Discussion Papers
Department of Economics
University of Copenhagen

No. 19-09

Uniform Consistency of Marked and Weighted Empirical Distributions of Residuals

by

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)

Uniform Consistency of Marked and Weighted Empirical Distributions of Residuals

Vanessa Berenguer-Rico*, Søren Johansen† & Bent Nielsen‡

24 May 2019

Abstract

A uniform weak consistency theory is presented for the marked and weighted empirical distribution function of residuals. New and weaker sufficient conditions for uniform consistency are derived. The theory allows for a wide variety of regressors and error distributions. We apply the theory to 1-step Huber-skip estimators. These estimators describe the widespread practice of removing outlying observations from an intial estimation of the model of interest and updating the estimation in a second step by applying least squares to the selected observations. Two results are presented. First, we give new and weaker conditions for consistency of the estimators. Second, we analyze the gauge, which is the rate of false detection of outliers, and which can be used to decide the cut-off in the rule for selecting outliers.

Keywords: 1-step Huber skip, Asymptotic theory, Empirical processes, Gauge, Marked and Weighted Empirical processes, Non-stationarity, Robust Statistics, Stationarity.

JEL classification: C01, C22

1 Introduction

We study the uniform consistency of marked and weighted empirical distribution functions of estimated residuals from a linear time series regression. This can be used to show consistency of various procedures appearing in robust statistics. A common empirical strategy is to estimate a regression equation by least squares or by a robust estimator, select observations with small residuals, and then re-estimate the regression for the selected observations using least squares. With the derived results, we can show consistency of the updated estimator, which is an example of a 1-step Huber-skip estimator. We can also evaluate the gauge of

*Department of Economics, University of Oxford. E-mail: vanessa.berenguer-rico@economics.ox.ac.uk.

†Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. E-mail: soren.johansen@econ.ku.dk.

‡Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk.

the procedure, which is defined in terms of the falsely detected outliers when in fact there are no outliers. The gauge is useful for choosing the cut-off for selection in the procedure. These results can also be used to study more complicated algorithms that involve iteration of 1-step Huber-skip estimators.

We consider a linear time series regression model

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

where the regressors x_i can be stationary, non-stationary, deterministically or even explosively trending and the scaled errors ε_i/σ are assumed to be independent, identically distributed and independent of the regressors. Let $\tilde{\beta}$ and $\tilde{\sigma}$ be initial estimators of β and σ, respectively, and define the estimated residuals by $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$.

The marked and weighted empirical distribution function of interest is

$$\hat{\mathsf{F}}_n^{w,p}(c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\varepsilon_i^p 1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)}, \tag{1.1}$$

where w_{in} are the weights, ε_i^p is the mark and $c \in \mathbb{R}$. The main result is an asymptotic expansion of the empirical distribution function $\hat{\mathsf{F}}_n^{w,p}(c)$ uniformly in c. The proof of the result has two ingredients. First, we show that, asymptotically, the empirical distribution function $\hat{\mathsf{F}}_n^{w,p}(c)$ does not depend on estimation errors. In order words, the empirical distribution function of residuals $\hat{\mathsf{F}}_n^{w,p}(c)$ is close to the empirical distribution function of the true errors

$$\mathsf{F}_n^{w,p}(c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c)}.$$

Second, we derive a Glivenko-Cantelli theorem for the empirical distribution function $\mathsf{F}_n^{w,p}(c)$. These two results are combined to derive uniform consistency of $\hat{\mathsf{F}}_n^{w,p}(c)$.

The problem does not appear to have been studied much, apart from some non-uniform results in Johansen and Nielsen (2009). Rather, the literature has focused on the empirical process formed from the empirical distribution function, that is,

$$\hat{\mathbb{F}}_n^{w,p}(c) = n^{1/2}\{\hat{\mathsf{F}}_n^{w,p}(c) - \bar{\mathsf{F}}_n^{w,p}(c)\},$$

where $\bar{\mathsf{F}}_n^{w,p}(c)$ is a suitable compensator. A variety of results for this empirical process exist in the literature. Billingsley (1968) studied the case without weights, marks and estimation error. Koul and Ossiander (1994), see also Koul (2002), considered the weighted empirical processes, in which the scale is known and $p = 0$, so that there are no marks. Johansen and Nielsen (2016a) and Berenguer-Rico, Johansen and Nielsen (2019) considered the general case with weights and marks. The marked empirical process of Koul and Stute (1999), and 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 considered here.

By focusing on the empirical distribution function $\hat{\mathsf{F}}_n^{w,p}(c)$ rather than the empirical process $\hat{\mathbb{F}}_n^{w,p}(c)$, we achieve simpler regularity conditions and a simpler proof. Plainly, to obtain uniform consistency from uniform weak convergence imposes too strong assumptions with a too complicated proof – the same way as deriving a law of large numbers result from a central limit theorem would do. Thus, the technical contributions are two-fold. First, the results for $\hat{\mathbb{F}}_n^{w,p}(c)$ require certain moment conditions on the errors and regressors and impose regularity assumptions on the density of ε_i/σ, whereas the results for $\hat{\mathsf{F}}_n^{w,p}(c)$ require weaker moment conditions and impose regularity assumptions on the distribution function of ε_i/σ. Hence, the distribution function need not be differentiable everywhere when analyzing $\hat{\mathsf{F}}_n^{w,p}(c)$ and the amount of moments required is substantially lower. Second, in both cases the proofs require chaining arguments. The chaining argument is considerably simpler in the case of the empirical distribution functions. Having, said that, since the marks are unbounded, we still need to use the iterated exponential martingale inequality of Johansen and Nielsen (2016a).

We apply the theory to study 1-step Huber-skip estimators. These estimators cover a widespread practice in applied work. It is well known that outliers can severely affect the results of a regression analysis. For this reason, applied researches often remove atypical values from the data, which have been previously identified via the residuals of an initial estimation of the model. Once the outliers have been removed, the model is re-estimated with the selected observations. This procedure can be described by what is known in the literature as 1-step Huber-skip estimators –see Rupert and Carroll (1980), Welsh and Ronchetti (2002), Johansen and Nielsen (2009, 2016b) or Jiao and Nielsen (2017). These estimators also appear as building blocks in iterative outlier detection algorithms such as the Forward Search, see Atkinson, Riani and Cerioli (2010), and Impulse Indicator Saturation within Autometrics by Doornik (2009), see also Pretis, Reade and Sucarrat (2018). Specifically, the 1-step Huber-skip estimators of β and σ^2 are

$$\hat{\beta}_c = \{\textstyle\sum_{i=1}^n x_i x_i' 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)}\}^{-1} \textstyle\sum_{i=1}^n x_i y_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)},$$
$$\hat{\sigma}_c^2 = (\tau_0^c/\tau_2^c)\{\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)}\}^{-1} \textstyle\sum_{i=1}^n (y_i - x_i'\hat{\beta})^2 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)},$$

for initial estimators $\tilde{\beta}, \tilde{\sigma}^2$, residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ and where $\tau_k^c = \mathsf{E}\{(\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i/\sigma| \leq c)}\}$. The sums in $\hat{\beta}_c$ and $\hat{\sigma}_c^2$ can be written in terms of marked and weighted empirical distribution functions of residuals. For instance, introducing first the model for y_i in $\hat{\beta}_c$ and normalizations $x_{in} = N' x_i$ for a deterministic normalization matrix N, we can write

$$n^{-1/2} N^{-1}(\hat{\beta}_c - \beta) = \{n^{-1} \textstyle\sum_{i=1}^n n x_{in} x_{in}' 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)}\}^{-1} n^{-1} \textstyle\sum_{i=1}^n n^{1/2} x_{in} \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)}. \tag{1.2}$$

The numerator and denominator can be written in terms of the marked and weighted empirical distribution functions in (1.1) as follows. Let $w_{in} = n^{1/2} N' x_i (n^{1/2} N' x_i)' = n x_{in} x_{in}'$ for the denominator and $w_{in} = n^{1/2} N' x_i = n^{1/2} x_{in}$ for the numerator. Then, noting that the indicator functions in (1.2) are two sided, we can write

$$n^{-1/2} N^{-1}(\hat{\beta}_c - \beta) = \{\hat{\mathsf{F}}_n^{xx,0}(c) - \hat{\mathsf{F}}_n^{xx,0}(-c^-)\}^{-1} \{\hat{\mathsf{F}}_n^{x,1}(c) - \hat{\mathsf{F}}_n^{x,1}(-c^-)\},$$

where $\hat{\mathsf{F}}_n^{w,p}(-c^-) = \lim_{h \to 0} \hat{\mathsf{F}}_n^{w,p}(-c-h)$. Hence, the theory presented in this paper allows us to consider various aspects of these estimators.

First, we derive sufficient conditions for the uniform consistency of $\hat{\beta}_c$ and $\hat{\sigma}_c^2$. Supposing that the initial estimators are consistent, the following features arise. In models with intercept, the estimators of intercept and scale are inconsistent when $\tau_1^c \neq 0$, in particular, when the innovation distribution is asymmetric. The slope coefficients are always consistent, but standard inference can be misleading due to the inconsistency of the scale estimator $\hat{\sigma}_c^2$. Correct inference requires a bias correction of $\hat{\sigma}_c^2$. In models without intercept and $\tau_1^c \neq 0$, the consistency of $\hat{\beta}_c$ depends on the type of regressors. It is inconsistent for stationary regressors with a non-zero mean, but consistent for random walk regressors. However, when $\tau_1^c \neq 0$, the scale estimator $\hat{\sigma}_c$ is inconsistent regardless of the stationarity properties of the regressors. Hence, a bias correction is required.

Second, we study the consistency properties of the empirical gauge, which is defined in terms of the number of falsely detected outliers, when no outliers are present. The notion of gauge was introduced by Hendry and Santos (2010), and it is based on an idea of Hoover and Perez (1999) in the context of variable selection –see also Hendry and Doornik (2014) and Johansen and Nielsen (2016b). The idea, in an outlier detection context, is that finding outliers in a sample is a multiple testing problem. Algorithms for that purpose involve tuning parameters, such as the cut-off c in the case of 1-step Huber-skip estimators, for instance. These tuning parameters can be chosen so as to control type I errors, that is the false detection of outliers. Controlling the size of the algorithms is one such approach, that is, controlling the probability of making no false detections at all. The size will inevitably be fragile with respect to the number of decisions. In the extreme, if we make n independent decisions with individual size p, then the overall size is $1-(1-p)^n$, which depends on n. In this example, the gauge or frequency of falsely detected outliers is simply the average of the individual sizes, that is p. Hence, it is independent of the number of decisions.

The empirical gauge for the 1-step Huber-skip estimator is defined as

$$\hat{\gamma}_c = n^{-1}\textstyle\sum_{i=1}^n 1_{(\tilde{\varepsilon}_i > \tilde{\sigma} c)}.$$

We obtain sufficient conditions under which $\hat{\gamma}_c$ converges to $\gamma_c = \mathsf{P}(|\varepsilon_1/\sigma| > c)$ uniformly in c. This result allows the investigator to set a level for the population gauge, say $\mathsf{P}(|\varepsilon_1/\sigma| > c) = 0.01$, which defines a cut-off value c. In particular, in a sample with $n = 100$ observations, a gauge of 1% corresponds to, on average, falsely declaring one observation as outlier. This calculation is feasible when the innovations are normal or follow some other known reference distribution. The practical consequence is that we can choose the cut-off c so as to control the gauge in an uncontaminated sample, just as the critical value for a standard test is chosen under the hypothesis. The asymptotic properties of the gauge have been studied by Johansen and Nielsen (2016b) and Jiao and Nielsen (2017) using theory for the process $\mathbb{F}_n$. With the new results for $\hat{\mathsf{F}}_n$, we can present significantly weaker assumptions.

The paper is organized as follows. In §2, the model and tools related to the empirical distribution function of residuals are described. In §3, the uniform weak consistency results for the marked and weighted empirical distribution function are presented. Applications to 1-step Huber-skip estimators and the gauge are given in §4. Most proofs are collected in the Appendix.

2 Model and main tools

For $i = 1, \ldots, n$ consider the multiple regression model

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

with possibly stochastic regressors x_i, unknown parameters β and scale σ. We assume that the scaled innovations ε_i/σ are independent, identically distributed with density f and distribution function $\mathsf{F}(c) = \mathsf{P}(\varepsilon_i/\sigma \leq c)$, as well as independent of x_i, $x_{i-1}, x_{i-2}, \ldots$, see Assumption 3.1.

Let $\tilde{\beta}, \tilde{\sigma}$ be estimators of the unknown β, σ. From $\tilde{\beta}$, we compute the residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$. Let N be a deterministic normalization matrix and define the normalized regressors and estimation errors

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

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, say $x_i = 2^i$, 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).

Normalized estimation errors, $\tilde{a}, \tilde{b}$, and regressors, x_{in}, will be entering the marked and weighted empirical distribution function of residuals as follows

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

with weight w_{in} and mark ε_i^p. Some relevant examples of the weights in applications are $w_{in} = 1$ and $w_{in} = n^{1/2}N'x_i = n^{1/2}x_{in}$ and $w_{in} = n^{1/2}N'x_i(n^{1/2}N'x_i)' = nx_{in}x_{in}'$, see, for instance, the 1-step Huber-skip estimator in (1.2).

We use three main tools when deriving a uniform law of large numbers for $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ in (2.3). The *first tool* allows us to replace $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ by

$$\mathsf{F}_n^{w,p}(a, b, c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\varepsilon_i^p 1_{(\varepsilon_i/\sigma \leq c + n^{-1/2}ac + x_{in}'b)}, \tag{2.4}$$

and study $\mathsf{F}_n^{w,p}(a, b, c)$ uniformly over a, b varying in expanding compact sets depending on n and $c \in \mathbb{R}$. Specifically, we make use of Lemma 3.1 in Berenguer-Rico, Johansen and Nielsen (2019). It states that if $\lim_{n\to\infty} \mathsf{P}(\tilde{\theta} \in \Theta) > 1 - \epsilon$ for some estimator $\tilde{\theta}$ in a compact set Θ and $\epsilon > 0$, then for any function $F_n(\theta, c)$ of $\theta \in \Theta$ and $c \in \mathbb{R}$, we have that

$$\mathsf{P}\{|F_n(\tilde{\theta}, c)| > \epsilon\} \leq \mathsf{P}\{\sup_{\theta \in \Theta} |F_n(\theta, c)| > \epsilon\} + \epsilon. \tag{2.5}$$

The *second tool* is a chaining argument which allows us to derive the required uniformity results over a, b, c. The argument is as follows, see also Berenguer-Rico, Johansen and Nielsen (2019). Consider the process $F_n(\theta, c)$ where $\theta \in \Theta$ and $c \in \mathbb{R}$. Introduce K grid points c_k and cover the set Θ by M balls with centres θ_m with a small radius δ. The 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 oscillation term.

The *third tool* is the lemma quoted below, which uses the iterated exponential martingale inequality by Johansen and Nielsen (2016a). In turn, it is based on the exponential martingale inequality for unbounded martingales by Bercu and Touati (2008), see also Bercu, Delyon and Rio (2015). Koul and Ossiander (1994) use the inequality by Freedman (1975) for bounded martingales in their analysis of the weighted empirical process of residuals. That approach is impractical with the inclusion of unbounded marks.

Lemma 2.1 (Johansen and Nielsen, 2016a, Lemma 4.2) *For $1 \leq \ell \leq L_n$, $1 \leq i \leq n$ 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

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

Lemma 2.1 simplifies when $z_{n\ell i}$ is an indicator function.

Lemma 2.2 *For $1 \leq \ell \leq L_n$, $1 \leq i \leq n$ let $z_{n\ell i}$ be $\mathcal{F}_{in}$ adapted indicator function. Let $D_n = \max_{1\leq \ell \leq L_n} \sum_{i=1}^n \mathsf{E}_{i-1} z_{n\ell i}$. Suppose, for some $\lambda > 0$, $\varsigma \geq 0$, that $L_n = \mathrm{O}(n^\lambda)$ and $\mathcal{E}_n = \mathsf{E}D_n = \mathrm{O}(n^\varsigma)$. If, in addition, $\upsilon > \varsigma/2$, then*

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

Proof. Since $|z_{n\ell i}|$ is an indicator function, then $z_{n\ell i}^{2^r} = |z_{n\ell i}|$ for any $r \in \mathbb{N}$. Therefore, we can apply Lemma 2.1 with r chosen so large that condition (ii) is satisfied. ∎

3 Uniform consistency results

In order to find a uniform Law of Large Numbers for $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$, see (2.3), we decompose

$$\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) = \mathsf{F}_n^{w,p}(0, 0, c) + \{\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) - \mathsf{F}_n^{w,p}(0, 0, c)\}. \tag{3.1}$$

The first term, $\mathsf{F}_n^{w,p}(0, 0, c)$, has no estimation error. For a fixed c, it is analyzed using a martingale Law of Large Numbers. For a varying c, Theorem 3.2 below can be used to get uniform convergence in c. The second term, $\{\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) - \mathsf{F}_n^{w,p}(0, 0, c)\}$, vanishes uniformly in c due to Theorem 3.1 below combined with the first tool (2.5) described above. This result shows that, asymptotically, $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$ does not depend on the estimation errors $\tilde{a}, \tilde{b}$. The main result, Theorem 3.3 below, combines Theorems 3.1, the inequality (2.5) and Theorem 3.2 to show the uniform consistency of $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c)$.

First, we prove the uniform convergence of $\mathsf{F}_n^{w,p}(a, b, c)$ under the following assumptions.

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 continuous distribution function F.*

Definition 3.1 *A distribution function* F *is*
(i) ***additively Lipschitz*** *if* $\exists C_L > 0 : \forall b, c \in \mathbb{R}$ *it holds that* $|\mathsf{F}(c+b) - \mathsf{F}(c)| \leq C_L|b|$;
(ii) ***multiplicatively Lipschitz*** *if* $\exists C_L, a_0 > 0 : \forall |a| \leq a_0, c \in \mathbb{R}$ *it holds that* $|\mathsf{F}(c+ca) - \mathsf{F}(c)| \leq C_L|a|$.

Remark 3.1 *Examples of Definition 3.1 include the normal, the triangular and the uniform distribution. More generally, if* F *has support on an open interval* $\mathcal{S}$ *and derivative* f *on* $\mathcal{S}$ *satisfying* $\sup_{c\in\mathbb{R}}(1+|c|)\mathsf{f}(c) < \infty$, *then* F *is additively and multiplicatively Lipschitz. To see this, first note that by the Mean Value Theorem,* $\mathsf{F}(c+b) - \mathsf{F}(c) = b\mathsf{f}(c^*)$ *for some* c^* *so* $|c^* - c| \leq b$. *Here,* $\mathsf{f}(c^*)$ *is bounded by assumption. Second, replacing* b *by* ca *gives that* $|\mathsf{F}(c+ca) - \mathsf{F}(c)| = |a||c|\mathsf{f}(c^*)$ *for* $|c^*-c| \leq |a||c|$. *By the triangle inequality,* $|c| \leq |c^*| + |c^* - c|$ *so that* $|c| \leq |c^*| + |a||c|$ *and* $|c|(1-|a|) \leq |c^*|$. *Thus, if* $|a| \leq a_0 = 1/2$ *then* $|c| \leq 2|c^*|$, *so that* $|\mathsf{F}(c+ca) - \mathsf{F}(c)| \leq 2|a||c^*|\mathsf{f}(c^*)$, *where* $|c^*|\mathsf{f}(c^*)$ *is bounded by assumption.*

Assumption 3.2 *Suppose, for* $p \in \mathbb{N}_0$ *and some* $\psi > 1$, *that*
(i) *innovations* ε_i/σ *satisfy*
  (a) *moments:* $\mathsf{E}|\varepsilon_i|^{p\psi} < \infty$;
  (b) F *is additively and multiplicatively Lipschitz (Definition 3.1);*
(ii) *regressors* x_{in} *satisfy* $\mathsf{E}n^{-1}\sum_{i=1}^n n^{1/2}|x_{in}| = \mathrm{O}(1)$;
(iii) *weights* w_{in} *satisfy* $\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|^{\psi} = \mathrm{O}(1)$.

Theorem 3.1 *Let Assumptions 3.1, 3.2 hold. Then, for any* $B > 0$, $\psi > 1$, $0 < \zeta < 1$,

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

The proof of Theorem 3.1 exploits that, with the inbuilt n-normalization in $\mathsf{F}_n^{w,p}$, the marks and weights can be separated from the indicators using Hölder's inequality. Thus, initially, we focus on the special case of Theorem 3.1 without marks and weights. This result is presented in Lemma A.4 in the Appendix. Its proof relies on two other intermediate lemmas. First, in Lemma A.2 we set $a = 0$ and chain over c while bounding the influence of b. Second, in Lemma A.3 we set $b = 0$ and deal with the scale estimation error a. Since a is as a multiplicative distortion of the quantile c, we chain over two quantiles simultaneously. These two results are then used in proving Lemma A.4.

Next, we establish an *in probability* version of the Glivenko-Cantelli theorem.

Theorem 3.2 *Let Assumptions 3.1, 3.2(ia, iii) hold with* $\psi = 2$. *Then, for any* $\omega > 0$,

$$\sup_{c\in\mathbb{R}} |\mathsf{F}_n^{w,p}(0,0,c) - (n^{-1}\textstyle\sum_{i=1}^n w_{in})\mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)}| = \mathrm{O}_\mathsf{P}(n^{\omega-1/3}).$$

Finally, we combine the theorems above to derive the main result, which analyzes the terms in the decomposition (3.1). We consider normalized estimators $\tilde{a}, \tilde{b}$ of order $\mathrm{O}_\mathsf{P}(n^{1/2-\zeta})$ for $0 < \zeta < 2/3$. In standard models $\zeta = 1/2$. However, when proving the following main result, the combination of Theorems 3.1 and 3.2 allows for $0 < \zeta < 2/3$.

Theorem 3.3 ***Uniform consistency result***. *Let Assumptions 3.1, 3.2 hold with $\psi = 2$ and suppose the estimators satisfy $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma = \mathsf{O}_\mathsf{P}(n^{1/2-\zeta})$ and $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)/\sigma = \mathsf{O}_\mathsf{P}(n^{1/2-\zeta})$ for some $0 < \zeta < 2/3$. Then*

$$\sup_{c\in\mathbb{R}} |\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) - (n^{-1}\textstyle\sum_{i=1}^n w_{in})\mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)}| = \mathsf{O}_\mathsf{P}(n^{-\zeta/2}).$$

Proof of Theorem 3.3. Since $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)$ and $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)$ are $\mathsf{O}_\mathsf{P}(n^{1/2-\zeta})$, we can use (2.5) and consider $R_n(a, b, c) = |\mathsf{F}_n^{w,p}(a, b, c) - (n^{-1}\sum_{i=1}^n w_{in})\mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)}|$ uniformly in $|a|, |b| \leq n^{1/2-\zeta}B$, $c \in \mathbb{R}$. Applying the triangle inequality and Theorems 3.1, 3.2 using Assumptions 3.1, 3.2 with $\psi = 2$ shows that $R_n(a, b, c) = \mathsf{O}_\mathsf{P}(n^{-\zeta/2}) + \mathsf{O}_\mathsf{P}(n^{\omega-1/3})$ uniformly in a, b, c for any $\omega > 0$ and where $0 < \zeta < 1$. Note that if $\zeta = 2/3 - \epsilon$ for some $\epsilon > 0$, then we can always find an $\omega < \epsilon/2$ so that the first term dominates. Hence, the first term dominates for $0 < \zeta < 2/3$. In turn, $R_n(a, b, c) = \mathsf{O}_\mathsf{P}(n^{-\zeta/2})$ for $0 < \zeta < 2/3$ as desired. ■

Theorem 3.3 provides a stochastic expansion in terms of $n^{-1}\sum_{i=1}^n w_{in}$. Depending on the nature of the weights, this will be deterministic or random.

Example 3.1 *Weights w_{in} are i.i.d. with mean μ. Then, $n^{-1}\sum_{i=1}^n w_{in} \overset{\mathsf{P}}{\to} \mu$, so that $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) = \mu\mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)} + \mathsf{o}_\mathsf{P}(1)$ uniformly in c.*

Example 3.2 *Weights w_{in} are normalized random walks $w_{in} = n^{-1/2}\sum_{j=1}^i \eta_j$ for i.i.d., zero mean η_j with unit variance. Then, $n^{-1}\sum_{i=1}^n w_{in} \overset{\mathsf{D}}{\to} \int_0^1 W_u du$ for a standard Brownian motion W_u so that $\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) \overset{\mathsf{D}}{\to} (\int_0^1 W_u du)\mathsf{E}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)}$ uniformly in c.*

4 Application to 1-step Huber-skip estimators

We now apply the above results to show, first, the uniform consistency of 1-step Huber-skip estimators and, second, the uniform consistency of its associated gauge.

4.1 Uniform consistency of the estimators

We define the 1-step Huber-skip estimators as follows. Let $\tilde{\beta}$ and $\tilde{\sigma}$ be initial estimators from which we can form the residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$. Observations with large scaled residuals $\tilde{\varepsilon}_i/\tilde{\sigma}$, that is, satisfying $|\tilde{\varepsilon}_i/\tilde{\sigma}| > c$ for a cut-off value c set up by the investigator, are removed from the sample and least squares regression is applied to the new sample giving the 1-step estimators

$$\hat{\beta}_c = \{\textstyle\sum_{i=1}^n x_i x_i' 1_{(|\tilde{\varepsilon}_i|\leq\tilde{\sigma}c)}\}^{-1}\sum_{i=1}^n x_i y_i 1_{(|\tilde{\varepsilon}_i|\leq\tilde{\sigma}c)}, \tag{4.1}$$

$$\hat{\sigma}_c^2 = (\tau_0^c/\tau_2^c)\{\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i|\leq\tilde{\sigma}c)}\}^{-1}\sum_{i=1}^n (y_i - x_i'\hat{\beta})^2 1_{(|\tilde{\varepsilon}_i|\leq\tilde{\sigma}c)}, \tag{4.2}$$

where

$$\tau_k^c = \textstyle\int_{-c}^c v^k \mathsf{f}(v)dv = \mathsf{E}\{(\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i/\sigma|\leq c)}\}.$$

We apply the above results to analyze the uniform weak consistency properties of $\hat{\beta}_c$, $\hat{\sigma}_c^2$.

The 1-step Huber skip estimators mathematically describe the following widely used practice in applied work; namely, the procedure where a researcher estimates a model using some initial estimator and, based on this initial estimate, removes atypical values from the sample. Then, in a second stage the model is re-estimated by least squares using only the remaining observations. This practice has been termed by Welsh and Ronchetti (2002) the *data analytic strategy.* The asymptotic properties of this procedure have been previously analyzed by Ruppert and Carroll (1980), Johansen and Nielsen (2009, 2013, 2016a, 2016b) and Jiao and Nielsen (2017). These papers have derived asymptotic expansions for the rescaled 1-step Huber-skip estimator. In this section, we focus on uniform consistency only and achieve weaker regularity conditions.

Assumption 4.1 *Part I. Suppose that*
(i) *innovations* ε_i/σ *satisfy*
  (a) *moments:* $\mathsf{E}|\varepsilon_i|^2 < \infty$;
  (b) F *is additively and multiplicatively Lipschitz (Definition 3.1);*
(ii) *regressors* x_{in} *satisfy*
  (a) $n^{-1}\mathsf{E}\sum_{i=1}^n |n^{1/2}x_{in}|^4 = \mathrm{O}(1)$;
  (b) $\hat{\Sigma}_n = \sum_{i=1}^n x_{in}x_{in}' \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0$ *and* $\hat{\mu}_n = n^{-1/2}\sum_{i=1}^n x_{in} \overset{\mathsf{D}}{\to} \mu$;
(iii) *the initial estimation errors* $N^{-1}(\tilde{\beta} - \beta)$, $n^{1/2}(\tilde{\sigma} - \sigma)$ *are* $\mathrm{O}_{\mathsf{P}}(n^{1/2-\zeta})$ *for some* $\zeta > 0$.
Part II. Suppose $\mathsf{E}|\varepsilon_i|^4 < \infty$ *and* $\tau_2^{c_0} > 0$ *for some* $c_0 > 0$.

We note that the limiting quantities Σ and μ in Assumption 4.1(ii), (b) will be deterministic in stationary settings whereas they can be stochastic in non-stationary models.

Theorem 4.1 *Suppose Assumptions 3.1, 4.1,I are satisfied. Choose* $c_0 > 0$ *so that* $\tau_0^{c_0} > 0$. *Then, uniformly in* $c \geq c_0$,

$$n^{-1/2}N^{-1}(\hat{\beta}_c - \beta) = \sigma\hat{\Sigma}_n^{-1}\hat{\mu}_n(\tau_1^c/\tau_0^c) + \mathrm{o}_{\mathsf{P}}(1) \overset{\mathsf{D}}{\to} \sigma\Sigma^{-1}\mu(\tau_1^c/\tau_0^c).$$

If, in addition Assumption 4.1,II is satisfied, then uniformly in $c \geq c_0$,

$$\hat{\sigma}_c^2 - \sigma^2 = -\sigma^2\hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n(\tau_1^c)^2/(\tau_2^c\tau_0^c) + \mathrm{o}_{\mathsf{P}}(1) \overset{\mathsf{D}}{\to} -\sigma^2\mu'\Sigma^{-1}\mu(\tau_1^c)^2/(\tau_2^c\tau_0^c).$$

Remark 4.1 *In a model with an intercept, the bias expressions simplify. To see this, suppose* $x_i = (1, z_i')'$ *and* $x_{in} = (n^{-1/2}, z_{in}')'$. *We then get* $\hat{\mu}_n' = (1,0)\Sigma_n$ *so that* $\hat{\mu}_n'\Sigma_n^{-1} = (1,0)$ *and* $\hat{\mu}_n'\Sigma_n^{-1}\hat{\mu}_n = 1$, *even with non-stationary regressors. The consistency results are then*

$$n^{-1/2}N^{-1}(\hat{\beta}_c - \beta) \overset{\mathsf{P}}{\to} \sigma\begin{pmatrix}1\\0\end{pmatrix}\frac{\tau_1^c}{\tau_0^c}, \qquad \hat{\sigma}_c^2 - \sigma^2 \overset{\mathsf{P}}{\to} -\sigma^2\frac{(\tau_1^c)^2}{\tau_2^c\tau_0^c}.$$

We see that the intercept and variance estimators can be biased, while the slope estimators are not biased. This feature is shared with, for instance, quantile regression, see Koenker and Bassett (1978). Moreover, the bias only depends on the τ_k^c *coefficients that are derived from the innovation distribution. When* $\tau_1^c \neq 0$ *the intercept is biased and, more critically, standard inference on both intercept and slopes is distorted because of the bias in* $\hat{\sigma}_c^2$. *Hence, a bias correction should be implemented. These results match the finding from least squares theory that a necessary condition for consistency of* β *is that* ε_i/σ *has mean zero, which ensures that the intercept is identified.*

Remark 4.2 *In a model without intercept, the bias on β depends on the properties of the regressors. First, for a stationary regressor we choose $N^{-1} = n^{1/2}$ so that $n^{-1/2}N^{-1}$ is the identity. In that case,*

$$\hat{\beta}_c - \beta = \sigma\hat{\Sigma}_n^{-1}\hat{\mu}_n(\tau_1^c/\tau_0^c) + \mathrm{o}_{\mathsf{P}}(1) \xrightarrow{\mathsf{P}} \sigma\Sigma^{-1}\mu(\tau_1^c/\tau_0^c).$$

The bias is present when $\mu\tau_1^c \neq 0$. If the stationary regressor has zero mean, then $\mu = 0$. If the stationary regressor has non-zero mean, then $\mu \neq 0$ and the bias occurs if $\tau_1^c \neq 0$. Second, for random walk regressors, we choose $N^{-1} = n$ and $n^{-1/2}N^{-1} = n^{1/2}$ so that $\hat{\beta}_c - \beta = \mathrm{o}_{\mathsf{P}}(1)$. Thus, in this case, $\hat{\beta}_c$ is consistent for all the coeficients even when $\tau_1^c \neq 0$. Nonetheless, if $\tau_1^c \neq 0$, the bias in $\hat{\sigma}_c^2$ is present regardless of the stationarity properties of the regressors. Therefore, a bias correction for $\hat{\sigma}_c^2$ is needed to conduct correct inferences.

Remark 4.3 *The biases are proportional to the coefficient τ_1^c. For a symmetric distribution, $\tau_1^c = 0$. For a non-symmetric distribution, τ_1^c can be non-zero. If ε_i/σ follows a distribution that is standard normal in the middle and unspecified in the tails, we can use the result to detect which part of the distribution is normal. To be precise, suppose ε_i/σ has a density $\mathsf{f}(x)$ that is standard normal for $|x| \leq \psi$ for some $\psi \in \mathbb{R}_0$ and asymmetric for $|x| > \psi$. In the context of a model with an intercept, Theorem 4.1 shows that the process of 1-step Huber-skip intercept estimators $(1,0)(\hat{\beta}_c - \beta)$ will converge to a function that is zero for $c \leq \psi$ whereas, depending on the form of the asymmetry, it is non-zero for $c > \psi$.*

Remark 4.4 *The result in Theorem 4.1 gives sufficient conditions for consistency that are weaker than conditions previously derived in the literature. Ruppert and Carroll (1980) derive an asymptotic expansion for $n^{1/2}(\hat{\beta}_c - \beta)$ in the case of fixed regressors, from which they obtain pointwise consistency for $\hat{\beta}_c$. Jiao and Nielsen (2017) derive an asymptotic expansion for $N^{-1}(\hat{\beta}_c - \beta)$ in a general setting, similar to the setup in this paper, but with stronger conditions on the error term and the regressors. Specifically, the error term, on the one hand, is assumed to have a density that is symmetric, satisfies certain smoothness conditions and has moments which depend on the dimension of the regressors. The regressors, on the other hand, are assumed to satisfy a maximal condition imposing in practice the existence of many more moments than those required here. Overall, as shown in Theorem 4.1 above, the conditions to derive uniform consistency can be substantially weakened.*

4.2 Uniform consistency of the gauge

An important question in relation to the 1-step Huber-skip estimator is how does the investigator choose the cut-off value c? The 1-step Huber-skip *algorithm* to detect outliers is defined as follows. Let $\tilde{\beta}$ and $\tilde{\sigma}$ be initial estimators from which we form the residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$. Choose a cut-off $c > 0$. The 1-step Huber-skip estimators $\hat{\beta}_c, \hat{\sigma}_c^2$ are the least squares estimators on observations satisfying $|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c$ as described in (4.1), (4.2). Hence, the observations satisfying $|\tilde{\varepsilon}_i/\tilde{\sigma}| > c$ are declared outliers. The empirical gauge is the frequency of declared outliers, that is

$$\hat{\gamma}_c = n^{-1}\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}|>c)}, \tag{4.3}$$

as introduced by Hendry and Santos (2010) and based on an idea of Hoover and Perez (1999). The population gauge, say γ_c, is the probability limit of $\hat{\gamma}_c$, when there are no outliers. We can choose the cut-off c indirectly, from a given value of the population gauge γ_c.

The asymptotic properties of the gauge in the context of 1-step Huber-skip estimators have been analyzed by Johansen and Nielsen (2016b) and Xiao and Nielsen (2017). Their analysis is based on expansions of the corresponding marked and weighted empirical process. Hence, their derived sufficient conditions can be weakened by using the results above. Note that in this case, $p = 0$ and $w_{in} = 1$ in the marked and weighted empirical distribution function of interest. Hence, the assumptions to Theorem 3.3 simplify as follows.

Assumption 4.2 *Suppose that*
(i) innovations ε_i/σ satisfy: F *is additively and multiplicatively Lipschitz (Definition 3.1);*
(ii) regressors x_{in} satisfy $n^{-1}\mathsf{E}\sum_{i=1}^{n}|n^{1/2}x_{in}| = \mathrm{O}(1)$;
(iii) the initial estimation errors $N^{-1}(\tilde{\beta} - \beta)$, $n^{1/2}(\tilde{\sigma} - \sigma)$ are $\mathrm{O}_{\mathsf{P}}(n^{1/2-\zeta})$ for some $\zeta > 0$.

Theorem 4.2 *Suppose Assumptions 3.1, 4.2 are satisfied. Choose $c_0 > 0$ so that $\tau_0^{c_0} = \mathsf{P}(|\varepsilon_i/\sigma| \leq c_0) > 0$. Then, uniformly in $c \geq c_0$,*

$$\hat{\gamma}_c \xrightarrow{\mathsf{P}} (1 - \tau_0^c).$$

Remark 4.5 *Theorem 4.2 shows that the empirical gauge converges, uniformly in c, to the population gauge, γ_c, so that*

$$\gamma_c = (1 - \tau_0^c) = \mathsf{P}(|\varepsilon_1/\sigma| > c).$$

This gives a way of establishing an overall rate of false rejection when deciding whether observation i, for $i = 1, ..., n$, is an outlier or not. In particular, assuming a distribution for the error term and choosing $\mathsf{P}(|\varepsilon_1/\sigma| > c)$ to be a certain value, one gets a cut-off value c. For instance, under normality, if γ_c is set to be 0.01, then $c = 2.57$. In this case, the gauge coincides with the size of the individual tests. See also Johansen and Nielsen (2016b) for a discussion on the idea of gauge and its relation to the false discovery rate for multiple test of Benjamini and Hochberg (1995).

Remark 4.6 *Theorem 7 in Johansen and Nielsen (2016b) derives pointwise consistency of the gauge under the assumption of normal errors and a list of conditions on the normalized regressors such as convergence of the first and second empirical moments, boundedness of the maximum over i and the existence of 9th moments. Theorem 4.2 above derives uniform consistency under the much weaker assumption 4.2.*

5 Concluding remarks

The main result is Theorem 3.3, which gives a uniform consistency result of the weighted and marked empirical distribution function of estimated residuals. The empirical distribution is consistent for the limit of the product of the average weight times the expectation of the truncated mark.

The result simplifies previous work on asymptotic expansions of the empirical distribution function and on the associated empirical process. Both the regularity conditions and the proofs are simplified in the same way as the Law of Large Number is simpler than the Central Limit Theorem. In particular, the consistency result requires fewer moments than the asymptotic expansions, and it requires continuity and smoothness of the distribution function rather than of the density function.

The result was 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. In particular, we show consistency of the estimators and the associated gauge under weaker conditions than in the previous literature.

A Appendix

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.

A.1 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 K disjoint intervals with end points

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

The definitions $c_0 = -\infty$ and $c_K = \infty$ are convenient even when the support is finite. In Johansen and Nielsen (2016a) and Berenguer-Rico, Johansen and Nielsen (2019), these chaining points are chosen using the function

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

for a given $r = 0, 1, \ldots$. Here, we will need $r = 0$ and $K \sim n^{1/2}$ when proving Theorem 3.1, while $r = 1$ and $K \sim n^{1/3-\omega}$ when proving Theorem 3.2. The function H_r is increasing in c. It is bounded when

$$H_r = \mathsf{H}_r(\infty) = \mathsf{E}(1 + |\varepsilon_i/\sigma|^{2^r p}) < \infty. \tag{A.3}$$

The points c_k are chosen so that

$$\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \qquad \text{for } k = 0, 1, \ldots, K. \tag{A.4}$$

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 $c^\dagger$.

A.2 A preliminary lemma

The following Lemma, based on Lemma A.10 in Berenguer-Rico, Johansen and Nielsen (2019), is used in the proof of the main results when chaining over a, c. On the one hand, it bounds the effect of multiplicative perturbations using the H_r distance. On the other hand, it gives an estimate of the number of c_k intervals that are needed to cover the perturbation. The result applies for general distance functions, say H.

Lemma A.1 *Let $c_a = c(1 + n^{-1/2}a)$ so that $c_0 = c$. Let $\mathsf{H}(c)$ be non-decreasing and multiplicatively Lipschitz with $H = \mathsf{H}(\infty) - \mathsf{H}(-\infty) < \infty$. Choose grid points c_k as in (A.1) so that $\mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = H/K$ for all k. Then:*
(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

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

Proof of Lemma A.1. (a) *The distance $\mathcal{H} = \mathsf{H}(c_a) - \mathsf{H}(c)$.* Since $|n^{-1/2}a| \leq n^{-\zeta}$, which vanishes for large n, then for any $a_0 > 0$ we have $|n^{-1/2}a| \leq a_0$ for large n. The multiplicative Lipschitz assumption then shows $|\mathcal{H}| \leq Cn^{-\zeta}$ 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

$$\mathcal{H}^* \leq |\mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}(c_a)| + |\mathsf{H}\{c_{k(c_0)}\} - \mathsf{H}(c)| + |\mathsf{H}(c_a) - \mathsf{H}(c)|. \tag{A.6}$$

Each of the first two terms in (A.6) are bounded by H/K. Indeed, since $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$ and, noting that $c_0 = c$,

$$0 \leq \mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}(c_a) \leq \mathsf{H}\{c_{k(c_a)}\} - \mathsf{H}\{c_{k(c_a)-1}\} = H/K.$$

The third term in (A.6) 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. ∎

A.3 Auxiliary results

Theorem 3.1 gives a uniform Law of Large Numbers for $\mathsf{F}_n^{w,p}(a, b, c)$. We start by considering the case without weights and marks, that is $\mathsf{F}_n^{1,0}(a, b, c)$. First, we analyze the special case where $a = 0$. Second, we consider the case where $b = 0$. Finally, both cases are combined to study the general case in which a, b vary.

Lemma A.2 *Let Assumption 3.1 hold and suppose*
(i) F *is additively Lipschitz (Definition 3.1);*
(ii) $\mathsf{E}n^{-1}\sum_{i=1}^{n} n^{1/2}|x_{in}| = \mathrm{O}(1)$.
Then, for any $B > 0$, $0 < \zeta < 1$, *we get*

$$\sup_{c\in\mathbb{R}} \sup_{|b|\le n^{1/2-\zeta}B} n^{-1}\textstyle\sum_{i=1}^{n}|1_{(\varepsilon_i/\sigma\le c+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\le c)}| = \mathrm{O}_{\mathsf{P}}(n^{-\zeta}).$$

Proof of Lemma A.2. Let $R_n(b,c) = n^{-1}\sum_{i=1}^{n}|1_{(\varepsilon_i/\sigma\le c+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\le c)}|$ and $\mathcal{R}_n = \sup_{c\in\mathbb{R}}\sup_{|b|\le n^{1/2-\zeta}B} R_n(b,c)$. We show $\mathcal{R}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

1. *Partition the support.* Because there are no marks, we take $p = 0$ and $r = 0$ so that (A.2) and (A.3) reduce to $\mathsf{H}_r(c) = 2\mathsf{F}(c)$ and $H_r = 2$. Partition the axis as laid out in (A.1) with $K = \mathrm{int}(H_r n^{\zeta})$. Thus, $\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \sim n^{-\zeta}$.

2. *Assign c to the partitioned support.* For each c there exists a $k = k(c)$ and grid points c_{k-1}, c_k so that $c_{k-1} < c \le c_k$.

3. *Bound* $R_n(b,c)$. Since $|b| \le n^{1/2-\zeta}B$ and $c_{k-1} < c \le c_k$,

$$|1_{(\varepsilon_i/\sigma\le c+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\le c)}| \le 1_{(\varepsilon_i/\sigma\le c_k+|x_{in}|n^{1/2-\zeta}B)} - 1_{(\varepsilon_i/\sigma\le c_{k-1}-|x_{in}|n^{1/2-\zeta}B)} = z_{ik}.$$

Thus, $0 \le R_n(b,c) \le \tilde{R}_{nk} = n^{-1}\sum_{i=1}^{n} z_{ik}$, which does not depend on b. In turn $\mathcal{R}_n \le \tilde{\mathcal{R}}_n$ where $\tilde{\mathcal{R}}_n = \max_{1\le k\le K}\tilde{R}_{nk}$. It suffices to show that $\tilde{\mathcal{R}}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

4. *Martingale decomposition.* Write $\tilde{R}_{nk} = \tilde{M}_{nk} + \bar{M}_{nk}$ where

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

It suffices to show that $\tilde{\mathcal{M}}_n = \max_{1\le k\le K}|\tilde{M}_{nk}|$ and $\bar{\mathcal{M}}_n = \max_{1\le k\le K}\bar{M}_{nk}$ are $\mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

5. *Conditional mean of* z_{ik}. The indicator function z_{ik} is $\mathcal{F}_{in}$ adapted. We find

$$\mathsf{E}_{i-1}z_{ik} = \mathsf{F}(c_k + |x_{in}|n^{1/2-\zeta}B) - \mathsf{F}(c_{k-1} - |x_{in}|n^{1/2-\zeta}B).$$

This can rewritten as

$$\mathsf{E}_{i-1}z_{ik} = \{\mathsf{F}(c_k + |x_{in}|n^{1/2-\zeta}B) - \mathsf{F}(c_k)\} + \{\mathsf{F}(c_k) - \mathsf{F}(c_{k-1})\} + \{\mathsf{F}(c_{k-1}) - \mathsf{F}(c_{k-1} - |x_{in}|n^{1/2-\zeta}B)\}.$$

The second term equals $1/K$ by construction. By the additive Lipschitz assumption (i), the first and the third term are each bounded by $C_L|x_{in}|n^{1/2-\zeta}B$. Overall, we get $\mathsf{E}_{i-1}z_{ik} \le 1/K + C|x_{in}|n^{1/2-\zeta}$, for some $C > 0$. In turn,

$$\mathcal{E}_n = \mathsf{E}\max_{1\le k\le K}\textstyle\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ik} \le n/K + Cn^{1/2-\zeta}\mathsf{E}\sum_{i=1}^{n}|x_{in}|.$$

Since $K \sim n^{\zeta}$ and $\mathsf{E}\sum_{i=1}^{n}|x_{in}| = \mathrm{O}(n^{1/2})$ by assumption (ii), we get $\mathcal{E}_n = \mathrm{O}(n^{1-\zeta})$.

6. *The compensator* $\bar{\mathcal{M}}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$. The Markov inequality implies that $\mathsf{P}(n^{\zeta}|\bar{\mathcal{M}}_n| > C) \le \mathsf{E}n_n^{\zeta}|\bar{\mathcal{M}}_n|/C$ for some $C > 0$. Note that $\mathsf{E}\bar{\mathcal{M}}_n = n^{-1}\mathcal{E}_n$. Thus, item 5 shows that $\mathsf{E}\bar{\mathcal{M}}_n = \mathrm{O}(n^{-\zeta})$ so that $\bar{\mathcal{M}}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

7. *The martingale* $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^{-\zeta})$. We use Lemma 2.2 for

$$n\tilde{\mathcal{M}}_n = \max_{1\le k\le K}|\textstyle\sum_{i=1}^{n}(z_{ik} - \mathsf{E}_{i-1}z_{ik})|,$$

with index $\ell = k$ so that $z_{i\ell} = z_{ik}$, the count $L_n = K$ and parameters $\lambda = \zeta > 0$ and $\upsilon = \varsigma = 1 - \zeta$. We verify the conditions of Lemma 2.2. In item 5, it was established that $z_{i\ell} = z_{ik}$ is $\mathcal{F}_{in}$ adapted. Further, item 5 shows that $\mathcal{E}_n = \mathrm{O}(n^{1-\zeta}) = \mathrm{O}(n^{\varsigma})$. Since $0 < \zeta < 1$, then $\varsigma = 1 - \zeta > 0$ and $\upsilon = \varsigma > \varsigma/2$. Hence, applying Lemma 2.2, we see that,

$$n\tilde{\mathcal{M}}_n = n \max_{1 \le k \le K} |\tilde{M}_{nk}| = \max_{1 \le k \le K} |\textstyle\sum_{i=1}^n (z_{ik} - \mathsf{E}_{i-1} z_{ik})| = \mathrm{o}_{\mathsf{P}}(n^{\upsilon}).$$

In particular, $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^{\upsilon - 1}) = \mathrm{o}_{\mathsf{P}}(n^{-\zeta})$. ∎

Lemma A.3 *Let Assumption 3.1 hold and suppose* F *is multiplicatively Lipschitz (Definition 3.1). Then, for any* $B > 0$, $0 < \zeta < 1$, *we get*

$$\sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/2-\zeta} B} n^{-1} \textstyle\sum_{i=1}^n |1_{(\varepsilon_i/\sigma \le c + n^{-1/2} ac)} - 1_{(\varepsilon_i/\sigma \le c)}| = \mathrm{O}_{\mathsf{P}}(n^{-\zeta}).$$

Proof of Lemma A.3. Let $c_a = c + n^{-1/2} ac$ so that $c_0 = c$. Define $R_n(c, c_a) = n^{-1} \sum_{i=1}^n |1_{(\varepsilon_i/\sigma \le c_a)} - 1_{(\varepsilon_i/\sigma \le c)}|$. We show that $\mathcal{R}_n = \sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/2-\zeta} B} |R_n(c, c_a)|$ is $\mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

1. *Partition the support.* Because there are no marks, we take $p = 0$ and $r = 0$ so that (A.2) and (A.3) reduce to $\mathsf{H}_r(c) = 2\mathsf{F}(c)$ and $H_r = 2$. Partition the axis as laid out in (A.1) with $K = \mathrm{int}(H_r n^{\zeta})$. Thus, $\mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-1}) = H_r/K \sim n^{-\zeta}$.

2. *Assign* c *and* c_a *to the partitioned support.* For each c_a there exists a $k(c_a)$ and grid points $c_{k(c_a)-1}, c_{k(c_a)}$ so that $c_{k(c_a)-1} < c_a \le c_{k(c_a)}$. Lemma A.1, using the multiplicative Lipschitz assumption shows that, for some $C > 0$ and some $D > 2$,

$$\sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/2-\zeta} B} |k(c_a) - k(c)| \le 2 + C n^{-\zeta} K/H \le D. \tag{A.7}$$

3. *Bound* $R_n(c, c_a)$ *and* $\mathcal{R}_n$. Add and subtract $1_{\{\varepsilon_i/\sigma \le c_{k(c_a)}\}}$ and $1_{\{\varepsilon_i/\sigma \le c_{k(c)}\}}$ so that

$$\begin{aligned} R_n(c, c_a) = n^{-1} \textstyle\sum_{i=1}^n |&1_{(\varepsilon_i/\sigma \le c_a)} - 1_{\{\varepsilon_i/\sigma \le c_{k(c_a)}\}} \\ &+ 1_{\{\varepsilon_i/\sigma \le c_{k(c_a)}\}} - 1_{\{\varepsilon_i/\sigma \le c_{k(c)}\}} - 1_{(\varepsilon_i/\sigma \le c)} + 1_{\{\varepsilon_i/\sigma \le c_{k(c)}\}}|. \end{aligned}$$

By the triangle inequality we get

$$R_n(c, c_a) \le R_n\{c_a, c_{k(c_a)}\} + R_n\{c_{k(c)}, c_{k(c_a)}\} + R_n\{c, c_{k(c)}\}. \tag{A.8}$$

Accordingly, the triangle inequality gives $\mathcal{R}_n \le \sum_{j=1}^3 \mathcal{R}_{jn}$ where

$$\begin{aligned} \mathcal{R}_{1n} &= \sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/2-\zeta} B} R_n\{c_a, c_{k(c_a)}\}, \\ \mathcal{R}_{2n} &= \sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/2-\zeta} B} |R_n\{c_{k(c)}, c_{k(c_a)}\}|, \\ \mathcal{R}_{3n} &= \sup_{c \in \mathbb{R}} R_n\{c, c_{k(c)}\}. \end{aligned}$$

The quantities $\mathcal{R}_{jn}$ involve suprema over a, c. The next step is to replace the suprema with a maximum over a finite number of grid point combinations. This maximum can, in turn, be analyzed using the iterated martingale inequality in Lemma 2.2. Thus, we start by arguing that $\mathcal{R}_{jn} \leq \mathcal{T}_n$ for each j, where

$$\mathcal{T}_n = \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} n^{-1}\textstyle\sum_{i=1}^n \{1_{(\varepsilon_i/\sigma\leq c_{k^*})} - 1_{(\varepsilon_i/\sigma\leq c_k)}\}.$$

Note that the third term $\mathcal{R}_{3n}$ is a special case of the first term $\mathcal{R}_{1n}$ with $a = 0$ so that $\mathcal{R}_{3n} \leq \mathcal{R}_{1n}$. Accordingly, the triangle inequality gives $\mathcal{R}_n \leq 2\mathcal{R}_{1n} + \mathcal{R}_{2n}$.

3.1. *The term* $\mathcal{R}_{1n}$: For this term, note that $c_{k(c_a)-1} < c_a \leq c_{k(c_a)}$ so that

$$R_n\{c_a, c_{k(c_a)}\} = n^{-1}\textstyle\sum_{i=1}^n [1_{\{\varepsilon_i/\sigma\leq c_{k(c_a)}\}} - 1_{\{\varepsilon_i/\sigma\leq c_a\}}] \leq R_n\{c_{k(c_a)-1}, c_{k(c_a)}\},$$

where the bound involves grid points that are one interval apart. It is, therefore, bounded by $\mathcal{T}_n$ uniformly in a, c since $\mathcal{T}_n$ takes maximum over pairs of grid points that are at most D steps apart where $D > 2$. Hence, $\mathcal{R}_{1n} \leq \mathcal{T}_n$.

3.2. *The term* $\mathcal{R}_{2n}$: This term involves grid points $k(c)$, $k(c_a)$ that may be more than one point apart. Indeed, by (A.7) we have $|k(c_a) - k(c)| \leq D$ uniformly in a, c. Thus, for any a, c let $k = \min\{k(c_a), k(c)\}$ and $k^* = \max\{k(c_a), k(c)\}$ with the property that $k \leq k^* \leq k + D$. As a consequence, we can bound

$$|R_n\{c_{k(c)}, c_{k(c_a)}\}| \leq \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} R_n(c_k, c_{k^*}) = \mathcal{T}_n,$$

uniformly in a, c. Hence, taking supremum over a, c we get $\mathcal{R}_{2n} \leq \mathcal{T}_n$.

3.3. *Combine items 3.1 and 3.2* to get $\mathcal{R}_n \leq 2\mathcal{R}_{1n} + \mathcal{R}_{2n} \leq 3\mathcal{T}_n$.

4. *Martingale decomposition of* $\mathcal{T}_n$. Let $z_{ikk^*} = 1_{(\varepsilon_i/\sigma\leq c_{k^*})} - 1_{(\varepsilon_i/\sigma\leq c_k)}$ and let $R_n(c_k, c_{k^*}) = \tilde{M}_n(c_k, c_{k^*}) + \bar{M}_n(c_k, c_{k^*})$ where

$$\tilde{M}_n(c_k, c_{k^*}) = n^{-1}\textstyle\sum_{i=1}^n (z_{ikk^*} - \mathsf{E}_{i-1} z_{ikk^*}), \qquad \bar{M}_n(c_k, c_{k^*}) = n^{-1}\textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{ikk^*}.$$

It suffices to show that $\tilde{\mathcal{M}}_n = \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} |\tilde{M}_n(c_k, c_{k^*})|$, for the martingale, and $\bar{\mathcal{M}}_n = \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} \bar{M}_n(c_k, c_{k^*})$, for the compensator, are $\mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

5. *Conditional mean of* z_{ikk^*}. Note that z_{ikk^*} is an $\mathcal{F}_{in}$ adapted indicator function with

$$\mathsf{E}_{i-1} z_{ikk^*} = \mathsf{F}(c_{k^*}) - \mathsf{F}(c_k) = (1/2)\{\mathsf{H}_r(c_{k^*}) - \mathsf{H}_r(c_k)\}.$$

Since $k \leq k^* \leq k + D$, we bound $\mathsf{E}_{i-1} z_{ikk^*} \leq (1/2)\{\mathsf{H}_r(c_{k+D}) - \mathsf{H}_r(c_k)\}$. Recalling the construction (A.4) and that $H_r = 2$ by item 1, we have $\mathsf{E}_{i-1} z_{ikk^*} \leq (D/2)(H_r/K) = D/K$, uniformly in i, k. Therefore,

$$\mathcal{E}_n = \mathsf{E} \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{ikk^*} \leq nD/K.$$

Since D is a constant, see (A.7), and $K \sim n^{\zeta}$ we get $\mathcal{E}_n = \mathrm{O}(n^{1-\zeta})$.

6. *The compensator satisfies* $\bar{\mathcal{M}}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$. The Markov inequality implies that $\mathsf{P}(n^{\zeta}\bar{\mathcal{M}}_n > C) \leq \mathsf{E}n^{\zeta}\bar{\mathcal{M}}_n/C$ for some $C > 0$. Note that $\mathsf{E}\bar{\mathcal{M}}_n = n^{-1}\mathcal{E}_n$. Thus, item 5 shows that $\mathsf{E}\bar{\mathcal{M}}_n = \mathrm{O}(n^{-\zeta})$ so that $\bar{\mathcal{M}}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

7. *The martingale* $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^{-\zeta})$. We use Lemma 2.2 for

$$n\tilde{\mathcal{M}}_n = \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D} |\textstyle\sum_{i=1}^n (z_{ikk^*} - \mathsf{E}_{i-1} z_{ikk^*})|,$$

with index $\ell = (k, k^*)$ so that $z_{i\ell} = z_{ikk^*}$, $\upsilon = 1-\zeta$, parameters $L_n = KD \sim n^\zeta$ and $\lambda = \zeta > 0$ while $\upsilon = \varsigma = 1 - \zeta$. We verify the conditions of Lemma 2.2. In item 5, it was established that $z_{i\ell} = z_{ikk^*}$ is $\mathcal{F}_{in}$ adapted. Further, item 5 shows that $\mathcal{E}_n = \mathrm{O}(n^{1-\zeta}) = \mathrm{O}(n^\varsigma)$. Since $0 < \zeta < 1$, then $\varsigma = 1 - \zeta > 0$ and $\upsilon = \varsigma > \varsigma/2$. Hence, applying Lemma 2.2, we see that, $n\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^\upsilon) = \mathrm{o}_{\mathsf{P}}(n^{1-\zeta})$ so that $\tilde{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(n^{-\zeta})$. ∎

Lemma A.4 *Let Assumption 3.1 hold and suppose*
(i) F *is additively and multiplicatively Lipschitz (Definition 3.1);*
(ii) $\mathsf{E} n^{-1} \sum_{i=1}^n n^{1/2} |x_{in}| = \mathrm{O}(1)$.
Then, for any $B > 0$, $0 < \zeta < 1$,

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/2-\zeta}B} n^{-1}\sum_{i=1}^n |1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}| = \mathrm{O}_{\mathsf{P}}(n^{-\zeta}).$$

Proof of Lemma A.4. Let $v_i(a,b,c) = |1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}|$ and $V_n(a,b,c) = n^{-1}\sum_{i=1}^n v_i(a,b,c)$. We show $\mathcal{V}_n = \sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/2-\zeta}B} V_n(a,b,c)$ is $\mathrm{O}_{\mathsf{P}}(n^{-\zeta})$.

Let $c_a = c+n^{-1/2}ac$. Add and subtract $1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac)} = 1_{(\varepsilon_i/\sigma\leq c_a)}$ and apply the triangle inequality to get

$$v_i(a,b,c) \leq |1_{(\varepsilon_i/\sigma\leq c_a+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c_a)}| + |1_{(\varepsilon_i/\sigma\leq c_a)} - 1_{(\varepsilon_i/\sigma\leq c)}| = v_i(0,b,c_a) + v_i(a,0,c).$$

Correspondingly, $V_n(a,b,c) = V_n(0,b,c_a) + V_n(a,0,c)$. Taking supremum for each term gives

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

Lemmas A.2, A.3, using assumptions (i, ii), show that $\mathcal{V}_{1n}$ and $\mathcal{V}_{2n}$ are both $\mathrm{O}_{\mathsf{P}}(n^{-\zeta})$. Thus, by the triangle inequality we also get $\mathcal{V}_n = \mathrm{O}_{\mathsf{P}}(n^{-\zeta})$. ∎

A.4 Proof of main results

Proof of Theorem 3.1. Define $V_n(a,b,c) = |\mathsf{F}_n^{w,p}(a,b,c) - \mathsf{F}_n^{w,p}(0,0,c)|$. We show that $\mathcal{V}_n = \sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/2-\zeta}B} V_n(a,b,c)$ is $\mathrm{O}_{\mathsf{P}}\{n^{-\zeta(\psi-1)/\psi}\} = \mathrm{o}_{\mathsf{P}}(1)$. By the triangle inequality,

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

Apply the Hölder inequality to get, for $\psi = 1+\chi > 1$,

$$V_n(a,b,c) \leq (n^{-1}\sum_{i=1}^n |w_{in}\varepsilon_i^p|^\psi)^{1/\psi} (n^{-1}\sum_{i=1}^n |1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}|^{\psi/\chi})^{\chi/\psi}. \quad \text{(A.9)}$$

Notice that $|1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}|^{\psi/\chi} = |1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}|$ and by Lemma A.4 using Assumptions 3.1, 3.2(ib, ii), for $0 < \zeta < 1$,

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/2-\zeta}B} n^{-1}\sum_{i=1}^n |1_{(\varepsilon_i/\sigma\leq c+n^{-1/2}ac+x_{in}'b)} - 1_{(\varepsilon_i/\sigma\leq c)}| = \mathrm{O}_{\mathsf{P}}(n^{-\zeta}). \quad \text{(A.10)}$$

We show that $n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi = \mathsf{O}_\mathsf{P}(1)$. By the Markov inequality $\mathsf{P}(n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi > C) \leq \mathsf{E}(n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi)/C$. Using the independence of ε_i and $\mathcal{F}_{i-1,n}$ in Assumption 3.1 gives $\mathsf{E}(n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi) = n^{-1}\sum_{i=1}^{n}\mathsf{E}|w_{in}|^\psi\mathsf{E}|\varepsilon_i^p|^\psi$. Since ε_i are i.i.d., $\mathsf{E}|\varepsilon_i^p|^\psi < \infty$ and $n^{-1}\sum_{i=1}^{n}\mathsf{E}|w_{in}|^\psi = \mathrm{O}(1)$ by Assumption 3.2(ia, iii), then $\mathsf{E}(n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi) = \mathrm{O}(1)$ and $n^{-1}\sum_{i=1}^{n}|w_{in}\varepsilon_i^p|^\psi = \mathsf{O}_\mathsf{P}(1)$ by the Markov inequality. Insert this bound and (A.10) into (A.9) to get $\mathcal{V}_n = \{\mathsf{O}_\mathsf{P}(1)\}^{1/\psi}\{\mathsf{O}_\mathsf{P}(n^{-\zeta})\}^{\chi/\psi} = \mathsf{O}_\mathsf{P}(n^{-\zeta\chi/\psi}) = \mathsf{o}_\mathsf{P}(1)$, as desired. ■

Proof of Theorem 3.2. Let $V_n(c) = n^{-1}\sum_{i=1}^{n}w_{in}\{\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)} - \mathsf{E}_{i-1}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c)}\}$. We want to prove, for all $\omega > 0$, that $\mathcal{V}_n = \sup_{c\in\mathbb{R}}|V_n(c)| = \mathsf{O}_\mathsf{P}(n^{\omega-1/3})$. Note that it suffices to prove the result for all $0 < \omega < 1/3$.

We choose $r = 1$ noting $H_1 < \infty$ by Assumption 3.2(ia) with $\psi = 2$. Partition the axis as laid out in (A.1) with $K = \mathrm{int}(H_1 n^{1/3-\omega})$. Thus, $\mathsf{H}_1(c_k) - \mathsf{H}_1(c_{k-1}) = H_1/K \sim n^{\omega-1/3}$. Thus, for each c there exists c_{k-1}, c_k so $c_{k-1} < c \leq c_k$.

Decompose $V_n(c) = V_n(c_k) + \{V_n(c) - V_n(c_k)\}$ where $V_n(c_k)$ is a discrete point term and $V_n(c) - V_n(c_k)$ is an oscillation term. By the triangle inequality $\mathcal{V}_n \leq \mathcal{V}_{1n} + \mathcal{V}_{2n}$, where

$$\mathcal{V}_{1n} = \max_{1\leq k\leq K}|V_n(c_k)|, \qquad \mathcal{V}_{2n} = \max_{1\leq k\leq K}\sup_{c_{k-1}\leq c\leq c_k}|V_n(c) - V_n(c_k)|.$$

It suffices to show that $\mathcal{V}_{1n}$ and $\mathcal{V}_{2n}$ are $\mathsf{O}_\mathsf{P}(n^{\omega-1/3})$.

1. *The term* $\mathcal{V}_{1n}$ is $\mathsf{o}_\mathsf{P}(n^{\omega-1/3})$. We use Lemma 2.1 for $nV_n(c_k)$ with index $\ell = k$ so that $z_{i\ell} = z_{ik} = w_{in}\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c_k)}$, $\upsilon = 2/3+\omega$, parameters $L_n = K$ and $\lambda = 1/3-\omega$ and $\varsigma = 1$ while $r = 1$. We verify the conditions of Lemma 2.1. Note that $z_{i\ell}$ is $\mathcal{F}_{in}$ adapted while $\mathsf{E}z_{i\ell}^2 < \infty$, since w_{in} and ε_i^p are independent with second moments by Assumptions 3.1, 3.2(ia) with $\psi = 2$.

The parameter $\lambda = 1/3 - \omega$. The set of indices ℓ has size $L_n = K \sim n^{1/3-\omega} \sim n^\lambda$ where $\lambda > 0$ for $0 < \omega < 1/3$.

The parameter $\varsigma = 1$. Since w_{in} and ε_i are independent while $\mathsf{E}\{\varepsilon_i^p 1_{(\varepsilon_i/\sigma\leq c_k)}\}^2 \leq \mathsf{E}\varepsilon_i^{2p} < \infty$ uniformly in i, k by Assumption *3.2(ia)* with $\psi = 2$ then

$$\mathcal{E}_{n1} = \mathsf{E}\max_{1\leq k\leq K}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ik}^2 = \mathsf{E}\max_{1\leq k\leq K}\sum_{i=1}^{n}w_{in}^2\mathsf{E}_{i-1}\varepsilon_i^{2p}1_{(\varepsilon_i/\sigma\leq c_k)} \leq C\mathsf{E}\sum_{i=1}^{n}w_{in}^2 = \mathrm{O}(n),$$

by Assumption 3.2(iii) with $\psi = 2$. We check the conditions (i, ii) of Lemma 2.1.

Condition (i) is that $\varsigma < 2\upsilon$. This holds since $\varsigma = 1 < 4/3 + 2\omega = 2\upsilon$ when $\omega > 0$.

Condition (ii) is that $\varsigma + \lambda < \upsilon 2^r$ with $r = 1$. We have $\varsigma + \lambda = 4/3 - \omega$, while $\upsilon 2^r = (2/3 + \omega)2 = 4/3 + 2\omega$.

Hence, Lemma 2.1 shows that $nV_n(c_k) = \mathsf{o}_\mathsf{P}(n^{2/3+\omega})$ so that $V_n(c_k) = \mathsf{o}_\mathsf{P}(n^{\omega-1/3})$.

2. *The term* $\mathcal{V}_{2n}$. Let $V_n(c) - V_n(c_k) = B_{1n}(c, c_k) - B_{2n}(c, c_k)$ where

$$B_{1n}(c, c_k) = n^{-1}\sum_{i=1}^{n}w_{in}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c)} - 1_{(\varepsilon_i/\sigma\leq c_k)}\},$$
$$B_{2n}(c, c_k) = n^{-1}\sum_{i=1}^{n}w_{in}\mathsf{E}_{i-1}\varepsilon_i^p\{1_{(\varepsilon_i/\sigma\leq c)} - 1_{(\varepsilon_i/\sigma\leq c_k)}\}.$$

Let $z_{ik} = |w_{in}||\varepsilon_i^p|\{1_{(\varepsilon_i/\sigma\leq c_k)} - 1_{(\varepsilon_i/\sigma\leq c_{k-1})}\}$. Using the triangle and Jensen inequalities and noting $c_{k-1} < c \leq c_k$, we get

$$|B_{1n}(c, c_k)| \leq n^{-1}\sum_{i=1}^{n}z_{ik}, \qquad |B_{2n}(c, c_k)| \leq n^{-1}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ik}.$$

A martingale decomposition gives a further bound $|V_n(c) - V_n(c_k)| \le \tilde{M}_{nk} + 2\bar{M}_{nk}$, where

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

Let $\tilde{\mathcal{M}}_n = \max_{1\le k\le K}\tilde{M}_{nk}$ and $\bar{\mathcal{M}}_n = \max_{1\le k\le K}\bar{M}_{nk}$. It suffices to show that $\tilde{\mathcal{M}}_n, \bar{\mathcal{M}}_n$ are $\mathrm{O}_\mathsf{P}(n^{\omega-1/3})$ for all $0 < \omega < 1/3$.

3. *Conditional moments of* z_{ik}. Note that z_{ik} is $\mathcal{F}_{in}$ adapted while $\mathsf{E}z_{ik}^2 < \infty$ since w_{in} and ε_i^p are independent with second moments by Assumptions 3.1, 3.2(ia, iii) with $\psi = 2$. In light of (A.4), (A.5), for $q = 0, 1$ and uniformly in k,

$$\mathsf{E}_{i-1}z_{ik}^{2q} \le |w_{in}|^{2q}\{\mathsf{H}_1(c_k) - \mathsf{H}_1(c_{k-1})\} = |w_{in}|^{2q}H_1/K.$$

In turn, we get using Assumption 3.2(iii) with $\psi = 2$ that, for $q = 0, 1$,

$$\mathcal{E}_{nq} = \mathsf{E}\max_{1\le k\le K}\textstyle\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ik}^{2q} \le (H_1/K)\mathsf{E}\textstyle\sum_{i=1}^{n}|w_{in}|^{2q} = \mathrm{O}(n/K).$$

Finally, since $K \sim n^{1/3-\omega}$ we get that $\mathcal{E}_{nq} = \mathrm{O}(n^{2/3+\omega})$.

4. *The compensator* $\bar{\mathcal{M}}_n$ is $\mathrm{O}_\mathsf{P}(n^{\omega-1/3})$. The Markov inequality implies that $\mathsf{P}(n^{1/3-\omega}|\bar{\mathcal{M}}_n| > C) \le \mathsf{E}n^{1/3-\omega}|\bar{\mathcal{M}}_n|/C$ for some $C > 0$. Note that $\mathsf{E}|\bar{\mathcal{M}}_n| = n^{-1}\mathcal{E}_{n0} = \mathrm{O}(n^{\omega-1/3})$ by item 3. Thus, $n^{1/3-\omega}\bar{\mathcal{M}}_n = \mathrm{O}_\mathsf{P}(1)$ and $\bar{\mathcal{M}}_n = \mathrm{O}_\mathsf{P}(n^{\omega-1/3})$.

5. *The martingale* $\tilde{\mathcal{M}}_n$ is $\mathrm{O}_\mathsf{P}(n^{\omega-1/3})$. We use Lemma 2.1 for $n\tilde{\mathcal{M}}_n$ with index $\ell = k$ so that $z_{i\ell} = z_{ik}$, $\upsilon = 2/3 + \omega$, parameters $L_n = K$ and $\lambda = 1/3 - \omega$ and $\varsigma = 2/3 + \omega$ while $r = 1$. We verify the conditions of Lemma 2.1. In item 3 it was established that $z_{i\ell}$ is $\mathcal{F}_{in}$ adapted and $\mathsf{E}z_{i\ell}^2 < \infty$.

The parameter $\lambda = 1/3 - \omega > 0$, since $\omega < 1/3$. The set of indices ℓ has size $L_n = K \sim n^\lambda$.

The parameter $\varsigma = 2/3 + \omega \ge 0$, since $\omega > 0$. Item 3 gives that $\mathcal{E}_{n1} = \mathrm{O}(n^{2/3+\omega}) = \mathrm{O}(n^\varsigma)$.

We check the conditions (i, ii) of Lemma 2.1.

Condition (i) is that $\varsigma < 2\upsilon$. This holds since $\varsigma = \upsilon = 2/3 + \omega < 2\upsilon$ when $\omega > 0$.

Condition (ii) is that $\varsigma + \lambda < \upsilon 2^r$ with $r = 1$. We have $\varsigma + \lambda = 1$, while $\upsilon 2^r = (2/3 + \omega)2 = 4/3 + 2\omega$.

Hence, Lemma 2.1 shows that $n\tilde{\mathcal{M}}_n = \mathrm{o}_\mathsf{P}(n^{2/3+\omega})$ so that $\tilde{\mathcal{M}}_n = \mathrm{o}_\mathsf{P}(n^{\omega-1/3})$. ∎

A.5 Proofs of application results

Introduce the two-sided empirical distribution function as follows

$$\mathsf{G}_n^{w,p}(\tilde{a}, \tilde{b}, c) = n^{-1}\textstyle\sum_{i=1}^{n}w_{in}\varepsilon_i^p 1_{(|\tilde{\varepsilon}_i|\le\tilde{\sigma}c)} = \mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, c) - \lim_{h\downarrow 0}\mathsf{F}_n^{w,p}(\tilde{a}, \tilde{b}, -c-h),$$

where we recall $\tilde{a}, \tilde{b}$ are the rescaled initial estimation errors for σ and β, respectively.

We will be interested in normalized least squares statistics $\mathsf{G}_n^{xx,0}(\cdot)$, $\mathsf{G}_n^{x,1}(\cdot)$, $\mathsf{G}_n^{1,0}(\cdot)$, $\mathsf{G}_n^{1,2}(\cdot)$, where the weight indices "xx", "x" and "1" are short hand for $nx_{in}x_{in}'$, $n^{1/2}x_{in}$ and 1, respectively. All G_n functions have argument $\tilde{a}, \tilde{b}, c$. In this way,

$$\mathsf{G}_n^{xx,0}(\cdot) = \mathsf{G}_n^{xx,0}(\tilde{a}, \tilde{b}, c) = n^{-1}\textstyle\sum_{i=1}^{n}n^{1/2}Nx_i(x_iNn^{1/2})'1_{(|\tilde{\varepsilon}_i|\le\tilde{\sigma}c)},$$

and the normalized least squares statistics (4.1)-(4.2) are

$$n^{-1/2}N^{-1}(\hat{\beta}_c - \beta) = \{\mathsf{G}_n^{xx,0}(\cdot)\}^{-1}\{\mathsf{G}_n^{x,1}(\cdot)\}, \tag{A.11}$$

$$\hat{\sigma}_c^2 - \sigma^2 = \frac{\tau_0^c}{\tau_2^c}\{\mathsf{G}_n^{1,0}(\cdot)\}^{-1}[\mathsf{G}_n^{1,2}(\cdot) - \sigma^2\frac{\tau_2^c}{\tau_0^c}\mathsf{G}_n^{1,0}(\cdot) - \mathsf{G}_n^{x',1}(\cdot)\{\mathsf{G}_n^{xx,0}(\cdot)\}^{-1}\mathsf{G}_n^{x,1}(\cdot)]. \tag{A.12}$$

Proof of Theorem 4.1. *Consistency of* $\hat{\beta}_c$: We apply Theorem 3.3 to $\mathsf{G}_n^{xx,0}$ and $\mathsf{G}_n^{x,1}$, noting that Assumption 4.1,I implies Assumptions 3.2 with $\psi = 2$ and $w_{in} = nx_{in}x_{in}'$ and $p = 0$ for $\mathsf{G}_n^{xx,0}$ and $w_{in} = n^{1/2}x_{in}$ and $p = 1$ for $\mathsf{G}_n^{x,1}$. We get

$$\mathsf{G}_n^{xx,0}(\cdot) = \tau_0^c\hat{\Sigma}_n + \mathsf{o}_\mathsf{P}(1), \qquad \mathsf{G}_n^{x,1}(\cdot) = \sigma\tau_1^c\hat{\mu}_n + \mathsf{o}_\mathsf{P}(1).$$

The denominator matrix, $\mathsf{G}_n^{xx,0}(\cdot)$, is invertible for large n by the assumptions on $\hat{\Sigma}_n$. Also $c \geq c_0$ so that τ_0^c is bounded away from zero. Therefore,

$$\mathsf{G}_n^{xx,0}(\cdot) = \tau_0^c\hat{\Sigma}_n\{1 + \mathsf{o}_\mathsf{P}(1)\}, \tag{A.13}$$

uniformly in $c \geq c_0 > 0$. Inserting the two expansions into (A.11) gives

$$n^{-1/2}N^{-1}(\hat{\beta}_c - \beta) = (\tau_0^c\hat{\Sigma}_n)^{-1}\{\sigma\tau_1^c\hat{\mu}_n + \mathsf{o}_\mathsf{P}(1)\}\{1 + \mathsf{o}_\mathsf{P}(1)\} = (\tau_0^c\hat{\Sigma}_n)^{-1}\sigma\tau_1^c\hat{\mu}_n + \mathsf{o}_\mathsf{P}(1).$$

By Assumption 4.1,I,$(ii),(b)$ then $n^{-1/2}N^{-1}(\hat{\beta}_c - \beta) \overset{\mathsf{D}}{\to} (\tau_0^c\Sigma)^{-1}\sigma\tau_1^c\mu$.

Consistency of $\hat{\sigma}_c^2$: We apply Theorem 3.3 to $\mathsf{G}_n^{1,0}$ and $\mathsf{G}_n^{1,2}$ noting that Assumption 4.1,I,II imply Assumptions 3.2 with $\psi = 2$. We get

$$\mathsf{G}_n^{1,0}(\cdot) = \tau_0^c + \mathsf{o}_\mathsf{P}(1), \qquad \mathsf{G}_n^{1,2}(\cdot) = \sigma^2\tau_2^c + \mathsf{o}_\mathsf{P}(1).$$

Note that $\mathsf{G}_n^{1,2}(\cdot) - \sigma^2(\tau_2^c/\tau_0^c)\mathsf{G}_n^{1,0}(\cdot) = \mathsf{o}_\mathsf{P}(1)$, while

$$\mathsf{G}_n^{x',1}(\cdot)\{\mathsf{G}_n^{xx,0}(\cdot)\}^{-1}\mathsf{G}_n^{x,1}(\cdot) = \{\sigma\tau_1^c\hat{\mu}_n' + \mathsf{o}_\mathsf{P}(1)\}(\tau_0^c\hat{\Sigma}_n)^{-1}\{\sigma\tau_1^c\hat{\mu}_n + \mathsf{o}_\mathsf{P}(1)\}\{1 + \mathsf{o}_\mathsf{P}(1)\},$$

reducing to $\sigma^2(\hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n)(\tau_1^c)^2/\tau_0^c + \mathsf{o}_\mathsf{P}(1)$. Combine with (A.12) to get the desired result. ■

Proof of Theorem 4.2. Write $\hat{\gamma}_c = 1 - \mathsf{G}_n^{1,0}(\cdot)$ and apply Theorem 3.3 with $w_{in} = 1$ and $p = 0$ using Assumption 4.2 to get $\hat{\gamma}_c \overset{\mathsf{P}}{\to} (1 - \tau_0^c)$. ■

References

Atkinson, A.C., Riani, M. and Cerioli, A. (2010) The forward search: Theory and data analysis (with discussion). *Journal of the Korean Statistical Society* 39, 117–134.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 289–300.

Bercu, B., Delyon, B. and Rio, E. (2015) *Concentration Inequalities for Sums and Martingales.* Cham: Springer.

Bercu, B. and Touati, A. (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848–1869.

Berenguer-Rico V., Johansen, S. and Nielsen, B. (2019) The analysis of marked and weighted empirical processes of estimated residuals. *Working paper, University of Copenhagen, Economics Department, 19-05.*

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

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

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

Freedman, D.A. (1975) On tail probabilities for martingales. *Annals of Probability* 3, 100–118.

Hendry, D.F. (1999) An econometric analysis of US food expenditure 1931–1989. In Magnus, J.R. and Morgan, M.S. (eds) *Methodology & Tacit Knowledge: Two Experiments in Econometrics*, Chichester: Wiley. pp 341–362.

Hendry, D.F. and Doornik, J.A. (2014) *Empirical Model Discovery and Theory Evaluation.* Cambridge MA, MIT Press.

Hendry, D.F. and Santos, C. (2010) An automatic test of super exogeneity. In Volatility and time series econometrics: essays in honor of Robert F. Engle (eds Bollerslev, T., Russell, J. R. & Watson, M. W.), 164–193. Oxford University Press, Oxford.

Hoover, K.D. and Perez, S.J. (1999) Data mining reconsidered: encompassing and the general-to-specific approach to specification search (with discussion). Econometrics Journal 2, 167–191.

Jiao, X. and Nielsen, B. (2017) Asymptotic analysis of iterated 1-step Huber-skip M-estimators with varying cut-offs. In Analytic Methods in Statistics, (J. Antoch, J. Jurečková, M. Maciak and M. Pešta, eds.). Springer Proc. Math. Stat. 193, 23–55.

Johansen, S. and Nielsen, B. (2009) An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J.L. and Shephard, N. (Eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.* Oxford: Oxford University Press. pp. 1-36.

Johansen, S. and Nielsen, B. (2013) Asymptotic theory for iterated one-step Huber-skip estimators. *Econometrics* 1, 53–70.

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

Johansen, S. and Nielsen, B. (2016b) Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics* 43, 321–381.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica* 46, 33–50.

Koul, H.L. (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models.* 2nd ed. Springer Lecture Notes in Statistics, 166.

Koul, H.L. and Ossiander, M. (1994) Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540–582.

Koul, H.L. and Stute, W. (1999) Nonparametric model checks for time series. *Annals of Statistics* 27, 204–236.

Pretis, F., Reade, J.J. and Sucarrat, G. (2018) General-to-Specific (GETS) regression modeling and Indicator Saturation for outliers and structural breaks. *Journal of Statistical Softward* 86(3).

Ruppert, D. and Carroll, R.J. (1980) Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828–838.

Welsh, A.H. and Ronchetti, E. (2002) A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference* 103, 287–310.