Discussion Papers
Department of Economics
University of Copenhagen

No. 19-05

The analysis of marked and weighted empirical processes of estimated residuals

Vanessa Berenguer-Rico, Søren Johansen and Bent Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

The analysis of marked and weighted empirical processes of estimated residuals

Vanessa Berenguer-Rico
Department of Economics; University of Oxford; Oxford; OX1 3UQ; UK and Mansfield College; Oxford; OX1 3TF
vanessa.berenguer-rico@economics.ox.ac.uk

Søren Johansen
Department of Economics; University of Copenhagen and CREATES; Department of Economics and Business; Aarhus University
soren.johansen@econ.ku.dk

Bent Nielsen[1]
Department of Economics; University of Oxford; Oxford; OX1 3UQ; UK and Nuffield College; Oxford; OX1 1NF; U.K.
bent.nielsen@nuffield.ox.ac.uk

29 April 2019

Summary An extended and improved theory is presented for marked and weighted empirical processes of residuals of time series regressions. The theory is motivated by 1-step Huber-skip estimators, where a set of good observations are selected using an initial estimator and an updated estimator is found by applying least squares to the selected observations. In this case, the weights and marks represent powers of the regressors and the regression errors, respectively. The inclusion of marks is a non-trivial extention to previous theory and requires refined martingale arguments.
Keywords 1-step Huber-skip; Non-stationarity; Robust Statistics; Stationarity.

1 Introduction

We consider marked and weighted empirical processes of residuals from a linear time series regression. Such processes are sums of products of an adapted weight, a mark that is a power of the innovations and an indicator for the residuals belonging to a half line. They have previously been studied by Johansen & Nielsen (2016a) - JN16 henceforth - generalising results by Koul & Ossiander (1994) and Koul (2002) for processes without marks. The results presented extend and improve upon expansions previously given in JN16, while correcting a mistake in the argument, simplifying proofs and allowing weaker conditions on the innovation distribution and regressors.

1.1 The setup

The results in this paper are aimed at analysis of 1-step Huber-skip estimators that are popular in the robust literature and used extensively in applied work without reference

[1]Support from Programme for Economic Modelling, Oxford gratefully acknowledged.

to the robust literature. While such estimators have been analyzed before, the present purpose is to improve conditions and proofs of the underlying empirical process results.

Consider the linear time series model

$$y_i = x_i'\beta + \varepsilon_i \qquad i = 1, \ldots, n,$$

where the innovations ε_i/σ have distribution function F for some scale parameter σ and where the regressors x_i can be stationary, deterministic or stochastically trending. Let $\tilde{\beta}, \tilde{\sigma}$ be initial estimators for the unknown β, σ. In applied work it is very common to use full sample least squares estimators, although Welsh & Ronchetti (2006) recommend to use robust estimators. In any case, we can construct initial residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$. Observations satisfying $|\tilde{\varepsilon}_i| > \tilde{\sigma}c$, for a certain cut-off value set up by the investigator, are declared outliers and removed. A new regression is then run with the selected observations satisfying $|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c$ giving an updated estimator, called the 1-step Huber-skib estimator,

$$\hat{\beta} = \{\textstyle\sum_{i=1}^n x_i x_i' 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}\}^{-1} \textstyle\sum_{i=1}^n x_i y_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}. \tag{1.1}$$

Asymptotic expansions for $N^{-1}(\hat{\beta} - \beta)$ are of interest, where N^{-1} is a deterministic normalizing matrix for the regressors. In particular, the generality of the normalization N^{-1} allows us to consider a variety of regressors including stationary and non-stationary variables. The normalized estimation error satisfies

$$N^{-1}(\hat{\beta} - \beta) = \{n^{-1}\textstyle\sum_{i=1}^n N' x_i x_i' N 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}\}^{-1} n^{-1} \textstyle\sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}. \tag{1.2}$$

Both numerator and denominator in (1.2) are examples of marked and weighted empirical distribution functions of the residuals $\tilde{\varepsilon}_i$, which are of the form

$$n^{-1}\textstyle\sum_{i=1}^n w_{in} \varepsilon_i^p 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}. \tag{1.3}$$

Specifically, the numerator and denominator have weights $w_{in} = N'x_i$ and $w_{in} = N'x_i x_i' N$, respectively, and marks ε_i^p with $p = 1$ and $p = 0$. Note that the mark is allowed to be unbounded. Therefore, the empirical process techniques derived in this paper can be used to obtain asymptotic expansions for $N^{-1}(\hat{\beta} - \beta)$.

1.2 Marked and weighted empirical processes of residuals

This paper provides an improved analysis of weighted and marked empirical distribution functions of the form (1.3). The proofs involve a number of steps.

First, in the residuals

$$\tilde{\varepsilon}_i = \varepsilon_i - x_i'(\tilde{\beta} - \beta) = \varepsilon_i - x_i' N N^{-1}(\tilde{\beta} - \beta),$$

the normalized random estimation error $N^{-1}(\tilde{\beta} - \beta)/\sigma$ is replaced by a deterministic quantity b. This requires that results are established uniformly over a compact set for b. Similarly, the normalized estimation error for scale $n^{1/2}(\tilde{\sigma} - \sigma)$ is replaced by a deterministic scale error a as in Jiao & Nielsen (2017). If instead results are uniform over a sequence of expanding compact sets it is possible to allow diverging normalized estimation errors. An example is the Least Median of Squares estimator of Rousseeuw (1984) which is $n^{1/3}$-consistent.

Second, weighted and marked empirical distribution functions are turned into normalized sums of martingale differences by subtracting their compensators.

Third, uniform analysis over estimation errors a, b and the quantile c is carried out by a chaining argument. This requires handling of tail probabilities of a family of martingales for which we use the iterated exponential martingale inequality of JN16, see Lemma A.1. When there are no marks, this is based on the Freedman (1975) inequality used by Koul & Ossiander (1994), but in general it uses the Bercu & Touati (2008) inequality.

Fourth, distances of two quantiles c_1 and c_2 are measured through a distance function $\mathsf{H}_r(c) = \int_{-\infty}^{c}(1 + x^{2^r p})\mathsf{f}(x)dx$ with derivative $\dot{\mathsf{H}}_r(c)$ for a suitable power r and where f is the density of ε_i. The derivative is assumed to be Lipschitz and bounded from above and below by two proportional unimodal functions. At the same time, the density f can have finite support. Examples include densities f, that are normal as well as uniform or triangular. In Lemmas A.8, A.9 we present improved inequalities for differences of these functions evaluated at two points: $\mathsf{H}_r(c_2) - \mathsf{H}_r(c_1)$ and $\dot{\mathsf{H}}_r(c_2) - \dot{\mathsf{H}}_r(c_1)$.

The regularity conditions are simpler than in JN16 since $\dot{\mathsf{H}}_r(c)$ is assumed to be Lipschitz rather than differentiable. The assumption of weakly unimodal bounds is equivalent, but more accessible, than a condition in JN16, see Lemma A.2. It is clearified that it suffices that f has support on an open interval. The class of functions with weakly unimodal bounds is shown to be closed under addition and multiplication, see Lemma A.3. It includes the normal, triangular and uniform distributions as well as mixtures thereof.

For the weights w_{in} and $x_{in} = N'x_i$ we require certain moment conditions. JN16 had the additional assumption that $\max_{1\leq i\leq n}|x_{in}|$ vanishes in probability. With the improved proof, this condition is no longer needed and the range of regressors extends from stationary and random walk-type regressors as in JN16 to explosive regressors.

1.3 Applications

The 1-step Huber-skip estimator $\hat{\beta}$ is popular in the robust literature. It is used extensively in applied work without reference to robust statistics. With the present results it is possible to update existing results to have simpler assumptions. The estimator $\hat{\beta}$ is a 1-step version of the skip-estimator of Huber (1964). Due to the hard rejection of outlying residuals, the estimator differs from the scoring-type 1-step estimator of Bickel (1975), see also (Jurečková et al., 2013, §7.4). It has various names in the literature: the Trimmed Least Squares Estimation by Ruppert & Carroll (1980); the Weighted Least Squares by (Rousseeuw & Leroy, 1987, p. 17, 153); and the Data Analytic Strategy by Welsh & Ronchetti (2006). A variant of the 1-step Huber-skip estimator can be used for scale estimation, when the regression parameter is estimated by the Least Trimmed Squares estimator of Rousseeuw (1984), see (Rousseeuw & Leroy, 1987, p. 17), Croux & Rousseeuw (1992), Johansen & Nielsen (2016b).

Least squares steps similar to the 1-step Huber-skip estimator are used in computer-intensive iterative procedures such as the Forward Search and the Impulse Indicator

Saturation. The Forward Search is an iterative algorithm for avoiding outliers in regression analysis suggested by Hadi & Simonoff (1993) and developed further by Atkinson & Riani (2000) and Atkinson et al. (2010). The algorithm starts with the selection of a subset of "good" observations. In the iteration step, a variant of the 1-step Huber-skip estimator is used and the size of the subset of "good" observations is increased by one. A related iterative algorithm is Impulse Indicator Saturation, based on an idea of Hendry (1999), see also (Hendry & Doornik, 2014, §15). It is implemented in Ox, see Doornik (2009) and R, see Pretis et al. (2018). A stylized version of the algorithm is the split-half algorithm suggested by Hendry et al. (2008). The idea is to split the sample into two, compute the least squares estimator in each sample and then use the estimator from one sub-sample to detect outliers in the other sub-sample. This gives rise to 1-step Huber-skip estimators. Johansen & Nielsen (2016b) review the available asymptotic theory for these algorithms. This includes a budding theory for chosing the cut-off values from the frequency of false discoveries, also called the gauge. A feature of this theory is that it is developed under the hypothesis of no outliers, where the reference distribution F is nice. The empirical process results presented here allows for more irregular distributions, which brings us closer to the analysis of these algorithms under contamination.

The results generalize previous work on the residual empirical distribution function for autoregressions by Engler & Nielsen (2009). This, in turn, builds on separate proofs of Lee & Wei (1999) and Koul & Leventhal (1989) for non-explosive and explosive cases, respectively. The present proof has a unified argument for those cases. The marked empirical processes of Koul & Stute (1999), Escanciano (2007) arise when the weights are $w_{in} = n^{-1/2}1_{(x_i \leq d)}$ and the present indicators $1_{(\varepsilon_i \leq \sigma c)}$ are set to unity. Their expansions are uniform in d, which is not allowed here.

1.4 Outline

The paper is organized as follows. In Section 2, the model and definitions related to the residual empirical processes are presented. The asymptotic analysis follows in Section 4. At first, we improve a result in JN16 concerning estimation error for location. The main results are presented in four theorems: First, the marked and weighted empirical process of residuals is shown to be asymptotically equivalent to the corresponding process of the true innovations; Second, the bias coming from the compensator is derived; Third, the tightness of the empirical process of the true errors is presented; Fourth, the previous three results are combined to give an asymptotic expansion of the marked and weighted empirical process of residuals. All proofs are collected in the Appendix.

2 The model and the empirical distribution function

We assume that (y_i, x_i) for $i = 1, \ldots, n$ satisfy the multiple regression equation

$$y_i = x_i'\beta + \varepsilon_i, \tag{2.1}$$

with scale σ, regressors x_i and parameter β, both of dimension $\dim x$. The scaled innovations ε_i/σ are independent and identically distributed with density f and distribution function $\mathsf{F}(c) = \mathsf{P}(\varepsilon_i/\sigma \leq c)$. In practice, the distribution F will often be standard normal. For each i the innovation ε_i is independent of the regressor x_i.

Suppose we have an initial estimator $\tilde{\beta}$ for the regression parameter β, residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ and an estimator $\tilde{\sigma}$ of the scale σ. Define, for some deterministic normalization matrix N, normalized estimators and regressors

$$\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma, \qquad \tilde{b} = N^{-1}(\tilde{\beta} - \beta)/\sigma, \qquad x_{in} = N'x_i, \tag{2.2}$$

so that $x_i'(\tilde{\beta} - \beta) = x_{in}'\tilde{b}\sigma$. In most situations, the normalization N is chosen so that $\sum_{i=1}^{n} x_{in}x_{in}'$ has a positive definite limit. In this way, we can choose $N = n^{-1/2}$ for stationary regressors and $N = n^{-1}$ for random walk regressors. If the regressors are $x_i = (1, i)$, we normalize them so that $x_{in} = (n^{-1/2}, n^{-3/2}i)$. If the regressors are explosive so that $x_i = 2^i$, we let $N = 2^{-n}$ so that $x_{in} = 2^{i-n}$. In the asymptotic analysis, we consider triangular arrays to accommodate the normalization built into x_{in}. This means that we also cover certain types of infill asymptotics. Suppose, in the context of model (2.1), that $x_i = 1_{(i \leq n^\dagger)}$ for some $n^\dagger \leq n$. The asymptotic constraint $n^\dagger/n = \tau$ for some $0 < \tau < 1$ can be accommodated by choosing $N^{-1} = n^{1/2}$ and $x_{in} = n^{-1/2}1_{(i \leq \tau n)}$ in (2.2).

The theory does, however, leave the possibility of choosing N through a tradeoff between two conditions. First, N should be so small that $\mathsf{E}\sum_{i=1}^{n} x_{in}x_{in}' = \mathrm{O}(1)$, which allows for non-convergence or convergence to zero. Second, N^{-1} should be so small that $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)$ is bounded by $n^{1/4-\eta}B$ for some $0 < \eta \leq 1/4$. This could potentially be useful in irregular situations, where asymptotic theory is less developed.

The marked and weighted empirical distribution functions of residuals are defined as

$$\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) = n^{-1}\sum_{i=1}^{n} w_{in}\varepsilon_i^p 1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)} = n^{-1}\sum_{i=1}^{n} w_{in}\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c + n^{-1/2}\tilde{a}c + x_{in}'\tilde{b})}, \tag{2.3}$$

where ε_i^p is the mark and w_{in} is a weight function that could be matrix valued and satisfies $\mathsf{E}\sum_{i=1}^{n} w_{in} = \mathrm{O}(n)$. Examples include $w_{in} = 1$, $w_{in} = n^{1/2}x_{in}$ and $w_{in} = nx_{in}x_{in}'$.

3 Techniques for analysis of empirical processes

The primary challenge in the asymptotic analysis of $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ in (2.3) is the estimation errors $\tilde{a}, \tilde{b}$, that is, to move from the empirical distribution function of residuals $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ to the empirical distribution function of innovations $\mathsf{F}_n^{w,p}(0, 0, c)$. For this purpose, we replace the normalized estimation errors $\tilde{a}$ and $\tilde{b}$ in (2.2) with deterministic terms a and b varying in an apropriate compact set which depends on n. We assume $\tilde{a}$ and $\tilde{b}$ are $\mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$ for $0 < \eta \leq 1/4$, so that $n^{\eta-1/4}\tilde{a}$ and $n^{\eta-1/4}\tilde{b}$ vary in compact sets with large probability. Thus, due to the following lemma, $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ can be analysed by studying $\mathsf{F}_n^{w,p}(a, b, c)$ uniformly over a large compact set for $n^{\eta-1/4}a$ and $n^{\eta-1/4}b$.

Lemma 3.1. *Let $\epsilon > 0$. Suppose a compact set Θ exists so $\lim_{n\to\infty} \mathsf{P}(\tilde{\theta} \in \Theta) > 1 - \epsilon$. Let $F_n(\theta, c)$ be some function of $\theta \in \Theta$ and $c \in \mathbb{R}$. Then,*

$$\mathsf{P}\{|F_n(\tilde{\theta}, c)| > \epsilon\} \leq \mathsf{P}\{\sup_{\theta\in\Theta} |F_n(\theta, c)| > \epsilon\} + \epsilon.$$

Proof. Since $\mathsf{P}(\mathcal{A}) \leq \mathsf{P}(\mathcal{A} \cap \mathcal{B}) + \mathsf{P}(\mathcal{B}^c)$ for events $\mathcal{A}, \mathcal{B}$, then

$$\mathsf{P}\{|F_n(\tilde{\theta}, c)| > \epsilon\} \leq \mathsf{P}\{|F_n(\tilde{\theta}, c)| > \epsilon, \tilde{\theta} \in \Theta\} + \mathsf{P}(\tilde{\theta} \notin \Theta).$$

The first term is bounded by considering the largest possible outcome of $|F_n(\theta, c)|$ for $\theta \in \Theta$. The second term vanishes by assumption. $\square$

The process $\mathsf{F}_n^{w,p}(a, b, c)$ is analyzed under the following triangular array assumption to the innovations ε_i, the regressors x_{in} and weights w_{in}.

Assumption 3.1. *Let $\mathcal{F}_{in}$ be an array of increasing sequences of σ-fields so that $\mathcal{F}_{i-1,n} \subset \mathcal{F}_{in}$ where $\varepsilon_{i-1}, x_{in}, w_{in}$ are $\mathcal{F}_{i-1,n}$ measurable and ε_i/σ is independent of $\mathcal{F}_{i-1,n}$ with density f, which is continuous on its support $\mathcal{S}$, which is an open interval $]\underline{c}, \overline{c}[$ with $-\infty \leq \underline{c} < \overline{c} \leq \infty$.*

Under Assumption 3.1 we apply a martingale decomposition to $\mathsf{F}_n^{w,p}(a, b, c)$ as follows. For a given n, let $\mathsf{E}_{i-1}(\cdot)$ denote the conditional expectation given $\mathcal{F}_{i-1,n}$. Thus, the compensator is the following sum of conditional expectations

$$\overline{\mathsf{F}}_n^{w,p}(a, b, c) = n^{-1} \textstyle\sum_{i=1}^n w_{in} \mathsf{E}_{i-1}\{\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c + n^{-1/2}ac + x'_{in}b)}\}. \tag{3.1}$$

From this, we form the marked and weighted empirical process

$$\mathbb{F}_n^{w,p}(a, b, c) = n^{1/2}\{\mathsf{F}_n^{w,p}(a, b, c) - \overline{\mathsf{F}}_n^{w,p}(a, b, c)\}, \tag{3.2}$$

which is a normalized sum of martingale differences, where the summands depend on n. This gives the martingale decomposition $\mathsf{F}_n^{w,p}(a, b, c) = \overline{\mathsf{F}}_n^{w,p}(a, b, c) + n^{-1/2}\mathbb{F}_n^{w,p}(a, b, c)$.

In the asymptotic theory, uniform results over a, b, c are proved using chaining arguments. This requires a compactification of the quantile axis for $c \in \mathbb{R}$, which is done by using the function $\mathsf{H}_r(c) = \mathsf{E}(1 + |\varepsilon_1/\sigma|^{2^r p}) 1_{(\varepsilon_1 \leq \sigma c)}$, see also §A.2.

Two somewhat different types of chaining arguments are used. To illustrate the first type of chaining technique, consider a generic empirical process $F_n(\theta, c)$ where $\theta \in \Theta$ and $c \in \mathbb{R}$. To set up the chaining in this case, introduce K gridpoints c_k so that $\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1})$ are constant in k and proportional to $1/K$. Then, cover the set Θ by M balls with centres θ_m with a small radius δ. The first chaining argument is

$$\begin{aligned}\sup_{\theta \in \Theta} \sup_{c \in \mathbb{R}} |F_n(\theta, c)| \leq \max_{1 \leq m \leq M} \max_{1 \leq k \leq K} |F_n(\theta_m, c_k)| \\ + \max_{1 \leq m \leq M} \max_{1 \leq k \leq K} \sup_{|\theta - \theta_m| \leq \delta} \sup_{c_{k-1} < c \leq c_k} |F_n(\theta, c) - F_n(\theta_m, c_k)|.\end{aligned}$$

The two bounding terms are denoted the discrete point term and the perturbation term.

The second chaining argument is used in the proof of the tightness of the empirical process $\mathbb{F}_n^{w,p}(0, 0, c)$ without estimation errors. The proof uses chaining over dyadic rational numbers on the set $\mathsf{H}_r(\mathbb{R})$. A result of this type is given in Theorem 4.3.

4 Uniform expansions of empirical processes

The following results are concerned with a uniform Central Limit Theorem for the empirical distribution function $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$. The analysis starts with the decomposition

$$\begin{aligned}
&n^{1/2}\{\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}\\
&=n^{1/2}\{\mathsf{F}_n^{w,p}(0,0,c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}+\mathcal{B}_n^{w,p}(\tilde{a},\tilde{b},c)\\
&+n^{1/2}\{\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(\tilde{a},\tilde{b},c)-\mathsf{F}_n^{w,p}(0,0,c)+\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}\\
&+n^{1/2}\{\bar{\mathsf{F}}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}-\mathcal{B}_n^{w,p}(\tilde{a},\tilde{b},c),
\end{aligned}\tag{4.1}$$

where $\mathcal{B}_n^{w,p}(a,b,c)$ is a bias term, which is linear in a, b. It is defined in (4.4) below. Thus, using the notation $\mathbb{F}_n^{w,p}$ defined in (3.2) we have

$$\begin{aligned}
n^{1/2}\{\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}&=\mathbb{F}_n^{w,p}(0,0,c)+\mathcal{B}_n^{w,p}(\tilde{a},\tilde{b},c)\\
+\{\mathbb{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\mathbb{F}_n^{w,p}(0,0,c)\}&+[n^{1/2}\{\bar{\mathsf{F}}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}-\mathcal{B}_n^{w,p}(\tilde{a},\tilde{b},c)].
\end{aligned}$$

The first term $\mathbb{F}_n^{w,p}(0,0,c)$ is a standard marked and weighted empirical process without estimation error. For a fixed c, it is analyzed using a martingale Central Limit Theorem. Viewed as a process, the tightness is shown in Theorem 4.3, which originates from JN16, whereas, for instance, Billingsley (1968) considers the special case without marks and weights and Koul & Ossiander (1994) consider the special case without marks. The third and fourth terms vanish by Theorems 4.1, 4.2 below. Thus, uniformly in c,

$$n^{1/2}\{\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\}=\mathbb{F}_n^{w,p}(0,0,c)+\mathcal{B}_n^{w,p}(\tilde{a},\tilde{b},c)+\mathsf{o}_{\mathsf{P}}(1).$$

4.1 Location estimation error and the empirical process

The first result requires some regularity of $\mathsf{h}(c)=(1+|c|^{2^r p})\mathsf{f}(c)$ for some r to be chosen.

Definition 4.1. *Let $\mathsf{h}(c)\geq 0$ have support $\mathcal{S}=]\underline{c},\overline{c}[$ where $-\infty\leq\underline{c}<\overline{c}\leq\infty$:*
(i) h *is **Lipschitz** if $\exists C_L>0$: $\forall c,c^\dagger\in\mathcal{S}$ then $|\mathsf{h}(c)-\mathsf{h}(c^\dagger)|\leq C_L|c-c^\dagger|$;*
(ii) h *has **weakly unimodal bounds** if a constant $C_u\geq 1$ and a function u exist so that $\forall c\in\mathbb{R}$: $u(c)\leq\mathsf{h}(c)\leq C_u u(c)$, where u has finite mode at $c_{mode}\in\mathcal{S}$, so that $u(c)$ is non-increasing for $c>c_{mode}$ and non-decreasing $c<c_{mode}$.*

Assumption 4.1. *Let $p\in\mathbb{N}_0$, $r\in\mathbb{N}$, $0<\eta\leq 1/4$ be given so that $r\geq 2$ and*

$$2^{r-1}>1+(1/4-\eta)(1+\dim x).\tag{4.2}$$

(i) innovations ε_i/σ. Suppose $\mathsf{h}(c)=(1+|c|^{2^r p})\mathsf{f}(c)$ is (a) integrable and (b) Lipschitz with weakly unimodal bounds (Definition 4.1);
(ii) regressors x_{in} and weights w_{in}, where w_{in} may be matrix valued, satisfy
$\mathsf{E}n^{-1}\sum_{i=1}^n(1+|w_{in}|^{2^r})(1+|n^{1/2}x_{in}|^2)=\mathrm{O}(1)$.

Lemma 4.1. *Suppose Assumptions 3.1, 4.1 hold. Let $0<\eta\leq 1/4$. Then, $\forall B>0$,*

$$\sup_{c\in\mathbb{R}}\sup_{|b|\leq n^{1/4-\eta}B}|\mathbb{F}_n^{w,p}(0,b,c)-\mathbb{F}_n^{w,p}(0,0,c)|=\mathsf{o}_{\mathsf{P}}(1).$$

We now give some remarks and some examples in relation to Assumption 4.1

Remark 4.1. *In stationary models,* $N = n^{-1/2}$ *so that* $\sum_{i=1}^{n} x_{in}x_{in}' = n^{-1}\sum_{i=1}^{n} x_i x_i'$ *converges. Standard estimators satisfy* $\tilde{b} = \mathsf{O}_{\mathsf{P}}(1)$ *so that* $\eta = 1/4$ *and* $r = 2$ *in (4.2). For non-standard estimators* $\tilde{b}$ *may diverge so that the required number of moments for* ε_i *grows linearly with the dimension of the regressor. This would be relevant for the* $n^{1/3}$*-consistent least median of squares regression estimator* $\tilde{\beta}_{LMS}$ *by Rousseeuw (1984). In that case, we get* $\eta = 1/12$ *since* $n^{1/2}(\tilde{\beta}_{LMS} - \beta) = \mathsf{O}_{\mathsf{P}}(n^{1/2-1/3}) = \mathsf{O}_{\mathsf{P}}(n^{1/4-1/12})$.

Remark 4.2. *We compare Assumption 4.1 with Assumptions 3.1, 4.1 in JN16.*
(*a*) *The coefficient* r *in (4.2) here satisfies a slighty weaker constraint in that a term* $\kappa(1 + \dim x)$ *has fallen away from the lower bound. One implication is that when the normalized estimators are bounded in probability,* $\tilde{a}, \tilde{b} = \mathsf{O}_{\mathsf{P}}(1)$ *so that* $\eta = 1/4$*, then we can choose* $r = 2$ *for regressors* x_{in} *of any dimension.*
(*b*) *Part* (*i*) *is simpler than the corresponding part in JN16. It is made clear that the support can be finite. It suffices that the function* h *is Lipschitz on the support rather than differentiable. The property of having weakly unimodal bounds is equivalent to the smoothness condition in JN16, but easier to apply, see Lemma A.2. JN16 required boundedness for certain functions of the density* f*. These conditions are now found to be consequences of other conditions due to Lemmas A.4, A.5 in the Appendix.*
(*c*) *The regressors satisfy moment conditions here without requiring that* $\max_{1\le i\le n} x_{in}$ *vanishes in contrast to earlier papers including Koul & Ossiander (1994), Engler & Nielsen (2009) and JN16. Thus, the results cover the explosive regressors. An example is* $x_i = 2^i$ *normalized as* $x_{in} = 2^{i-n}$*. The normalized estimator* $\sum_{i=1}^{n} x_{in}\varepsilon_i / \sum_{i=1}^{n} x_{in}^2$ *converges in distribution when* ε_i *is iid and the sum of squares* $\sum_{i=1}^{n} x_{in}^2$ *converges, but* $\max_{1\le i\le n} x_{in} = x_{nn} = 1$ *is not vanishing.*

Example 4.1. *Suppose that* $\mathsf{h}(c) = (1 + |c|^a)\mathsf{f}(c)$*, for some* $a > 0$*. We demonstrate that* h *satisfies Assumption 4.1 for uniform, triangular, normal and mixture densities.*
(*a*) *Suppose* f *is the uniform density or the triangular density,* $\Delta(c) = 1 - |c|$ *for* $|c| \le 1$*. Because* $\mathsf{h}(c) \ge \mathsf{f}(c)$ *and the uniform and triangular have bounded support* $\mathcal{S}$*, we can choose* $u(c) = \mathsf{f}(c)$*, and* $C_u = \max_{c\in\mathcal{S}}(1 + |c|^a)$*, so that* h *has weakly unimodal bounds as in* (*ib*)*. Moreover, the densities have bounded right and left derivatives, so the* h *functions are Lipschitz and satisfy* (*ic*).
(*b*) *Suppose* $\mathsf{f} = \varphi$ *is standard normal. There exists* $c_0 > 0$ *such that* $(1 + |c|^a)\varphi(c)$ *is decreasing for* $c \ge c_0$ *and increasing for* $c \le -c_0$*. Let* $u_0 = \min_{|c|\le c_0}(1 + |c|^a)\varphi(c)$*. Then condition (A.6) holds with*

$$u(c) = \min\{u_0, (1 + |c|^a)\varphi(c)\}, \qquad C_u = \max_{|c|\le c_0}(1 + c^a)\varphi(c)/u_0. \tag{4.3}$$

(*c*) *Mixture densities* $\mathsf{f}(c) = (1 - \epsilon)\mathsf{f}_1(c) + \epsilon\mathsf{f}_2(c)$*. The class of function with weakly unimodal bounds and locally Lipschitz is closed to addition, see Lemma A.3. Thus,* (*ib*) *is satisfied.*

Example 4.2. *The Lipschitz and weak unimodal bounds conditions in Assumption 4.1(ib) are supplementary:*
(*a*) *The density proportional to* $1 - |c|^{1/2}$ *for* $|c| \le 1$ *is unimodal, but not Lipschitz.*

(b) The following oscillating function is Lipschitz, but it does not have unimodal bounds. For $c \geq 1$ and $m-1 \in \mathbb{N}$ let $\tilde{m} = m + (m-1)^{-2} - m^{-2}$ and define, for $m < c \leq m+1$,

$$f(c) = m^{-2} + (\tilde{m} - c)1_{(m \leq c \leq \tilde{m})} + (m^{-1} - |c - m - 1/2|)1_{(|c-m-1/2| \leq m^{-1})}.$$

The function is Lipschitz with Lipschitz coefficient 1. A decreasing lower bound must satisfy $u(m+1/2) \leq u(m+1/4)$ and $u(m+1/4) \leq \mathsf{f}(m+1/4)$ so that

$$\frac{f(m+1/2)}{u(m+1/2)} \geq \frac{f(m+1/2)}{f(m+1/4)} \geq \frac{1/m}{1/m^2} = m,$$

which is unbounded for large m, so that f does not have weakly unimodal bounds.

4.2 Further intermediate results

Some further results are needed before the expansion (4.1) of $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ can be analyzed. In parallel with the previous Lemma 4.1 for the location estimation error the next Lemma is concerned with scale estimation error. It simplifies proof and assumptions of Theorem 5 in Jiao & Nielsen (2017).

Lemma 4.2. *Suppose Assumptions 3.1, 4.1(i, ii) hold with only $r = 2$. Let $0 < \eta \leq 1/4$. Then, $\forall B > 0$,*

$$\sup_{c \in \mathbb{R}} \sup_{|a| \leq n^{1/4-\eta}B} |\mathbb{F}_n^{w,p}(a, 0, c) - \mathbb{F}_n^{w,p}(0, 0, c)| = \mathrm{o}_{\mathsf{P}}(1).$$

Combining the Lemmas 4.1, 4.2 leads to the following result.

Theorem 4.1. *Suppose Assumptions 3.1, 4.1 hold. Let $0 < \eta \leq 1/4$. Then, $\forall B > 0$,*

$$\sup_{c \in \mathbb{R}} \sup_{|a|,|b| \leq n^{1/4-\eta}B} |\mathbb{F}_n^{w,p}(a, b, c) - \mathbb{F}_n^{w,p}(0, 0, c)| = \mathrm{o}_{\mathsf{P}}(1).$$

Next, we linearize the compensator. The result generalizes Jiao & Nielsen (2017, Theorem 8) by replacing a differentiability assumption with a Lipschitz condition.

Assumption 4.2. *Suppose, for $p \in \mathbb{N}_0$,*
(i) innovations ε_i/σ satisfy
- *(a) moments: $\mathsf{E}|\varepsilon_i|^p < \infty$;*
- *(b) smoothness: $|c|^q \mathsf{f}(c)$ is Lipschitz for $q = p, p+1, p+2$ (Definition 4.1);*
- *(c) boundedness: $\sup_{c \in \mathcal{S}}(1 + |c|)|c|^p \mathsf{f}(c) < \infty$;*

(ii) weights and regressors: $n^{-1}\sum_{i=1}^n |w_{in}|(1 + |n^{1/2}x_{in}|^2) = \mathrm{O}_{\mathsf{P}}(1)$.

Theorem 4.2. *Suppose Assumptions 3.1, 4.2 hold. Define*

$$\mathcal{B}_n^{w,p}(a, b, c) = \sigma^p c^p \mathsf{f}(c) n^{-1/2} \sum_{i=1}^n w_{in}(n^{-1/2}ac + x_{in}'b). \tag{4.4}$$

Let $0 < \eta \leq 1/4$. Then, $\forall B > 0$,

$$\sup_{c \in \mathbb{R}} \sup_{|a|,|b| \leq n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{F}}_n^{w,p}(a, b, c) - \overline{\mathsf{F}}_n^{w,p}(0, 0, c)\} - \mathcal{B}_n^{w,p}(a, b, c)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}).$$

Finally, we quote a tightness result proved under the following assumptions.

Assumption 4.3. *Suppose, for $p \in \mathbb{N}_0$,*
(i) innovations: $\mathsf{E}|\varepsilon_i|^q < \infty$ *for some* $q > 4p$*;*
(ii) weights and regressors: $\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|^4(1+|n^{1/2}x_{in}|) = \mathrm{O}(1)$.

Theorem 4.3 (JN16, Theorem 4.4)**.** *Suppose Assumptions 3.1, 4.3 hold. Then,* $\forall \epsilon > 0$,

$$\lim_{\phi\downarrow 0}\limsup_{n\to\infty}\mathsf{P}\{\sup_{c,c^\dagger\in\mathbb{R}:|\mathsf{F}(c^\dagger)-\mathsf{F}(c)|\leq\phi}|\mathbb{F}_n^{w,p}(0,0,c^\dagger)-\mathbb{F}_n^{w,p}(0,0,c)|>\epsilon\}=0$$

4.3 Expansion of the empirical distribution function

We can now analyze the expansion of $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$ in (4.1). This gives an asymptotic uniform expansion of $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$ that is linear in the estimation errors $\tilde{a},\tilde{b}$.

Theorem 4.4. *Stochastic expansion.* *Suppose Assumptions 3.1, 4.1, 4.2(i), 4.3(i) hold for some* $0 < \eta \leq 1/4$ *so that* $\tilde{a} = n^{1/2}(\tilde{\sigma}-\sigma)/\sigma$ *and* $\tilde{b} = N^{-1}(\tilde{\beta}-\beta)/\sigma$ *are* $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$*. Then, the process* $\mathbb{F}_n^{w,p}(0,0,c)$ *is tight and uniformly in* $c \in \mathbb{R}$,

$$\begin{aligned} n^{1/2}\{&\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)-\bar{\mathsf{F}}_n^{w,p}(0,0,c)\} \\ &= \mathbb{F}_n^{w,p}(0,0,c)+\sigma^p c^p \mathsf{f}(c) n^{-1/2}\sum_{i=1}^n w_{in}(n^{-1/2}c\tilde{a}+x_{in}'\tilde{b})+\mathrm{o}_\mathsf{P}(1). \end{aligned} \tag{4.5}$$

Proof. Since $\tilde{a}$ and $\tilde{b}$ are $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$, the expansion follows by Lemma 3.1 and Theorems 4.1, 4.2 and the tightness follows by Theorem 4.3. □

The set of Assumptions in Theorem 4.4 simplify when distinguishing between the three cases where $p \in \mathbb{N}$ with either $r = 2$ or $r > 2$ and where $p = 0$. The latter case was also studied by Koul & Ossiander (1994).

Corollary 4.1. *Suppose Assumption 3.1 holds and consider the special case where* $r = 2$ *and* $p \in \mathbb{N}$*. Let* $\tilde{a} = n^{1/2}(\tilde{\sigma}-\sigma)/\sigma$ *and* $\tilde{b} = N^{-1}(\tilde{\beta}-\beta)/\sigma$ *be* $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$*. Suppose*
(i) η *satisfies* $0 < \eta \leq 1/4$ *and* $2 > 1 + (1/4-\eta)(1+\dim x)$*;*
(ii) $\mathsf{h}(c) = (1+|c|^{4p})\mathsf{f}(c)$ *is integrable and Lipschitz with weakly unimodal bounds;*
(iii) $\mathsf{E}n^{-1}\sum_{i=1}^n(1+|w_{in}|^4)(1+|n^{1/2}x_{in}|^2) = \mathrm{O}(1)$*;*
(iv) $c^q\mathsf{f}(c)$ *is integrable for some* $q > 4p$*.*
Then, the process $\mathbb{F}_n^{w,p}(0,0,c)$ *is tight and the expansion of* $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$ *in (4.5) holds.*

Proof. We verify Assumptions 4.1, 4.2(i), 4.3(i). Assumption 4.1 matches conditions (i), (ii), (iii). Assumption 4.3(i) matches condition (iv). For Assumption 4.2(i) part (a) requires integrability of $|c|^p\mathsf{f}(c)$, which follows from condition (ii). Part (b) requires Lipschitz of $c^p\mathsf{f}(c)$, $c^{p+1}\mathsf{f}(c)$, $c^{p+2}\mathsf{f}(c)$. This follows from condition (ii) due to Lemma A.6. Part (c) requires $(1+|c|)|c|^p\mathsf{f}(c)$ is bounded. This function is bounded by $\mathsf{h}(c) = (1+|c|^{4p})\mathsf{f}(c)$, when $p \in \mathbb{N}$. In turn, $\mathsf{h}(c)$ is bounded under condition (ii) due to Lemma A.4. □

Corollary 4.2. *Suppose Assumption 3.1 holds and consider the special case where* $r > 2$ *and* $p \in \mathbb{N}$*. Let* $\tilde{a} = n^{1/2}(\tilde{\sigma}-\sigma)/\sigma$ *and* $\tilde{b} = N^{-1}(\tilde{\beta}-\beta)/\sigma$ *be* $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$*. Suppose*
(i) η *satisfies* $0 < \eta \leq 1/4$ *and* $2^{r-1} > 1 + (1/4-\eta)(1+\dim x)$*;*

(ii) $\mathsf{h}(c) = (1 + |c|^{2^r p})\mathsf{f}(c)$ *is integrable and Lipschitz with weakly unimodal bounds;*
(iii) $\mathsf{E}n^{-1}\sum_{i=1}^n (1 + |w_{in}|^{2^r})(1 + |n^{1/2}x_{in}|^2) = \mathrm{O}(1)$.
Then, the process $\mathbb{F}_n^{w,p}(0,0,c)$ *is tight and the expansion of* $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$ *in (4.5) holds.*

Proof. The proof follows that of Corollary 4.2, except for one step. Assumption 4.2(ia) requires integrability of $|c|^q\mathsf{f}(c)$ for some $q > 4p$. This now follows from the integrability of $|c|^{2^r p}\mathsf{f}(c)$ in condition (ii) since $2^r p > 4p$ for $r > 2$. □

Corollary 4.3. *Suppose Assumption 3.1 holds and consider the special case where* $r \geq 2$ *and* $p = 0$. *Let* $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$ *and* $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)/\sigma$ *be* $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$. *Suppose*
(i) η *satisfies* $0 < \eta \leq 1/4$ *and* $2^{r-1} > 1 + (1/4 - \eta)(1 + \dim x)$*;*
(ii) $|c|^q\mathsf{f}(c)$ *is integrable for some* $q > 0$*;* $(1 + c^2)\mathsf{f}(c)$ *is Lipschitz; and* $\mathsf{f}(c)$ *has weakly unimodal bounds;*
(iii) $\mathsf{E}n^{-1}\sum_{i=1}^n (1 + |w_{in}|^{2^r})(1 + |n^{1/2}x_{in}|^2) = \mathrm{O}(1)$.
Then, the process $\mathbb{F}_n^{w,p}(0,0,c)$ *is tight and the expansion of* $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$ *in (4.5) holds.*

Proof. We need to verify Assumptions 4.1, 4.2(i), 4.3(i). For Assumption 4.1 note that part (i) follows from condition (ii), since it implies that f is integrable and Lipschitz with weakly unimodal bounds due to Lemma A.6. The remaining parts of Assumption 4.1 correspond to condition (i, iii). Assumption 4.3(i) requires integrability of $|c|^q\mathsf{f}(c)$ for some $q > 0$, which is stated in condition (ii). For Assumption 4.2(i) part (a) is trivially satisfied when $p = 0$. Part (b) requires Lipschitz of $\mathsf{f}(c)$, $c\mathsf{f}(c)$, $c^2\mathsf{f}(c)$, which follows from condition (ii) due to Lemma A.6. Part (c) requires that $\mathsf{f}(c)$ and $|c|\mathsf{f}(c)$ are bounded. This follows from Lemmas A.4, A.5, respectively, under condition (ii). □

Theorem 4.4 provides a stochastic expansion of $\mathsf{F}_n^{w,p}(\tilde{a},\tilde{b},c)$. When the weights are stationary or deterministic the limit will be a Gaussian process. This can be proved using the Central Limit Theorem for martingale difference arrays of Dvoretzky (1972).

Example 4.3. *Let the weights* w_{in} *satisfy* $\hat{\Sigma}_w = n^{-1}\sum_{i=1}^n w_{in}^2 \to \Sigma_w$ *in probability and* $\mathsf{E}\hat{\Sigma}_w \to \Sigma_w$. *Let* $\omega_{c,p}^2 = \mathsf{Var}\{\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c)}\}$. *Then* $\mathbb{F}_n^{w,p}(0,0,c)$ *converges to a Gaussian process with variance* $\Sigma_w \omega_{c,p}^2$. *The Dvoretzky (1972) result requires a Lindeberg condition, which is satisfied with fourth moments as in Assumption 4.3.*

Non-Gaussian limiting processes arise with random walk-type weights.

Example 4.4. *Let the weights* $w_{in} = n^{-1/2}\sum_{j=1}^i \eta_j$ *be normalized random walks where* η_j *are i.i.d. with zero mean and unit variance. For* $u \in [0,1]$ *let* $W_u, B_u^{(c)}$ *be independent standard Brownian motions. Let* $\mathrm{int}(nu)$ *denote the integer part of* nu*. Then* $w_{\mathrm{int}(nu),n}$ *and* $n^{-1/2}\sum_{i=1}^{\mathrm{int}(nu)}\{\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c)} - \mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c)}\}$ *converge in distribution to* $W_u, B_u^{(c)}\omega_{c,p}^2$*, for fixed c. Thus,* $\mathbb{F}_n^{w,p}(0,0,c)$ *converges to a process which, for each c, can be expressed as the stochastic integral* $\int_0^1 W_u dB_u^{(c)}\omega_{c,p}^2$*, see Chan & Wei (1988).*

5 Concluding remarks

The main result is Theorem 4.4, which gives an asymptotic uniformly linear expansion of the weighted and marked empirical distribution function of estimated residuals. The expansion has three terms. First, the compensator of the weighted and marked empirical distribution function applied to the true innovations. Second, the empirical process defined from weighted and marked empirical distribution function applied to the true innovations. Third, a bias term that is linear in the normalized estimation error.

The result generalizes previous work of Billingsley (1968) for empirical distribution functions of the true innovations and of Koul and Ossiander for the case without marks. The new proof corrects a mistake in the proof of JN16. In the process of writing the new proof the conditions have been made more accessible. In particular, the necessary smothness conditions have been formulated as a combination of a Lipschitz property and unimodal bounds. The conditions to the regressors have been simplified so that the present result also covers explosive regressors. This unifies separate proofs for explosive and non-explosive cases by Koul & Leventhal (1989) and Lee & Wei (1999).

The result can be used to analyze 1-step Huber-skip estimators which appear in various robust statistical procedures. They also appear implicitly in the common data analytic strategy of first estimating a least squares regression, dropping observations with large residuals and then reestimating a regression on the selected observations. Johansen & Nielsen (2016b) review these results.

A Proofs

For sequences s_n, t_n we say $s_n \sim t_n$ if $s_n = \mathrm{O}(t_n)$ and $t_n = \mathrm{O}(s_n)$. The weights w_{in} may be matrix valued. To show that the resulting matrix of empirical processes vanishes, it suffices to show this for each element. Thus, we proceed in this appendix as if w_{in} is scalar. Throughout the rest of the Appendix we denote by C a generic constant, which need not be the same in different expressions. Let $\mathrm{int}(x)$ denote the integer part of x.

A.1 Iterated martingale inequality

In the proofs we will make frequent use of the following iterated exponential martingale inequality which builds on the exponential martingale inequality by Bercu & Touati (2008), see also Bercu et al. (2015), Bercu & Touati (2018).

Lemma A.1 (JN16, Lemma 4.2)**.** *For $1 \leq \ell \leq L_n$ and $1 \leq i \leq n$ let $z_{n\ell i}$ be $\mathcal{F}_{in}$ adapted and $\mathsf{E} z_{n\ell i}^{2^r} < \infty$ for some $r \in \mathbb{N}$. Let $D_{nq} = \max_{1\leq\ell\leq L_n} \sum_{i=1}^n \mathsf{E}_{i-1} z_{n\ell i}^{2^q}$ for $1 \leq q \leq r$. Suppose, for some $\lambda > 0$, $\varsigma \geq 0$, that $L_n = \mathrm{O}(n^\lambda)$ and $\mathcal{E}_{nq} = \mathsf{E} D_{nq} = \mathrm{O}(n^\varsigma)$ for $1 \leq q \leq r$. Then, if $\upsilon > 0$ is chosen such that*

(i) $\varsigma < 2\upsilon$,   (ii) $\varsigma + \lambda < \upsilon 2^r$,

it holds that $\max_{1\leq\ell\leq L_n} |\sum_{i=1}^n (z_{n\ell i} - \mathsf{E}_{i-1} z_{n\ell i})| = \mathrm{o}_{\mathsf{P}}(n^\upsilon)$.

A.2 Metric and cover

The chaining argument is based on a finite number of points $c_k \in \mathbb{R}$, $k = 0, 1, \ldots, K$, which define a cover of $\mathbb{R}$ by the K intervals

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty. \tag{A.1}$$

The definitions of c_0 and c_K are convenient even when the support is finite. In JN16 these chaining points are chosen using the function

$$\mathsf{H}_r(c) = \int_{-\infty}^{c}(1 + |u|^{2^r p})\mathsf{f}(u)du = \mathsf{E}(1 + |\varepsilon_i/\sigma|^{2^r p})1_{(\varepsilon_i/\sigma \le c)}, \tag{A.2}$$

for a given $r = 0, 1, \ldots,$ such that the intervals have the same size when measured by the increments of H_r, that is,

$$\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \qquad \text{for } k = 0, 1, \ldots, K. \tag{A.3}$$

The function H_r is increasing in c and bounded when

$$H_r = \mathsf{H}_r(\infty) = \int_{-\infty}^{\infty}(1 + |u|^{2^r p})\mathsf{f}(u)du = \mathsf{E}(1 + |\varepsilon_i/\sigma|^{2^r p}) < \infty. \tag{A.4}$$

The inequality $|\varepsilon^s| < 1 + |\varepsilon|^r$ for $0 \le s \le r$ implies that, for $c \le c^\dagger$,

$$\mathsf{E}\{|\varepsilon_i/\sigma|1_{(c<\varepsilon_i/\sigma\le c^\dagger)}\}^{2^s p} \le \mathsf{E}(1 + |\varepsilon_i/\sigma|^{2^r p})1_{(c<\varepsilon_i/\sigma\le c^\dagger)} = \mathsf{H}_r(c^\dagger) - \mathsf{H}_r(c). \tag{A.5}$$

We refer to $\mathsf{H}_r(c^\dagger) - \mathsf{H}_r(c)$ as the H_r-distance between c and $c^\dagger$.

The count K will be a function of n. In the different proofs, the count K and the power r will be chosen differently. In the chaining argument, we compare H_r evaluated at c and at $c + n^{-1/2}ac + x_{in}'b$. The proofs consider the additive perturbation $c + x_{in}'b$ for $a = 0$, in Lemma A.9 and the multiplicative perturbation $c + n^{-1/2}ac = c(1 + n^{-1/2}a)$ for $b = 0$ in Lemma A.10. These will be used in the subsequent proofs of Lemmas 4.1, 4.2, when chaining over a, c and b, c respectively.

A.3 Weakly unimodal functions

We consider a reformulation of the condition in JN16, §B.5.

Lemma A.2. *Let $\mathsf{h}(c) \ge 0$ have support on an open interval $\mathcal{S} \subset \mathbb{R}$. Then, $\mathsf{h}(c)$ has a weakly unimodal bound $u(c)$ (Definition 4.1) if and only if h has the bound $\underline{\mathsf{h}}(c) \le \mathsf{h}(c) \le \overline{\mathsf{h}}(c) \le C_{\mathsf{h}}\underline{\mathsf{h}}(c)$ for all $c \in \mathcal{S}$ and a constant $C_{\mathsf{h}} \ge 1$ and where for a point c_{h}*

$$\underline{\mathsf{h}}(c) = \{\inf_{c_{\mathsf{h}} \le d \le c} \mathsf{h}(d)\}1_{(c\ge0)} + \{\inf_{-c\le d\le c_{\mathsf{h}}} \mathsf{h}(d)\}1_{(c<0)},$$
$$\overline{\mathsf{h}}(c) = \{\sup_{c\le d<\overline{c}} \mathsf{h}(d)\}1_{(c\ge0)} + \{\sup_{\underline{c}<d\le -c} \mathsf{h}(d)\}1_{(c<0)}.$$

Proof. " $\Leftarrow$ " : The functions $\underline{\mathsf{h}}$ and $\overline{\mathsf{h}}$ are weakly unimodal.

" $\Rightarrow$ " : Choose $c_{\mathsf{h}} = c_{\text{mode}}$. Then, $\underline{\mathsf{h}}$ is the largest weakly unimodal function less than h, while $\overline{\mathsf{h}}$ is the smallest weakly unimodal function larger than h. Therefore, $u(c) \le \underline{\mathsf{h}}(c) \le \mathsf{h}(c)$ and $\mathsf{h}(c) \le \overline{\mathsf{h}}(c) \le C_u u(c)$. Since $u(c) \le \underline{\mathsf{h}}(c)$ we can choose $C_{\mathsf{h}} = C_u$. □

We note that if $\mathsf{h}(c) = 0$ then the weakly unimodal bound is also zero, that is $u(c) = 0$. Further, the class of functions with weakly unimodal bound is closed under multiplication with a positive constant. Thus, if h has a weakly unimodal bound so has $C\mathsf{h}$ for any $C > 0$.

We show that the class of non-negative functions with weakly unimodal bound is closed to taking minimum, maximum, addition and multiplication.

Lemma A.3. *Let* $\mathsf{h}_1, \mathsf{h}_2$ *have support on open intervals* $\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{R}$ *and suppose* $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \emptyset$. *Let* f *be a function from* $\mathbb{R}^2$ *to* $[0, \infty[$ *that can represent addition, multiplication or taking maximum or minimum. If each function* h_j *has a weakly unimodal bound, see Definition 4.1, then* $f\{h_1(x), h_2(x)\}$ *also has a weakly unimodal bound.*

Proof. By assumption there exists $\bar{x} \in \mathcal{S}_1 \cap \mathcal{S}_2$. Let $\bar{v} = \min\{v_1(\bar{x}), v_2(\bar{x})\}$, so that $\bar{v}_i(x) = v_i(x) \wedge \bar{x}$ are two weakly unimodal functions. The function f is non-decreasing in its arguments. Therefore, $f\{\bar{v}_1(x), \bar{v}_2(x)\}$ is weakly unimodal because $\bar{v}_i(x)$ is non-decreasing for $x \leq \bar{x}$, and non-increasing for $x \geq \bar{x}$, such that the same holds for $f\{\bar{v}_1(x), \bar{v}_2(x)\}$.

To see that $f\{\bar{v}_1(x), \bar{v}_2(x)\}$ provides weakly unimodal bounds, recall $v_i(c) \leq \mathsf{h}(c) \leq C_i v_i(x)$ and let $v_i^{\max} = \max_{x \in \mathcal{S}_i} v_i(x)$ to get the lower bound

$$f\{\bar{v}_1(x), \bar{v}_2(x)\} \leq f\{v_1(x), v_2(x)\} \leq f\{\mathsf{h}_1(x), \mathsf{h}_2(x)\},$$

and the upper bound, for $C = \max(C_1 v_1^{\max}, C_2 v_2^{\max})/\bar{v}$,

$$\begin{aligned} f\{\mathsf{h}_1(x), \mathsf{h}_2(x)\} \leq f\{C_1 v_1(x), C_2 v_2(x)\} &\leq f\{C_1 \frac{v_1^{\max}}{\bar{v}} \bar{v}_1(x), C_2 \frac{v_2^{\max}}{\bar{v}} \bar{v}_2(x)\} \\ &\leq f\{C\bar{v}_1(x), C\bar{v}_2(x)\} \leq f(C, C) f\{\bar{v}_1(x), \bar{v}_2(x)\}, \end{aligned}$$

where the last inequality uses the assumed functional form of f. $\square$

A.4 Bounds on the distance function

In the following lemmas, we provide uniform bounds for $\dot{\mathsf{H}}_r$ and of the increment for $\dot{\mathsf{H}}_r$ over two points c_1, c_2. Lemma A.8, extracts the main argument in the proof of Lemma B.1 in JN16 with a simplified proof and weaker conditions replacing differentiability with Lipschitz continuity and allowing a finite support. The results are derived for a general function $\dot{\mathsf{H}}$, as described in Assumption A.1 below. This general result will then be applied to the particular function $\dot{\mathsf{H}}_r$ in (A.2).

Assumption A.1. *Let* $\mathsf{H}(c) = \int_{-\infty}^{c} \dot{\mathsf{H}}(x) dx$, *where* $\dot{\mathsf{H}}(x) \geq 0$. *Suppose*
(i) *the support of* $\dot{\mathsf{H}}$, *that is* $\mathcal{S} = \{c : \dot{\mathsf{H}}(c) > 0\}$, *is an interval with endpoints* $\underline{c}, \bar{c}$ *so that* $-\infty \leq \underline{c} < \bar{c} \leq \infty$;
(ii) $H = \mathsf{H}(\infty) = \int_{\mathbb{R}} \dot{\mathsf{H}}(c) dc < \infty$;
(iii) $\dot{\mathsf{H}}$ *is Lipschitz (Definition 4.1);*
(iv) $\dot{\mathsf{H}}$ *has weakly unimodal bounds (Definition 4.1):* $\exists C_u \geq 1, u(c), \forall c \in \mathbb{R}$:

$$0 \leq u(c) \leq \dot{\mathsf{H}}(c) \leq C_u u(c). \tag{A.6}$$

Lemma A.4. *Suppose Assumption A.1*(i)-(iii) *is satisfied. Then,* $\dot{\mathsf{H}}$ *is bounded. Further, if* $\bar{c} = \infty$, *then* $\dot{\mathsf{H}}(c) \to 0$, *for* $c \to \infty$ *and if* $\underline{c} = -\infty$, *then* $\dot{\mathsf{H}}(c) \to 0$, *for* $c \to -\infty$.

Proof. Let c_0 be an interior point of $\mathcal{S}$.

If $\bar{c} < \infty$, since $\dot{\mathsf{H}}$ is Lipschitz continuous by Assumption A.1(iii), it has a continuous extension to $\bar{c}$ so we can define $\dot{\mathsf{H}}(\bar{c})$, and it is therefore bounded on $[c_0, \bar{c}]$. The same argument shows that $\dot{\mathsf{H}}$ can be extended to $\underline{c}$ and is bounded on $[\underline{c}, c_0]$ if $\underline{c} > -\infty$.

If $\bar{c} = \infty$ and $c_0 \leq c_1 < c$ then $\infty \geq c_1 + \dot{\mathsf{H}}(c_1)/C_L$ and $c_0 \leq c_1$. Further, the Lipschitz Assumption A.1(iii) implies $\dot{\mathsf{H}}(c) \geq C_L \max\{0, c_1 + \dot{\mathsf{H}}(c_1)/C_L - c\}$. We find that

$$\int_{c_0}^{\infty} \dot{\mathsf{H}}(c)dc \geq \int_{c_1}^{c_1+\dot{\mathsf{H}}(c_1)/C_L} C_L\{c_1 + \dot{\mathsf{H}}(c_1)/C_L - c\}dc = \{\dot{\mathsf{H}}(c_1)\}^2/(2C_L).$$

It then follows from Assumption A.1(ii), that $\max_{c_1 \geq c_0}\{\dot{\mathsf{H}}(c_1)\}^2 \leq 2C_L \int_{c_0}^{\infty} \dot{\mathsf{H}}(c)dc < \infty$, so that $\dot{\mathsf{H}}$ is bounded on $[c_0, \infty[$ and $\dot{\mathsf{H}}(c_0) \to 0$ for $c_0 \to \infty$. A similar argument for $\underline{c} = -\infty$ shows that $\dot{\mathsf{H}}$ is bounded on $]-\infty, c_0]$ and that $\dot{\mathsf{H}}(c_0) \to 0$ for $c_0 \to -\infty$. $\square$

Lemma A.5. *Suppose Assumption A.1 is satisfied. Then* $|c|\dot{\mathsf{H}}(c)$ *is bounded.*

Proof. Assumption A.1(iv) implies that $\dot{\mathsf{H}}(c) \leq C_u u(c)$ where $u(c)$ is non-negative and non-increasing on $c \geq c_{\text{mode}}$ so that

$$\int_{c_{\text{mode}}}^{\infty} u(x)dx \geq \int_{c_{\text{mode}}}^{c} u(x)dx \geq \int_{c_{\text{mode}}}^{c} u(c)dx = (c - c_{\text{mode}})u(c).$$

Rearranging the inequality and using that u is non-increasing gives

$$cu(c) \leq \int_{c_{\text{mode}}}^{\infty} u(x)dx + c_{\text{mode}}u(c) \leq \int_{c_{\text{mode}}}^{\infty} u(x)dx + c_{\text{mode}}u(c_{\text{mode}}) < \infty.$$

The bound is uniform in c since, first, the integral is finite by Assumption A.1(ii, iv), second, the mode is finite and, third, $u(c_{\text{mode}})$ is finite due to Lemma A.4 using Assumption A.1(i)-(iii). Hence, $c\dot{\mathsf{H}}(c) \leq C_u cu(c) < \infty$. A similar argument can be made for $c \leq c_{\text{mode}}$, which combining gives $|c|\dot{\mathsf{H}}(c) < \infty$. $\square$

The following results are used when seeking to simplify the various Lipschitz and boundedness conditions.

Lemma A.6. *Let* $\mathsf{h}(c) = (1 + |c|^q)\mathsf{f}(c)$ *be bounded and Lipschitz (Definition 4.1) for some* $q > 0$. *Then,* $|c|^p\mathsf{f}(c)$ *is Lipschitz for any* $p \in \mathbb{N}_0$ *such that* $p \leq q$.

Proof. Consider $||c|^p\mathsf{f}(c) - |c^\dagger|^p\mathsf{f}(c^\dagger)|$. Write $|c|^p\mathsf{f}(c) = \{|c|^p/(1 + |c|^q)\}\mathsf{h}(c)$ and add and subtract $\{|c^\dagger|^p/(1 + |c^\dagger|^q)\}\mathsf{h}(c)$ so that, $\forall c, c^\dagger \in \mathcal{S}$,

$$||c|^p\mathsf{f}(c) - |c^\dagger|^p\mathsf{f}(c^\dagger)| \leq |\frac{|c|^p}{1 + |c|^q} - \frac{|c^\dagger|^p}{1 + |c^\dagger|^q}|\mathsf{h}(c) + \frac{|c^\dagger|^p}{1 + |c^\dagger|^q}|\mathsf{h}(c) - \mathsf{h}(c^\dagger)|.$$

For the first term use that $|c|^p/(1 + |c|^q)$ is Lipschitz as it has right and left bounded derivatives for all c, while $\mathsf{h}(c)$ is bounded by Assumption. For the second term use that $|c^\dagger|^p/(1 + |c^\dagger|^q)$ is bounded by unity, while $\mathsf{h}(c)$ is Lipschitz by Assumption. $\square$

Lemma A.7. *Let* $\mathsf{h}_q(c) = (1 + |c|^q)\mathsf{f}(c)$ *satisfy Assumption A.1 for some* $q \in \mathbb{N}_0$. *Then,* $\mathsf{h}_m(c) = (1 + |c|^m)\mathsf{f}(c)$ *satisfies Assumption A.1 for any* $m \in \mathbb{N}_0$ *such that* $m < q$.

Proof. We check the four parts to Assumption A.1 for $\mathsf{h}_m(c)$.
Part (i): the functions $\mathsf{h}_q(c)$ and $\mathsf{h}_m(c)$ have the same support.
Part (ii): integrability of $\mathsf{h}_q(c)$ implies integrability of $\mathsf{h}_m(c)$ as $m < q$.
Part (iii): by Lemma A.6 using Assumption A.1(i)-(iii), we have that $\mathsf{f}(c)$ and $|c|^m\mathsf{f}(c)$ are Lipschitz and so is their sum.

Part (iv): Write $\mathsf{h}_m(c) = v(|c|)\mathsf{h}_q(c)$ where $v(x) = (1+x^m)/(1+x^q)$ for $x \geq 0$. We first argue that $v(x)$ is strictly decreasing in $x \geq 1$ for $m < q$. The derivative is

$$\dot{v}(x) = \frac{mx^{m-1}(1+x^q) - (1+x^m)qx^{q-1}}{(1+x^q)^2} = \frac{mqx^{m+q-1}}{(1+x^q)^2}\{(1+x^{-q})/q - (1+x^{-m})/m\},$$

which is negative since $1 + x^{-q} < 1 + x^{-m}$ for $m < q$ and $x \geq 1$. Since $v(x)$ is strictly decreasing for $x \geq 1$ and continuous on $\mathbb{R}$ then $v(|c|)$ must have a mode in some $|c| \leq 1$. Thus, $v(|c|)$ is weakly unimodal, see Definition 4.1. The product $\mathsf{h}_m(c) = v(|c|)\mathsf{h}_q(c)$ of the weakly unimodal functions is weakly unimodal, see Lemma A.3. $\square$

Lemma A.8. *Let $\dot{\mathsf{H}}$ satisfy Assumption A.1. Then there exist constants C_{H}, $K_0 > 0$, so that for any $K > K_0$ and any $c_1 < c_2$ so that $\mathsf{H}(c_2) - \mathsf{H}(c_1) \leq H/K$, then*

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq C_{\mathsf{H}}/K^{1/2}. \tag{A.7}$$

Proof. 1. Tail behavior of $\dot{\mathsf{H}}$. Recall $\mathcal{S}$ is the support of $\dot{\mathsf{H}}$ with endpoints $\underline{c} < \overline{c}$. We have three types of behaviour of $\dot{\mathsf{H}}$ around $\overline{c}$, and a similar situation around $\underline{c}$:

1.1: $\overline{c} < \infty$ and $\dot{\mathsf{H}}$ is not continuous at $\overline{c}$, that is, $\lim_{c\uparrow\overline{c}} \dot{\mathsf{H}}(c) > 0 = \lim_{c\downarrow\overline{c}} \dot{\mathsf{H}}(c)$;

1.2: $\overline{c} < \infty$ and $\dot{\mathsf{H}}$ is continuous at $\overline{c}$, that is, $\dot{\mathsf{H}}(\overline{c}) = 0$;

1.3: $\overline{c} = \infty$, and $\dot{\mathsf{H}}(c) \to 0, c \to \infty$.

Examples are uniform, triangular and normal densities, respectively, see Example 4.1.
2. *Existence of c_+, c_-.* Let c_{mode} be a mode of the unimodal bound u. Let $c_{\mathsf{median}} = \mathsf{H}^{-1}(H/2)$, noting that H is strictly increasing on $\mathcal{S}$. We first show that, in the three cases 1.1-1.3 above, for large K, a c_+ exists so that $c_{\mathrm{mode}} \leq c_+ \leq \overline{c}$ and

$$\dot{\mathsf{H}}(c_+) \geq H/K^{1/2} \qquad \text{and} \qquad \mathsf{H}(c_+) \geq \mathsf{H}(c_{\mathsf{median}}) + H/K. \tag{A.8}$$

Similarly, a c_- exists so that $\underline{c} \leq c_- \leq c_{\mathrm{mode}}$ and

$$\dot{\mathsf{H}}(c_-) \geq H/K^{1/2} \qquad \text{and} \qquad \mathsf{H}(c_-) \leq \mathsf{H}(c_{\mathsf{median}}) - H/K. \tag{A.9}$$

For later use, it is relevant to point out that, given the c_{median} conditions (A.8) and (A.9), c_+ and c_- are separated by at least $2H/K$ while, by assumption, c_1 and c_2 are separated by at most H/K. We start by showing the existence of c_+ in cases 1.1-1.3.

Case 1.1: Define $c_{\mathrm{mode}} \leq c_+ = \overline{c}$. Since $\lim_{c\uparrow\overline{c}} \dot{\mathsf{H}}(c) > 0$ we can choose K so large that (A.8) is satisfied.

Case 1.2: Due to the continuity of $\dot{\mathsf{H}}$ at $\overline{c}$ then, for large K, there exist $c_+ > c_{\mathrm{mode}}$ so that (A.8) is satisfied with equality, that is,

$$\dot{\mathsf{H}}(c_+) = H/K^{1/2}. \tag{A.10}$$

Case 1.3: By Lemma A.4, $\dot{\mathsf{H}}$ vanishes in the tails. Thus, for large K there exist c_+ so that $c_{\mathrm{mode}} < c_+$ and (A.8) and (A.10) are satisfied.

In all three cases, 1.1-1.3, c_+ is in the tail of the distribution, hence, the median condition in (A.8) is satisfied. The derivation for c_- is analogous.
3. An inequality for $|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)|$. By the Lipschitz condition in Assumption A.1(*iii*)

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq C_L(c_2 - c_1), \tag{A.11}$$

while the definition of $c_1 < c_2$ and the mean value theorem applied to H give

$$H/K \geq \mathsf{H}(c_2) - \mathsf{H}(c_1) = (c_2 - c_1)\dot{\mathsf{H}}(c^*), \tag{A.12}$$

for c^* satisfying $c_1 < c^* < c_2$. Eliminating $c_2 - c_1$ from (A.11), (A.12) shows that

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq C_L H/\{K\dot{\mathsf{H}}(c^*)\}. \tag{A.13}$$

4. *Bounding* $|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)|$. The desired results follow by showing that

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq C_L C_u/K^{1/2} + 2C_u H/K^{1/2} = (C_L + 2H)C_u/K^{1/2}, \tag{A.14}$$

which we do next. The proof depends on the position of c_1, c_2 relative to c_-, c_+. First, note that the case $c_1 < c_- < c_+ < c_2$ is ruled out by the construction in item 2, since c_+ and c_- are at least two H/K intervals apart while c_1 and c_2 are at most one H/K interval apart. Hence, there are three possible situations:

4.1: $c_- \leq c_1 < c_2 \leq c_+$; 4.2: $c_+ \leq c_1 < c_2$; 4.3: $c_- \leq c_1 \leq c_+ \leq c_2$,

while the cases $c_1 \leq c_- \leq c_2 \leq c_+$ and $c_1 < c_2 \leq c_+$ are symmetric to 4.3 and 4.2.

Case 4.1: $c_- \leq c_1 < c^* < c_2 \leq c_+$. We first show $\dot{\mathsf{H}}(c^*) \geq C_u^{-1}H/K^{1/2}$ when $c_{\text{mode}} \leq c^*$. The bound in Assumption A.1(iv) shows that $\dot{\mathsf{H}}(c^*) \geq u(c^*)$. The unimodality of u implies $u(c^*) \geq \lim_{c\uparrow c_+} u(c)$ so that $\dot{\mathsf{H}}(c^*) \geq \lim_{c\uparrow c_+} u(c)$. The ordering in (A.6) gives $u(c) \geq C_u^{-1}\dot{\mathsf{H}}(c)$ so that $\dot{\mathsf{H}}(c^*) \geq C_u^{-1}\lim_{c\uparrow c_+}\dot{\mathsf{H}}(c)$. The construction of c_+ in (A.8) and (A.10) shows that $\lim_{c\uparrow c_+}\dot{\mathsf{H}}(c) \geq H/K^{1/2}$ for cases 1.1-1.3 so that $\dot{\mathsf{H}}(c^*) \geq C_u^{-1}H/K^{1/2}$.

If $c^* \leq c_{\text{mode}}$, we show $\dot{\mathsf{H}}(c^*) \geq C_u^{-1}H/K^{1/2}$ by comparing c^* with c_- instead of c_+.

Now, insert $\dot{\mathsf{H}}(c^*) \geq C_u^{-1}H/K^{1/2}$ in (A.13)

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq C_L H/\{KC_u^{-1}H/K^{1/2}\} = C_L C_u/K^{1/2}, \tag{A.15}$$

which is the first term of (A.14).

Case 4.2: $c_+ \leq c_1 < c_2$. For the case 1.1, where $\overline{c} < \infty$ and $\dot{\mathsf{H}}$ non-continuous, we have that $c_+ = \overline{c}$ so that $\mathsf{H}(c_+) = \mathsf{H}(c_1) = \mathsf{H}(c_2) = \mathsf{H}(\infty) = H$, and hence $|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| = 0$.

For the cases 1.2, 1.3 use the triangle inequality and the bound (A.6) to get

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq \dot{\mathsf{H}}(c_2) + \dot{\mathsf{H}}(c_1) \leq C_u u(c_2) + C_u u(c_1).$$

Noting $c_{\text{mode}} \leq c_+ \leq c_1 < c_2$, the weak unimodality for u, the ordering (A.6) and the construction (A.10) with equality give

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq 2C_u u(c_+) \leq 2C_u\dot{\mathsf{H}}(c_+) = 2C_u H/K^{1/2}. \tag{A.16}$$

This is the second term of (A.14).

Case 4.3: $c_- \leq c_1 \leq c_+ \leq c_2$. Add and subtract $\dot{\mathsf{H}}(c_+)$ so that

$$|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_1)| \leq |\dot{\mathsf{H}}(c_+) - \dot{\mathsf{H}}(c_1)| + |\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_+)|,$$

by the triangle inequality. The first term is an example of case 4.1, so that (A.15) shows that $|\dot{\mathsf{H}}(c_+) - \dot{\mathsf{H}}(c_1)| \leq C_L C_u/K^{1/2}$, which is the first term in (A.14). The second term is an example of case 4.2, so that (A.16) shows that $|\dot{\mathsf{H}}(c_2) - \dot{\mathsf{H}}(c_+)| \leq 2C_u H/K^{1/2}$. Combine these results to complete the proof of (A.14). □

We can now establish a covering for local variations in b and c. This will be used in the subsequent proof of Lemma 4.1, when chaining over b, c. The result replaces Lemma B.2 in JN16 and clarifies how the covering depends on x_{in}. For an additive perturbation $c + x_{in}'b$, as a basis for the chaining argument, we choose balls with centers b_m and radius δ and construct a new covering of $\mathbb{R}$ based on intervals $[\underline{c}_{kmi}, \overline{c}_{kmi}]$, see (A.17). We then find a bound on the H_r distance of these intervals. The Lemma does not exploit the particular structure of the H_r-function so we formulate it for a general function H.

Lemma A.9. *Consider k, m, x_{in} and $\delta > 0$. Suppose $\mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = H/K$, $\forall k$. Let*

$$\underline{c}_{kmi} = c_{k-1} + x_{in}'b_m - |x_{in}|\delta, \qquad \overline{c}_{kmi} = c_k + x_{in}'b_m + |x_{in}|\delta.$$

Then, $\forall b, c$ so that $|b - b_m| \leq \delta$ and $c_{k-1} < c \leq c_k$,

$$\underline{c}_{kmi} < c + x_{in}'b \leq \overline{c}_{kmi}. \tag{A.17}$$

Suppose in addition that $\dot{\mathsf{H}}$ satisfies Assumption A.1. Then, $\exists C > 0$, $\forall m, x_{in}$,

$$\max_{0<k\leq K} |\mathsf{H}(\overline{c}_{kmi}) - \mathsf{H}(\underline{c}_{kmi})| \leq C(K^{-1} + |x_{in}'b_m|K^{-1/2} + |x_{in}|\delta + |x_{in}'b_m|^2). \tag{A.18}$$

Proof. 1. *Proof of (A.17).* Since $c_{k-1} < c \leq c_k$ and $|b - b_m| \leq \delta$,

$$c_{k-1} - |x_{in}|\delta < c + x_{in}'(b - b_m) \leq c_k + |x_{in}|\delta.$$

It follows by adding $x_{in}'b_m$ that $\underline{c}_{kmi} < c + x_{in}'b \leq \overline{c}_{kmi}$.
2. *Proof of (A.18).* Note H is well-defined by Assumption A.1(ii). Let

$$\mathcal{H}_{kmi} = \mathsf{H}(\overline{c}_{kmi}) - \mathsf{H}(\underline{c}_{kmi}) = \mathsf{H}(c_k + x_{in}'b_m + |x_{in}|\delta) - \mathsf{H}(c_{k-1} + x_{in}'b_m - |x_{in}|\delta). \tag{A.19}$$

Using the mean value theorem, we get for intermediate points c^*, c_* that

$$\mathsf{H}(c_k + x_{in}'b_m + |x_{in}|\delta) = \mathsf{H}(c_k + x_{in}'b_m) + |x_{in}|\delta\dot{\mathsf{H}}(c^*), \tag{A.20}$$

$$\mathsf{H}(c_{k-1} + x_{in}'b_m - |x_{in}|\delta) = \mathsf{H}(c_{k-1} + x_{in}'b_m) - |x_{in}|\delta\dot{\mathsf{H}}(c_*). \tag{A.21}$$

Using the mean value theorem once again, we get for intermediate points c^{**}, c_{**} that

$$\mathsf{H}(c_k + x_{in}'b_m) = \mathsf{H}(c_k) + x_{in}'b_m\dot{\mathsf{H}}(c_k) + x_{in}'b_m\{\dot{\mathsf{H}}(c^{**}) - \dot{\mathsf{H}}(c_k)\}, \tag{A.22}$$

$$\mathsf{H}(c_{k-1} + x_{in}'b_m) = \mathsf{H}(c_{k-1}) + x_{in}'b_m\dot{\mathsf{H}}(c_{k-1}) + x_{in}'b_m\{\dot{\mathsf{H}}(c_{**}) - \dot{\mathsf{H}}(c_{k-1})\}. \tag{A.23}$$

The Lipschitz condition in Assumption A.1(iii) implies that

$$|\dot{\mathsf{H}}(c^{**}) - \dot{\mathsf{H}}(c_k)| \leq C_L|c^{**} - c_k| \leq C_L|x_{in}'b_m|.$$

Likewise $|\dot{\mathsf{H}}(c_{**}) - \dot{\mathsf{H}}(c_{k-1})| \leq C_L|x_{in}'b_m|$. Inserting the expressions (A.20)-(A.23) and the Lipschitz bounds in the equation (A.19) for $\mathcal{H}_{kmi}$ gives

$$|\mathcal{H}_{kmi}| \leq |\mathsf{H}(c_k) - \mathsf{H}(c_{k-1})| + |x_{in}'b_m||\dot{\mathsf{H}}(c_k) - \dot{\mathsf{H}}(c_{k-1})| + (x_{in}'b_m)^2C_L + |x_{in}|\delta|\dot{\mathsf{H}}(c^*) + \dot{\mathsf{H}}(c_*)|.$$

For the first term, note $\mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = H/K$ by assumption. For the second term, apply $\dot{\mathsf{H}}(c_k) - \dot{\mathsf{H}}(c_{k-1}) \leq C_H/K^{1/2}$ uniformly in k by Lemma A.8 using Assumption A.1. For the third term, apply that $C_L < \infty$. For the fourth term, note that $\sup_{c\in\mathbb{R}} \dot{\mathsf{H}}(c) < \infty$ by Lemma A.4. Thus, the desired result follows, uniformly in k. $\square$

The following Lemma bounds H_r distances of multiplicative perturbations. It also gives an estimate of the number of c_k intervals, that are needed to cover the perturbation. This is used in the proof of Lemma 4.2, when chaining over a, c. The result does not use the particular structure of H_r and applies for general distance functions.

Lemma A.10. *Let $c_a = c(1 + n^{-1/2}a)$ so that $c_0 = c$. Suppose $\mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = H/K$ for all k, that $\dot{\mathsf{H}}$ is continuous on its support $\mathcal{S}$ and $\sup_{c\in\mathbb{R}} |c|\dot{\mathsf{H}}(c) < \infty$.*
(a) A constant $C > 0$ exists so that for all $\zeta > 0$,

$$\sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/2-\zeta}B} |\mathsf{H}(c_a) - \mathsf{H}(c)| \leq Cn^{-\zeta}.$$

(b) Choose an index $k(c_a)$ and grid points $c_{k(c_a)}$ so that $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$. Then, the number of grid points between c_a and c satisfies

$$\sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/2-\zeta}B} |k(c_a) - k(c_0)| \leq 2 + Cn^{-\zeta}K/H.$$

Proof. (a) *The distance $\mathcal{H} = \mathsf{H}(c_a) - \mathsf{H}(c)$.* Because $c_a - c = n^{-1/2}ac$, the mean value theorem gives $\mathcal{H} = n^{-1/2}ac\dot{\mathsf{H}}(\tilde{c})$ for an intermediate point $\tilde{c}$, so $|\tilde{c} - c| \leq n^{-1/2}|ac|$. This implies that $|c| \leq |\tilde{c}| + |c - \tilde{c}| \leq |\tilde{c}| + n^{-1/2}|a||c|$. Solving for $|c|$, we get

$$|c| \leq \frac{|\tilde{c}|}{1 - n^{-1/2}|a|} \leq 2|\tilde{c}|,$$

since $n^{-1/2}|a| \leq 1/2$ for large n. This gives $|\mathcal{H}| \leq |n^{-1/2}a|2|\tilde{c}|\dot{\mathsf{H}}(\tilde{c})$. By Assumption, $|\tilde{c}|\dot{\mathsf{H}}(\tilde{c})$ is bounded uniformly in c, a, while $|a| \leq n^{1/2-\zeta}B$ so that $|\mathcal{H}|\leq Cn^{-\zeta}$ as desired.

(b) *Translating the distance $\mathcal{H}$ in item (a) into a number of grid points.* We start by bounding $\mathcal{H}^* = |\mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}\{c_{k(c_0)}\}|$. Add and subtract $\mathsf{H}(c_a)$ and $\mathsf{H}(c)$ and apply the triangle inequality to get

$$\mathcal{H}^* \leq |\mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}(c_a)| + |\mathsf{H}\{c_{k(c_0)}\} - \mathsf{H}(c)| + |\mathsf{H}(c_a) - \mathsf{H}(c)|. \tag{A.24}$$

Each of the first two terms in (A.24) are bounded by H/K. Indeed, since $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$ and, noting that $c_0 = c$,

$$|\mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}(c_a)| \leq |\mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}\{c_{k(c_a)-1}\}| = H/K.$$

The third term in (A.24) equals $|\mathcal{H}|$ and satisfies $|\mathcal{H}|\leq Cn^{-\zeta}$ as shown in part (a). Overall

$$\mathcal{H}^* \leq 2H/K + Cn^{-\zeta} = (2 + Cn^{-\zeta}K/H)H/K,$$

implying that $|k(c_a) - k(c_0)| \leq 2 + Cn^{-\zeta}K/H$ uniformly in a, c. $\square$

B Proof of empirical process results

Proof of Lemma 4.1. Let $z_{ibc} = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i\leq c+\sigma^{-1}x_{in}'b)} - 1_{(\varepsilon_i\leq c)}\}$ and write

$$R_n(b, c) = \mathbb{F}_n^{w,p}(0, b, c) - \mathbb{F}_n^{w,p}(0, 0, c) = n^{-1/2}\textstyle\sum_{i=1}^n(z_{ibc} - \mathsf{E}_{i-1}z_{ibc}).$$

We show $\mathcal{R}_n = \sup_{c\in\mathbb{R}} \sup_{|b|\leq n^{1/4-\eta}B} |R_n(b,c)| = \mathsf{o}_\mathsf{P}(1)$.

1. *The chaining setup.*

1.1. *Choice of covering radius.* We chain over b and c. Later, in item 3.6, we choose a small $\delta > 0$ which will be used to form the covering balls for b and c in items 1.2 and 1.3. We note that $\delta > 0$ will be chosen independently of the sample size n.

1.2. *Construct c-balls.* We choose r so that $2^{r-1} > 1 + (1/4 - \eta)(1 + \dim x)$ as given in (4.2) in Assumption 4.1. Consider the $\delta > 0$ mentioned in item 1.1. For $\delta, n > 0$, partition the range of c as laid out in (A.1) with $K = \mathrm{int}(H_r n^{1/2}/\delta)$ using Assumption 4.1(ia) so that $H_r < \infty$. Thus, $\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \sim \delta/n^{1/2}$. Then, for each c there exists c_{k-1}, c_k so $c_{k-1} < c \leq c_k$. We note that $\mathsf{F}_n^{w,p}$ and $\bar{\mathsf{F}}_n^{w,p}$ and hence $\mathbb{F}_n^{w,p}$ and R_n are well-defined at $c = \infty$.

1.3. *Construct b-balls.* Cover the set $|b| \leq n^{1/4-\eta}B$ with balls of radius δ and centers b_m. The number of balls is $M \sim (n^{1/4-\eta}B/\delta)^{\dim x} \sim n^{(1/4-\eta)\dim x}/\delta^{\dim x}$. Thus, for any b there exists a b_m so that $|b - b_m| \leq \delta$.

1.4. *Chaining.* Write $R_n(b,c) = R_n(b_m,c_k) + \{R_n(b,c) - R_n(b_m,c_k)\}$, where $R_n(b_m,c_k)$ is a discrete point term and $R_n(b,c) - R_n(b_m,c_k)$ is a local perturbation term. By the triangle inequality $\mathcal{R}_n \leq \mathcal{R}_{n1} + \mathcal{R}_{n2}$ where

$$\begin{aligned}
\mathcal{R}_{n1} &= \max_{1\leq k\leq K} \max_{1\leq m\leq M} |R_n(b_m,c_k)|, \\
\mathcal{R}_{n2} &= \max_{1\leq k\leq K} \max_{1\leq m\leq M} \sup_{c_{k-1}<c\leq c_k} \sup_{|b-b_m|\leq\delta} |R_n(b,c) - R_n(b_m,c_k)|.
\end{aligned}$$

It suffices to show that $\mathcal{R}_{nj} = \mathsf{o}_\mathsf{P}(1)$ for $j = 1, 2$.

2. *The discrete point term $\mathcal{R}_{n,1}$.* Write $R_n(b_m,c_k) = n^{-1/2}\sum_{i=1}^n (z_{ikm} - \mathsf{E}_{i-1}z_{ikm})$ where

$$z_{ikm} = z_{i,b_m,c_k} = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c_k + x_{in}'b_m)} - 1_{(\varepsilon_i/\sigma\leq c_k)}\}. \tag{B.1}$$

We use Lemma A.1 for $n^{1/2}R_n(b_m,c_k)$ with $\upsilon = 1/2$, index $\ell = (k,m)$ so that $z_{i\ell} = z_{ikm}$, parameters $L_n = KM$ and $\lambda = 1/2 + (1/4 - \eta)\dim x$ and $\varsigma = 3/4 - \eta$ and r is given in item 1.2. We verify the conditions of Lemma A.1. Note that $z_{i\ell}$ is $\mathcal{F}_{in}$ adapted and $\mathsf{E}z_{i\ell}^{2^r} < \infty$ since bounding the difference of indicator functions by unity and using independence of ε_i and $\mathcal{F}_{i-1,n}$ gives $\mathsf{E}z_{i\ell}^{2^r} \leq \mathsf{E}w_{in}^{2^r}\varepsilon_i^{p2^r} = \mathsf{E}w_{in}^{2^r}\mathsf{E}\varepsilon_i^{p2^r}$, which is finite by Assumption 4.1(ia, ii).

The parameter λ. The set of indices ℓ has size $L_n = KM$. Since $K \sim n^{1/2}/\delta$ and $M \sim n^{(1/4-\eta)\dim x}$ while δ is fixed, $L_n \sim n^\lambda$ where $\lambda = 1/2 + (1/4 - \eta)\dim x > 0$.

The parameter ς. Since $|1_{(\varepsilon_i/\sigma\leq c_k + x_{in}'b_m)} - 1_{(\varepsilon_i/\sigma\leq c_k)}| \leq 1_{(c_k - |x_{in}||b_m| < \varepsilon_i/\sigma\leq c_k + |x_{in}||b_m|)}$ and given the inequality (A.5), we find for $1 \leq q \leq r$ that

$$\mathcal{S}_i = \mathsf{E}_{i-1}[\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c_k + x_{in}'b_m)} - 1_{(\varepsilon_i/\sigma\leq c_k)}\}]^{2^q} \leq \mathsf{H}_r(c_k + |x_{in}||b_m|) - \mathsf{H}_r(c_k - |x_{in}||b_m|).$$

Applying the mean value theorem to the bound gives $\mathcal{S}_i \leq 2|x_{in}||b_m|\dot{\mathsf{H}}_r(c^*)$ for an intermediate point c^* so $c_k - |x_{in}||b_m| \leq c^* \leq c_k + |x_{in}||b_m|$. Since $|b_m| \leq n^{1/4-\eta}B$, while $\sup_{v\in\mathbb{R}} \dot{\mathsf{H}}_r(v) < \infty$ by Assumption 3.1, 4.1(i) and Lemma A.4, we find, uniformly in ℓ,

$$\mathcal{S}_i \leq Cn^{1/4-\eta}|x_{in}|. \tag{B.2}$$

Since $z_{i\ell} = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i \le c_k + \sigma^{-1}x'_{in}b_m)} - 1_{(\varepsilon_i \le c_k)}\}$ and w_{in} is $\mathcal{F}_{i-1,n}$ adapted we get that $\mathsf{E}_{i-1}(z_{i\ell})^{2^q} = |w_{in}|^{2^q}\mathcal{S}_i$. Inserting the bound to $\mathcal{S}_i$ in (B.2) gives the bound $\mathsf{E}_{i-1}(z_{i\ell})^{2^q} \le C|w_{in}|^{2^q}n^{1/4-\eta}|x_{in}|$ and therefore, writing $n^{1/4-\eta} = n^{-1}n^{1/2}n^{3/4-\eta}$,

$$D_{nq} = \max_{1\le\ell\le L_n} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1}(z_{i\ell})^{2^q} \le Cn^{-1}\textstyle\sum_{i=1}^n |w_{in}|^{2^q}|n^{1/2}x_{in}|n^{3/4-\eta}.$$

Since $|w_{in}|^{2^q} \le 1 + |w_{in}|^{2^r}$ then $|w_{in}|^{2^q}|n^{1/2}x_{in}| \le (1+|w_{in}|^{2^r})|n^{1/2}x_{in}|$. Thus,

$$\mathcal{E}_{nq} = \mathsf{E}D_{nq} \le Cn^{3/4-\eta}\mathsf{E}n^{-1}\textstyle\sum_{i=1}^n(1+|w_{in}|^{2^r})|n^{1/2}x_{in}| = \mathrm{O}(n^{\varsigma}),$$

where $\varsigma = 3/4 - \eta \ge 0$ since the expectation of the average is $\mathrm{O}(1)$ by Assumption 4.1(ii).

Condition (i) of Lemma A.1 is that $\varsigma < 2\upsilon$. This holds since $\varsigma = 3/4 - \eta < 1 = 2\upsilon$.

Condition (ii) of Lemma A.1 is that $\varsigma + \lambda < \upsilon 2^r$. We have

$$\varsigma + \lambda = 3/4 - \eta + 1/2 + (1/4 - \eta)\dim x = 1 + (1/4-\eta)(1+\dim x).$$

By (4.2) in Assumption 4.1 r is chosen so that $1 + (1/4-\eta)(1+\dim x) < \upsilon 2^r = 2^{r-1}$.

Hence, Lemma A.1 shows that $n^{1/2}\mathcal{R}_{n1} = \max_{k,m}|n^{1/2}R_n(b_m,c_k)| = \mathrm{o}_{\mathsf{P}}(n^{1/2})$ which in turn implies $\mathcal{R}_{n1} = \max_{k,m}|R_n(b_m,c_k)| = \mathrm{o}_{\mathsf{P}}(1)$.

3. *The perturbation term* $\mathcal{R}_{n2}$ *is* $\mathrm{o}_{\mathsf{P}}(1)$. 3.1. *A first bound for* $\mathcal{R}_{n2}$. Write

$$R_n(b,c) - R_n(b_m,c_k) = n^{-1/2}\textstyle\sum_{i=1}^n\{r_i(b,b_m,c,c_k) - \mathsf{E}_{i-1}r_i(b,b_m,c,c_k)\} \tag{B.3}$$

with $r_i(b,b_m,c,c_k) = w_{in}\varepsilon_i^p[\{1_{(\varepsilon_i/\sigma \le c + x'_{in}b)} - 1_{(\varepsilon_i/\sigma\le c)}\} - \{1_{(\varepsilon_i/\sigma\le c_k + x'_{in}b_m)} - 1_{(\varepsilon_i/\sigma \le c_k)}\}]$. Pairing the second and fourth and the first and third indicators we get $r_i(b,b_m,c,c_k) = s_i(c_k,0,c,0) - s_i(c_k,b_m,c,b)$ where

$$s_i(c_k,b_m,c,b) = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\le c_k+x'_{in}b_m)} - 1_{(\varepsilon_i/\sigma\le c+x'_{in}b)}\}.$$

Correspondingly, let $S_n(c_k,b_m,c,b) = n^{-1/2}\sum_{i=1}^n\{s_i(c_k,b_m,c,b) - \mathsf{E}_{i-1}s_i(c_k,b_m,c,b)\}$ so that we can write $R_n(b,c) - R_n(b_m,c_k) = S_n(c_k,0,c,0) - S_n(c_k,b_m,c,b)$.

By the triangle inequality $\mathcal{R}_{n2} \le \mathcal{S}_{n1} + \mathcal{S}_{n2}$ where

$$\mathcal{S}_{n1} = \max_{1\le k\le K}\sup_{c_{k-1}<c\le c_k}|S_n(c_k,0,c,0)|,$$
$$\mathcal{S}_{n2} = \max_{1\le k\le K}\max_{1\le m\le M}\sup_{c_{k-1}<c\le c_k}\sup_{|b-b_m|\le\delta}|S_n(c_k,b_m,c,b)|.$$

Note that $\mathcal{S}_{n1} \le \mathcal{S}_{n2}$ since we can choose one of the b-centers to be 0, say $b_{m'} = 0$ for some m' and choose $b=0$ so that $S_n(c_k,0,c,0) = S_n(c_k,b_{m'},c,b)$ can be a term in $\mathcal{S}_{n2}$. Thus, to show $\mathcal{R}_{n2} = \mathrm{o}_{\mathsf{P}}(1)$ it suffices to show that $\mathcal{S}_{n2} = \mathrm{o}_{\mathsf{P}}(1)$.

3.2. *Bounding the function* $s_i(c_k,b_m,c,b) = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\le c_k+x'_{in}b_m)} - 1_{(\varepsilon_i/\sigma\le c+x'_{in}b)}\}$. By (A.17) in Lemma A.9 we have $\underline{c}_{kmi} < c + x'_{in}b \le \overline{c}_{kmi}$ where

$$\underline{c}_{kmi} = c_{k-1} + x'_{in}b_m - |x_{in}|\delta, \qquad \overline{c}_{kmi} = c_k + x'_{in}b_m + |x_{in}|\delta. \tag{B.4}$$

As a consequence, we get, uniformly in b,c,

$$|s_i(c_k,b_m,c,b)| \le \tilde{z}_{ikm} = |w_{in}||\varepsilon_i|^p\{1_{(\varepsilon_i/\sigma\le\overline{c}_{kmi})} - 1_{(\varepsilon_i/\sigma\le\underline{c}_{kmi})}\}. \tag{B.5}$$

3.3. *Bounding the function* $|S_n(c_k, b_m, c, b)|$. The triangle inequality gives

$$|S_n(c_k, b_m, c, b)| \leq n^{-1/2} \sum_{i=1}^n \{|s_i(c_k, b_m, c, b)| + \mathsf{E}_{i-1}|s_i(c_k, b_m, c, b)|\}.$$

Using the bound $|s_i(c_k, b_m, c, b)| \leq \tilde{z}_{ikm}$ in (B.5) in item 3.2 we get the further bound

$$|S_n(c_k, b_m, c, b)| \leq M_{nkm} = n^{-1/2} \sum_{i=1}^n (\tilde{z}_{ikm} + \mathsf{E}_{i-1}\tilde{z}_{ikm}),$$

uniformly in b, c. Thus, $\mathcal{R}_{n2}$ and $\mathcal{S}_{n2}$ are $\mathrm{o}_{\mathsf{P}}(1)$ if $M_{nkm} = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in k, m.
3.4. *Martingale decomposition.* Define

$$\tilde{M}_{nkm} = n^{-1/2} \sum_{i=1}^n (\tilde{z}_{ikm} - \mathsf{E}_{i-1}\tilde{z}_{ikm}), \qquad \bar{M}_{nkm} = n^{-1/2} \sum_{i=1}^n \mathsf{E}_{i-1}\tilde{z}_{ikm}. \tag{B.6}$$

Add and subtract $\bar{M}_{nkm}$ to get $M_{nkm} = \tilde{M}_{nkm} + 2\bar{M}_{nkm}$. Thus, $M_{nkm} = \mathrm{o}_{\mathsf{P}}(1)$ if

$$\tilde{\mathcal{M}}_n = \max_{1\leq k\leq K} \max_{1\leq m\leq M} \tilde{M}_{nkm} = \mathrm{o}_{\mathsf{P}}(1), \qquad \bar{\mathcal{M}}_n = \max_{1\leq k\leq K} \max_{1\leq m\leq M} \bar{M}_{nkm} = \mathrm{o}_{\mathsf{P}}(1). \tag{B.7}$$

3.5. *The conditional moments* $\mathcal{D}_{ikmq} = \mathsf{E}_{i-1}(\tilde{z}_{ikm}^{2^q})$. To show that $\tilde{z}_{ikm}$ has finite moments note that by Assumption 3.1 the independence of ε_i and $\mathcal{F}_{i-1,n}$ gives $\mathsf{E}\tilde{z}_{ikm}^{2^r} \leq \mathsf{E}|w_{in}|^{2^r}\mathsf{E}|\varepsilon_i|^{2^r p}$, which is finite by Assumption 4.1(ia, ii).

Recall from (B.4) that $\underline{c}_{kmi} = c_{k-1} + x_{in}'b_m - |x_{in}|\delta$ and $\bar{c}_{kmi} = c_k + x_{in}'b_m + |x_{in}|\delta$. Since w_{in} is $\mathcal{F}_{i-1,n}$ adapted then, for $0 \leq q \leq r$,

$$\mathcal{D}_{ikmq} = \mathsf{E}_{i-1}(\tilde{z}_{ikm}^{2^q}) = |w_{in}|^{2^q}\mathsf{E}_{i-1}|\varepsilon_i|^{p2^q}\{1_{(\varepsilon_i/\sigma\leq\bar{c}_{kmi})} - 1_{(\varepsilon_i/\sigma\leq\underline{c}_{kmi})}\}.$$

The inequality (A.5) implies $\mathcal{D}_{ikmq} \leq |w_{in}|^{2^q}\{\mathsf{H}_r(\bar{c}_{kmi}) - \mathsf{H}_r(\underline{c}_{kmi})\}$. Thus, Lemma A.9 requiring Assumption 4.1(i) gives, for some constant C not depending on k, m, x_{in}, that

$$\mathcal{D}_{ikmq} \leq C|w_{in}|^{2^q}(K^{-1} + |x_{in}'b_m|K^{-1/2} + \delta|x_{in}| + |x_{in}'b_m|^2).$$

Using that $|b_m| \leq Bn^{1/4-\eta}$ and $K^{-1} \sim \delta n^{-1/2}$ we get the further bound

$$\mathcal{D}_{ikmq} \leq \mathcal{D}_{iq} = C|w_{in}|^{2^q}(\delta n^{-1/2} + \delta^{1/2}n^{-1/4}n^{1/4-\eta}|x_{in}| + \delta|x_{in}| + n^{1/2-2\eta}|x_{in}|^2).$$

This reduces as

$$\mathcal{D}_{iq} \leq \delta Cn^{-1/2}|w_{in}|^{2^q}(1 + |n^{1/2}x_{in}|) + n^{-1/2-\eta}|w_{in}|^{2^q}(1 + \delta^{1/2})(1 + |n^{1/2}x_{in}|^2).$$

In turn, since $\mathsf{E}\sum_{i=1}^n(1 + |w_{in}|^{2^r})(1 + |n^{1/2}x_{in}|^2) = \mathrm{O}(n)$ by Assumption 4.1(ii), we get

$$\mathcal{E}_{nq} = \mathsf{E} \max_{1\leq k\leq K} \max_{1\leq m\leq M} \sum_{i=1}^n \mathsf{E}_{i-1}(\tilde{z}_{ikm})^{2^q} \leq \mathsf{E}\sum_{i=1}^n \mathcal{D}_{iq} = \delta\mathrm{O}(n^{1/2}) + (1 + \delta^{1/2})\mathrm{O}(n^{1/2-\eta}),$$

where the order terms are uniform in k, m, x_{in}.
3.6. *The compensator is* $\bar{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$. Note that $\mathsf{E}\bar{\mathcal{M}}_n = n^{-1/2}\mathcal{E}_{n0}$. Thus, item 3.5 shows that $\mathsf{E}\bar{\mathcal{M}}_n = \delta\mathrm{O}(1)$. The Markov inequality then shows $\bar{\mathcal{M}}_n = \delta\mathrm{O}_{\mathsf{P}}(1)$ so that $\forall\epsilon > 0$, $\exists C > 0$ so that $\mathsf{P}(\bar{\mathcal{M}}_n \geq \delta C) \leq \epsilon$. We are still free to choose δ which will be exploited now. For any $\gamma > 0$ we can choose $\delta = \gamma/C$ so that $\mathsf{P}(\bar{\mathcal{M}}_n \geq \gamma) \leq \epsilon$. Hence, $\bar{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$.

3.7. *The martingale is* $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$. Recall from (B.6), (B.5) that

$$\tilde{M}_{nkm} = n^{-1/2}\textstyle\sum_{i=1}^n(\tilde{z}_{ikm} - \mathsf{E}_{i-1}\tilde{z}_{ikm}), \quad \tilde{z}_{ikm} = |w_{in}||\varepsilon_i|^p\{1_{(\varepsilon_i/\sigma\leq\bar{c}_{kmi})} - 1_{(\varepsilon_i/\sigma\leq\underline{c}_{kmi})}\}.$$

We use Lemma A.1 for $n^{1/2}\tilde{M}_{nkm}$ with $\upsilon = 1/2$, index $\ell = (k,m)$ so that $z_{i\ell} = \tilde{z}_{ikm}$, $L_n = KM$, $\lambda = 1/2 + (1/4-\eta)\dim x$ and $\varsigma = 3/4-\eta$, while r satisfies $2^{r-1} > 1+(1/4-\eta)(1+\dim x)$ as given in (4.2) in Assumption 4.1. We verify the conditions of Lemma A.1. In item 3.5 it was established that $z_{i\ell} = \tilde{z}_{ikm}$ is $\mathcal{F}_{in}$ adapted and $\mathsf{E}z_{i\ell}^{2^r} < \infty$.

The parameter $\lambda = 1/2 + (1/4-\eta)\dim x$ as in item 2.

The parameter $\varsigma = 1/2$. Apply item 3.5 for $1 \leq q \leq r$ to see that $\mathcal{E}_{nq} = \delta\mathrm{O}(n^{1/2}) = \mathrm{O}(n^{1/2})$ since δ is fixed. Since $\varsigma = 1/2 \leq 3/4-\eta$ we can use the same argument as in item 2, so that *Conditions* $(i),(ii)$ are satisfied.

Hence, Lemma A.1 shows that $n^{1/2}\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^{1/2})$ so that $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$. Since also $\bar{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$ as shown in item 3.5, we have that $\mathcal{R}_{n2}$, $\mathcal{S}_{n2}$, M_{nkm}, M_{nkm} are $\mathrm{o}_{\mathsf{P}}(1)$. In item 2 it was shown that $\mathcal{R}_{n1} = \mathrm{o}_{\mathsf{P}}(1)$. In combination $\mathcal{R}_n = \mathrm{o}_{\mathsf{P}}(1)$. □

Remark B.1. *Lemma 4.1 corrects Theorem 4.1 in JN16. The issue arises in the analysis of the oscillation terms in item 6 of that proof. The oscillation terms correspond to (B.3) in the present proof. It has to be argued, that these are small uniformly in grid points indexed by k,m and in small deviations therefrom indexed by b,c. The idea is to find a bound that is uniform in b,c and then deal with k,m using the iterated martingale inequality in Lemma A.1, while keeping track of the regressors x_i. The bound $(C4)$ in JN16 established using their Lemma B.2 depends on x_i, which ruins the uniformity. Here we replace that lemma with Lemma A.9, which allows us to keep track of x_i.*

Proof of Lemma 4.2. Let $c_a = c + n^{-1/2}ac$ so that $c_0 = c$ and define the summands $z_i(c_a,c) = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c_a)} - 1_{(\varepsilon_i/\sigma\leq c)}\}$. Let

$$Z_n(c_a,c) = n^{1/2}\{\mathbb{F}_n^{w,p}(a,0,c) - \mathbb{F}_n^{w,p}(0,0,c)\} = \textstyle\sum_{i=1}^n\{z_i(c_a,c) - \mathsf{E}_{i-1}z_i(c_a,c)\}. \quad \text{(B.8)}$$

We want to prove that $\mathcal{Z}_n = \sup_{c\in\mathbb{R}}\sup_{|a|\leq n^{1/4-\eta}B}|Z_n(c_a,c)| = \mathrm{o}_{\mathsf{P}}(n^{1/2})$.

1. *Partition the support.* We choose $r=2$, noting that H_r is finite by Assumption 4.1(ia). Consider a small $\delta > 0$, which will be chosen in item 4. Partition the axis as laid out in (A.1) with $K = \mathrm{int}(H_r n^{1/2}/\delta)$. Thus, $\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \sim n^{-1/2}$.

2. *Assign c_a and $c_0 = c$ to the partitioned support.* For each c_a there exists an integer $k(c_a)$ and grid points $c_{k(c_a)-1}, c_{k(c_a)}$ so that $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$. Assumption 4.1$(i)$ is the same as Assumption A.1, which by Lemma A.5 implies that $\sup_{c\in\mathbb{R}}|c|(1+c^{4p})\mathsf{f}(c) < \infty$. With this property and Assumption 3.1, Lemma A.10 applies. Used with $1/2-\zeta = 1/4-\eta$ and $K \sim n^{1/2}/\delta$ it gives, for some $C > 0$,

$$\sup_{c\in\mathbb{R}}\sup_{|a|\leq n^{1/4-\eta}B}|k(c_a) - k(c)| \leq 2 + Cn^{-1/4-\eta}K = D_n. \quad \text{(B.9)}$$

3. *Bound $Z_n(c_a,c)$.* Add and subtract $1_{\{\varepsilon_i/\sigma\leq c_{k(c_a)}\}}$ and $1_{\{\varepsilon_i/\sigma\leq c_{k(c)}\}}$ so that

$$\begin{aligned} z_i(c_a,c) = w_{in}\varepsilon_i^p[&1_{(\varepsilon_i/\sigma\leq c_a)} - 1_{\{\varepsilon_i/\sigma\leq c_{k(c_a)}\}} \\ &+ 1_{\{\varepsilon_i/\sigma\leq c_{k(c_a)}\}} - 1_{\{\varepsilon_i/\sigma\leq c_{k(c)}\}} - 1_{(\varepsilon_i/\sigma\leq c)} + 1_{\{\varepsilon_i/\sigma\leq c_{k(c)}\}}]. \end{aligned}$$

By the triangle inequality we get

$$|Z_n(c_a, c)| \leq |Z_n\{c_a, c_{k(c_a)}\}| + |Z_n\{c_{k(c_a)}, c_{k(c)}\}| + |Z_n\{c, c_{k(c)}\}|. \tag{B.10}$$

The third term is a special case of the first term with $a = 0$ since $c = c_0$. Accordingly, the triangle inequality gives $\mathcal{Z}_n \leq 2\mathcal{Z}_{1n} + \mathcal{Z}_{2n}$ where

$$\mathcal{Z}_{1n} = \sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}B} |Z_n\{c_a, c_{k(c_a)}\}|, \qquad \mathcal{Z}_{2n} = \sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}B} |Z_n\{c_{k(c)}, c_{k(c_a)}\}|.$$

3.1. *The term* $\mathcal{Z}_{1n}$: The summands are $z_i\{c_a, c_{k(c_a)}\} = w_{in}\varepsilon_i^p[1_{\{\varepsilon_i/\sigma\leq c_{k(c_a)}\}} - 1_{\{\varepsilon_i/\sigma\leq c_a\}}]$ where $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$. They satisfy

$$|z_i\{c_a, c_{k(c_a)}\}| \leq |w_{in}||\varepsilon_i|^p|1_{(\varepsilon_i/\sigma\leq c_{k(c_a)})} - 1_{(\varepsilon_i/\sigma\leq c_{k(c_a)-1})}| = z_{i,k(c_a)-1,k(c_a)},$$

where the bound only depends on a, c through grid points that are one interval apart. Hence, using the triangle inequality

$$|Z_n\{c_a, c_{k(c_a)}\}| \leq \textstyle\sum_{i=1}^n z_{i,k(c_a)-1,k(c_a)} + \sum_{i=1}^n \mathsf{E}_{i-1} z_{i,k(c_a)-1,k(c_a)},$$

so that we can bound, uniformly in a, c,

$$|Z_n\{c_a, c_{k(c_a)}\}| \leq \mathcal{N}_n = \max_{1\leq k\leq K} \textstyle\sum_{i=1}^n z_{i,k-1,k} + \max_{1\leq k\leq K} \sum_{i=1}^n \mathsf{E}_{i-1} z_{i,k-1,k}.$$

Adding and subtracting $\mathsf{E}_{i-1} z_{i,k-1,k}$ and applying the triangle inequality gives that $\mathcal{N}_n \leq \widetilde{\mathcal{N}}_n + 2\overline{\mathcal{N}}_n$, where $\widetilde{\mathcal{N}}_n = \max_{1\leq k\leq K} |\tilde{N}_{nk}|$ and $\overline{\mathcal{N}}_n = \max_{1\leq k\leq K} \bar{N}_{nk}$ with

$$\tilde{N}_{nk} = \textstyle\sum_{i=1}^n (z_{i,k-1,k} - \mathsf{E}_{i-1} z_{i,k-1,k}), \qquad \bar{N}_{nk} = \sum_{i=1}^n \mathsf{E}_{i-1} z_{i,k-1,k}.$$

Hence,

$$\mathcal{Z}_{1n} = \sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}B} |Z_n\{c_a, c_{k(c_a)}\}| \leq \mathcal{N}_n \leq \widetilde{\mathcal{N}}_n + 2\overline{\mathcal{N}}_n.$$

3.2. *The term* $\mathcal{Z}_{2n}$: In this case $z_i\{c_{k(c_a)}, c_{k(c)}\}$ involves grid points $k(c_a), k(c)$ that may be more than one point apart. Indeed, by (B.9) we have $|k(c_a) - k(c)| \leq D_n$ uniformly in a, c. As a consequence, we can bound

$$\mathcal{Z}_{2n} = \sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}B} |Z_n\{c_{k(c_a)}, c_{k(c)}\}| \leq \mathcal{M}_n = \max_{1\leq k\leq K} \max_{k\leq k^*\leq k+D_n} |Z_n(c_k, c_{k^*})|,$$

with $z_i(c_k, c_{k^*}) = w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c_{k^*})} - 1_{(\varepsilon_i/\sigma\leq c_k)}\} = z_{ikk^*}$.

3.3. *Combine items 3.1-3.2.* We note that $\widetilde{\mathcal{N}}_n \leq \mathcal{M}_n$, as the former involves a maximum over terms that are one grid point apart, whereas the second involves a maximum over terms that are up to $D_n > 2$ grid points apart. In summary, we get

$$\mathcal{Z}_n \leq 2\mathcal{Z}_{1n} + \mathcal{Z}_{2n} \leq 2\widetilde{\mathcal{N}}_n + 4\overline{\mathcal{N}}_n + \mathcal{M}_n \leq 4\overline{\mathcal{N}}_n + 3\mathcal{M}_n,$$

so that it suffices to argue that $\overline{\mathcal{N}}_n, \mathcal{M}_n = \mathrm{o}_{\mathsf{P}}(n^{1/2})$.

4. *The compensator* $\overline{\mathcal{N}}_n = \mathrm{o}_{\mathsf{P}}(n^{1/2})$. Note that $z_{i,k-1,k}$ is $\mathcal{F}_{in}$ adapted and $\mathsf{E}(z_{i,k-1,k}) < \infty$, since bounding the difference of indicator functions by unity and using independence

of ε_i and $\mathcal{F}_{i-1,n}$ gives $\mathsf{E}z_{i,k-1,k} \leq \mathsf{E}|w_{in}||\varepsilon_i|^p = \mathsf{E}|w_{in}|\mathsf{E}|\varepsilon_i|^p$, which is finite by Assumption 3.1,4.1(ii). In light of (A.5) we get

$$\mathsf{E}_{i-1}z_{i,k-1,k} \leq |w_{in}|\{\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1})\} = |w_{in}|H_r/K.$$

In turn, using Assumption 4.1(ii) we get

$$\mathsf{E}\overline{\mathcal{N}}_n = \mathsf{E}\max_{1\leq k\leq K}\textstyle\sum_{i=1}^n \mathsf{E}_{i-1}|z_{i,k-1,k}|^{2^q} \leq (H_r/K)\mathsf{E}\textstyle\sum_{i=1}^n |w_{in}|^{2^q} = \mathrm{O}(n/K).$$

Finally, since $K \sim n^{1/2}/\delta$ we get $\mathsf{E}\overline{\mathcal{N}}_n = \delta\mathrm{O}(n^{1/2})$. Since we are free to choose δ we get that $\overline{\mathcal{N}}_n = \mathrm{o}_\mathsf{P}(n^{1/2})$ following the argument in item 3.7 of the proof of Lemma 4.1.
5. *Conditional moments of* z_{ikk^*}. Note that z_{ikk^*} is $\mathcal{F}_{in}$ adapted and that $\mathsf{E}z_{ikk^*}^4$ is finite by an argument similar to that in item 4. In light of (A.3), (A.5) and since $k \leq k^* \leq k + D_n$ by (B.9) we get, for $q = 1, 2$,

$$\mathsf{E}_{i-1}z_{ikk^*}^{2^q} \leq |w_{in}|^{2^q}\{\mathsf{H}_r(c_{k^*}) - \mathsf{H}_r(c_k)\} = |w_{in}|^{2^q}(k^* - k)H_r/K \leq |w_{in}|^{2^q}D_nH_r/K.$$

In turn, we get using Assumption 4.1(ii) that, for $q = 1, 2$,

$$\mathcal{E}_{nq} = \mathsf{E}\max_{1\leq k\leq K}\max_{k^*:k\leq k^*\leq k+D}\textstyle\sum_{i=1}^n \mathsf{E}_{i-1}z_{ikk^*}^{2^q} \leq (D_nH_r/K)\mathsf{E}\textstyle\sum_{i=1}^n |w_{in}|^{2^q} = \mathrm{O}(nD_n/K).$$

Since $D_n = 2 + Cn^{-1/4-\eta}K$ and $K \sim n^{1/2}/\delta$ where δ is fixed, then $\mathcal{E}_{nq} = \mathrm{O}(n^{3/4-\eta})$.
6. *The martingale* $\mathcal{M}_n = \mathrm{o}_\mathsf{P}(n^{1/2})$. We use Lemma A.1 for $\mathcal{M}_n$ with $\upsilon = 1/2$, index $\ell = (k, k^*)$ so that $z_{i\ell} = z_i(c_{k^*}, c_k)$, parameters $L_n = KD_n \sim n^{3/4-\eta}$ and $\lambda = \varsigma = 3/4 - \eta > 0$ while $r = 2$. We verify the conditions of Lemma A.1. In item 5 it was established that $z_{i\ell} = z_{ikk^*}$ is $\mathcal{F}_{in}$ adapted and $\mathsf{E}z_{ikk^*}^4 < \infty$.
The parameter $\lambda = 3/4 - \eta$. The set of indices ℓ has size $L_n = KD_n \sim n^{3/4-\eta} \sim n^\lambda$.
The parameter $\varsigma = 3/4 - \eta$. Apply item 5 to see that $\mathcal{E}_{nq} = \mathrm{O}(n^{3/4-\eta}) = \mathrm{O}(n^\varsigma)$.
Condition (i) is that $\varsigma < 2\upsilon$. This holds since $0 < \eta$, so that $\varsigma = 3/4 - \eta < 2\upsilon = 1$.
Condition (ii) is that $\varsigma + \lambda < \upsilon 2^r$ with $r = 2$. We have $\varsigma + \lambda = 3/2 - 2\eta$, while $\upsilon 2^r = (1/2)4 = 2$.

Hence, Lemma A.1 shows that $\mathcal{M}_n = \mathrm{o}_\mathsf{P}(n^{1/2})$. As $\overline{\mathcal{N}}_n = \mathrm{o}_\mathsf{P}(n^{1/2})$ by item 4 we get $\mathcal{Z}_n = \mathrm{o}_\mathsf{P}(n^{1/2})$ as noted in item 3.3. □

*Proof of Theorem **4.1**.* Let $V_n(a,b,c) = n^{-1/2}\sum_{i=1}^n\{v_i(a,b,c) - \mathsf{E}_{i-1}v_i(a,b,c)\}$, where

$$v_i(a,b,c) = w_{in}(\varepsilon_i/\sigma)^p\{1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}\}.$$

We want to prove $\mathcal{V}_n = \sup_{c\in\mathbb{R}}\sup_{|a|,|b|\leq n^{1/4-\eta}B}|V_n(a,b,c)| = \mathrm{o}_\mathsf{P}(1)$.

Let $c_a = c + n^{-1/2}ac$. Adding and subtracting $1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac)} = 1_{(\varepsilon_i/\sigma\leq c_a)}$ we get $v_i(a,b,c) = v_i(0,b,c_a) + v_i(a,0,c)$, so that $V_n(a,b,c) = V_n(0,b,c_a) + V_n(a,0,c)$. Taking supremum for each term we see that, for $0 < \eta \leq 1/4$,

$$\sup_{c\in\mathbb{R}}\sup_{|a|,|b|\leq n^{1/4-\eta}B}|V_n(0,b,c_a)| = \sup_{c\in\mathbb{R}}\sup_{|b|\leq n^{1/4-\eta}B}|V_n(0,b,c)| = \mathcal{V}_{1n},$$
$$\sup_{c\in\mathbb{R}}\sup_{|a|,|b|\leq n^{1/4-\eta}B}|V_n(a,0,c)| = \sup_{c\in\mathbb{R}}\sup_{|a|\leq n^{1/4-\eta}B}|V_n(a,0,c)| = \mathcal{V}_{2n}.$$

Thus, using the triangle inequality we see that $\mathcal{V}_n = \mathrm{o}_\mathsf{P}(1)$ if $\mathcal{V}_{1n}, \mathcal{V}_{2n} = \mathrm{o}_\mathsf{P}(1)$. This follows from the Lemmas 4.1, 4.2 using Assumptions 3.1, 4.1. □

Proof of Theorem 4.2. Let $C_n(a,b,c) = n^{1/2}\{\overline{\mathsf{F}}_n^{w,p}(a,b,c) - \overline{\mathsf{F}}_n^{w,p}(0,0,c)\}$. The definition of $\overline{\mathsf{F}}_n^{w,p}(a,b,c)$, see (3.1), shows that

$$C_n(a,b,c) = \sigma^p n^{-1/2}\textstyle\sum_{i=1}^n w_{in}\mathsf{E}_{i-1}(\varepsilon_i/\sigma)^p\{1_{(\varepsilon_i/\sigma\le c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\le c)}\},$$

noting that $\mathsf{E}|\varepsilon_i|^p < \infty$ by Assumption 4.2(ia). As a consequence we can write

$$C_n(a,b,c) = \sigma^p n^{-1/2}\textstyle\sum_{i=1}^n w_{in}\mathcal{I}_i \qquad \text{where} \qquad \mathcal{I}_i = \int_c^{c+n^{-1/2}ac+x_{in}'b} u^p\mathsf{f}(u)du. \tag{B.11}$$

Recall from (4.4) that $\mathcal{B}_n^{w,p}(a,b,c) = \sigma^p c^p\mathsf{f}(c)n^{-1/2}\sum_{i=1}^n w_{in}(n^{-1/2}ac + x_{in}'b)$ and define

$$\mathcal{E} = C_n(a,b,c) - \mathcal{B}_n^{w,p}(a,b,c) = \sigma^p n^{-1/2}\textstyle\sum_{i=1}^n w_{in}\mathcal{D}_i, \tag{B.12}$$

where $\mathcal{D}_i = \mathcal{I}_i - (n^{-1/2}ac + x_{in}'b)c^p\mathsf{f}(c)$. We argue that $\mathcal{E}_i = \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$ uniformly in a,b,c.
1. *The case* $|c| \le 1$. Write $\mathcal{I}_i$ as $\int_{-\infty}^{c+n^{-1/2}ac+x_{in}'b} u^p\mathsf{f}(u)du$ minus $\int_{-\infty}^{c} u^p\mathsf{f}(u)du$. The latter integral has derivative $c^p\mathsf{f}(c)$, so that the mean value theorem gives

$$\mathcal{I}_i = (n^{-1/2}ac + x_{in}'b)c_{*i}^p\mathsf{f}(c_{*i}),$$

for an intermediate point c_{*i}, so that $|c - c_{*i}| \le |n^{-1/2}ac + x_{in}'b|$. In turn,

$$\mathcal{D}_i = \mathcal{I}_i - (n^{-1/2}ac + x_{in}'b)c^p\mathsf{f}(c) = (n^{-1/2}ac + x_{in}'b)\{c_{*i}^p\mathsf{f}(c_{*i}) - c^p\mathsf{f}(c)\}. \tag{B.13}$$

Since $c^p\mathsf{f}(c)$ is Lipschitz by Assumption 4.2(ib) while $|c - c_{*i}| \le |n^{-1/2}ac + x_{in}'b|$, then $|\mathcal{D}_i| \le C_L(n^{-1/2}ac + x_{in}'b)^2$. Use first that $|c| \le 1$ and then the bounds $|a|,|b| \le Bn^{1/4-\eta}$ and $(1+x)^2 \le 2(1+x^2)$ to get

$$|\mathcal{D}_i| \le C_L(|n^{-1/2}a| + |x_{in}'b|)^2 \le C_L B^2 n^{-1/2-2\eta}2\{1 + (n^{1/2}|x_{in}|)^2\}. \tag{B.14}$$

Insert the bound for $\mathcal{D}_i$ in (B.12), while applying the triangle inequality to get

$$|\mathcal{E}| = |C_n(a,b,c) - \mathcal{B}_n^{w,p}(a,b,c)| \le \sigma^p 2C_L B^2 n^{-2\eta}n^{-1}\textstyle\sum_{i=1}^n |w_{in}|\{1 + (n^{1/2}|x_{in}|)^2\}.$$

The sum is $\mathrm{O}_{\mathsf{P}}(n)$ by Assumption 4.2(ii) so that $\mathcal{E} = \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$ as desired.
2. *The case* $|c| > 1$ requires a more careful expansion. Rewrite $\mathcal{I}_i = \mathcal{I}^a + \mathcal{I}_i^b$ where

$$\mathcal{I}^a = \textstyle\int_c^{c_a} u^p\mathsf{f}(u)du, \qquad \mathcal{I}_i^b = \int_{c_a}^{c_a+x_{in}'b} u^p\mathsf{f}(u)du,$$

by inserting the division point $c_a = c + n^{-1/2}ac$. Accordingly write $\mathcal{D}_i = \mathcal{D}^a + \mathcal{D}_i^b$ where

$$\mathcal{D}^a = \mathcal{I}^a - n^{-1/2}ac^{p+1}\mathsf{f}(c), \qquad \mathcal{D}_i^b = \mathcal{I}_i^b - x_{in}'bc^p\mathsf{f}(c).$$

We expand these terms separately.
3. *The term* $\mathcal{D}^a$. Apply the mean value theorem to get $\mathcal{I}^a = (n^{-1/2}ac)c_*^p\mathsf{f}(c_*)$ for an intermediate point c_* so that $|c - c_*| < |n^{-1/2}ac|$. Add and subtract $(n^{-1/2}ac)c^p\mathsf{f}(c)$ and $(n^{-1/2}a)c^{-1}c_*^{p+2}\mathsf{f}(c)$, which is well-defined for $|c| > 1$, to get

$$\mathcal{D}^a = \mathcal{I}^a - n^{-1/2}ac^{p+1}\mathsf{f}(c) = n^{-1/2}a[c^{-1}\{c_*^{p+2}\mathsf{f}(c_*) - c^{p+2}\mathsf{f}(c)\} + c^{-1}(c^2 - c_*^2)c_*^p\mathsf{f}(c_*)].$$

We analyse the two summands of the square bracket, $\mathcal{D}_1^a$ and $\mathcal{D}_2^a$ say, separately.

For $\mathcal{D}_1^a = c^{-1}\{c_*^{p+2}\mathsf{f}(c_*) - c^{p+2}\mathsf{f}(c)\}$, the Lipschitz Assumption 4.2(ib) for $c^{p+2}\mathsf{f}(c)$ and the bounds $|c - c_*| < |n^{-1/2}ac|$ and $|c^{-1}| < 1$ imply that $|\mathcal{D}_1^a| \le C_L|n^{-1/2}a|$.

For $\mathcal{D}_2^a = c^{-1}(c^2 - c_*^2)c_*^p\mathsf{f}(c_*)$, write $(c^2 - c_*^2) = (c - c_*)(c + c_*)$. Recall $|c - c_*| \le |n^{-1/2}a||c|$. Further, we argue $|c_* + c| \le 3|c_*|$. Indeed, since $|c| \le |c_*| + |c_* - c| \le |c_*| + |n^{-1/2}a||c|$ then $|c|(1 - |n^{-1/2}a|) \le |c_*|$. Since $|n^{-1/2}a| < 1/2$ for large n, then $|c| \le 2|c_*|$ and therefore $|c_* + c| \le 3|c_*|$. In combination, we get $|c^2 - c_*^2| \le 3|n^{-1/2}a||c||c_*|$ so that $\mathcal{D}_2^a \le 3|n^{-1/2}a|^2|c_*|^{p+1}\mathsf{f}(c_*)$. Since $|c_*|^{p+1}\mathsf{f}(c_*)$ is bounded by Assumption 4.2(ic), we get that $|\mathcal{D}_2^a| \le C|n^{-1/2}a|$ for some $C > 0$.

Taken together we get that $\mathcal{D}^a = n^{-1/2}a(\mathcal{D}_1^a + \mathcal{D}_2^a)$ satisfies $|\mathcal{D}^a| \le C|n^{-1/2}a|^2$.

4. *The term* $\mathcal{D}_i^b = \mathcal{I}_i^b - x_{in}'bc^p\mathsf{f}(c)$. The mean value theorem gives $\mathcal{I}_i^b = x_{in}'bc_*^p\mathsf{f}(c_*)$ where $|c_a - c_*| < |x_{in}'b|$. Add and subtract $x_{in}'b\{c^p\mathsf{f}(c) + c_a^p\mathsf{f}(c_a) + c^{-1}c_a^{p+1}\mathsf{f}(c)\}$ to get

$$\mathcal{D}_i^b = x_{in}'b[\{c_*^p\mathsf{f}(c_*) - c_a^p\mathsf{f}(c_a)\} + c^{-1}\{c_a^{p+1}\mathsf{f}(c_a) - c^{p+1}\mathsf{f}(c)\} + c^{-1}(c - c_a)c_a^p\mathsf{f}(c_a)].$$

We analyse the three summands of the square bracket, $\mathcal{D}_{i1}^b$, $\mathcal{D}_2^b$ and $\mathcal{D}_3^b$ say, separately.

For $\mathcal{D}_{i1}^b = c_*^p\mathsf{f}(c_*) - c_a^p\mathsf{f}(c_a)$, the Lipschitz Assumption 4.2(ib) for $c^p\mathsf{f}(c)$ and the bound $|c_a - c_*| < |x_{in}'b|$ imply that $|\mathcal{D}_{i1}^b| \le C_L|x_{in}'b|$.

For $\mathcal{D}_2^b = c^{-1}\{c_a^{p+1}\mathsf{f}(c_a) - c^{p+1}\mathsf{f}(c)\}$, the Lipschitz Assumption 4.2(ib) for $c^{p+1}\mathsf{f}(c)$ and the identity $c^{-1}(c_a - c) = n^{-1/2}a$ imply that $|\mathcal{D}_2^b| \le C_L|n^{-1/2}a|$

For $\mathcal{D}_3^b = c^{-1}(c - c_a)c_a^p\mathsf{f}(c_a)$, note that $c^{-1}(c_a - c) = n^{-1/2}a$ and that $|c_a|^p\mathsf{f}(c_a)$ is bounded by Assumption 4.2(ib) so that $|\mathcal{D}_3^b| \le C|n^{-1/2}a|$.

Taken together, $\mathcal{D}_i^b \le x_{in}'b(\mathcal{D}_{i1}^b + \mathcal{D}_2^b + \mathcal{D}_3^b)$ satisfies $|\mathcal{D}_i^b| \le C|x_{in}'b|(|x_{in}'b| + |n^{-1/2}a|)$.

5. *The term* $\mathcal{D}_i = \mathcal{D}^a + \mathcal{D}_i^b$. Combine the results above to get that $|\mathcal{D}_i| \le C(|n^{-1/2}a| + |x_{in}'b|)^2$ for some $C > 0$. This bound is of the same form as (B.14) so the remaining proof for $|c| \le 1$ applies. □

Acknowledgements

Comments from James Duffy and Xiyu Jiao are greatfully acknowledged.

References

Atkinson, A. & Riani, M. (2000). *Robust Diagnostic Regression Analysis.* New York: Springer.

Atkinson, A., Riani, M., & Cerioli, A. (2010). The forward search: Theory and data analysis (with discussion). *J. Korean Statist. Soc.*, *39*, 117–163.

Bercu, B., Delyon, B., & Rio, E. (2015). *Concentration inequalities for sums and martingales.* Cham: Springer.

Bercu, B. & Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Prob.*, *18*, 1848–1869.

Bercu, B. & Touati, A. (2018). New insights on concentration inequalities for self-normalized martingales with applications. Technical report, arXiv:1810.10590.

Bickel, P. (1975). One-step huber estimates in the linear model. *J. Amer. Statist. Assoc.*, *70*, 428–434.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley & Sons.

Chan, N. & Wei, C. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, *16*, 367–401.

Croux, C. & Rousseeuw, P. J. (1992). A class of high-breakdown scale estimators based on subranges. *Comm. Statist. Theory Methods*, *21*, 1935–1951.

Doornik, J. A. (2009). Autometrics. In J. L. Castle & N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry* (pp. 88–121). Oxford: Oxford University Press.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability*, *2*, 513–533.

Engler, E. & Nielsen, B. (2009). The empirical process of autoregressive residuals. *Econometrics Journal*, *12*, 367–381.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statist. Sinica*, *17*, 115–138.

Freedman, D. (1975). On tail probabilities for martingales. *Ann. Probab.*, *3*, 100–118.

Hadi, A. & Simonoff, J. (1993). Procedures for the identification of multiple outliers in linear models. *J. Amer. Statist. Assoc.*, *88*, 1264–1272.

Hendry, D. F. (1999). An econometric analysis of US food expenditure, 1931–1989. In J. R. Magnus & M. S. Morgan (Eds.), *Methodology and tacit knowledge: Two experiments in econometrics* (pp. 341–361). New York: John Wiley & Sons.

Hendry, D. F. & Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics.* London: MIT Press.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in fully saturated regression. *Computational Statistics*, *23*, 317–335.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*, 73–101.

Jiao, X. & Nielsen, B. (2017). Asymptotic analysis of iterated 1-step huber-skip m-estimators with varying cut-offs. In J. Antoch, J. Jurečková, M. Maciak, & M. Pešta (Eds.), *Analytic Methods in Statistics*, volume 193 of *Springer Proc. Math. Stat.* Springer.

Johansen, S. & Nielsen, B. (2016a). Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli*, *22*, 1131–1183.

Johansen, S. & Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scand. J. Stat.*, *43*, 321–81.

Jurečková, J., Sen, P., & Picek, J. (2013). *Methodology in Robust and Nonparametric Statistics*. Boca Raton, FL: CRC Press.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models* (2nd ed.). New York: Springer.

Koul, H. L. & Leventhal, S. (1989). Weak convergence of the residual empirical process in explosive autoregression. *Ann. Statist.*, *17*, 1784–1794.

Koul, H. L. & Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.*, *22*, 540–562.

Koul, H. L. & Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.*, *27*, 204–236.

Lee, S. & Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.*, *27*, 237–261.

Pretis, F., Reade, J., & Sucarrat, G. (2018). General-to-specific (gets) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software*, *86*.

Rousseeuw, P. J. (1984). Least median of squares regressions. *J. Amer. Statist. Assoc.*, *79*, 871–880.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Hoboken, NJ: John Wiley & Sons.

Ruppert, D. & Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.*, *75*, 828–838.

Welsh, A. H. & Ronchetti, E. (2006). A journey in single steps: robust one-step m-estimation in linear regression. *J. Statist. Plann. Inference*, *103*, 287–310.