\beta^0,\delta)} J_n(\beta) < J_n\left(\beta^0\right)\big) \to 0$ as $n \to +\infty$ implies $P\Big(\hat{\beta}_n^{(NLTS,h_n)} \in U(\beta^0,\delta)\Big) \to 1$ as $n \to +\infty$, that is, the consistency of $\hat{\beta}_n^{(NLTS,h_n)}$ (δ was an arbitrary positive number). To

verify $P\big(\inf_{\beta\in B-U(\beta^0,\delta)} J_n(\beta) < J_n\left(\beta^0\right)\big) \to 0$ note that

$$\begin{aligned}
&P\left(\inf_{\beta\in B-U(\beta^0,\delta)} J_n(\beta) < J_n\left(\beta^0\right)\right) = \\
&= P\left(\inf_{\beta\in B-U(\beta^0,\delta)} \left[J_n(\beta) - J(\beta) + J(\beta)\right] < J_n\left(\beta^0\right)\right) \\
&= P\left(\inf_{\beta\in B-U(\beta^0,\delta)} \left[J_n(\beta) - J(\beta)\right] < J_n(\beta^0) - \inf_{\beta\in B-U(\beta^0,\delta)} J(\beta)\right) \\
&\le P\left(\sup_{\beta\in B} |J_n(\beta) - J(\beta)| > \inf_{\beta\in B-U(\beta^0,\delta)} J(\beta) - J_n(\beta^0)\right).
\end{aligned}$$

Since $J_n(\beta^0) \to J(\beta^0)$ almost surely for $n \to \infty$ (see Assumption NC4 and remark 7) and the identification condition NC4 implies

$$(\forall\delta > 0)\,(\exists\alpha > 0)\left(\inf_{\beta\in B-U(\beta^0,\delta)} J(\beta) - J(\beta^0) > \alpha\right),$$

it immediately follows that

$$(\forall\delta > 0)\,(\exists\alpha > 0)\left(\lim_{n\to\infty}\left[\inf_{\beta\in B-U(\beta^0,\delta)} J(\beta) - J_n(\beta^0)\right] > \alpha\right)$$

almost surely for $n \to \infty$. Thus, it is enough to show that for all $\alpha > 0$

$$P\left(\sup_{\beta\in B} |J_n(\beta) - J(\beta)| > \alpha\right) \to 0 \text{ as } n \to +\infty.$$

This is indeed the result stated in Lemma 7 for function $t(x_i, \varepsilon_i; \beta) = r_i^2(\beta)$ ($r_i^2(\beta)$ is uniformly continuous in β on any compact subset of $\mathrm{supp}(x, \varepsilon)$ because of Assumption H1, and moreover, it satisfies Assumptions NC1–NC3). □

Next, let us recall that Assumptions NC are the sufficient conditions for the consistency of NLTS. However, if we enrich Assumptions NC to below stated Assumptions NN, we are able to prove even $\sqrt{n}$-consistence and asymptotic normality of NLTS. Also Assumptions NN correspond mostly to the assumptions required for the uniform law of large numbers in nonlinear models due to Andrews (1987), because key assumptions for asymptotic normality of NLS (and thus NLTS) are that sums $\frac{1}{n}\sum_{i=1}^n h'_{\beta_j}(x_i,\beta)h'_{\beta_k}(x_i,\beta)$, $\frac{1}{n}\sum_{i=1}^n h(x_i,\beta)h''_{\beta_j\beta_k}(x_i,\beta)$ and $\frac{1}{n}\sum_{i=1}^n \left(h''_{\beta_j\beta_k}(x_i,\beta)\right)^2$ converge to their expectations uniformly in β over $\beta \in B$ for all $j, k = 1, \ldots, p$; see, for example, Chapter 12 in Seber (1989).

Assumptions NN. Let Assumptions NC hold, and additionally, Assumptions NC1–NC3 are satisfied for functions

- $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta) \cdot I\left(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\right)$,
- $q(x_i, \varepsilon_i; \beta) = r_i(\beta) \cdot h''_{\beta_j\beta_k}(x_i, \beta) \cdot I\left(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\right)$,
- $q(x_i, \varepsilon_i; \beta) = h'_{\beta_j}(x_i, \beta) \cdot h'_{\beta_k}(x_i, \beta) \cdot I\left(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\right)$,

where $j, k = 1, \ldots, p$.

***Remark* 10** *The only assumption that seems to be more restrictive here, is the moment condition NC3 used for these functions. However, this does not limit us significantly since this assumption actually coincides with Assumptions H3 (boundedness of these functions with respect to β) and H4 (existence of certain moments of $h(x, \beta)$ and its derivatives). Hence, these are just additional regularity conditions.*

Finally, combining all the conditions stated so far, namely, D, H, and NN, we can prove the $\sqrt{n}$-consistence of NLTS.

Theorem 3 *Let Assumptions D, H, and NN hold. Then* $\hat{\beta}_n^{(NLTS,h_n)}$ *is* $\sqrt{n}$*-consistent, i.e.,*

$$\sqrt{n}\left(\hat{\beta}_n^{(NLTS,h_n)} - \beta^0\right) = \mathcal{O}_p(1)$$

as $n \to +\infty$.

To prove this theorem, we need again one additional assertion.

Proposition 2 *Let* x_1 *and* x_2 *be (mutually) independent random variables and* $c > 0$. *Then* $x_1 \cdot I(|x_1| < c)$ *and* $x_2 \cdot I(|x_2| < c)$ *are independent random variables as well.*

Proof: See Víšek (1999a), Assertion 1. □

Let us prove Theorem 3 now.

Proof: We already know that $\hat{\beta}_n^{(NLTS,h_n)}$ is consistent (Theorem 2). Now, we shall use the normal equations to derive $\sqrt{n}$-consistency.

First, take a look at the derivatives of the objective function that form the normal equations. Let us denote the objective function of NLTS by $S_n(x_i, \varepsilon_i; \beta) = \frac{1}{n}\sum_{i=1}^n r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big)$. We showed in Section 3.3.1 (see equation (11)) that

$$\frac{\partial S_n(x_i, \varepsilon_i; \beta)}{\partial \beta} = -\frac{2}{n}\sum_{i=1}^n r_i(\beta) \cdot h_\beta^{'}(x_i, \beta) \cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big)$$

almost surely, and by the same argument (Lemma 1), it follows that

$$\begin{aligned}\frac{\partial^2 S_n(x_i, \varepsilon_i; \beta)}{\partial \beta \partial \beta^T} &= \frac{2}{n}\sum_{i=1}^n h_\beta^{'}(x_i, \beta) \cdot h_\beta^{'}(x_i, \beta)^T \cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big) \\ &- \frac{2}{n}\sum_{i=1}^n r_i(\beta) \cdot h_{\beta\beta}^{''}(x_i, \beta) \cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big)\end{aligned}$$

almost surely. Using Assumptions NN and Lemma 7, we obtain for all three functions, $S_n(x_i, \varepsilon_i; \beta)$ and its first two derivatives, that

$$\begin{aligned}\sup_{\beta \in B}\left|S_n(x_i, \varepsilon_i; \beta) - \mathsf{E}\left[r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\right]\right| &\to 0, \\ \sup_{\beta \in B}\Bigg|\frac{\partial^2 S_n(x_i, \varepsilon_i; \beta)}{\partial \beta \partial \beta^T} - 2\,\mathsf{E}\left[h_\beta^{'}(x_i, \beta) \cdot h_\beta^{'}(x_i, \beta)^T \cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\right] & \\ +2\,\mathsf{E}\left[r_i(\beta) \cdot h_{\beta\beta}^{''}(x_i, \beta) \cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\right]\Bigg| &\to 0\end{aligned}$$

in probability for $n \to \infty$.

Next, $\hat{\beta}_n^{(NLTS,h_n)}$ is a solution of the normal equations $\frac{\partial S_n(x_i, \varepsilon_i; \beta)}{\partial \beta} = 0$. Thus, using Taylor's expansion theorem

$$0 = \frac{\partial S_n(x_i, \varepsilon_i; \hat{\beta}_n^{(NLTS,h_n)})}{\partial \beta} = \frac{\partial S_n(x_i, \varepsilon_i; \beta^0)}{\partial \beta} + \frac{\partial^2 S_n(x_i, \varepsilon_i; \xi_n)}{\partial \beta \partial \beta^T} \cdot (\beta - \beta^0), \tag{45}$$

where $\xi_n \in \left[\beta^0, \hat{\beta}_n^{(NLTS,h_n)}\right]_\varkappa$. Since $\hat{\beta}_n^{(NLTS,h_n)} \to \beta^0$, the same holds for $\xi_n \to \beta^0$ in probability. Moreover, $\frac{\partial^2 S_n(x_i, \varepsilon_i; \beta)}{\partial \beta \partial \beta^T}$ converges uniformly to a nonstochastic function in β:

$$\mathsf{E}\left[h_\beta^{'}(x_i, \beta) \cdot h_\beta^{'}(x_i, \beta)^T \cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big) - r_i(\beta) \cdot h_{\beta\beta}^{''}(x_i, \beta) \cdot I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)\right],$$

which is continuous (see the verification of Assumption A3 in Lemma 6) and equals to

$$\mathsf{E}\left[h_\beta^{'}(x_i,\beta)\cdot h_\beta^{'}(x_i,\beta)^T\cdot I\Big(r_i^2(\beta)\le G_\beta^{-1}(\lambda)\Big)\right]$$

at β^0 because

$$\mathsf{E}\left[r_i(\beta^0)\cdot h_{\beta\beta}^{''}\big(x_i,\beta^0\big)\cdot I\Big(r_i^2(\beta^0)\le G_{\beta^0}^{-1}(\lambda)\Big)\right]=\mathsf{E}\left[\varepsilon_i\cdot h_{\beta\beta}^{''}\big(x_i,\beta^0\big)\cdot I\Big(\varepsilon_i^2\le G_{\beta^0}^{-1}(\lambda)\Big)\right]=0$$

(see Assumption D2). Therefore,

$$\frac{\partial^2 S_n(x_i,\varepsilon_i;\xi_n)}{\partial\beta\partial\beta^T}\to\mathsf{E}\left[h_\beta^{'}\big(x_i,\beta^0\big)\cdot h_\beta^{'}\big(x_i,\beta^0\big)^T\cdot I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\right]=Q_h\cdot\lambda$$

in probability as $n\to\infty$, where Q_h is a non-singular positive definite matrix. Now, after rewriting (45) as

$$\sqrt{n}(\hat\beta_n^{(NLTS,h_n)}-\beta^0)=-\left[\frac{\partial^2 S_n(x_i,\varepsilon_i;\beta)}{\partial\beta\partial\beta^T}\right]^{-1}\left[\sqrt{n}\frac{\partial S_n(x_i,\varepsilon_i;\beta^0)}{\partial\beta}\right],$$

it is clearly sufficient to verify only that $\sqrt{n}\frac{\partial S_n(x_i,\varepsilon_i;\beta^0)}{\partial\beta}=\mathcal{O}_p(1)$ in order to prove that $\sqrt{n}(\hat\beta_n^{(NLTS,h_n)}-\beta^0)=\mathcal{O}_p(1)$ as $n\to\infty$.

So, let us analyze

$$\begin{aligned}\sqrt{n}\cdot\frac{1}{2}\frac{\partial S_n(x_i,\varepsilon_i;\beta^0)}{\partial\beta} &= -\frac{2}{n}\sum_{i=1}^n r_i(\beta^0)\cdot h_\beta^{'}\big(x_i,\beta^0\big)\cdot I\Big(r_i^2(\beta^0)\le r_{[h]}^2(\beta^0)\Big)\\ &= -\frac{2}{n}\sum_{i=1}^n\varepsilon_i\cdot h_\beta^{'}\big(x_i,\beta^0\big)\cdot I\Big(\varepsilon_i^2\le\varepsilon_{[h]}^2\Big)\end{aligned}$$

and show that it behaves as $\mathcal{O}_p(1)$ for $n\to+\infty$. Apparently,

$$\begin{aligned}\sqrt{n}\cdot\frac{1}{2}\frac{\partial S_n(x_i,\varepsilon_i;\beta^0)}{\partial\beta} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i\cdot h_\beta^{'}\big(x_i,\beta^0\big)\cdot I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i\cdot h_\beta^{'}\big(x_i,\beta^0\big)\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big] \quad (46)\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i\cdot h_\beta^{'}\big(x_i,\beta^0\big)\cdot I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big). \quad (47)\end{aligned}$$

First, keeping in mind that $\varepsilon_i^2\equiv r_i^2\big(\beta^0\big)$, we can employ once again the Chebyshev inequality for nonnegative random variables and write for the first term (46) and $j=1,\ldots,p$

$$\begin{aligned}&P\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^n h_{\beta_j}^{'}(x_i,\beta^0)\varepsilon_i\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big]\right|>K\right)\\ \le\ &\frac{1}{\sqrt{n}K}\mathsf{E}\left|\sum_{i=1}^n h_{\beta_j}^{'}(x_i,\beta^0)\varepsilon_i\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big]\right|\\ \le\ &\frac{\sqrt{n}}{K}\mathsf{E}\left|h_{\beta_j}^{'}(x_i,\beta^0)\varepsilon_i\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big]\right|\\ =\ &\frac{\sqrt{n}}{K}\mathsf{E}\left|h_{\beta_j}^{'}(x_i,\beta^0)\right|\mathsf{E}\left|\varepsilon_i\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big]\right|.\end{aligned}$$

Because $\mathsf{E}\left|h_{\beta_j}^{'}\big(x_i,\beta^0\big)\right|$ exists and is finite (see Assumption H4), we only have to prove that

$$\sqrt{n}\cdot\mathsf{E}\left|\varepsilon_i\cdot\Big[I\Big(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\Big)-I\big(\varepsilon_i^2\le G^{-1}(\lambda)\big)\Big]\right|=\mathcal{O}(1).$$

Using Corollary 1 and Lemma 4, we can write

$$\begin{aligned}
& \sqrt{n}\cdot \mathsf{E}\left|\varepsilon_i\cdot\left[I\left(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\right)-I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right]\right| \\
\le\ & \sqrt{n}\cdot \mathsf{E}\left\{\varepsilon_i\cdot\left|I\left(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\right)-I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right|\,\middle|\,\varepsilon_{[h_n]}^2\le 2G^{-1}(\lambda)\right\}+\sqrt{n}\cdot P\left(\varepsilon_{[h_n]}^2>2G^{-1}(\lambda)\right) \\
=\ & \sqrt{n}\cdot \mathsf{E}\left\{2G^{-1}(\lambda)\cdot\left|I\left(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\right)-I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right|\,\middle|\,\varepsilon_{[h_n]}^2\le 2G^{-1}(\lambda)\right\}+\mathcal{O}\left(n^{-\frac{1}{2}}\right) \\
=\ & 2G^{-1}(\lambda)\cdot\sqrt{n}\cdot P\left(\left|I\left(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\right)-I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right|\,\middle|\,\varepsilon_{[h_n]}^2\le 2G^{-1}(\lambda)\right)+\mathcal{O}\left(n^{-\frac{1}{2}}\right) \\
\le\ & 2G^{-1}(\lambda)\cdot\sqrt{n}\cdot P\left(I\left(\varepsilon_i^2\le\varepsilon_{[h_n]}^2\right)\ne I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right)/P\left(\varepsilon_{[h_n]}^2\le 2G^{-1}(\lambda)\right)+\mathcal{O}\left(n^{-\frac{1}{2}}\right) \\
\le\ & 2G^{-1}(\lambda)\cdot\sqrt{n}\cdot\mathcal{O}\left(n^{-\frac{1}{2}}\right)+\mathcal{O}\left(n^{-\frac{1}{2}}\right)=\mathcal{O}(1)
\end{aligned}$$

as $n\to\infty$. Hence, we can conclude that (46) is $\mathcal{O}_p(1)$ as $n\to+\infty$. Second, term (47) is bounded in probability as well: Lemma 2 allows us to use the Feller-Lindenberg central limit theorem for (47) since $\mathsf{E}\left[\varepsilon_i\cdot h_\beta'(x_i,\beta^0)\cdot I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right]=0$ and

$$\mathsf{var}\left[\varepsilon_i\cdot h_\beta'(x_i,\beta^0)\cdot I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right]=\mathsf{E}\left[\varepsilon_i\cdot h_\beta'(x_i,\beta^0)\cdot I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right]^2=\sigma^2\cdot\mathsf{E}\left(h_\beta'(x_i,\beta^0)\right)^2$$

(see Assumption D2 and H4). This in turn implies that (47) converges in distribution to a normally distributed random variable, and is, therefore, bounded in probability.

Thus, we have proved that $\sqrt{n}\frac{\partial S_n(x_i,\varepsilon_i;\beta^0)}{\partial\beta}=\mathcal{O}_p(1)$, $\frac{\partial^2 S_n(x_i,\varepsilon_i;\xi_n)}{\partial\beta\partial\beta^T}\to\lambda Q_h$, and consequently, $\sqrt{n}(\hat\beta_n^{(NLTS,h_n)}-\beta^0)=\mathcal{O}_p(1)$ as $n\to+\infty$. □

3.5 Asymptotic normality

Finally, the main result concerning the asymptotic normality of NLTS can be derived.

Theorem 4 *Let assumptions D, H, NCA, and NN are fulfilled and*

$$C_\lambda=\lambda-q_\lambda\cdot\left\{f\left(G^{-1}(\lambda)\right)+f\left(G^{-1}(\lambda)\right)\right\}\ne 0.$$

Then

$$\sqrt{n}\left(\hat\beta_n^{(NLTS,h_n)}-\beta^0\right)=n^{-\frac{1}{2}}Q_h^{-1}C_\lambda^{-1}\cdot\sum_{i=1}^{n}\left(y_i-h\left(x_i,\beta^0\right)\right)\cdot h_\beta'\left(x_i,\beta^0\right)\cdot I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)+o_p(1)$$

and $\hat\beta_n^{(NLTS,h_n)}$ is asymptotically normal with its expectation equal to β^0:

$$\sqrt{n}\left(\hat\beta_n^{(NLTS,h_n)}-\beta^0\right)\xrightarrow{F}N(0,V),$$

where $V=C_\lambda^{-2}\cdot Q_h^{-1}\,\mathsf{var}\left[\varepsilon_i\cdot h_\beta'(x_i,\beta^0)\cdot I\left(\varepsilon_i^2\le G^{-1}(\lambda)\right)\right]Q_h^{-1}$.

Proof: Due to Theorem 3, $t_n=\sqrt{n}\left(\hat\beta_n^{(NLTS,h_n)}-\beta^0\right)=\mathcal{O}_p(1)$ as $n\to+\infty$. Therefore, using Theorem 1, we can also write (once again, $q_\lambda=\sqrt{G^{-1}(\lambda)}$ is used for the simplicity of notation)

$$\begin{aligned}
& n^{-\frac{1}{4}}\left(S_n(t_n)+n^{\frac{1}{2}}Q_h t_n\cdot[\lambda-q_\lambda\cdot\{f(-q_\lambda)+f(q_\lambda)\}]\right) \\
=\ & n^{-\frac{1}{4}}\left(S_n\left(\sqrt{n}\left(\hat\beta_n^{(NLTS,h_n)}-\beta^0\right)\right)+\right. \\
& \left.+n^{\frac{1}{2}}Q_h\sqrt{n}\left(\hat\beta_n^{(NLTS,h_n)}-\beta^0\right)\cdot[\lambda-q_\lambda\cdot\{f(-q_\lambda)+f(q_\lambda)\}]\right) \\
=\ & \mathcal{O}_p(1).
\end{aligned}$$

Substituting for $S_n(t)$ yields

$$\begin{aligned} n^{-\frac{1}{2}} \quad & \sum_{i=1}^{n} \left[\left(y_i - h\left(x_i, \hat{\beta}_n^{(NLTS,h_n)}\right)\right) \cdot h_\beta^{'}\left(x_i, \hat{\beta}_n^{(NLTS,h_n)}\right) \cdot I\left(r_i^2 \left(\hat{\beta}_n^{(NLTS,h_n)}\right) \leq r_{[h_n]}^2 \left(\hat{\beta}_n^{(NLTS,h_n)}\right)\right)\right. \\ & \left. - \left(y_i - h\left(x_i, \beta^0\right)\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right)\right] + \\ + \quad & n^{\frac{1}{2}} Q_h \sqrt{n} \left(\hat{\beta}_n^{(NLTS,h_n)} - \beta^0\right) \cdot \left[\lambda - q_\lambda \cdot \{f(-q_\lambda) + f(q_\lambda)\}\right] \\ = \quad & o_p(1) \end{aligned}$$

and since the first summand in the previous equation is by definition of $\hat{\beta}_n^{(NLTS,h_n)}$ equal to zero and $C_\lambda = \lambda - q_\lambda \cdot \{f(-q_\lambda) + f(q_\lambda)\}$

$$\begin{aligned} \sqrt{n}\left(\hat{\beta}_n^{(NLTS,h_n)} - \beta^0\right) \quad &= \quad n^{-\frac{1}{2}} Q_h^{-1} C_\lambda^{-1} \cdot \sum_{i=1}^{n} \left(y_i - h\left(x_i, \beta^0\right)\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) \\ & \qquad + o_p(1) \\ &= \quad n^{-\frac{1}{2}} Q_h^{-1} C_\lambda^{-1} \cdot \sum_{i=1}^{n} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right) + o_p(1) \\ &+ \quad n^{-\frac{1}{2}} Q_h^{-1} C_\lambda^{-1} \cdot \sum_{i=1}^{n} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \times \qquad (48) \\ & \qquad \times \left[I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right]. \end{aligned}$$

First, we will show that term (48) is negligible in probability. As

$$\begin{aligned} & \mathsf{E}\left\{ n^{\frac{1}{4}} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \left[I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right]\right\} \\ = \quad & \mathsf{E}\left\{ n^{\frac{1}{4}} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \mathsf{E}\left(\left| I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right| \Big| x_i\right)\right\} \\ = \quad & \mathcal{O}\left(n^{-\frac{1}{4}}\right) \cdot \mathsf{E}\, r_i(\beta^0) \cdot \mathsf{E}\, h_\beta^{'}\left(x_i, \beta^0\right) = o(1) \end{aligned}$$

($\varepsilon_i \equiv r_i(\beta^0)$ and x_i are independent random variables) and

$$\begin{aligned} & \mathsf{var}\left\{ n^{\frac{1}{4}} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \left[I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right]\right\} \\ \leq \quad & n^{\frac{1}{2}}\, \mathsf{E}\left\{ r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \mathsf{var}\left(\left| I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right| \Big| x_i\right)\right\} \\ + \quad & n^{\frac{1}{2}}\, \mathsf{var}\left\{ r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \mathsf{E}\left(\left| I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right| \Big| x_i\right)\right\} \\ \leq \quad & \mathcal{O}(1)\left\{ \mathsf{E}\left[r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right)\right] + \mathsf{var}\left[r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right)\right]\right\} \\ = \quad & \mathcal{O}(1), \end{aligned}$$

(see Lemma 6), we can conclude that

$$n^{-\frac{1}{4}} \sum_{i=1}^{n} r_i(\beta^0) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot \left[I\left(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right)\right]$$

is asymptotically normally distributed. Hence, term (48) is negligible in probability $o_p(1)$. Given this result,

$$\sqrt{n}\left(\hat{\beta}_n^{(NLTS,h_n)} - \beta^0\right) = n^{-\frac{1}{2}} Q_h^{-1} C_\lambda^{-1} \sum_{i=1}^{n} \left(y_i - h\left(x_i, \beta^0\right)\right) \cdot h_\beta^{'}\left(x_i, \beta^0\right) \cdot I\left(r_i^2(\beta^0) \leq G^{-1}(\lambda)\right) + o_p(1), \qquad (49)$$

which is the first assertion of the theorem. Using the Feller-Lindenberg central limit theorem for the right hand side of equation (49), we obtain the second assertion of the theorem, namely, asymptotic linearity of $\hat{\beta}_n^{(NLTS,h_n)}$. □

4 Conclusion

In this paper, I have introduced the nonlinear least trimmed squares estimator and derived its asymptotic properties. Thus, the applicability of the LTS procedure is extended to various intrinsically nonlinear models. Maybe more important is the adaptation of the methodology used in proofs, which should presumably allow us an extension of the trimmed-estimation procedures in the area of nonlinear models with a more complicated error structure, such as limited-dependent-variable models. Finally, given the rather theoretical character of the paper, it remains to be seen whether the existing computational procedures designed for LTS in the linear regression model suit well various nonlinear settings.

A Proofs of lemmas and other auxiliary propositions

In this appendix, I present first the proofs of all lemmas used in this paper, see Section A.1. Some fundamental results needed to prove these lemmas are formulated and derived in Section A.2.

A.1 Proofs of lemmas and propositions

Proposition 1 *Let $x_1, x_2, \ldots$ be a sequence of independent identically distributed random variables with an absolutely continuous distribution function $F(x)$. Let $b(x)$ be a lower bound for $F(x)$ in a neighborhood U_1 of $+\infty$. If $b(x)$ can be chosen as $1 - \frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \max_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$. Analogously, let $c(x)$ be an upper bound for $F(x)$ in a neighborhood U_2 of $-\infty$. If $c(x)$ can be chosen as $\frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \min_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$.*

Proof: We prove the lemma just for the case of the lower bound, $b(x)$, the other case can be derived similarly. The cumulative distribution function of $x_{max} = \max_{i=1,\ldots,n} x_i$ is $F_n(x) = F^n(x)$. We want to show that for any $\varepsilon > 0$ there is $K > 0$ such that $P(x_{max} > K\sqrt[4]{n}) = 1 - F_n(K\sqrt[4]{n}) < \varepsilon$. This is equivalent to the assertion that $F_n(K\sqrt[4]{n}) \to 1$ as $K \to +\infty$ uniformly for $n > n_0$ and some n_0. Because $b(x) < F(x)$, it also holds $b^n(x) < F^n(x) = F_n(x)$ and thus it is enough to verify that $b^n(K\sqrt[4]{n}) \to 1$ as $K \to +\infty$ uniformly for $n > n_0$. In general, $P_4(x) = a_1x^4 + a_2x^3 + a_3x^2 + a_4x + a_5$ and its leading coefficient a_1 has to be positive—otherwise, $b(x) > 1$ for x large enough and it could not be a lower bound to a distribution function, which is at most equal to one. So, let us assume without loss of generality that $P_4(x) = x^4$ and $b(x) = 1 - \frac{1}{x^4}$. Hence,

$$b^n\left(K\sqrt[4]{n}\right) = \left(1 - \frac{1}{Kn}\right)^n = \left[\left(1 - \frac{1}{Kn}\right)^{Kn}\right]^{\frac{1}{K}} \to \left(\frac{1}{e}\right)^{\frac{1}{K}} = \sqrt[K]{\frac{1}{e}},$$

$b^n\left(K\sqrt[4]{n}\right)$ converges monotonically to a positive number smaller than one for a fixed $K > 0$; moreover, this number $\frac{1}{\sqrt[K]{e}}$ as well as $b^n\left(K\sqrt[4]{n}\right)$ increase with K. Therefore, we can find $n_0 > 0$ such that $b^n\left(K\sqrt[4]{n}\right) > \sqrt[K]{\frac{1}{3}}$ for all $n > n_0$ and $K > 1$. Since $\sqrt[K]{\frac{1}{3}} \to 1$ for $K \to \infty$, also $b^n(K\sqrt[4]{n}) \to 1$ as $K \to +\infty$ uniformly for $n > n_0$. This closes the proof. □

Lemma 1 *Let $n \in \mathbb{N}$ and $k_h(\beta) : \mathbb{R}^p \to \{1, \ldots, n\}$ be a function that represents an index of an observation such that $r^2_{k_h(\beta)}(\beta) = r^2_{[h]}(\beta)$, $h \in \{1, \ldots, n\}$. Under Assumptions D and H, there exists a set Ω_1, $P(\Omega_1) = 1$, such that for every $\omega \in \Omega_1$ there is some neighborhood $U(\beta^0, \varepsilon(\omega))$ of β^0 such that the function $k_h(\beta)$ is constant on $U(\beta^0, \varepsilon(\omega))$ for all $h \in \{1, \ldots, n\}$.*

Proof: Given our distributional assumptions about $r_i(\beta^0) \equiv \varepsilon_i$ and an arbitrary fixed n, the probability that any two of residuals $r_i(\beta^0), i = 1, \ldots, n$, have the same value is equal to zero (residuals $r_1(\beta^0), \ldots, r_n(\beta^0)$ are independent identically distributed random variables that are continuously distributed). In other words, the set of events $\omega \in \Omega^n$ for which some residuals are equal at β^0 has probability zero—$P(\Omega_0 = \{\omega \in \Omega^n : \exists i, j \in \{1, \ldots, n\}, i \neq j, r_i(\beta^0, \omega) = r_j(\beta^0, \omega)\}) = 0$. Moreover, there is a $\delta' > 0$ such that $r_i(\beta)$ is continuous on $\bar{U}(\beta^0, \delta')$, and therefore also uniformly continuous on $\bar{U}(\beta^0, \delta')$. Therefore, for any given $\omega \notin \Omega_0$ and $\kappa(\omega) = \frac{1}{2} \min_{i,j=1,\ldots,n} \left|r_j(\beta^0, \omega) - r_i(\beta^0, \omega)\right| > 0$ we can find an $\varepsilon(\omega) > 0$ such that, for any $\beta \in U(\beta^0, \varepsilon(\omega))$, it holds that $\left|r_i(\beta, \omega) - r_i(\beta^0, \omega)\right| < \kappa(\omega)$ for all $i = 1, \ldots, n$. Consequently, mapping $k_h(\beta)$ is constant on $U(\beta^0, \varepsilon(\omega))$ for any $\omega \notin \Omega_0$ because the ordering of residuals $r_i^2(\beta)$ is independent of β on this set. The set $\Omega_1 = \Omega - \Omega_0$. □

Lemma 2 $P\left(\left\{\omega = (\omega_1, \ldots, \omega_n) \in \Omega^n : r_i^2(\beta, \omega_i) = r^2_{[h]}(\beta, \omega)\right\}\right) = \frac{1}{n}$ *for any $n \in \mathbb{N}$, $i, h \in \{1, \ldots, n\}$, and $\beta \in B$.*

Proof: As $r_i^2(\beta,\omega_i), i = 1,\ldots,n$, form a vector of independent identically distributed random variables for given n and β,

$$\begin{aligned} &P\Big(\Big\{\omega = (\omega_1,\ldots,\omega_n) \in \Omega^n : r_i^2(\beta,\omega_i) = r_{[h]}^2(\beta,\omega)\Big\}\Big) = \\ &= \quad P\Big(\Big\{\omega = (\omega_1,\ldots,\omega_n) \in \Omega^n : r_j^2(\beta,\omega_j) = r_{[h]}^2(\beta,\omega)\Big\}\Big) \end{aligned}$$

for any $i, j \in \{1,\ldots,n\}$ and a fixed $h \in \{1,\ldots,n\}$. Moreover,

$$\sum_{i=1}^{n} P\Big(\Big\{\omega = (\omega_1,\ldots,\omega_n) \in \Omega^n : r_i^2(\beta,\omega_i) = r_{[h]}^2(\beta,\omega)\Big\}\Big) = 1$$

since $P(\{\omega \in \Omega^n : \exists i \neq j \in \{1,\ldots,n\}, r_i(\beta,\omega_i) = r_j(\beta,\omega_j)\}) = 0$. Putting the last two equations together, we immediately get for any $i = 1,\ldots,n$

$$P\Big(\Big\{\omega = (\omega_1,\ldots,\omega_n) \in \Omega^n : r_i^2(\beta,\omega_i) = r_{[h]}^2(\beta,\omega)\Big\}\Big) = \frac{1}{n},$$

which closes the proof. □

Lemma 3 *Let $1/2 < \lambda < 1$ and $0 < c < G_\beta^{-1}(\lambda)$ be a real constant, where G_β represents the distribution function of $r_i^2(\beta)$, $\beta \in B$. Then, under Assumptions D, $P\Big(r_{[h_n]}^2(\beta) \leq c\Big) = \mathcal{O}\big(n^{-k}\big)$ for any $k \in \mathbb{N}$ as $n \to +\infty$.*

Proof: As $\beta \in B$ is a fixed parameter vector, let us for the sake of simplicity denote the distribution function $G_\beta(x)$ by $G(x)$ throughout the proof (overshadowing the till now used definition $G \equiv G_{\beta^0}$). The distribution function of $r_{[h_n]}^2(\beta)$ is then given by

$$G_{h_n}(x) = \sum_{i=h_n}^{n} P_i(x), \quad P_i(x) = \binom{n}{i} G(x)^i (1-G(x))^{n-i}. \tag{50}$$

Let $\lambda' = G(c) < \lambda < 1$ and let M_{h_n} be an upper bound for $G_{h_n}(x)$ on the interval $\langle 0, c\rangle$. We show that $\lim_{n\to\infty} n^k M_{h_n} = 0$ for any $k \in \mathbb{N}$.

First, let us bring attention to one fundamental property of $G(x)^i(1-G(x))^{n-i}$, the main element of $P_i(x)$ in (50). For any $a \in (\lambda, 1)$

$$\begin{aligned} \Big(G(x)^a\,(1-G(x))^{1-a}\Big)' &= aG(x)^{a-1}g(x)(1-G(x))^{1-a} - (1-a)G(x)^a(1-G(x))^{-a\lambda}g(x) \\ &= G(x)^{a-1}g(x)(1-G(x))^{-a}\,(a(1-G(x)) - (1-a)G(x)) \\ &\geq 0, \quad \forall 0 < x \leq c < G^{-1}(\lambda) \leq G^{-1}(a) \end{aligned}$$

($G(x)$ is monotonic). Therefore, $\left(\frac{G(x)}{a}\right)^a \left(\frac{1-G(x)}{1-a}\right)^{1-a} < \left(\frac{\lambda'}{a}\right)^a \left(\frac{1-\lambda'}{1-a}\right)^{1-a} = C(a) < 1$ as $x \in \langle 0, c\rangle$. Since $\lambda' < a$, the derivative of $\left(\frac{\lambda'}{a}\right)^a \left(\frac{1-\lambda'}{1-a}\right)^{1-a}$ with respect to a is negative:

$$\begin{aligned} \left[\left(\frac{\lambda'}{a}\right)^a \left(\frac{1-\lambda'}{1-a}\right)^{1-a}\right]' &= \left[e^{a\ln\frac{\lambda'}{a} + (1-a)\ln\frac{1-\lambda'}{1-a}}\right]' = \left[e^{-a\ln a - (1-a)\ln(1-\alpha)}\right]' \\ &= (-\ln a - 1 + \ln(1-a) + 1)\cdot e^{-a\ln a - (1-a)\ln(1-\alpha)} \\ &= \ln\frac{1-a}{a} \cdot \left(\frac{\lambda'}{a}\right)^a \left(\frac{1-\lambda'}{1-a}\right)^{1-a} < 0, \end{aligned}$$

because $a \in (\lambda, 1) \subseteq (\frac{1}{2}, 1)$ and $0 < \frac{1-a}{a} < 1$. Hence, $C(a) < C(\frac{1}{2}(\lambda + \lambda')) = C < 1$. We show the usefulness of this result in a moment.

Now, we analyze the function $G_{h_n}(x)$ itself. Taking into account $h_n/n \to \lambda$ (h_n is defined as $[\lambda n]$), it follows that we can write $h_n/n = \lambda + a_n$, where $|a_n| < \frac{1}{n}$. Moreover, notice that $\frac{i}{n} > \lambda'' = \frac{1}{2}(\lambda + \lambda')$ for any n sufficiently high and $i \geq h_n$. Let us take some $0 \leq x \leq c$ (therefore, $G(x) \neq 1$). Using the Stirling formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \left(1 + \mathcal{O}\left(\frac{1}{12n}\right)\right),$$

we get for $n^{k+1/2} \cdot G_{h_n}(x)$

$$\begin{aligned}
&\lim_{n\to\infty} n^{k+1/2} \sum_{i=h_n}^{n} P_i(x) = \lim_{n\to\infty} n^{k+1/2} \sum_{i=h_n}^{n} \binom{n}{i-1} G(x)^i (1-G(x))^{n-i} \\
= &\lim_{n\to\infty} n^{k+1/2} \sum_{i=h_n}^{n} \frac{n!}{i!(n-i)!} G(x)^i [1-G(x)]^{n-i} \\
= &\lim_{n\to\infty} n^{k+1/2} \sum_{i=h_n}^{n} \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \left(1+\mathcal{O}\left(\frac{1}{12n}\right)\right) \times G(x)^i [1-G(x)]^{n-i}}{\sqrt{2\pi i}\left(\frac{i}{e}\right)^i \cdot \left(1+\mathcal{O}\left(\frac{1}{12i}\right)\right) \times \sqrt{2\pi(n-i)}\left(\frac{n-i}{e}\right)^{n-i} \cdot \left(1+\mathcal{O}\left(\frac{1}{12(n-i)}\right)\right)} \\
= &\lim_{n\to\infty} n^{k} \sum_{i=h_n}^{n} \frac{\frac{1}{e}\sqrt{2\pi} \cdot \left(1+\mathcal{O}\left(\frac{1}{12n}\right)\right) \times G(x)^i [1-G(x)]^{n-i}}{\sqrt{2\pi\frac{i}{n}}\left(\frac{i}{n}\right)^i \cdot \left(1+\mathcal{O}\left(\frac{1}{12i}\right)\right) \times \sqrt{2\pi\frac{n-i}{n}}\left(\frac{n-i}{n}\right)^{n-i} \cdot \left(1+\mathcal{O}\left(\frac{1}{12(n-i)}\right)\right)} \\
= &\lim_{n\to\infty} n^{k} \sum_{i=h_n}^{n} \left(\frac{G(x)}{\frac{i}{n}}\right)^i \left(\frac{1-G(x)}{\frac{n-i}{n}}\right)^{n-i} \cdot \frac{\frac{1}{e}\sqrt{2\pi} \cdot \left(1+\mathcal{O}\left(\frac{1}{12n}\right)\right)}{\sqrt{2\pi\frac{i}{n}} \cdot \left(1+\mathcal{O}\left(\frac{1}{12i}\right)\right) \times \sqrt{2\pi\frac{n-i}{n}} \cdot \left(1+\mathcal{O}\left(\frac{1}{12(n-i)}\right)\right)} \\
= &\lim_{n\to\infty} n^{k} \sum_{i=h_n}^{n} \left(\frac{G(x)}{\frac{i}{n}}\right)^i \left(\frac{1-G(x)}{\frac{n-i}{n}}\right)^{n-i} \cdot \mathcal{O}(1) \\
= &\lim_{n\to\infty} n^{k} \sum_{i=h_n}^{n} \left[\left(\frac{G(x)}{\frac{i}{n}}\right)^{\frac{i}{n}} \left(\frac{1-G(x)}{\frac{n-i}{n}}\right)^{\frac{n-i}{n}}\right]^n \cdot \mathcal{O}(1) \\
\leq &\lim_{n\to\infty} n^{k} \sum_{i=h_n}^{n} C^n \cdot \mathcal{O}(1) \\
= &\lim_{n\to\infty} n^{k+1} C^n \cdot \mathcal{O}(1) \\
= &\ 0.
\end{aligned}$$

Therefore, we have proved that $\lim_{n\to\infty} n^k \cdot \sup_{x\in\langle 0,c\rangle} G_{h_n}(x) = 0$, which closes the proof as $P(r^2_{[h_n]}(\beta) \leq c) = G_{h_n}(c)$. □

Corollary 1 *Analogously, it is possible under Assumptions D and H to show that for real constants* $1/2 < \lambda < 1$ *and* $G^{-1}_{\beta}(\lambda) < c < \infty$ *it holds under conditions D that* $P\left(r^2_{[h_n]}(\beta) \geq c\right) = \mathcal{O}\left(n^{-k}\right)$ *for any* $k \in \mathbb{N}$ *as* $n \to +\infty$.

Corollary 2 *Let* $1/2 < \lambda < 1$ *and* $0 < c < G^{-1}(\lambda) < c' < \infty$ *be real constants. Under assumptions D and H, it holds that*

$$P\left(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : r^2_{[h_n]}(\beta) \notin (c, c')\right) = \mathcal{O}\left(n^{-k}\right) \tag{51}$$

for any $k \in \mathbb{N}$ *as* $n \to +\infty$.

Proof: First, note that $r^2_i(\beta, \omega) \to r^2_i(\beta^0, \omega)$ for $\beta \to \beta^0$ and any $\omega \in \Omega$ (convergence almost surely). So, for $\beta \to \beta^0$ and $G_\beta(x)$ being the cumulative distribution function of $r^2_i(\beta)$, it holds that $G_\beta(x) \to G_{\beta^0}(x) \equiv G(x)$ for all $x \in \mathbb{R}$ (convergence in distribution) because $G(x)$ is an absolutely continuous distribution function. The same is valid about convergence of $r_i(\beta, \omega)$ to $r_i(\beta^0, \omega)$ and their corresponding distribution functions $F_\beta(x)$ and $F_{\beta^0}(x) \equiv F(x)$. Now, we show

that this convergence of distribution functions $F_\beta(x) \to F_{\beta^0}(x) \equiv F(x)$ (and consequently of $G_\beta(x) \to G_{\beta^0}(x) \equiv G(x)$—$G_\beta(x) = F_\beta(\sqrt{x}) - F_\beta(-\sqrt{x})$) is uniform over all sequences $\beta_n \to \beta^0$ such that $\beta_n \in U(\beta^0, n^{-\frac{1}{2}}M)$. The reason is that $r_i(\beta,\omega) = r_i(\beta^0,\omega) - h'_\beta(x_i,\xi)\cdot\left(\beta-\beta^0\right)$, where the second term $h'_\beta(x_i,\xi)\cdot\left(\beta-\beta^0\right)$ can be bounded by a random variable $h_x(x_i)\cdot K\cdot\left(\beta-\beta^0\right) = \mathcal{O}_p\left(n^{\frac{1}{4}}\right)\cdot\mathcal{O}\left(n^{-\frac{1}{2}}\right) = \mathcal{O}_p\left(n^{-\frac{1}{4}}\right)$ independently of β (see Assumption H3).

Now, let $\varepsilon > 0$ be chosen so that $(\lambda-2\varepsilon,\lambda+2\varepsilon) \subset (G(c),G(c'))$ and $(G^{-1}(\lambda)-2\varepsilon, G^{-1}(\lambda)+2\varepsilon) \subset (c,c')$ (remember, $G(c) < \lambda < G(c')$). Moreover, because of the mentioned uniform convergence, there exists $n_0 \in \mathbb{N}$ such that $G^{-1}_{\beta_n}(\lambda) \in (G^{-1}(\lambda)-\varepsilon, G^{-1}(\lambda)+\varepsilon)$ for any $\beta_n \in U(\beta^0, n^{-\frac{1}{2}}M)$ and $n > n_0$. Hence, $(G^{-1}_{\beta_n}(\lambda)-\varepsilon, G^{-1}_{\beta_n}(\lambda)+\varepsilon) \subset (c,c')$, and $|G_{\beta_n}(c)-\lambda| > \varepsilon$ and $|G_{\beta_n}(c')-\lambda| > \varepsilon$ for all $\beta_n \in U(\beta^0, n^{-\frac{1}{2}}M)$ and $n > n_0$. Hence, the constant λ' in the proof of Lemma 3 can be chosen equal to $\lambda-\varepsilon$ independently of $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. Thus, we can follow the same steps as in the proof of Lemma 3 to derive (51). □

Corollary 3 *Let $1/2 < \lambda < 1$ and $\varepsilon > 0$ be a sufficiently small real constants. Under Assumptions D and NC, it holds that*

$$P\left(\exists\beta \in B : r^2_{[h_n]}(\beta) \notin \left(G^{-1}_\beta(\lambda)-\varepsilon, G^{-1}_\beta(\lambda)+\varepsilon\right)\right) = \mathcal{O}\left(n^{-k}\right) \tag{52}$$

for any $k \in \mathbb{N}$ as $n \to +\infty$.

Proof: Let $\lambda'_\beta = G_\beta\left(G^{-1}_\beta(\lambda)-\varepsilon\right) < \lambda$ and $a \in (\lambda,1)$. We know that $\left(\frac{G(x)}{a}\right)^a\left(\frac{1-G(x)}{1-a}\right)^{1-a} < \left(\frac{\lambda'_\beta}{a}\right)^a\left(\frac{1-\lambda'_\beta}{1-a}\right)^{1-a} = C(a,\beta) < 1$ as $x \in \left\langle 0, G^{-1}_\beta(\lambda)-\varepsilon\right\rangle$ (see Lemma 3). Furthermore, $C(a,\beta) < C(\frac{1}{2}(\lambda+\lambda'_\beta)) = C < 1$. To prove the result of this corollary, we can follow the same steps as in Lemma 3 as long as we show that $\sup_{\beta\in B} C(a,\beta) < 1$, which is equivalent to

$$\Lambda = \sup_{\beta\in B}\lambda'_\beta = \sup_{\beta\in B} G_\beta\left(G^{-1}_\beta(\lambda)-\varepsilon\right) < \lambda.$$

Then we can choose $C(a,\beta) < C(\frac{1}{2}(\lambda+\Lambda)) = C < 1$ and complete the proof in the same way as the proof of Lemma 3. Because

$$G_\beta\left(G^{-1}_\beta(\lambda)-\varepsilon\right) = \lambda - g_\beta(\xi)\varepsilon,$$

it is sufficient to know that there is some $\delta > 0$ such that

$$\inf_{\beta\in B}\inf_{z\in(-\delta,\delta)} g_\beta\left(G^{-1}_\beta(\lambda)+z\right) > 0.$$

However, this is one of Assumptions NC, which closes the proof of the corollary. □

Lemma 4 *Under Assumptions D and H, for any fixed $i \in \mathbb{N}$ and $n \geq i$*

$$P\left(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\left(r_i^2(\beta) \leq r^2_{[h_n]}(\beta)\right) \neq I\left(r_i^2(\beta) \leq G^{-1}_\beta(\lambda)\right)\right) = \mathcal{O}\left(n^{-\frac{1}{2}}\right)$$

as $n \to +\infty$, or analogously

$$P\left(\sup_{\beta\in U(\beta^0, n^{-\frac{1}{2}}M)}\left|I\left(r_i^2(\beta) \leq r^2_{[h_n]}(\beta)\right) - I\left(r_i^2(\beta) \leq G^{-1}_\beta(\lambda)\right)\right| \neq 0\right) = \mathcal{O}\left(n^{-\frac{1}{2}}\right)$$

as $n \to +\infty$.

Proof: Let us introduce a bit of notation first: $\upsilon_{in}(\beta) = I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big) - I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big)$. We have to derive, in fact, an upper bound for

$$\mathsf{E} \sup_{\beta \in U(\beta^0, n^{-\frac{1}{2}}M)} |\upsilon_{in}(\beta)| = P\left(\sup_{\beta \in U(\beta^0, n^{-\frac{1}{2}}M)} |\upsilon_{in}(\beta)| = 1 \right).$$

For the sake of simplicity, we will omit in what follows the specification of the set across which the supremum is considered and write simply $P(\exists \beta : |\upsilon_{in}(\beta)| = 1) = P\big(\sup_\beta |\upsilon_{in}(\beta)| = 1\big)$ keeping in mind that we mean $P\Big(\sup_{\beta \in U(\beta^0, n^{-\frac{1}{2}}M)} |\upsilon_{in}(\beta)| = 1\Big)$ and $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. Without loss of generality, we derive only $P(\exists \beta : \upsilon_{in}(\beta) = -1)$, i.e.,

$$P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big) = 0 \wedge I\Big(r_i^2(\beta) \le G_\beta^{-1}(\lambda)\Big) = 1\Big);$$

the other case can be analyzed analogously.

Before we start with the derivation, notice that the distribution function of an order statistics $r_{[h]}^2(\beta)$ for a given $h, 1 \le h \le n$, is given by (presuming that G_β and g_β are the c.d.f. and p.d.f. of $r_i^2(\beta)$)

$$G_{\beta,h}(x) = \sum_{j=h}^{n} P_j(x,\beta), \quad P_j(x,\beta) = \binom{n}{j} G_\beta(x)^j (1 - G_\beta(x))^{n-j}$$

and the corresponding probability density function is given by (for $n \ge 2$)

$$g_{\beta,h}(x) = n \binom{n-1}{h-1} g_\beta(x) G_\beta(x)^{h-1} (1 - G_\beta(x))^{n-h}.$$

Throughout this proof, we use notation G_{β^0} and $G_{\beta^0,h}$ instead of G and G_h to make it consistent with frequent use of G_β and $G_{\beta,h}$. The same applies for g_{β^0}.

Now, let us consider an $\omega = (\omega_1, \ldots, \omega_n) \in \Omega^n$ and assume without loss of generality that $i = 1$. Given $\omega' = (\omega_2, \ldots, \omega_n) \in \Omega^{n-1}$ and $\big(r_2^2(\beta, \omega_2), \ldots, r_n^2(\beta, \omega_n)\big)$

$$r_{[h]}^2(\beta,\omega) = \begin{cases} r_{[h-1]}^2(\beta,\omega') & \text{if } r_1^2(\beta,\omega_1) < r_{[h-1]}^2(\beta,\omega') \\ r_1^2(\beta,\omega_1) & \text{if } r_{[h-1]}^2(\beta,\omega') \le r_1^2(\beta,\omega_1) \le r_{[h]}^2(\beta,\omega') \\ r_{[h]}^2(\beta,\omega') & \text{if } r_{[h]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \end{cases} \tag{53}$$

Denoting Ω_1, Ω_2, and Ω_3 subsets of Ω^n corresponding to the three (disjoint) cases in (53), we can write

$$\begin{aligned} P(\{\omega \in \Omega^n | \exists \beta : \upsilon_{1n}(\beta) = -1\}) &= P(\{\omega \in \Omega_1 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) \\ &+ P(\{\omega \in \Omega_2 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) \\ &+ P(\{\omega \in \Omega_3 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) \end{aligned}$$

and analyze this sum one by one.

1. $P(\{\omega \in \Omega_1 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) \le P\Big(\exists \beta : r_{[h_n]}^2(\beta,\omega) < r_1^2(\beta,\omega_1) < r_{[h_n]}^2(\beta,\omega)\Big) = 0.$

2. $P(\{\omega \in \Omega_2 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) = P\Big(\exists \beta : r_{[h_n-1]}^2(\beta,\omega') \le r_1^2(\beta,\omega_1) = r_{[h_n]}^2(\beta,\omega) \le G_\beta^{-1}(\lambda)\Big)$ can be analyzed in exactly the same way as $P(\{\omega \in \Omega_3 | \exists \beta : \upsilon_{1n} = -1\})$, see point 3.

3. $P(\{\omega \in \Omega_3 | \exists \beta : \upsilon_{1n}(\beta) = -1\}) = P\Big(\exists \beta : r_{[h_n]}^2(\beta,\omega') = r_{[h_n]}^2(\beta,\omega) < r_1^2(\beta,\omega_1) \le G_\beta^{-1}(\lambda)\Big)$. We can structure this last term in the following way ($1 \gg \varepsilon > 0$ is an arbitrary, but fixed real number; the choice of ε will be discussed later):

$$\begin{aligned} P\Big(r_{[h_n]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \le G_\beta^{-1}(\lambda)\Big) &\le P\Big(r_{[h_n]}^2(\beta,\omega') \le G_{\beta^0}^{-1}(\lambda) - \varepsilon/2\Big) \\ &+ P\Big(G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r_{[h_n]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \le G_\beta^{-1}(\lambda)\Big). \end{aligned} \tag{54}$$

Please note that $G_\beta^{-1}(\lambda) \in (G_{\beta^0}^{-1}(\lambda) - \varepsilon/4, G_{\beta^0}^{-1}(\lambda) + \varepsilon/4)$ for n larger than a certain n_0 because $G_\beta(x) \to G_{\beta^0}(x)$ for all $x \in \mathbb{R}$ (remember, $G \equiv G_{\beta^0}$). Since Corollary 2 implies $P\Big(r^2_{[h_n]}(\beta,\omega) \le G_{\beta^0}^{-1}(\lambda) - \varepsilon/2\Big) = o(\frac{1}{n})$ as $n \to +\infty$, we have to analyze just the second term on the right hand side of the equality:

$$P\Big(G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r^2_{[h_n]}(\beta,\omega) < r_1^2(\beta,\omega_1) \le G^{-1}(\lambda)\Big) = \tag{55}$$
$$= \int_{\omega' \in \Omega^{n-1}} \int_{\omega_1 \in \Omega} I\Big(G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r^2_{[h_n]}(\beta,\omega') < r_1^2(\beta,\omega_1) \le G_\beta^{-1}(\lambda)\Big) dP(\omega_1) dP(\omega').$$

Let $1 \gg \varepsilon > 0$ be a fixed real number small enough and $n > n_0$ large enough so that (see Assumption D3 and related notation in Section 3.2 for the definition of functions G_β and g_β)

(a) $G_{\beta^0}^{-1}(\lambda) - 2\varepsilon > 0$, and consequently $G_\beta^{-1}(\lambda) - \varepsilon > 0$ for $n > n_0$; hence, the boundedness of $f(x)$ (and thus $f_\beta(x)$ as well) indicates that there is a $M_g > 0$ such that $g_\beta(x) < M_g$ for all $G_{\beta^0}^{-1}(\lambda) - \varepsilon \le x$,

(b) $G_{\beta^0}(G_{\beta^0}^{-1}(\lambda) - \varepsilon) = \gamma > 0$ and $G_\beta(G_{\beta^0}^{-1}(\lambda) - \varepsilon) > \gamma/2 > 0$,

(c) $g_\beta(x) > m_g > 0$ almost surely for all $G_{\beta^0}^{-1}(\lambda) - 2\varepsilon \le x \le G_{\beta^0}^{-1}(\lambda) + 2\varepsilon$, where $m_g > 0$ is a real constant (this is possible because $g_\beta(G_\beta^{-1}(\lambda)) > 0$ for any $\lambda \in (0,1)$, and hence $g_\beta(G_{\beta^0}^{-1}(\lambda)) > 0$ for any $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ and $n > n_0$),

(d) $M_g\varepsilon/\lambda < 1$ and $M_g\varepsilon/(1-\lambda) < 1$,

(e) $m_g^2/2 > \left|\frac{\lambda-1}{\lambda}\right| (1 - M_g\varepsilon/\lambda)^{\lambda-2} M_g^3\varepsilon + \left|\frac{\lambda}{1-\lambda}\right| M_g^3\varepsilon + \left|\frac{\lambda}{1-\lambda}\right| \left|\frac{\lambda-1}{\lambda}\right| (1 - M_g\varepsilon/\lambda)^{\lambda-2} M_g^4\varepsilon^2$,

(f) $G_\beta(x) = G_\beta(G_\beta^{-1}(\lambda)) + g_\beta(\xi)(x - G_\beta^{-1}(\lambda)) = \lambda + g_\beta(\xi)(x - G_\beta^{-1}(\lambda))$ for all $x \in (G_{\beta^0}^{-1}(\lambda) - 2\varepsilon, G_{\beta^0}^{-1}(\lambda) + 2\varepsilon)$, where $\xi \in (x, G_\beta^{-1}(\lambda))$.

Suppose further that $n > n_0$, where n_0 is defined above. Then we can write (see equation (55))

$$\begin{aligned}
& P\Big(\exists \beta : G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r^2_{[h_n]}(\beta,\omega) < r_1^2(\beta,\omega_1) \le G_{\beta^0}^{-1}(\lambda)\Big) \le \\
= & \int_{\omega' \in \Omega^{n-1}} \int_{\omega_1 \in \Omega} I\Big(G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r^2_{[h_n]}(\beta,\omega') < r_1^2(\beta,\omega_1) \le G_\beta^{-1}(\lambda)\Big) dP(\omega_1) dP(\omega') \\
\le & \int_{\omega' \in \Omega^{n-1}} M_g \cdot \\
& \cdot \sup_\beta \left\{ \left| G_\beta^{-1}(\lambda) - r^2_{[h_n]}(\beta,\omega') \right| \cdot I\Big(G_{\beta^0}^{-1}(\lambda) - \varepsilon/2 < r^2_{[h_n]}(\beta,\omega') \le G_\beta^{-1}(\lambda)\Big) \right\} dP(\omega') \\
= & M_g \cdot \int_{G_{\beta^0}^{-1}(\lambda) - \varepsilon/2}^{G_\beta^{-1}(\lambda)} \sup_\beta \left\{ \left| G_\beta^{-1}(\lambda) - x \right| \cdot g_{\beta,h_n}(x) \right\} dx \qquad (56) \\
= & M_g \cdot \int_0^{G_\beta^{-1}(\lambda) - G_{\beta^0}^{-1}(\lambda) + \varepsilon/2} \sup_\beta \left\{ |y| \cdot g_{\beta,h_n}(G_\beta^{-1}(\lambda) - y) \right\} dy \\
\le & M_g \cdot \int_0^\varepsilon y \cdot \sup_\beta \left\{ g_{\beta,h_n}(G_\beta^{-1}(\lambda) - y) \right\} dy.
\end{aligned}$$

To see how this integral behaves, it is necessary to analyze function $g_{\beta,h_n}(\cdot)$ in a neighborhood of $G_\beta^{-1}(\lambda)$ given by 2ε for $n \to +\infty$. As for $x \in (G^{-1}(\lambda) - \varepsilon, G^{-1}(\lambda))$

$$\frac{g_{\beta,h_n}(x)}{\sqrt{n}} = \sqrt{n} \binom{n-1}{r-1} g_\beta(x) G_\beta(x)^{h-1} (1 - G_\beta(x))^{n-h} \le M_g \sqrt{n} \binom{n-1}{r-1} G_\beta(x)^{h-1} (1 - G_\beta(x))^{n-h},$$

it suffices to analyze the latter term. Using the Stirling formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \left(1 + \mathcal{O}\left(\frac{1}{12n}\right)\right)$$

and $h_n/n = \lambda + a_n$, where $|a_n| < \frac{1}{n}$, we can arrange the expression in question in the following way (notice that $G_\beta(x) \neq 0$ and $G_\beta(x) \neq 1$ for $x \in (G_\beta^{-1}(\lambda) - \varepsilon, G_\beta^{-1}(\lambda))$):

$$\begin{aligned}
& \sqrt{n}\binom{n-1}{r-1} G_\beta(x)^{h_n-1}(1-G_\beta(x))^{n-h_n} = \\
= \;& n^{1/2}\frac{(n-1)!}{(h_n-1)!(n-h_n+1)!} G_\beta(x)^{h_n-1}[1-G_\beta(x)]^{n-h_n+1} \\
= \;& n^{1/2}\frac{\sqrt{2\pi(n-1)}\left(\frac{n-1}{e}\right)^{n-1}\cdot\left(1+\mathcal{O}\left(\frac{1}{12(n-1)}\right)\right)}{\sqrt{2\pi(h_n-1)}\left(\frac{h_n-1}{e}\right)^{h_n-1}\cdot\left(1+\mathcal{O}\left(\frac{1}{12(h_n-1)}\right)\right)} \times \\
& \times\frac{G_\beta(x)^{h_n-1}[1-G_\beta(x)]^{n-h_n+1}}{\sqrt{2\pi(n-h_n+1)}\left(\frac{n-h_n+1}{e}\right)^{n-h_n+1}\cdot\left(1+\mathcal{O}\left(\frac{1}{12(n-h_n+1)}\right)\right)} \\
= \;& \frac{\frac{1}{e}\sqrt{2\pi}\cdot G_\beta(x)^{h_n-1}[1-G_\beta(x)]^{n-h_n+1}\cdot(1+o(1))}{\sqrt{2\pi\frac{h_n-1}{n-1}}\left(\frac{h_n-1}{n-1}\right)^{h_n-1}\cdot\sqrt{2\pi\frac{n-h_n+1}{n}}\left(\frac{n-h_n+1}{n-1}\right)^{n-h_n+1}} \\
= \;& \frac{(1+o(1))}{e\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left(\frac{G_\beta(x)}{\frac{h_n-1}{n-1}}\right)^{h_n-1}\left(\frac{1-G_\beta(x)}{\frac{n-h_n+1}{n-1}}\right)^{n-h_n+1} \\
= \;& \frac{(1+o(1))}{e\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left(\frac{G_\beta(x)}{\lambda}\right)^{h_n-1}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{n-h_n+1}\cdot\left(\frac{\lambda}{\frac{h_n-1}{n-1}}\right)^{h_n-1}\left(\frac{1-\lambda}{1-\frac{h_n-2}{n-1}}\right)^{n-h_n+1} \\
= \;& \frac{(1+o(1))}{e\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left(\frac{G_\beta(x)}{\lambda}\right)^{h_n-1}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{n-h_n+1}\cdot \\
& \cdot\left(1+\frac{\lambda+na_n-1}{\lambda(n-1)}\right)^{-h_n+1}\left(1+\frac{2-\lambda-na_n}{(1-\lambda)(n-1)}\right)^{-n+h_n-1} \\
\geq \;& \frac{(1+o(1))}{e\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left(\frac{G_\beta(x)}{\lambda}\right)^{h_n-1}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{n-h_n+1} \\
& \cdot\left(1+\frac{1}{\lambda(n-1)}\right)^{-\lambda(n-1)}\left(1+\frac{3}{(1-\lambda)(n-1)}\right)^{-(1-\lambda)(n-1)} \\
= \;& \frac{(1+o(1))}{e^5\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left(\frac{G_\beta(x)}{\lambda}\right)^{h_n-1}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{n-h_n+1} \\
= \;& \frac{(1+o(1))}{e^5\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left[\left(\frac{G_\beta(x)}{\lambda}\right)^{\lambda}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{1-\lambda}\right]^n\cdot\left(\frac{\lambda}{G_\beta(x)}\right)^{a_n-\frac{1}{n}}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{-a_n+\frac{1}{n}} \\
= \;& \frac{(1+o(1))}{e^5\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left[\left(\frac{G_\beta(x)}{\lambda}\right)^{\lambda}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{1-\lambda}\right]^n.
\end{aligned}$$

Similarly, an upper bound for $\frac{g_{\beta,h_n}(x)}{\sqrt{n}}$ can be derived:

$$\frac{g_{\beta,h_n}(x)}{\sqrt{n}} \leq \frac{M_g(1+o(1))}{\sqrt{2\pi\lambda(1-\lambda)}}\cdot\left[\left(\frac{G_\beta(x)}{\lambda}\right)^{\lambda}\left(\frac{1-G_\beta(x)}{1-\lambda}\right)^{1-\lambda}\right]^n.$$

In the next step, we employ Taylor's expansion of functions $G_\beta(x)$ and $(1+x)^\lambda$ to analyze the behavior of $g_{\beta,h_n}(x)$ is a neighborhood of $G_\beta^{-1}(\lambda)$. The Taylor expansions $G_\beta(G_\beta^{-1}(\lambda) - x) =$

$\lambda - g_\beta(\xi)x$ and $(1+x)^\lambda = 1 + \lambda x + \frac{1}{2}\lambda(\lambda-1)\zeta$ ($\xi \in (G_\beta^{-1}(\lambda) - x, G_\beta^{-1}(\lambda))$ and $\zeta \in (0,x)$) allow us to modify the bounds for $\frac{g_{\beta,h_n}(x)}{\sqrt{n}}$ as follows ($x \in (G_\beta^{-1}(\lambda) - \varepsilon, G_\beta^{-1}(\lambda))$, $\frac{1}{2} < \lambda < 1$):

$$\begin{aligned}
&\left(\frac{G_\beta(G_\beta^{-1}(\lambda)-x)}{\lambda}\right)^\lambda \left(\frac{1-G_\beta(G_\beta^{-1}(\lambda)-x)}{1-\lambda}\right)^{1-\lambda} = \\
= &\left(1 - \frac{g_\beta(\xi)}{\lambda}x\right)^\lambda \left(1 + \frac{g_\beta(\xi)}{1-\lambda}x\right)^{1-\lambda} \\
= &\left(1 - g_\beta(\xi)x + \frac{1}{2}\frac{\lambda-1}{\lambda}(1-\zeta)^{\lambda-2}g_\beta^2(\xi)x^2\right)\left(1 + g_\beta(\xi)x + \frac{1}{2}\frac{\lambda}{1-\lambda}(1+\zeta')^{-\lambda-1}g_\beta^2(\xi)x^2\right) \\
= &\ 1 - g_\beta^2(\xi)x^2 + \frac{1}{2}g_\beta^2(\xi)x^2\left[\frac{\lambda-1}{\lambda}(1-\zeta)^{\lambda-2} + \frac{\lambda}{1-\lambda}(1+\zeta')^{-\lambda-1}\right] + \mathcal{O}\left(x^3\right) \\
\geq &\ 1 - g_\beta^2(\xi)x^2 + \mathcal{O}\left(x^3\right) \geq 1 - \frac{3}{2}g_\beta^2(\xi)x^2 \geq 1 - \frac{1}{2}M_g^2x^2
\end{aligned}$$

because of assumptions on ε, and similarly

$$\left(\frac{G_\beta(G_\beta^{-1}(\lambda)-x)}{\lambda}\right)^\lambda \left(\frac{1-G_\beta(G_\beta^{-1}(\lambda)-x)}{1-\lambda}\right)^{1-\lambda} \leq 1 - \frac{1}{2}g_\beta^2(\xi)x^2 \leq 1 - \frac{1}{2}m_g^2x^2,$$

where $\xi \in (G_\beta^{-1}(\lambda) - x, G_\beta^{-1}(\lambda))$ and $\lambda\zeta, (1-\lambda)\zeta' \in (0, g_\beta(\xi)x)$. Having these results in hand, we can estimate the last integral in (56) from above

$$M_g \cdot \int_0^\varepsilon y \cdot \sup_\beta g_{\beta,h_n}(G_\beta^{-1}(\lambda) - y)\mathrm{d}y \leq \sqrt{n}M_g^2\frac{(1+o(1))}{\sqrt{2\pi\lambda(1-\lambda)}} \cdot \int_0^\varepsilon y \cdot (1 - \frac{1}{2}m_g^2y^2)^n\mathrm{d}y$$

and similarly from below. As

$$\begin{aligned}
\int_0^\varepsilon y \cdot (1 - \frac{1}{2}m_g^2y^2)^n\mathrm{d}y &= \frac{1}{m_g^2}\int_{1-\frac{1}{2}m_g^2\varepsilon^2}^1 u^n\mathrm{d}u \\
&= \left[\frac{1}{m_g^2}\frac{u^{n+1}}{n+1}\right]_{1-\frac{1}{2}m_g^2\varepsilon^2}^1 \\
&= \frac{1}{m_g^2(n+1)}\left[1 - \left(1 - \frac{1}{2}m_g^2\varepsilon^2\right)^{n+1}\right],
\end{aligned}$$

it follows that

$$\begin{aligned}
P(\{\omega \in \Omega_3 | \exists\beta : \upsilon_{in}(\beta) = -1\}) &= P\left(\exists\beta : r_{[h_n]}^2(\beta,\omega') = r_{[h_n]}^2(\beta,\omega) < r_1^2(\beta,\omega_1) \leq G_\beta^{-1}(\lambda)\right) \\
&= \mathcal{O}\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

Thus, we finally get the result

$$\begin{aligned}
P(\{\omega \in \Omega^n | \exists\beta : \upsilon_{in}(\beta) = -1\}) &= P(\{\omega \in \Omega_1 | \exists\beta : \upsilon_{in}(\beta) = -1\}) \\
&+ P(\{\omega \in \Omega_2 | \exists\beta : \upsilon_{in}(\beta) = -1\}) + P(\{\omega \in \Omega_3 | \exists\beta : \upsilon_{in}(\beta) = -1\}) \\
&= \mathcal{O}\left(n^{-\frac{1}{2}}\right)
\end{aligned}$$

as $n \to +\infty$. □

Corollary 4 *Under Assumptions D and H,*

$$P\left(\exists\beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta)\right) \neq I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\right) \middle| x_i\right) = \mathcal{O}\left(n^{-\frac{1}{2}}\right)$$

as $n \to +\infty$.

Proof: Since the distribution functions of y_i and $r_i^2(\beta)$ conditional on x_i are absolutely continuous, the proof proceeds along the same lines as the proof of the previous Lemma 4. □

Corollary 5 *Under Assumptions D, H, and NC, for any fixed $i \in \mathbb{N}$ and $n \geq i$*

$$P\Big(\exists \beta \in B : I\Big(r_i^2(\beta) \leq r_{[h_n]}^2(\beta)\Big) \neq I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)\Big) = \mathcal{O}\Big(n^{-\frac{1}{2}}\Big)$$

as $n \to +\infty$, or analogously

$$P\left(\sup_{\beta \in B} \left| I\Big(r_i^2(\beta) \leq r_{[h_n]}^2(\beta)\Big) - I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)\right| \neq 0\right) = \mathcal{O}\Big(n^{-\frac{1}{2}}\Big)$$

as $n \to +\infty$.

Proof: To prove this result, we can follow the proof of the previous Lemma 4, but we have to make sure that all steps are uniformly valid for all $\beta \in B$.

First, equation (54) can be written as

$$\begin{aligned} P\Big(r_{[h_n]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \leq G_\beta^{-1}(\lambda)\Big) \quad &\leq \quad P\Big(r_{[h_n]}^2(\beta,\omega') \leq G_\beta^{-1}(\lambda) - \varepsilon/2\Big) \\ &+ \quad P\Big(G_\beta^{-1}(\lambda) - \varepsilon/2 < r_{[h_n]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \leq G_\beta^{-1}(\lambda)\Big). \end{aligned}$$

Then we have to find out more about two probabilities on the right side of the inequality. Due to Corollary 3, the first probability

$$P\Big(r_{[h_n]}^2(\beta,\omega') \leq G_\beta^{-1}(\lambda) - \varepsilon/2\Big) = \mathcal{O}\big(n^{-1}\big)$$

uniformly over $\beta \in B$ for $n \to \infty$. Next, the second probability can be expressed as

$$\begin{aligned} & P\Big(\exists \beta : G_\beta^{-1}(\lambda) - \varepsilon/2 < r_{[h_n]}^2(\beta,\omega) < r_1^2(\beta,\omega_1) \leq G_\beta^{-1}(\lambda)\Big) \leq \\ = \quad & \int_{\omega' \in \Omega^{n-1}} \int_{\omega_1 \in \Omega} I\Big(G_\beta^{-1}(\lambda) - \varepsilon/2 < r_{[h_n]}^2(\beta,\omega') < r_1^2(\beta,\omega_1) \leq G_\beta^{-1}(\lambda)\Big) dP(\omega_1) dP(\omega') \\ \leq \quad & \int_{\omega' \in \Omega^{n-1}} M_g \cdot \\ & \cdot \sup_\beta \left\{ \left| G_\beta^{-1}(\lambda) - r_{[h_n]}^2(\beta,\omega') \right| \cdot I\Big(G_\beta^{-1}(\lambda) - \varepsilon/2 < r_{[h_n]}^2(\beta,\omega') \leq G_\beta^{-1}(\lambda)\Big) \right\} dP(\omega') \\ = \quad & M_g \cdot \int_{G_\beta^{-1}(\lambda)-\varepsilon/2}^{G_\beta^{-1}(\lambda)} \sup_\beta \left\{ \left| G_\beta^{-1}(\lambda) - x \right| \cdot g_{\beta,h_n}(x) \right\} dx \\ = \quad & M_g \cdot \int_0^{\varepsilon/2} \sup_\beta \left\{ |y| \cdot g_{\beta,h_n}(G_\beta^{-1}(\lambda) - y) \right\} dy \\ \leq \quad & M_g \cdot \int_0^{\varepsilon} y \cdot \sup_\beta \left\{ g_{\beta,h_n}(G_\beta^{-1}(\lambda) - y) \right\} dy. \end{aligned}$$

This second term can be treated in the same way as in the proof of Lemma 4 as long as we are able to find $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that the requirements (a)–(f) on page 32 in Lemma 4 are satisfied uniformly for all $\beta \in B$.

1. Requirement (a) follows from Assumption NC5—$I_G = \inf_{\beta \in B} G_\beta^{-1}(\lambda) > 0$—and NC6—$g_\beta(z)$ is bounded by M_{gg} on (I_G, ∞) uniformly in β over B.

2. Requirement (b) follows from Assumption NC5—$m_{gg} = \inf_{\beta \in B} \inf_{z \in (-\delta,\delta)} g_\beta\left(G_\beta^{-1}(\lambda) + z\right) > 0$—because $G_\beta\left(G_\beta^{-1}(\lambda) - \varepsilon\right) = \lambda - g_\beta(\xi)\varepsilon$.

3. Requirement (c) is equivalent to Assumption NC5—$I_g = \inf_{\beta \in B} \inf_{z \in (-\delta,\delta)} g_\beta \left(G_\beta^{-1}(\lambda) + z \right) > 0$.

4. Requirement (d) and (e) are independent of β, only M_g and m_g are replaced by M_{gg} and m_{gg}.

5. Requirement (f) just requires the existence of the probability density function for any β, so it is satisfied as well.

Hence, the proof can follow along the same lines for all $\beta \in B$ and because the bounds are chosen independently of β, the result holds uniformly in $\beta \in B$. □

Lemma 5 *Let Assumptions D and H hold and $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. Then it holds*

1. *For the conditional probability*

$$\begin{aligned} & P\left(I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \ne I\left(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\right) \middle| x_i \right) \\ = & \left| h_\beta'(x_i, \beta^0)^T (\beta - \beta^0) \right| \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\} + \mathcal{O}_p\left(n^{-\frac{1}{2}}\right) \\ = & \mathcal{O}_p\left(n^{-\frac{1}{4}}\right) \end{aligned}$$

and

$$\begin{aligned} & \mathsf{E}\left(\operatorname{sgn} r_i(\beta^0) \cdot \left(I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) - I\left(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\right) \right) \middle| x_i \right) = \\ & = h_\beta'(x_i, \beta^0)^T (\beta - \beta^0) \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\} + \mathcal{O}_p\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

as $n \to +\infty$.

2. *For the corresponding unconditional probability*

$$\begin{aligned} & P\left(I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \ne I\left(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\right) \right) \\ = & \mathsf{E}_x \left| h_\beta'(x_i, \beta^0)^T (\beta - \beta^0) \right| \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\} + \mathcal{O}\left(n^{-\frac{1}{2}}\right) \\ = & \mathcal{O}\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

as $n \to +\infty$.

3. *For the conditional probability of supremum over β*

$$\begin{aligned} & P\left(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \ne I\left(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\right) \middle| x_i \right) \\ = & n^{-\frac{1}{2}} M \cdot \sum_{j=1}^{p} \left| h_{\beta_j}'(x_i, \beta^0) \right| \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\} + \mathcal{O}_p\left(n^{-\frac{1}{2}}\right) \\ = & \mathcal{O}_p\left(n^{-\frac{1}{4}}\right) \end{aligned}$$

as $n \to +\infty$.

4. *For the corresponding unconditional probability of supremum over β*

$$\begin{aligned} & P\left(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : I\left(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\right) \ne I\left(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\right) \right) \\ = & n^{-\frac{1}{2}} M \cdot \sum_{j=1}^{p} \mathsf{E}_x \left| h_{\beta_j}'(x_i, \beta^0) \right| \cdot \left\{ f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right) \right\} + \mathcal{O}\left(n^{-\frac{1}{2}}\right) \\ = & \mathcal{O}\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

as $n \to +\infty$.

Proof: Let us define

$$\nu_i(\beta) = I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big) - I\Big(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\Big).$$

We have to compute

$$P\Big(\Big|I\Big(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\Big) - I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big)\Big| = 1\Big|\, x_i\Big) = P(|\nu_i(\beta)| = 1|x_i)$$

and to prove that the corresponding unconditional probability is (asymptotically) linear in $\beta - \beta^0$. In addition to that, we shall estimate

$$P\left(\sup_{\beta \in U(\beta^0, n^{-\frac{1}{2}}M)} |\nu_i(\beta)| = 1 \,\middle|\, x_i\right) = P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : |\nu_i(\beta)| = 1\Big|\, x_i\Big)$$

(these two probabilities are equivalent because the supremum is always attained—$|\nu_i(\beta)|$ can be only zero or one).

1. First, let us compute $P(\nu_i(\beta) = -1|x_i)$. In the following derivations, it is necessary to keep in mind that we consider all $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ so that most of the results can be used later when $P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : \nu_i(\beta) = -1\Big|\, x_i\Big)$ is estimated. Apparently, $\nu_i(\beta) = -1$ if and only if

$$r_i^2(\beta) > r_{[h_n]}^2(\beta) \quad \text{and} \quad r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0).$$

It holds that

$$r_i^2\big(\beta^0\big) \le r_{[h_n]}^2(\beta^0) \Rightarrow r_i(\beta^0) \in \big(-r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0)\big) \tag{57}$$

and

$$r_i^2(\beta) > r_{[h_n]}^2(\beta) \Rightarrow r_i(\beta) \in \big(-\infty, -r_{[h_n]}(\beta)\big) \cup \big(r_{[h_n]}(\beta), +\infty\big)\,. \tag{58}$$

By means of the Taylor expansion we can write (for a given $\omega \in \Omega$)

$$\begin{aligned} r_i(\beta) &= (y_i - h(x_i, \beta)) \\ &= \big(y_i - h\big(x_i, \beta^0\big)\big) - h_\beta^{'}(x_i, \xi)^T(\beta - \beta^0) \\ &= r_i(\beta^0) - h_\beta^{'}(x_i, \xi)^T(\beta - \beta^0) \\ &= r_i(\beta^0) - \Delta_h(x_i, \beta) \end{aligned}$$

where $\xi \in \big[\beta^0, \beta\big]_{\mathscr{X}}$ and $\Delta_h(x_i, \beta) = h_\beta^{'}(x_i, \xi)^T(\beta - \beta^0)$ are introduced for the simplicity of notation ($[\cdot]_{\mathscr{X}}$ denotes a convex span, see Section 3.1). Taking this result into account, assertions (57) and (58) imply that

$$r_i(\beta^0) \in \big(-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta) + \Delta_h(x_i, \beta)\big) \cup \big(r_{[h_n]}(\beta) + \Delta_h(x_i, \beta), r_{[h_n]}(\beta^0)\big)\,, \tag{59}$$

where the convention $(a, b) = \emptyset$ if $b < a$ is used. For $\nu_i(\beta) = 1$, it is possible to derive analogously

$$r_i(\beta^0) \in \big(-r_{[h_n]}(\beta) + \Delta_h(x_i, \beta), -r_{[h_n]}(\beta^0)\big) \cup \big(r_{[h_n]}(\beta^0), r_{[h_n]}(\beta) + \Delta_h(x_i, \beta)\big)\,. \tag{60}$$

Given results (59) and (60), we can write $P(|\nu_i(\beta)| = 1|x_i)$ as

$$P\big(r_i(\beta^0) \in \big[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta) + \Delta_h(x_i, \beta)\big]_{\mathscr{X}} \cup \big[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta) + \Delta_h(x_i, \beta)\big]_{\mathscr{X}}\big|\, x_i\big). \tag{61}$$

Lemma 10 allows us to simplify this expression even further:

$$P\big(r_i(\beta^0) \in \big[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\big]_{\mathscr{X}} \cup \big[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\big]_{\mathscr{X}}\big|\, x_i\big) + o_p\Big(n^{-\frac{1}{2}}\Big) \tag{62}$$

as $n \to +\infty$. Please, notice that, conditionally on x_i, $\nu_i(\beta) \neq 0$ implies $\operatorname{sgn} r_i(\beta^0) \cdot \nu_i(\beta) = \operatorname{sgn} \Delta_h(x_i, \beta)$ with probability approaching 1 as $1 - \mathcal{O}\left(\frac{1}{n}\right)$ with $n \to \infty$. First, $\Delta_h(x_i, \beta)$ is given by x_i and β, and for a given x_i, it is bounded $\Delta_h(x_i, \beta) = h'_\beta(x_i, \xi)^T(\beta - \beta^0) \leq \mathcal{O}(1)(\beta - \beta^0)$ and converges to zero for $\beta \to \beta^0$. So we can choose, for example, $n_0 \in \mathbb{N}$ such that $\Delta_h(x_i, \beta) < \sqrt{\frac{1}{2}G^{-1}(\lambda)}$ for all $n > n_0$ (remember, $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$). Second, Lemma 3 implies that $P\left(r^2_{[h_n]}(\beta^0) < \frac{1}{2}G^{-1}(\lambda)\right) = \mathcal{O}\left(\frac{1}{n}\right)$, and consequently, $P\left(\left|r_{[h_n]}(\beta^0)\right| < \sqrt{\frac{1}{2}G^{-1}(\lambda)}\right) = \mathcal{O}\left(\frac{1}{n}\right)$ as $n \to \infty$. Therefore, we can write with probability higher than $1 - \mathcal{O}\left(\frac{1}{n}\right)$ that for $\Delta_h(x_i, \beta) > 0$ (see (59) and (60))

- $\nu_i(\beta) = 1$ corresponds to $r_i(\beta^0) \in \left(r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right) \subset (0, +\infty)$, thus $v_i(\beta) > 0$ if $r_i(\beta^0) > 0$.
- $\nu_i(\beta) = -1$ corresponds to $r_i(\beta^0) \in \left(-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right) \subset (-\infty, 0)$, thus $v_i(\beta) < 0$ if $r_i(\beta^0) < 0$.

Similarly for the case of $\Delta_h(x_i, \beta) < 0$.

Let us remind that $r_i(\beta^0) \equiv \varepsilon_i$. Moreover, its probability density function $f(x)$ is bounded from above by a positive constant, let us say M_f, and is differentiable due to assumption D3. Therefore, we can write using Lemma 8 ($q_\lambda = \sqrt{G^{-1}(\lambda)}$ is introduced for the sake of simplicity):

$$\begin{aligned} &P\left(r_i(\beta^0) \in \left[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right]_{\not\asymp} \cup \left[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right]_{\not\asymp} \middle| x_i\right) \\ &\quad = P\left(r_i(\beta^0) \in \left[-q_\lambda - \xi_1, -q_\lambda - \xi_1 + \Delta_h(x_i, \beta)\right]_{\not\asymp} \cup \left[q_\lambda + \xi_1, q_\lambda + \xi_1 + \Delta_h(x_i, \beta)\right]_{\not\asymp} \middle| x_i\right), \end{aligned}$$

where ξ_1 and ξ_2 are random variables behaving like $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Taylor's expansion for the distribution function of ε_i further implies (remember, $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$)

$$\begin{aligned} & P\left(r_i(\beta^0) \in \left[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right]_{\not\asymp} \cup \left[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i, \beta)\right]_{\not\asymp} \middle| x_i\right) \\ = \;& \left|\Delta_h(x_i, \beta)\right| \cdot \left\{f_\varepsilon\left(-q_\lambda\right) + f_\varepsilon\left(q_\lambda\right) + f'_\varepsilon(\xi_3) \cdot \left(\Delta_h(x_i, \beta) + \xi_1\right) + f'_\varepsilon(\xi_4) \cdot \left(\Delta_h(x_i, \beta) + \xi_2\right)\right\} \\ = \;& \left|h'_\beta(x_i, \xi)^T(\beta - \beta^0)\right| \cdot \left\{f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right)\right\} + \mathcal{O}_p\left(n^{-1}\right) \\ = \;& \left|h'_\beta(x_i, \beta^0)^T(\beta - \beta^0)\right| \cdot \left\{f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right)\right\} + \mathcal{O}_p\left(n^{-\frac{1}{2}}\right), \end{aligned}$$

where the last step uses Taylor's expansion of the first derivative of $h(x, \beta)$ at point β^0:

$$h'_\beta(x_i, \xi) = h'_\beta\left(x_i, \beta^0\right) + h''_{\beta\beta}(x_i, \zeta)(\xi - \beta^0).$$

Hence, the first assertion of part 1 is proved—the inequality

$$P\left(I\left(r_i^2\left(\beta^0\right) \leq r^2_{[h_n]}(\beta^0)\right) \neq I\left(r_i^2(\beta) \leq r^2_{[h_n]}(\beta)\right) \middle| x_i\right) \leq \mathcal{O}_p\left(n^{-\frac{1}{4}}\right)$$

follows from assumptions D1 ($\max_i \|x_i\| = \mathcal{O}_p\left(n^{\frac{1}{4}}\right)$), H3, and the fact that $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. The second assertion follows immediately from the note explaining that $\operatorname{sgn} r_i(\beta^0) \cdot \nu_i(\beta) = \operatorname{sgn} \Delta_h(x_i, \beta)$ with probability higher than $1 - \mathcal{O}\left(\frac{1}{n}\right)$.

2. Next, we shall evaluate the corresponding unconditional probability, that is the expectation of $P(|\nu_i(\beta)| = 1|x_i)$ over x_i, and check its asymptotic linearity in $\beta - \beta^0$. As

$$\begin{aligned} & P\left(I\left(r_i^2\left(\beta^0\right) \leq r^2_{[h_n]}(\beta^0)\right) \neq I\left(r_i^2(\beta) \leq r^2_{[h_n]}(\beta)\right)\right) \\ = \;& \mathsf{E}_x\, P(|\nu_i(\beta)| = 1|x_i) \\ = \;& \mathsf{E}_x \left|h'_\beta\left(x_i, \beta^0\right)^T(\beta - \beta^0)\right| \cdot \left\{f\left(-\sqrt{G^{-1}(\lambda)}\right) + f\left(\sqrt{G^{-1}(\lambda)}\right)\right\} + \mathcal{O}\left(n^{-\frac{1}{2}}\right), \end{aligned}$$

the result is apparent once we take into account that $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$.

3. We have derived in part 1 of this proof that

$$\begin{aligned} P(|\nu_i(\beta)| = 1|x_i) &= P\big(r_i(\beta^0) \in \big[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i,\beta)\big]_{\varkappa} \cup \\ &\quad \cup \big[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i,\beta)\big]_{\varkappa} \big| \, x_i\big) + o_p\Big(n^{-\frac{1}{2}}\Big) \end{aligned} \tag{63}$$

as $n \to \infty$, where $o_p\Big(n^{-\frac{1}{2}}\Big)$ holds uniformly over all $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$. The length of the intervals in (63) is a function of $\beta - \beta^0$. Further, notice that the lower bound of the interval and $r_i(\beta^0)$ itself does not depend on β, only the length of the interval is β-dependent, and this length converges to zero as $\beta \to \beta^0$ with increasing n. Now, the crucial point here is that the set of events $\omega \in \Omega$ such that a continuously distributed random variable $r_i(\beta^0) \equiv \varepsilon_i$ belongs to such intervals depends purely on the lower and upper bounds of the intervals, and consequently, only on their lengths $h'_\beta(x_i,\xi)^T(\beta - \beta^0)$ in our case. Therefore, the set of events $\omega \in \Omega$ such that there exists $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ for which the continuously distributed random variable ε_i belongs to the interval specified in (63) and the probability of this set reduce to finding the supremum of the length of the interval over all $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$.

Hence, we can write

$$\begin{aligned} & P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : r_i(\beta^0) \in \big[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta^0) + \Delta_h(x_i,\beta)\big]_{\varkappa} \cup \\ & \qquad \cup \big[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta^0) + \Delta_h(x_i,\beta)\big]_{\varkappa} \big| \, x_i\Big) \\ = & \sup_{\beta \in U(\beta^0, n^{-\frac{1}{2}}M)} \Big|h'_\beta\big(x_i,\beta^0\big)^T(\beta - \beta^0)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) \\ = & \; n^{-\frac{1}{2}}M \cdot \sum_{j=1}^{p} \Big|h'_{\beta_j}\big(x_i,\beta^0\big)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) \end{aligned}$$

as $n \to +\infty$. Thus, the third assertion is verified using the same argument as in part 1 of the proof.

4. Finally, we should find the corresponding unconditional probability, that is the expectation of $P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : |\nu_i(\beta)| = 1\Big)$. The assertion is a direct consequence of the fact that

$$\begin{aligned} & \mathsf{E}_x \, P\Big(\exists \beta \in U(\beta^0, n^{-\frac{1}{2}}M) : |\nu_i(\beta)| = 1 \Big| \, x_i\Big) \\ = & \; n^{-\frac{1}{2}}M \cdot \sum_{j=1}^{p} \mathsf{E}_x \Big|h'_{\beta_j}\big(x_i,\beta^0\big)\Big| \cdot \Big\{f\Big(-\sqrt{G^{-1}(\lambda)}\Big) + f\Big(\sqrt{G^{-1}(\lambda)}\Big)\Big\} + \mathcal{O}\Big(n^{-\frac{1}{2}}\Big) \end{aligned}$$

(note that $\mathsf{E}_x \, \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big) = \mathcal{O}\Big(n^{-\frac{1}{2}}\Big)$ because we compute here an expectation of probability, that is, an expectation of a random variable with a bounded domain). □

Corollary 6 *Assume that $\lambda \in (\frac{1}{2}, 1\rangle$ and there exists a $\beta \in U(\beta^0, n^{-\frac{1}{2}}M)$ such that*

$$I\Big(r_i^2(\beta^0) \le r_{[h_n]}^2(\beta^0)\Big) \neq I\Big(r_i^2(\beta) \le r_{[h_n]}^2(\beta)\Big).$$

Then for any $\varepsilon > 0$ we can find a constant $K > 0$ such that

$$\Big|\big|r_i(\beta^0)\big| - \sqrt{G^{-1}(\lambda)}\Big| = \Big|r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot \sqrt{G^{-1}(\lambda)}\Big| < n^{-\frac{1}{4}}K$$

for all n large enough with probability larger than $1 - \varepsilon$, that is $r_i(\beta^0) - \operatorname{sgn} r_i(\beta^0) \cdot \sqrt{G^{-1}(\lambda)} = \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big)$ as $n \to +\infty$..

Proof: In the proof of Lemma 5 (see (59)–(62)), we have shown that

$$\nu_i(\beta) = I\Big(r_i^2(\beta) \leq r_{[h_n]}^2(\beta)\Big) - I\Big(r_i^2\big(\beta^0\big) \leq r_{[h_n]}^2(\beta^0)\Big)$$

can be non-zero for a given x_i if and only if

$$\begin{aligned} & r_i(\beta^0) \in \left[-r_{[h_n]}(\beta^0), -r_{[h_n]}(\beta) + \Delta_h(x_i,\beta)\right]_{\varkappa} \cup \left[r_{[h_n]}(\beta^0), r_{[h_n]}(\beta) + \Delta_h(x_i,\beta)\right]_{\varkappa} \Longleftrightarrow \\ & \Longleftrightarrow r_i(\beta^0) \in \left[-\sqrt{G^{-1}(\lambda)} - \xi_1, -q_\lambda - \xi_1 + \Delta_h(x_i,\beta)\right]_{\varkappa} \cup \left[\sqrt{G^{-1}(\lambda)} + \xi_1, q_\lambda + \xi_1 + \Delta_h(x_i,\beta)\right]_{\varkappa}, \end{aligned}$$

where ξ_1 and ξ_2 are random variables behaving like $\mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big)$ and $\Delta_h(x_i,\beta) = h'_\beta(x_i,\xi)^T(\beta-\beta^0) = \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big)$ due to assumption D2 and H3. Hence,

$$\left|\left|r_i(\beta^0)\right| - \sqrt{G^{-1}(\lambda)}\right| = \mathcal{O}_p\Big(n^{-\frac{1}{4}}\Big)$$

as $n \to +\infty$. □

Lemma 6 *Let Assumptions D and H hold and assume that $t(x,\varepsilon;\beta)$ is a real-valued function continuous in β uniformly in x and ε over any compact subset of the support of (x,ε). Moreover, assume that Assumptions NC hold for $t(x,\varepsilon;\beta)$. Finally, let $r_i(\beta) = \varepsilon_i + h\big(x_i,\beta^0\big) - h(x_i,\beta)$ and G_β denote the distribution function of $r_i^2(\beta)$ (for any $\beta \in B$). Then*

$$\sup_{\beta\in B}\left|\frac{1}{n}\sum_{i=1}^{n}\Big[t(x_i,\varepsilon_i;\beta)\cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)\Big] - \mathsf{E}\Big[t(x_i,\varepsilon_i;\beta)\cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)\Big]\right| \to 0$$

as $n \to +\infty$ almost surely.

Proof: This result is nothing but an application of the uniform law of large numbers for nonlinear models and we use here its variant due to Andrews (1987). Therefore, we just have to verify that the assumptions of the uniform law of large numbers are satisfied. We verify here assumptions A1, B1, B2, and A3, and employ them together with Corollary 1 (see Andrews (1987)). To do so, let us follow the notation used in Andrews(1987) and denote

$$q(x,\varepsilon;\beta) = t(x,\varepsilon;\beta)\cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big) = t(x,\varepsilon;\beta)\cdot I\Big(\big[\varepsilon + h\big(x,\beta^0\big) - h(x,\beta)\big]^2 \leq G_\beta^{-1}(\lambda)\Big)$$

and

$$\begin{aligned} q_*(x,\varepsilon;\beta,\rho) &= \inf_{\beta'\in U(\beta,\rho)} q(x,\varepsilon;\beta'), \\ q^*(x,\varepsilon;\beta,\rho) &= \sup_{\beta'\in U(\beta,\rho)} q(x,\varepsilon;\beta'), \\ q(x,\varepsilon) &= \sup_{\beta\in B}\left|q(x,\varepsilon;\beta)\right|. \end{aligned}$$

Assumption A1 B is compact metric space: this is satisfied because of Assumption NC1.

Assumption B1 (x_i,ε_i) should be a sequence of strongly mixing random variables with mixing numbers $\alpha(s), s = 1,2,\ldots$, that satisfy $\alpha(s) = o(s^{-\alpha/(\alpha-1)})$ as $s \to \infty$ for some $\alpha \geq 1$: this condition of asymptotic weak dependence is satisfied for $\alpha = 1$, because (x_i,ε_i) are independent random vectors in our case (Assumption D1).

Assumption B2, part a $q^*(x_i,\varepsilon_i;\beta,\rho), q_*(x_i,\varepsilon_i;\beta,\rho)$, and $q(x_i,\varepsilon_i)$ are random variables and $q^*(\cdot,\cdot;\beta,\rho), q_*(\cdot,\cdot;\beta,\rho)$ are measurable functions for all $i \in \mathbb{N}$, all $\beta \in B$, and all ρ sufficiently small: this follows from Assumption NC2 and the fact that $(x_i,\varepsilon_i), i = 1,\ldots,n$, is a sequence of identically distributed random variables.

Assumption B2, part b $\mathsf{E}\, q(x_i,\varepsilon_i)^{1+\delta} < \infty$ for some $\delta > 0$: this follows from Assumption NC3 and the fact that (x_i,ε_i) is a sequence of identically distributed random variables.

Assumption A3 For all $\beta \in B$,

$$\lim_{\rho\to 0} |\mathsf{E}\, q^*(x_i,\varepsilon_i;\beta,\rho) - \mathsf{E}\, q(x_i,\varepsilon_i;\beta)| = 0 \quad \text{and} \quad \lim_{\rho\to 0} |\mathsf{E}\, q_*(x_i,\varepsilon_i;\beta,\rho) - \mathsf{E}\, q(x_i,\varepsilon_i;\beta)| = 0. \tag{64}$$

Without loss of generality, we will prove this result for $i = 1$ and only for supremum q^*(the other part can be proved analogously). By definition of $q^*(x_1,\varepsilon_1;\beta,\rho)$,

$$q^*(x_1,\varepsilon_1;\beta,\rho) = \sup_{\beta'\in U(\beta,\rho)} t(x_1,\varepsilon_1;\beta')\cdot I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) \tag{65}$$

$$= t(x_1,\varepsilon_1;\beta)\cdot I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big) \tag{66}$$

$$+ \sup_{\beta'\in U(\beta,\rho)} t(x_1,\varepsilon_1;\beta)\cdot\Big[I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) - I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big)\Big] \tag{67}$$

$$+ \sup_{\beta'\in U(\beta,\rho)} [t(x_1,\varepsilon_1;\beta') - t(x_1,\varepsilon_1;\beta)]\cdot I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big). \tag{68}$$

Hence, to verify (64) we just need to show that the expectations of (67) and (68) converge to zero for $\rho \to 0$.

1. Let us start with (67). First, observe that

$$\sup_{\beta\in B}\left\{t(x_1,\varepsilon_1;\beta)\cdot\Big[I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) - I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big)\Big]\right\} \le$$

$$\le \sup_{\beta\in B}|t(x_1,\varepsilon_1;\beta)|\cdot\sup_{\beta'\in U(\beta,\rho)}\left|I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) - I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big)\right|,$$

where $\sup_{\beta\in B}|t(x_1,\varepsilon_1;\beta)|$ is a function independent of β and with a finite expectation. Because the difference $\left|I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) - I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big)\right|$ is always lower or equal to one, (67) has an integrable majorant independent of β. Therefore, if we show that the probability

$$\lim_{\rho\to 0} P\left(\sup_{\beta'\in U(\beta,\rho)}\left|I\Big(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\Big) - I\Big(r_1^2(\beta) \le G_{\beta}^{-1}(\lambda)\Big)\right| = 1\right) = 0,$$

it implies, that the expectation of (67) converges to zero for $\rho \to 0$ as well.

Second, let us derive an intermediate result regarding the convergence of distribution function $G_{\beta'}$ to G_β. Note that $r_1^2(\beta',\omega) \to r_1^2(\beta,\omega)$ for $\beta' \to \beta$ and any $\omega \in \Omega$ (convergence almost surely). So, for $\beta' \to \beta$ and $G_\beta(x)$ being the cumulative distribution function of $r_1^2(\beta)$, it holds that $G_{\beta'}(x) \to G_\beta(x)$ for all $x \in \mathbb{R}$ (convergence in distribution) because $G_\beta(x)$ is an absolutely continuous distribution function. The same is valid about convergence of $r_1(\beta',\omega)$ to $r_1(\beta,\omega)$ and their corresponding distribution functions $F_{\beta'}(x)$ and $F_\beta(x)$. Now, we show that this convergence of distribution functions $F_{\beta'}(x) \to F_\beta(x)$ (and consequently of $G_{\beta'}(x) \to G_\beta(x)$—$G_\beta(x) = F_\beta(\sqrt{x}) - F_\beta(-\sqrt{x})$) is uniform over all sequences $\beta_n^{'} \to \beta$ such that $\beta_n^{'} \in U(\beta,\rho_n)$, where ρ_n is a sequence of positive numbers such that $\rho_n \to 0$. The reason is that $r_1(\beta',\omega) = r_1(\beta,\omega) - h_\beta^{'}(x_1,\xi)\cdot(\beta'-\beta)$, where the second term $h_\beta^{'}(x_1,\xi)\cdot(\beta'-\beta)$ can be bounded by a random variable $h_x(x_1)\cdot K\cdot(\beta'-\beta) = \mathcal{O}_p(1)\cdot\mathcal{O}\Big(n^{-\frac{1}{2}}\Big) = \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big)$ (see Assumption H3). Because G_β is absolutely continuous, the same is true about convergence of $G_{\beta'}^{-1}(\lambda)$ to $G_\beta^{-1}(\lambda)$—it is uniform over all sequences $\beta_n^{'} \to \beta$ such that $\beta_n^{'} \in U(\beta,\rho_n)$.

Third, let us choose now an arbitrary, but fixed $\varepsilon > 0$. Then we can find $n_1 \in \mathbb{N}$ such that $\left|G_{\beta'}^{-1}(\lambda) - G_\beta^{-1}(\lambda)\right| < \frac{\varepsilon}{8M_{gg}}$ for any $\beta' \in U(\beta,\rho^1)$, where $\rho^1 = \rho_{n_1}$ and M_{gg} is the uniform

upper bound for the probability density functions of $r_1^2(\beta)$ over all $\beta \in B$ (see Assumption NC6). Further, $r_1^2(\beta') = r_1^2(\beta) + r_1(\beta) \cdot h_\beta^{'}(x_1, \xi) \cdot (\beta' - \beta)$, where $\xi \in [\beta, \beta']_\varkappa$. So, we can find $n_2 \in \mathbb{N}$ and $\rho^2 = \rho_{n_2}$ such that $r_1(\beta) \cdot h_\beta^{'}(x_1, \xi) \cdot (\beta' - \beta) < \frac{\varepsilon}{8M_{gg}}$ with probability greater than $1 - \frac{\varepsilon}{2}$ ($r_1(\beta) \cdot h_\beta^{'}(x_1, \xi) = \mathcal{O}_p(1)$). Hence, setting $\rho^\varepsilon = \min\{\rho^1, \rho^2\}$,

$$
\begin{aligned}
& P\left(\sup_{\beta' \in U(\beta, \rho^\varepsilon)} \left| I\left(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right) \right| = 1 \right) \\
& \le \frac{\varepsilon}{2} + P\left(r_1^2(\beta) \in \left(G_\beta^{-1}(\lambda) - \frac{\varepsilon}{4M_{gg}}, G^{-1}(\lambda) + \frac{\varepsilon}{4M_{gg}} \right) \right) \\
& \le \frac{\varepsilon}{2} + \frac{2\varepsilon}{4M_{gg}} \cdot M_{gg} = \varepsilon,
\end{aligned}
$$

because M_{gg} is the uniform upper bound for the probability density functions of $r_1^2(\beta)$ over all $\beta \in B$. Thus, we have shown that for any $\varepsilon > 0$ we can find $\rho^\varepsilon > 0$ such that

$$
P\left(\sup_{\beta' \in U(\beta, \rho^\varepsilon)} \left| I\left(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right) \right| = 1 \right) < \varepsilon
$$

and thus

$$
\lim_{\rho \to 0} P\left(\sup_{\beta' \in U(\beta, \rho)} \left| I\left(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right) \right| = 1 \right) = 0.
$$

We have verified that the expectation of (67) converges to zero for $\rho \to 0$.
2. We should deal now with (68) and prove that

$$
\begin{aligned}
\left| \lim_{\rho \to 0} \mathsf{E} \left\{ \sup_{\beta' \in U(\beta, \rho)} [t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)] \cdot I\left(r_1^2(\beta') \le G_{\beta'}^{-1}(\lambda)\right) \right\} \right| & \le \\
\lim_{\rho \to 0} \mathsf{E} \left\{ \sup_{\beta' \in U(\beta, \rho)} |t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)| \right\} & = 0.
\end{aligned}
$$

First, note that the difference

$$
|t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)| \le |t(x_1, \varepsilon_1; \beta')| + |t(x_1, \varepsilon_1; \beta)| \le 2 \sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)|
$$

can be bounded from above by a function independent of β and having a finite expectation. Let $2\,\mathsf{E} \sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)| = U_E$.
Second, for an arbitrary fixed $\varepsilon > 0$, we can find a compact subset A_ε of the support of (x_1, ε_1) (and its complement $\overline{A_\varepsilon}$) such that $P((x_1, \varepsilon_1) \in A_\varepsilon) > 1 - \frac{\varepsilon}{2U_E}$ (both x_1 and ε_1 are random variables with finite second moments) and $2 \int_{\overline{A_\varepsilon}} \sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)| < \frac{\varepsilon}{2}$. Given this set A_ε and $\beta \in B$, we can employ continuity of $t(x_1, \varepsilon_1; \beta)$ in β (uniform over all $(x_1, \varepsilon_1) \in A_\varepsilon$) and find an $\rho^\varepsilon > 0$ such that $\sup_{(x_1, \varepsilon_1) \in A_\varepsilon} \sup_{\beta' \in U(\beta, \rho)} |t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)| < \frac{\varepsilon}{2}$. Hence,

$$
\begin{aligned}
\mathsf{E} \left\{ \sup_{\beta' \in U(\beta, \rho)} |t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)| \right\} & \le \int_{\overline{A_\varepsilon}} 2 \sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)| \, \mathrm{d}F_x(x_1) \mathrm{d}F_\varepsilon(\varepsilon_1) \\
& \quad + \int_{A_\varepsilon} \frac{\varepsilon}{2} \mathrm{d}F_x(x_1) \mathrm{d}F_\varepsilon(\varepsilon_1) \\
& \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}
$$

and consequently,

$$\lim_{\rho \to 0} \mathsf{E} \left\{ \sup_{\beta' \in U(\beta, \rho)} |t(x_1, \varepsilon_1; \beta') - t(x_1, \varepsilon_1; \beta)| \right\} = 0.$$

We have verified that the expectation of (68) converges to zero for $\rho \to 0$. Thus, assumption A3, Andrews (1987), is valid as well.

Since we have verified all assumptions needed for the uniform law of large numbers, we can use it for $\frac{1}{n} \sum_{i=1}^{n} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right]$ to get the result of the lemma. □

Lemma 7 *Let Assumptions D and H hold and assume that $t(x, \varepsilon; \beta)$ is a real-valued function continuous in β uniformly in x and ε over any compact subset of the support of (x, ε). Moreover, assume that Assumptions NC hold for $t(x, \varepsilon; \beta)$. Furthermore, let $r_i(\beta) = \varepsilon_i + h\left(x_i, \beta^0\right) - h(x_i, \beta)$ and G_β denote the distribution function of $r_i^2(\beta)$ (for any $\beta \in B$). Finally, let $h_n/n \to \lambda \in \left(\frac{1}{2}, 1\right)$. Then*

$$\sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) \right] - \mathsf{E} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right| \to 0$$

as $n \to +\infty$ in probability.

Proof: Using the result of Lemma 6 (the assumptions of this lemma and Lemma 6 are identical) and

$$\begin{aligned}
& \sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) \right] - \mathsf{E} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right| \\
\leq & \sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] - \mathsf{E} \left[t(x_i, \varepsilon_i; \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right| \\
+ & \sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} t(x_i, \varepsilon_i; \beta) \cdot \left[I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) - I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right|,
\end{aligned}$$

we just have to prove that

$$\sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} t(x_i, \varepsilon_i; \beta) \cdot \left[I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) - I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right| \to 0$$

in probability for $n \to \infty$. The Chebyshev inequality for non-negative random variables—$P(X \geq K) \leq \mathsf{E}\, X/K$—implies for a sequence of non-negative random variables X_n that if expectations $\mathsf{E}\, X_n$ converges to zero for $n \to \infty$, then the sequence X_n converges to 0 in probability. So, we will derive now that

$$\begin{aligned}
& \mathsf{E} \left\{ \sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} t(x_i, \varepsilon_i; \beta) \cdot \left[I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) - I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right] \right| \right\} \\
= & \mathsf{E} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sup_{\beta \in B} |t(x_i, \varepsilon_i; \beta)| \cdot \sup_{\beta \in B} \left| I\left(r_i^2(\beta) \leq r_{[h_n]}^2(\beta) \right) - I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right| \right\} \\
= & \mathsf{E} \left\{ \sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)| \cdot \sup_{\beta \in B} \left| I\left(r_1^2(\beta) \leq r_{[h_n]}^2(\beta) \right) - I\left(r_1^2(\beta) \leq G_\beta^{-1}(\lambda) \right) \right| \right\} \to 0
\end{aligned}$$

converges to zero for $n \to \infty$. Since we assume that $\sup_{\beta \in B} |t(x_1, \varepsilon_1; \beta)|$ has a finite first moment, all we have to actually prove is

$$\mathsf{E}\left\{\sup_{\beta \in B}\left|I\left(r_1^2(\beta) \le r_{[h_n]}^2(\beta)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right)\right|\right\} \quad =$$

$$P\left(\sup_{\beta \in B}\left|I\left(r_1^2(\beta) \le r_{[h_n]}^2(\beta)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right)\right| = 1\right) \quad \to \quad 0$$

for $n \to \infty$. But this is the claim of Corollary 5:

$$P\left(\sup_{\beta \in B}\left|I\left(r_1^2(\beta) \le r_{[h_n]}^2(\beta)\right) - I\left(r_1^2(\beta) \le G_\beta^{-1}(\lambda)\right)\right| = 1\right) = \mathcal{O}\left(n^{-\frac{1}{2}}\right)$$

as $n \to \infty$. □

A.2 Fundamental results for order statistics

The assertions presented in this section describe some fundamental properties of order statistics of regression residuals. The formulation of these assertions and their proofs resulted from personal communication with the author of Víšek (1999a) and I generalized it to the nonlinear setup used in this paper (the notation is kept close to that of Víšek).

Lemma 8 *Let $\lambda \in \left(\frac{1}{2}, 1\right\rangle$ and put $h_n = [\lambda n]$ for $n \in \mathbb{N}$. Under assumption D3 on the distribution function $G(x)$ of $r_i^2(\beta^0)$ and its probability density function $g(x)$, it holds that for some $\delta > 0$ there is $L_g > 0$ such that*

$$\inf_{x \in (G^{-1}(\lambda)-\delta, G^{-1}(\lambda)+\delta)} g(x) > L_g > 0.$$

Then for any $\varepsilon \in (0,1)$ there is $K_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that for all $n > n_\varepsilon$

$$P\left(\left|r_{[h_n]}^2\left(\beta^0\right) - G^{-1}(\lambda)\right| < n^{-\frac{1}{2}} \cdot K_\varepsilon\right) > 1 - \varepsilon,$$

that is, $\left|r_{[h_n]}^2\left(\beta^0\right) - G^{-1}(\lambda)\right| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ as $n \to +\infty$.

Proof: First, let us recall that $r_i^2(\beta^0) \equiv \varepsilon_i^2$. Let us take a fixed $\varepsilon \in (0,1)$ and set K_ε so that $P(X \in (-K_\varepsilon, K_\varepsilon)) > 1 - \frac{\varepsilon}{2}$, whereby the random variable X has the standard normal distribution. Then put $K_1 = K_\varepsilon \cdot L_g^{-1}$ and denote

$$v_{1i} = I\left(r_i^2(\beta^0) \le G^{-1}(\lambda) + n^{-\frac{1}{2}} K_1\right).$$

As for all i, $\mathsf{E}\, v_{1i} = P(v_{1i} = 1) = P\left(r_i^2(\beta^0) \le G^{-1}(\lambda) + n^{-\frac{1}{2}} K_1\right) \ge \lambda$—hence, $\mathsf{E}\, v_{1i} \in (\lambda, 1\rangle$ and $\mathsf{var}\, v_{1i} = \mathsf{E}\, v_{1i}(1 - \mathsf{E}\, v_{1i}) = a \in \langle 0, \lambda(1-\lambda))$. In case of $\mathsf{var}\, v_{1i} = 0$, $\sum_{i=1}^n v_{1i} = n \ge h_n$. In all other cases, it is easy to verify the Feller-Lindenberg conditions ($C_n^2 = \sum_{i=1}^n \mathsf{var}\, v_{1i} = an \le \lambda(1-\lambda)n \le \frac{1}{4}n$), so we obtain immediately

$$\mathcal{L}\left\{S_n = \frac{\sum_{i=1}^n (v_{1i} - \mathsf{E}\, v_{1i})}{C_n}\right\} \to N(0,1).$$

Moreover, according to the Glivenko theorem, this convergence in distribution is uniform on $\mathbb{R}$ (the distribution function of $N(0,1)$ is everywhere continuous), so there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$ the distance of distribution functions $\mathcal{L}(S_n)$ and Φ in the supremum norm is lower than $\frac{\varepsilon}{2}$. Consequently,

$$P\left(\left|\frac{\sum_{i=1}^n (v_{1i} - \mathsf{E}\, v_{1i})}{C_n}\right| \le K_\varepsilon\right) > 1 - \varepsilon.$$

Since $C_n \leq \frac{1}{2}\sqrt{n}$, it follows that with probability greater or equal to $1-\varepsilon$

$$-\frac{1}{2}\sqrt{n}\cdot K_\varepsilon + \sum_{i=1}^{n} \mathsf{E}\, v_{1i} \leq \sum_{i=1}^{n} v_{1i}. \tag{69}$$

Taking $n_1 > n_0$ such that $n^{-\frac{1}{2}}K_1 < \delta$ for all $n > n_1$, $\mathsf{E}\, v_{1i} > \lambda + n^{-\frac{1}{2}}K_1 \cdot L_g = \lambda + n^{-\frac{1}{2}}K_\varepsilon$. This result together with equation (69) implies that

$$n\lambda + \frac{1}{2}\sqrt{n}K_\varepsilon = -\frac{1}{2}\sqrt{n}\cdot K_\varepsilon + n\lambda + \sqrt{n}K_\varepsilon < -\frac{1}{2}\sqrt{n}\cdot K_\varepsilon + \sum_{i=1}^{n} \mathsf{E}\, v_{1i} \leq \sum_{i=1}^{n} v_{1i}.$$

But this means that at least $\lambda n \geq h_n$ of residuals $r_i^2(\beta^0)$ are smaller than $G^{-1}(\lambda) + n^{-\frac{1}{2}}K_\varepsilon$; in other words, $r^2_{[h_n]}(\beta^0) \leq G^{-1}(\lambda) + n^{-\frac{1}{2}}K_\varepsilon$. The corresponding lower inequality can be found by repeating these steps for

$$v_{2i} = I\Big(r_i^2(\beta^0) \geq G^{-1}(\lambda) - n^{-\frac{1}{2}}K_1\Big).$$

□

Lemma 9 *Let $\lambda \in \left(\frac{1}{2}, 1\right\rangle$ and put $h_n = [\lambda n]$ for $n \in \mathbb{N}$. Under assumption D3 on the distribution functions $F(x)$ and $G(x)$ of $r_i(\beta^0)$ and $r_i^2(\beta^0)$, respectively, and their probability density functions $f(x)$ and $g(x)$, it holds that for some $\delta > 0$ there is $L_g > 0$ and $U_f > 0$ such that*

$$\inf_{x \in (G^{-1}(\lambda)-\delta, G^{-1}(\lambda)+\delta)} g(x) > L_g > 0 \quad \textit{and} \quad \sup_{x \in (G^{-1}(\lambda)-\delta, G^{-1}(\lambda)+\delta)} f(x) < U_f < \infty.$$

Then for any $\varepsilon \in (0,1)$ there is K_ε and $n_\varepsilon \in \mathbb{N}$ such that uniformly in $t \in \mathcal{T}_M$ for all $n > n_\varepsilon$

$$P\Big(\Big|r^2_{[h_n]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) - G^{-1}(\lambda)\Big| < n^{-\frac{1}{2}}\cdot K_\varepsilon\Big) > 1-\varepsilon,$$

that is, $\Big|r^2_{[h_n]}\Big(\beta^0 - n^{-\frac{1}{2}}t\Big) - G^{-1}(\lambda)\Big| = \mathcal{O}_p\Big(n^{-\frac{1}{2}}\Big)$ uniformly in $t \in \mathcal{T}_M = \{t \in \mathbb{R}^p : \|t\| \leq M\}$ as $n \to +\infty$.

Proof: The proof of this lemma follow the same lines as the proof of Lemma 8. Thus, we will explain in details only those parts that are different from the previous proof.

First, let $n > n_0 = \left[M^2 \cdot \delta^{-2}\right]$ and denote $K_1 = \sup_{\xi \in \left[\beta^0, \beta^0 - n^{-\frac{1}{2}}t\right]_\varkappa} \mathsf{E}\left\|h'_\beta(x_i,\xi)\right\| < \infty$. Sticking to the same choice of an arbitrary, but fixed $\varepsilon \in (0,1)$ and K_ε as in the previous proof, let $K_2 = (\frac{3}{2}K_\varepsilon + M\cdot K_1 \cdot U_f)\cdot L_g^{-1}$ and $q_{\lambda,n} = G^{-1}(\lambda) + n^{-\frac{1}{2}}K_2$. Keeping in mind that (using Taylor's expansion)

$$r_i(\beta^0 - n^{-\frac{1}{2}}t) = r_i(\beta^0) + \Big(h(x_i, \beta^0 - n^{-\frac{1}{2}}t) - h(x_i,\beta^0)\Big) = r_i(\beta^0) + h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t, \tag{70}$$

where $\xi \in \left[\beta^0, \beta^0 - n^{-\frac{1}{2}}t\right]_\varkappa$, consider now

$$w_{1i} = I\Big(r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq G^{-1}(\lambda) + n^{-\frac{1}{2}}K_1\Big).$$

Then we find that for all i

$$\begin{aligned}
\mathsf{E}\, w_{1i} = P(w_{1i}=1) &= P\Big(r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq q_{\lambda,n}\Big) \\
&= P\Big(-\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t \leq e_i \leq +\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\Big) \\
&= \mathsf{E}_x\left(F\left(+\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right) - F\left(-\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right)\right)
\end{aligned}$$

and the Feller-Lindenberg theorem for w_i finally implies an analogy of (69):

$$-\frac{1}{2}\sqrt{n}\cdot K_\varepsilon + \sum_{i=1}^{n} \mathsf{E}\, w_{1i} \le \sum_{i=1}^{n} w_{1i} \tag{71}$$

with probability higher than $1-\varepsilon$.

The next step is to estimate the mean value $\mathsf{E}\, w_{1i} = \mathsf{E}_x\, \mathsf{E}(w_{1i}|x_i)$, once again employing Taylor's theorem:

$$\begin{aligned}
\mathsf{E}(w_{1i}|x_i) &= F\left(+\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right) - F\left(-\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right) \\
&\ge F\left(\sqrt{q_{\lambda,n}}\right) - U_f\cdot\left|h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right| - F\left(-\sqrt{q_{\lambda,n}} - h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right) \\
&\ge G(q_{\lambda,n}) - U_f\cdot\left|h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right| = G(G^{-1}(\lambda)+n^{-\frac{1}{2}}K_2) - U_f\cdot\left|h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right| \\
&\ge \lambda + n^{-\frac{1}{2}}\cdot K_2 L_g - U_f\cdot\left|h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right|,
\end{aligned}$$

hence

$$\mathsf{E}\, w_{1i} \ge \lambda + n^{-\frac{1}{2}}\cdot K_2 L_g - U_f\cdot \mathsf{E}\left|h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t\right|.$$

Returning to (71), we get

$$\begin{aligned}
\sum_{i=1}^{n} w_{1i} &\ge n\lambda + \sqrt{n}\left\{K_2 L_g - \frac{1}{2}K_\varepsilon - U_f\cdot\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left|h'_\beta(x_i,\xi)^T t\right|\right\} \\
&\ge n\lambda + \sqrt{n}\left\{K_2 L_g - \frac{1}{2}K_\varepsilon - U_f\cdot\frac{1}{n}\sum_{i=1}^{n} M\cdot\mathsf{E}\left\|h'_\beta(x_i,\xi)\right\|\right\} \\
&= n\lambda + \sqrt{n}\left\{K_2 L_g - \frac{1}{2}K_\varepsilon - U_f\cdot M\cdot K_1\right\} \\
&= n\lambda + \sqrt{n}K_\varepsilon,
\end{aligned}$$

which indicates again that at least $\lambda n \ge h_n$ of residuals $r_i(\beta^0 - n^{-\frac{1}{2}}t)$ are smaller than $G^{-1}(\lambda) + n^{-\frac{1}{2}}K_\varepsilon$. The corresponding lower inequality can be found by repeating these steps for

$$w_{2i} = I\left(r_i^2\left(\beta^0 - n^{-\frac{1}{2}}t\right) \ge G^{-1}(\lambda) - n^{-\frac{1}{2}}K_1\right).$$

□

Lemma 10 *Let assumption D3 be satisfied. Moreover, let* $\lambda \in \left(\frac{1}{2},1\right\rangle$, $\tau \in \left(\frac{1}{2},1\right)$, *and put* $h_n = [\lambda n]$ *for* $n \in \mathbb{N}$. *Then for any* $\varepsilon \in (0,1)$ *there is* $K_\varepsilon > 0$ *and* $n_\varepsilon \in \mathbb{N}$ *such that uniformly in* $t \in \mathcal{T}_M$ *for all* $n > n_\varepsilon$

$$P\left(\left|r^2_{[h_n]}\left(\beta^0 - n^{-\frac{1}{2}}t\right) - r^2_{[h_n]}\left(\beta^0\right)\right| < n^{-\tau}\cdot K_\varepsilon\right) > 1-\varepsilon,$$

that is, $\left|r^2_{[h_n]}\left(\beta^0 - n^{-\frac{1}{2}}t\right) - r^2_{[h_n]}\left(\beta^0\right)\right| = o_p\left(n^{-\frac{1}{2}}\right)$ *uniformly in* $t \in \mathcal{T}_M$ *as* $n \to +\infty$.

Proof: As the first and main step, we show that for any $\varepsilon \in (0,1)$ there exist K_ε and n_ε such that uniformly in $t \in \mathcal{T}_M$ for all $n > n_\varepsilon$

$$P\left(\left|r_{[h_n]}\left(\beta^0 - n^{-\frac{1}{2}}t\right) - r_{[h_n]}\left(\beta^0\right)\right| < n^{-\tau}\cdot K_\varepsilon\right) > 1-\varepsilon \tag{72}$$

(please, remember the convention introduced in Section 3.2 that $r_{[h]}(\beta) = \operatorname{sgn} r_{[h]}(\beta)\cdot\sqrt{r^2_{[h]}(\beta)}$, whereas the order statistics of residuals $r_i(\beta)$ is referred by $r_{(h)}(\beta)$). Please, notice also that we know due to equation (70)

$$r_i(\beta^0 - n^{-\frac{1}{2}}t) = r_i(\beta^0) + h'_\beta(x_i,\xi)^T\cdot n^{-\frac{1}{2}}t, \tag{73}$$

where $\xi \in \left[\beta^0, \beta^0 - n^{-\frac{1}{2}}t\right]$.

Now, all assertion in the following part of the proof are meant conditionally on values of x_i. Let us suppose that $h'_\beta(x_i,\xi)^T t \geq 0$ for a given i (the other case can be analyzed analogously). Then $r_i(\beta^0) + h'_\beta(x_i,\xi)^T n^{-\frac{1}{2}}t \geq r_i(\beta^0)$ which means that all such residuals $r_i(\beta^0 - n^{-\frac{1}{2}}t)$ are larger then residuals $r_i(\beta^0) \equiv \varepsilon_i$. In other words, some residuals evaluated at point $\beta^0 - n^{-\frac{1}{2}}t$ compared to β^0 are shifted out of interval $\left\langle -r_{[h_n]}\left(\beta^0\right), r_{[h_n]}\left(\beta^0\right)\right\rangle$ on its right hand side and some are shifted into it on its left hand side. The assertion (72) can be proved in the following way: considering a bit larger interval $\left\langle -r_{[h_n]}\left(\beta^0\right) - n^{-\tau}K_1, r_{[h_n]}\left(\beta^0\right) + n^{-\tau}K_1\right\rangle$, it is to be shown that such an interval contains at least h_n residuals $r_i(\beta^0 - n^{-\frac{1}{2}}t)$ for some sufficiently large constant K_1. To do so, we shall try to find a number m_1 of indices $i = 1,\ldots,n$ for which (with a probability close to 1)

$$r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0) \quad \text{and} \quad r_i(\beta^0 - n^{-\frac{1}{2}}t) \geq r_{[h_n]}\left(\beta^0\right) + n^{-\tau}K_1. \tag{74}$$

Such indices represent the observations that decrease the number of residuals inside the interval $\left\langle -r_{[h_n]}\left(\beta^0\right) - n^{-\tau}K_1, r_{[h_n]}\left(\beta^0\right) + n^{-\tau}K_1\right\rangle$. Similarly, we try to find a number m_2 of indices $i = 1,\ldots,n$ for which (with a probability close to 1)

$$r_i(\beta^0) \leq -r_{[h_n]}(\beta^0) \quad \text{and} \quad r_i(\beta^0 - n^{-\frac{1}{2}}t) \geq -r_{[h_n]}\left(\beta^0\right) - n^{-\tau}K_1. \tag{75}$$

These indices correspond to the observations that were not in the interval $\left\langle -r_{[h_n]}\left(\beta^0\right), r_{[h_n]}\left(\beta^0\right)\right\rangle$ before but they move inside the interval $\left\langle -r_{[h_n]}\left(\beta^0\right) - n^{-\tau}K_1, r_{[h_n]}\left(\beta^0\right) + n^{-\tau}K_1\right\rangle$, and thus, increase the number of residuals contained in it. Since there are just h_n indices among all $i = 1,\ldots,n$ satisfying $r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)$, the number of indices such that $r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq r_{[h_n]}^2\left(\beta^0\right) + n^{-\tau}K_1$ equals $h_n - m_1 + m_2$. Therefore, all we have to do is to verify that the difference $m_2 - m_1$ is positive with probability close to 1.

Using (73), case (74) is equivalent to

$$r_{[h_n]}(\beta^0) + n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t \leq e_i \leq r_{[h_n]}(\beta^0).$$

Similarly, (75) is valid if and only if

$$-r_{[h_n]}(\beta^0) - n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t \leq e_i \leq -r_{[h_n]}(\beta^0).$$

Thus, it seems to be helpful to study the probability of the events $z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t \leq e_i \leq z$ for some $z \in \mathbb{R}$. This probability can be expressed by means of the distribution function $F(x)$ (remember, everything till now is conditional on x_i):

$$F(z) - F\left(z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right) = \int_{z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t}^{z} f(t)dt.$$

Expanding the density in the integral, $f(t) = f(z) + f'(\zeta_t)t$, we get

$$\begin{aligned}
\int_{z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t}^{z} f(t)dt \quad &\geq \quad f(z)\left[\pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right] + \\
&\quad + L_f\left[\pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right]^2 \\
\int_{z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t}^{z} f(t)dt \quad &\leq \quad f(z)\left[\pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right] + \\
&\quad + U_f\left[\pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right]^2,
\end{aligned}$$

which results in

$$F(z) - F\left(z \pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right) = f(z)\left[\pm n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t\right] + \mathcal{O}\left(n^{-1}\right). \tag{76}$$

Having these results in hand, the same idea as in the previous Lemmas 8 and 9 can be used. Let us consider for a fixed $z \in \mathbb{R}$

$$w_{1i}(z) = I\Big(z + n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t \le e_i \le z\Big)$$

and

$$w_{2i}(z) = I\Big(-z - n^{-\tau}K_1 - h'_\beta(x_i,\xi)^T \cdot n^{-\frac{1}{2}}t \le e_i \le -z\Big).$$

Apparently, $m_2 - m_1 = \sum_{i=1}^n (w_{2i}(z) - w_{1i}(z))$ for $z = r_{[h_n]}(\beta^0)$. Let us denote $s_i(z) = w_{2i}(z) - w_{1i}(z)$. Employing (76), we obtain $\mathsf{E}\, s_i(z) = n^{-\tau}K_1 \cdot (f(z) + f(-z)) + \mathcal{O}\big(n^{-1}\big)$ and hence also $\mathsf{var}\, s_i(z) = n^{-\tau}K_1 \cdot (f(z) + f(-z)) + \mathcal{O}\big(n^{-1}\big)$. The Feller-Lindenberg conditions for the central limit theorem can be easily verified, and thus, two constants K_ε and n_ε such that for all $n \ge n_\varepsilon$

$$P\left(\left|\frac{\sum_{i=1}^n (s_i(z) - \mathsf{E}\, s_i(z))}{C_n}\right| < K_\varepsilon\right) > 1 - \varepsilon$$

can be found, where $C_n^2 = n^{1-\tau}K_1 \cdot (f(z) + f(-z)) + \mathcal{O}(1)$. Now, this result is independent of x_i, which means that it holds not only conditional on x_i but without conditioning as well. Further, it follows that for $n \ge n_\varepsilon$ with probability greater than $1 - \varepsilon$

$$\sum_{i=1}^n s_i(z) \ge -n^{\frac{1}{2}(1-\tau)}\sqrt{K_1 \cdot (f(z) + f(-z))}K_\varepsilon + n^{1-\tau}K_1 \cdot (f(z) + f(-z))\,. \tag{77}$$

As $t \in \big(\frac{1}{2}, 1\big)$, the last expression increases in n above all limits for a given K_1 because $f(z)$ is bounded from above and away from zero as well in a neighborhood of $G^{-1}(\lambda)$. Thus, we can find n_ε such that for all $n > n_\varepsilon$ the right hand side of (77) is positive, and consequently, the number of the residuals $r_i^2\left(\beta^0 - n^{-\frac{1}{2}}t\right)$ that fall to the interval $\left\langle -r_{[h_n]}\left(\beta^0\right) - n^{-\tau}K_1, r_{[h_n]}\left(\beta^0\right) + n^{-\tau}K_1\right\rangle$ is at least h_n with probability greater than $1 - \varepsilon$. We can conclude that for some $\infty > K_2 > 0$

$$r_{[h_n]}(\beta^0 - n^{-\frac{1}{2}}t) \le r_{[h_n]}(\beta^0) + n^{-\tau} \cdot K_2,$$

and analogously, the corresponding lower inequality can be derived.

Finally, Lemmas 8 and 9 imply that both $r_{[h_n]}(\beta^0 - n^{-\frac{1}{2}}t)$ and $r_{[h_n]}(\beta^0)$ are bounded in probability. Thus, utilizing equality $a^2 - b^2 = (a+b)(a-b)$, we obtain immediately the assertion of this lemma. □

References

[1] Amemiya, T. (1983): Non-linear regression models, in Griliches, Z., and Intriligator, M. D., eds., *Handbook of Econometrics*, Vol. 1, North-Holland, Amsterdam, pp. 333–389.

[2] Andrews, D. W. K. (1987): Consistency in Nonlinear Econometic Models: A Generic Uniform Law of Large Numbers, *Econometrica*, Vol. 55 / 6, pp. 1465–1471.

[3] Box, G. E. P. and Cox, D. (1964): An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B*, 211–264.

[4] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986): *Robust Statistics, The Approach Based on Influence Function*, Wiley series in Probability and Mathematical Statistics, New York.

[5] Ichimura, H. (1993): Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models, *Journal of Econometrics*, July 1993, pp. 71–120.

[6] Jurečková, J. (1984): Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika*, Vol. 20, pp. 345–357.

[7] Jurečková, J., Sen, P. K. (1989): Uniform Second Order Asymptotic Linearity of M-statistics, *Statistics & Decisions*, Vol. 7, pp. 263–276.

[8] Rousseeuw, P. J., Leroy, A. M. (1987): *Robust Regression and Outlier Detection*, Wiley, New York.

[9] ——— (1985): Multivariate Estimation With High Breakdown Point, in W. Grossman, G. Pflug, I. Vincze, and W. Wertz eds., *Mathematical Statistics and Applications, Vol. B*, Reidel, Dordrecht, The Netherlands, pp. 283–297.

[10] ———, Van Driessen, K. (1999): Computing LTS Regression for Large Data Sets, *Technical Report, University of Antwerp.*

[11] Seber, G. A. F., Wild, C. J. (1989): *Nonlinear regression*, John Wiley & Sons, USA.

[12] Tong, H. (1990): *Nonlinear time series: A Dynamical Systems Approach*, Oxford University Press, Oxford.

[13] Víšek, J. Á (1996a): Sensitivity Analysis of M-estimators, *Annals of the Institute of Statistical Mathematics*, Vol. 48, pp. 469–495.

[14] ——— (1999a): The least trimmed squares—random carriers, *Bulletin of the Czech Econometric Society*, Volume **10/1999**, 1–30.

[15] White, H. (1980): Nonlinear Regression on Cross-Section Data, *Econometrica*, Volume 48, No. 3, pp. 721–746.