

# Asymptotic Analysis of Iterated 1-step Huber-skip M-estimators with Varying Cut-offs

Xiyu Jiao and Bent Nielsen

August 25, 2016

**Abstract** We consider outlier detection algorithms for time series regression based on iterated 1-step Huber-skip M-estimators. This paper analyses the role of varying cut-offs in such algorithms. The argument involves an asymptotic theory for a new class of weighted and marked empirical processes allowing for estimation errors of the scale and the regression coefficient.

**Key words:** The iterated 1-step Huber-skip M-estimator; Tightness; A fixed point; Poisson approximation to gauge; Weighted and marked empirical processes.

## 1 Introduction

We consider outlier detection methods that are based on iterated 1-step Huber-skip M-estimators for linear regression models with regressors that are stationary or deterministically or stochastically trending. Each 1-step estimator relies on a cut-off value when classifying observations as outliers or not. In this paper we allow the cut-off value to vary with sample size and iteration step. To analyze this asymptotically, we generalize some recent results for residual empirical processes, which allow for variation in location, scale and quantile. The model is a linear regression

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

---

Xiyu Jiao

Department of Economics, University of Oxford and Mansfield College, Oxford OX1 3TF, UK  
e-mail: xiyu.jiao@mansfield.ox.ac.uk

Bent Nielsen

Department of Economics, University of Oxford and Nuffield College, Oxford OX1 1NF, UK  
e-mail: bent.nielsen@nuffield.ox.ac.uk

where  $\varepsilon_i/\sigma$  are independent of  $\mathcal{F}_{i-1} = \sigma(x_1, \dots, x_i, \varepsilon_1, \dots, \varepsilon_{i-1})$  with the common density  $f$ . Outliers are pairs of observations that do not conform with the model.

Iterated 1-step Huber-skip M-estimators mimic the Huber [14] skip estimator, which has criterion function  $\rho(t) = \min(t^2, c^2)/2$  as opposed to the Huber estimator with criterion function  $\rho(t) = t^2/2$  for  $|t| \leq c$  and  $\rho(t) = c|t| - c^2/2$  otherwise, see also [8, p. 104], [19, p. 175]. The 1-step Huber-skip M-estimator starts from an initial estimator  $(\tilde{\beta}, \tilde{\sigma}^2)$ . This is used to decide which observations are outlying through

$$v_i = 1_{(|y_i - x_i' \tilde{\beta}| \leq \tilde{\sigma} c)}, \quad (2)$$

where the choice of the cut-off  $c$  is related to the known reference density  $f$ . For those observations that are not outlying, we run a least squares regression and get the one-step Huber-skip estimator

$$\hat{\beta} = \left( \sum_{i=1}^n x_i x_i' v_i \right)^{-1} \left( \sum_{i=1}^n x_i y_i v_i \right), \quad (3)$$

$$\hat{\sigma}^2 = \varsigma^{-2} \left( \sum_{i=1}^n v_i \right)^{-1} \left\{ \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 v_i \right\}, \quad (4)$$

where  $\varsigma^2$  is the consistency factor as in (8). This step can be iterated. The iteration may be initiated by a robust estimator. More simply we get the Robustified Least Squares and the Impulse Indicator Saturation starting with the full or split sample least squares. The latter algorithm was introduced in the empirical work of US food expenditure by Hendry, see [9, 10].

Outlier detection algorithms have a positive probability to find outliers even when, in fact, the data generation process has no outliers. We evaluate the performance of such algorithms by the concept of a gauge, which is the expected retention rate of falsely discovered outliers. This is a measure of type I error and it gives us an indirect way of choosing the cut-off  $c$ . It is defined as follows. The algorithms assign stochastic indicators  $v_i$  to all observations such as in (2) so that  $v_i = 0$  when observation  $i$  is declared as an outlier, otherwise  $v_i = 1$ . When the model has no contamination, the sample and population gauge are

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^n (1 - v_i), \quad \mathbb{E} \hat{\gamma} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(1 - v_i). \quad (5)$$

Hoover and Perez [13] originally introduced the idea of a gauge in a simulation study of general-to-specific variable selection algorithms. The concept of a gauge was formally proposed by Hendry and Santos [12] as the expected retention rate of irrelevant regressors in the context of model selection algorithms. Comprehensive simulation studies on the gauge for the model selection algorithm Autometrics are presented in [6, 10]. an asymptotic analysis for the gauge of some outlier detection algorithms is presented in [18].

One-step estimators have been considered before in [2, 23]. The 1-step Huber-skip estimator was studied in [25]. Asymptotic distribution theory has been derived

for the location model in [11] and for the time series regression [15]. Iteration was investigated in [16]. An asymptotic expansion for the sample gauge was established in [18]. All these asymptotic analyses are restricted to the situation where the cut-off and the number of iterations are not both increasing.

The purpose of this paper is to build an asymptotic theory which can explore how variation in the cut-off affects the iterated 1-step Huber-skip M-estimator. In particular, we prove the tightness and fixed point theorems for the iterated 1-step M-estimator with the varying cut-off. Moreover, this paper demonstrates an asymptotic Poisson distribution to the gauge in a situation where the cut-off increases with the sample size while the number of iterations also increases.

The argument involves a theory for a new class of weighted and marked empirical processes. This is defined from the generalized empirical distribution function

$$\widehat{F}_n^{g,p}(a,b,c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + n^{-1/2} ac + x_{in}' b)}, \quad (6)$$

where the weights  $g_{in}$  are combinations of the normalized  $\mathcal{F}_{i-1}$  measurable regressors  $x_{in}$  and  $\varepsilon_i^p$  are the  $\mathcal{F}_i$  adapted marks, while  $a, b$  represent the normalized estimation errors for  $\sigma, \beta$ . When  $p = 0$  the mark is unity and we get the weighted empirical distribution function considered by for instance [20]. Processes of the type  $n^{-1/2} \sum_{i=1}^n \varepsilon_i 1_{(x_i \leq c)}$  are called marked processes, see [20, p. 43], but are not special cases of the weighted and marked empirical distribution functions.

We derive asymptotic expansions that are uniform in  $a, b, c$  and allow for a near  $n^{1/4}$  inefficiency in the estimation uncertainties  $a, b$ . This generalizes results by Koul and Ossiander, see [20, 21, 22], who allowed unbounded weights  $g_{in}$  but no marks  $\varepsilon_i^p$ . They used a truncation argument for  $\mathcal{F}_{i-1}$  measurable weights  $g_{in}$ . This together with the boundedness of the  $\mathcal{F}_i$  measurable indicator function meant that they could apply the Freedman [7] exponential inequality for bounded martingales. Here, we use the iterated martingale inequality of [18] reported as Lemma 3 in the appendix. This is based on the Bercu and Touati [1] exponential inequality for unbounded martingales, so that we can avoid the truncation argument and more easily allow the  $\mathcal{F}_i$  measurable product of the mark and indicator to be unbounded. The result also generalizes [15, 18] who did not allow joint variation of all of  $a, b, c$ .

The outline of this paper is the following. We first review the model and iterated 1-step Huber-skip M-estimator algorithm in §2. Then, the main results follows in §3. §4 provides theory for the weighted and marked empirical process with proofs in Appendix 1, 2 and 3. Proofs of the main theorems in §3 follow in Appendix 4.

## 2 Model and outlier detection algorithms

The regression model with some notations is described first. We review the iterated 1-step Huber-skip M-estimators including the Robustified Least Squares and the Impulse Indicator Saturation.

## 2.1 Model

Suppose we have data  $(y_i, x_i)$ ,  $i = 1, 2, \dots, n$ , where  $y_i$  is univariate and  $x_i$  is multivariate with dimension  $\dim x$ . Assume the data satisfies the regression equation

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

This setting can represent both classical regression and time series models. Moreover, regressors  $x_i$  can be a deterministic or stochastic trend. Innovations  $\varepsilon_i$  are independent of the filtration  $\mathcal{F}_{i-1}$  generated by  $(x_1, \dots, x_i, \varepsilon_1, \dots, \varepsilon_{i-1})$ , and are identically distributed with scale  $\sigma$  so that  $\varepsilon_i/\sigma$  has the known density  $f$  and distribution function  $F(c) = P(\varepsilon_i/\sigma \leq c)$ . In practice, the innovation distribution, characterised by  $f, F$ , will often be assumed to be standard normal or at least symmetric. Outlier detection algorithms use absolute residuals and then calculate robust least squares estimators from the non-outlying sample. This implicitly assumes symmetry, while non-symmetry leads to bias forms. We assume symmetry when analyzing the iterated 1-step Huber-skip M-estimator algorithm in §3, but not for the general empirical process results in §4.

For the absolute error  $|\varepsilon_i|/\sigma$  we denote the density by  $g$  and the distribution function by  $G(c) = P(|\varepsilon_i|/\sigma \leq c)$  for  $c > 0$ . Here we use  $c$  as notation for the quantile of the distribution  $G(c)$ . In the course of the analysis this will be linked to the cut-off of the 1-step estimator in (3) and the argument of the weighted and marked empirical distribution function in (6). Now, with a symmetry assumption,  $G(c) = 2F(c) - 1$  and  $g(c) = 2f(c)$ . Define  $\psi = G(c)$  so the probability of exceeding the cut-off  $c$  is  $\gamma = 1 - \psi$ . Suppose the  $k$ -th moment of the density  $f$  exists, then introduce

$$\tau_k = \int_{-\infty}^{\infty} u^k f(u) du, \quad \tau_k^c = \int_{-c}^c u^k f(u) du. \quad (7)$$

Thus  $\tau_0^c = \psi$ ,  $\tau_2 = 1$  while  $\tau_k = \tau_k^c = 0$  for odd  $k$  when assuming symmetry. Define the conditional variance of  $\varepsilon_i/\sigma$  given  $(|\varepsilon_i|/\sigma \leq c)$  as

$$\zeta_c^2 = \frac{\tau_2^c}{\psi} = \frac{\int_{-c}^c u^2 f(u) du}{P(|\varepsilon_i| \leq \sigma c)}. \quad (8)$$

This will be used as a bias correction factor for the variance estimate computed from the selected non-outlying sample. For a standard normal reference distribution, we have  $\tau_2^c = \psi - 2cf(c)$ ,  $\tau_4^c = 3\psi - 2c(c^2 + 3)f(c)$  and  $\tau_4 = 3$ .

## 2.2 The iterated 1-step Huber-skip M-estimator algorithm

We first define the iterated 1-step Huber-skip M-estimator algorithm. Specific examples include the Robustified Least Squares and the Impulse Indicator Saturation.

**Algorithm 1 Iterated 1-step Huber-skip M-estimator.** Choose a cut-off  $c > 0$ .

1. Choose initial estimators  $\hat{\beta}_c^{(0)}$ ,  $(\hat{\sigma}_c^{(0)})^2$  and let  $m = 0$ .

2. Define indicator variables for selecting non-outlying observations

$$v_{i,c}^{(m)} = \mathbf{1}_{(|y_i - x_i' \hat{\beta}_c^{(m)}| \leq \hat{\sigma}_c^{(m)} c)}. \quad (9)$$

3. Compute least squares estimators

$$\hat{\beta}_c^{(m+1)} = \left( \sum_{i=1}^n x_i x_i' v_{i,c}^{(m)} \right)^{-1} \left( \sum_{i=1}^n x_i y_i v_{i,c}^{(m)} \right), \quad (10)$$

$$(\hat{\sigma}_c^{(m+1)})^2 = \varsigma_c^{-2} \left( \sum_{i=1}^n v_{i,c}^{(m)} \right)^{-1} \left\{ \sum_{i=1}^n (y_i - x_i' \hat{\beta}_c^{(m+1)})^2 v_{i,c}^{(m)} \right\}. \quad (11)$$

4. Let  $m = m + 1$  and repeat 2 and 3.

In §3 we show how to choose the cut-off  $c$  indirectly from the gauge defined in (5). The algorithm could start with a robust estimator, while the Robustified Least Squares is initiated using the full sample least squares. The latter is not robust with respect to high leverage points in cross section data. Leverage points seem to be less of a problem in time series models when lagged variables are included as regressors.

Another example is the Impulse Indicator Saturation which was initially proposed in the empirical work [9]. The algorithm was studied comprehensively in [10, 11]. The idea is to divide full sample into two sub-samples and use regression estimates calculated from each sub-sample to detect outliers in the other sub-sample.

**Algorithm 2 Impulse Indicator Saturation.** Choose a cut-off  $c > 0$ .

1.1. Split full sample into two sets  $\mathcal{S}_j$ ,  $j = 1, 2$  of  $n_j$  observations where  $\sum_{j=1}^2 n_j = n$ .

1.2. Calculate least squares estimators based upon each sub-sample  $\mathcal{S}_j$  for  $j = 1, 2$

$$\hat{\beta}_j = \left( \sum_{i \in \mathcal{S}_j} x_i x_i' \right)^{-1} \left( \sum_{i \in \mathcal{S}_j} x_i y_i \right), \quad \hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i \in \mathcal{S}_j} (y_i - x_i' \hat{\beta}_j)^2. \quad (12)$$

1.3. Define the initial indicator variables for selecting non-outlying observations

$$v_{i,c}^{(-1)} = \mathbf{1}_{(i \in \mathcal{S}_1)} \mathbf{1}_{(|y_i - x_i' \hat{\beta}_2| \leq \hat{\sigma}_2 c)} + \mathbf{1}_{(i \in \mathcal{S}_2)} \mathbf{1}_{(|y_i - x_i' \hat{\beta}_1| \leq \hat{\sigma}_1 c)}. \quad (13)$$

1.4. Compute  $\hat{\beta}_c^{(0)}$ ,  $(\hat{\sigma}_c^{(0)})^2$  using (10), (11) with  $m = -1$ , and then let  $m = 0$ .

2. Follow the step 2,3,4 in Algorithm 1.

The Impulse Indicator Saturation is possibly more robust than the Robustified Least Squares when we have prior knowledge that outliers are located in a particular subset of the whole sample. The choice of the initial sets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  should be iterated since the location of contaminated observations is unknown in most practical situations, see [6].

### 3 The main results

We start by listing the assumptions. Then follows the new tightness and fixed point result for the iterated estimator defined in Algorithm 1. Finally the gauge of the iterated estimator is analysed. The result is uniform in the cut-off value, which generalises [15, 16] which set the threshold fixed. This allows us to analyse the gauge of the iterated estimator when the cut-off value is drifting.

#### 3.1 Assumptions

We list the sufficient assumptions for asymptotic theory of iterated 1-step Huber-skip M-estimators. These assumptions are somewhat stronger than they need to be. In section §4 on the one-sided empirical process, we will introduce some weaker assumptions. For instance, we will then abandon the symmetry assumption of  $f$ .

Innovations  $\varepsilon_i$  and regressors  $x_i$  must satisfy some moment conditions so as to carry out asymptotic analysis. Regressors  $x_i$  can be temporally dependent and trending deterministically or stochastically. We therefore need a normalisation matrix  $N$  that allows for different behaviour of the components of the regressor vector  $x_i$ . In the case of a stationary regressor we need a standard  $n^{-1/2}$  normalisation so that  $N$  must be proportional to the identity matrix of the same dimension as  $x_i$ , that is  $N = n^{-1/2}I_{\dim x}$ . Likewise, if  $x_i$  is a random walk we have  $N = n^{-1}I_{\dim x}$ . If the regressors are unbalanced as in  $x_i = (1, i)'$  we can choose  $N = \text{diag}(n^{-1/2}, n^{-3/2})$ .

**Assumption 1** *Let  $\mathcal{F}_i$  be an increasing sequence of  $\sigma$ -fields so  $\varepsilon_{i-1}$  and  $x_i$  are  $\mathcal{F}_{i-1}$  measurable and  $\varepsilon_i$  is independent of  $\mathcal{F}_{i-1}$ . Let  $\varepsilon_i/\sigma$  have a symmetric, continuously differentiable density  $f$  which is positive on  $\mathbf{R}$ . For some values of  $\kappa, \eta$  such that  $0 \leq \kappa < \eta \leq 1/4$ , choose an integer  $r \geq 2$  so*

$$2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim x). \quad (14)$$

Let  $q = 1 + 2^{r+1}$ . Denote  $c_0 > 0$  as a finite number. Suppose

- (i) the density  $f$  satisfies
  - (a)  $u^q f(u), |u^{q+1} \dot{f}(u)|$  are decreasing for large  $u$ ;
  - (b)  $f(u_n - n^{-1/4}A)/f(u_n) = O(1)$  as  $n \rightarrow \infty$  for some  $A > 0$  and all sequences  $u_n \rightarrow \infty$  so  $u_n = o(n^{1/4})$ ;
  - (c)  $f(u)/[u\{1 - F(u)\}] = O(1)$  for  $u \rightarrow \infty$ ;
- (ii) the regressors  $x_i$  satisfy
  - (a)  $\Sigma_n = \sum_{i=1}^n N' x_i x_i' N \xrightarrow{P} \Sigma > 0$ ;
  - (b)  $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = O_P(1)$ ;
  - (c)  $n^{-1} \mathbf{E} \sum_{i=1}^n |n^{1/2} N' x_i|^q = O(1)$ ;
- (iii) the initial estimator  $(\tilde{\beta}, \tilde{\sigma}^2)$  satisfies
  - (a)  $N^{-1}(\tilde{\beta} - \beta) = O_P(n^{1/4-\eta})$ ;
  - (b)  $n^{1/2}(\tilde{\sigma}^2 - \sigma^2) = O_P(n^{1/4-\eta})$ .

There is a trade-off between  $\kappa$ ,  $\eta$ , the dimension  $\dim x$  and the required number of moments  $r$ , see [17, Remark 3.1]. The conditions (i), (ii) are satisfied in a range of situations. In particular, condition (ia) is satisfied by the normal and t distribution, see [17, Example 3.1]; condition (ib, ic) is satisfied by the normal, see [18, Remark 2]; condition (ii) is satisfied by stationary, random walk and deterministically trending regressors, see [17, Example 3.2]. Condition (iii) allows the standardised estimation errors to diverge at a rate of  $n^{1/4-\eta}$  rather than being bounded in probability. In particular,  $\eta = 1/4$  can be chosen for estimators with standard convergence rates.

### 3.2 Properties of the iterated estimators

The first result is a stochastic expansion of the 1-step Huber-skip M-estimator in terms of the original estimator, a kernel and a small remainder term.

**Theorem 1.** *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 1. Suppose Assumption 1(ia, ii) holds, and that  $N^{-1}(\widehat{\beta}_c^{(m)} - \beta)$ ,  $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$  are  $O_{\mathbb{P}}(1)$ . Then uniformly in  $c \in [c_0, \infty)$  and as  $n \rightarrow \infty$*

$$\begin{aligned} N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) &= \frac{2cf(c)}{\psi} N^{-1}(\widehat{\beta}_c^{(m)} - \beta) + (\psi \Sigma)^{-1} \sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + o_{\mathbb{P}}(1), \\ n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) &= \frac{c(c^2 - \zeta_c^2)f(c)}{\tau_2^c} n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma) \\ &\quad + \frac{1}{2\sigma \tau_2^c} n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \zeta_c^2 \sigma^2) 1_{(|\varepsilon_i| \leq \sigma c)} + o_{\mathbb{P}}(1). \end{aligned}$$

Theorem 1 shows that the updated estimation error for  $\beta$  depends on the previous estimation error for  $\beta$ , but not on the estimation uncertainty for  $\sigma$ . The estimation error for  $\sigma$  has a similar property. This is a consequence of symmetry imposed on the density  $f$ . More complex situations can also be analyzed where the reference distribution  $f$  is non-symmetric and the cut-off  $c$  is chosen in a matching way, see [15]. The proof uses the empirical process theory in §4.

The next result shows that the iterated estimator is tight in iteration  $m \in [0, \infty)$  and in the cut-off value  $c \in [c_0, \infty)$ . This builds on [16].

**Theorem 2.** *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 1. Suppose Assumption 1(ia, ii, iii) holds with  $\eta = 1/4$ . Then as  $n \rightarrow \infty$*

$$\sup_{0 \leq m < \infty} \sup_{c_0 \leq c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = O_{\mathbb{P}}(1).$$

Assumption 1(iii) with  $\eta = 1/4$  corresponds to a standard convergence rate for the initial estimator. Theorem 1 provides the 1-step relationship between the updated estimator and the original estimator. Since  $\sup_{c_0 \leq c < \infty} |2cf(c)/\psi| < 1$  and

$\sup_{c_0 \leq c < \infty} |c(c^2 - \zeta_c^2)f(c)/\tau_c^c| < 1$  implied by Assumption 1 (ia), see [16, Theorem 3.5], a geometric argument and mathematical induction are used to show tightness.

The fixed point result can now be shown. Initially the tight estimator is assumed available. This is iterated through the one-step equation presented in Theorem 1.

**Theorem 3.** *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 1. Suppose Assumption 1 (ia, ii, iii) holds with  $\eta = 1/4$ . Then for all  $\varepsilon, \delta > 0$  a pair  $m_0, n_0 > 0$  exists, so for all  $m > m_0$  and  $n > n_0$*

$$\mathbb{P}\left\{ \sup_{c_0 \leq c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \widehat{\beta}_c^*)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \widehat{\sigma}_c^*)| > \delta \right\} < \varepsilon,$$

where

$$N^{-1}(\widehat{\beta}_c^* - \beta) = \frac{1}{\psi - 2cf(c)} \Sigma^{-1} \sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)},$$

$$n^{1/2}(\widehat{\sigma}_c^* - \sigma) = \frac{1}{2\sigma\{\tau_c^c - c(c^2 - \zeta_c^2)f(c)\}} n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \zeta_c^2 \sigma^2) 1_{(|\varepsilon_i| \leq \sigma c)}.$$

Based on Theorem 2, if the initial estimator is bounded in a large compact set with large probability, then any iterated estimator takes values in the same compact set no matter what value of the cut-off  $c$  is chosen in the interval  $[c_0, \infty)$ . The proof of Theorem 3 is to further argue the deviation between the  $m$ -fold iterated estimator and the fixed point is the sum of two terms vanishing exponentially and in probability respectively when  $m$  and  $n$  are sufficiently large.

The iterated 1-step Huber-skip M-estimator can be seen as a special case of iteratively reweighted least squares with binary weights. Dollinger and Staudte [5] applied an influence function argument to demonstrate convergence of iteratively reweighted least squares with smooth weights. Even if the spirit is similar, our proof is different due to binary weights. The idea of iterating one-step estimator can also be found in [4], which analyzed the first order autoregression with infinite variance.

### 3.3 Properties of the gauge

Johansen and Nielsen [18] proved the Poisson approximation to the gauge for the finite step Huber-skip M-estimator. But the iterated result was not established, since they did not have the empirical process theory which investigates the varying quantile  $c$  and estimation errors for  $\beta$  and  $\sigma$ . This paper shows the Poisson approximation to the gauge for the iterated 1-step Huber-skip M-estimator.

A Poisson exceedence theory arises in the scenario where the cut-off value  $c$  is set to allow the fixed number  $\lambda$  of outliers regardless of the sample size  $n$ . For some  $\lambda > 0$ , the cut-off value  $c_n$  is set so as to let

$$nP(|\varepsilon_i| > \sigma c_n) = \lambda. \quad (15)$$

$\lambda$	$x$								
	$c_{100}$	$c_{200}$	0	1	2	3	4	5	
5	1.960	2.241	0.01	0.04	0.12	0.27	0.44	0.62	
1	2.576	2.807	0.37	0.74	0.92	0.98	1.00		
0.5	2.807	3.023	0.61	0.91	0.98	1.00			
0.25	3.023	3.227	0.78	0.97	1.00				
0.1	3.291	3.481	0.90	1.00					

**Table 1** The probability of detecting at most  $x$  outliers approximated by a Poisson distribution for a given  $\lambda$ , and the cut-off  $c_n = \Phi^{-1}\{1 - \lambda/(2n)\}$  for  $n = 100, 200$ .

Notice that  $c_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Define  $v_{i,c_n}^{(m)}$ ,  $\widehat{\beta}_{c_n}^{(m+1)}$ ,  $(\widehat{\sigma}_{c_n}^{(m+1)})^2$  by replacing  $c$  by  $c_n$  in expressions (9), (10), (11). The corresponding sample gauge is

$$\widehat{\gamma}_{c_n}^{(m)} = \frac{1}{n} \sum_{i=1}^n (1 - v_{i,c_n}^{(m)}) = \frac{1}{n} \sum_{i=1}^n 1_{(|y_i - x_i \widehat{\beta}_{c_n}^{(m)}| > \widehat{\sigma}_{c_n}^{(m)} c_n)}. \quad (16)$$

Theorem 2, 3 shows that any iterated estimator is tight, so lower and upper bounds can be found for the indicators appearing in the gauge. By exploring these bounds, the following Poisson limit theorem arises.

**Theorem 4.** *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 1. Let  $c_n$  be defined from (15). Suppose Assumption 1 holds with  $\eta = 1/4$ . Then for all  $0 \leq m < \infty$  and as  $n \rightarrow \infty$ , the sample gauge in (16) satisfies*

$$n \widehat{\gamma}_{c_n}^{(m)} \xrightarrow{D} \text{Poisson}(\lambda).$$

Table 1 assumes that  $\varepsilon_i/\sigma$  follows a standard normal distribution. For a given  $\lambda$ , the cut-off in (15) satisfies  $c_n = \Phi^{-1}\{1 - \lambda/(2n)\}$ . Cut-off values are shown for  $n = 100, 200$ . The Poisson approximation gives the probability of finding at most  $x$  outliers. There is an increase from 62% to 90% for the probability of detecting at most  $x = \lambda$  outliers as  $\lambda$  declines from 5 to 0.1. The reason is due to the left skewness of the Poisson distribution. In particular, we focus on the case where  $\lambda = 1$  and  $n = 100$ . The cut-off is  $c_n = 2.58$  and the probability to find at most 1, 2 outliers are 0.74, 0.92. This means it regularly finds 2 outliers when there are none.

The Robustified Least Squares and Impulse Indicator Saturation are special versions of iterated 1-step Huber-skip M-estimators with different starting points. Their initial points do not depend on the cut-off, and thus satisfy the tightness property. Therefore, Theorems 1, 2, 3 and 4 apply for these algorithms.

## 4 Weighted and marked empirical process

Consider the weighted and marked empirical distribution function

$$\widehat{F}_n^{g,p}(a,b,c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + n^{-1/2} ac + x'_{in} b)},$$

with  $\mathcal{F}_{i-1}$  adapted weights  $g_{in}$  and  $\mathcal{F}_i$  measurable marks  $\varepsilon_i^p$ . Let  $a \in \mathbf{R}$ ,  $b \in \mathbf{R}^{\dim x}$  represent estimation errors  $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)$ ,  $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)$ , while  $c \in \mathbf{R}$  is the quantile. Define normalized regressors  $x_{in} = N'x_i$  so that  $\sum_{i=1}^n x_{in}x'_{in}$  converges. For example,  $N = n^{-1/2}I_{\dim x}$  if  $\{x_i\}_{i=1}^n$  is stationary, while  $N = n^{-1}I_{\dim x}$  for a random walk. Our interest focuses on weights  $g_{in}$  given as either of 1,  $n^{1/2}N'x_i$ ,  $nN'x_ix'_iN$  and  $p$  as either of 0, 1, 2. To form the empirical process, introduce the compensator

$$\bar{F}_n^{g,p}(a,b,c) = \frac{1}{n} \sum_{i=1}^n g_{in} E_{i-1} \varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + n^{-1/2} ac + x'_{in} b)}, \quad (17)$$

where  $E_{i-1}(\cdot) = E(\cdot | \mathcal{F}_{i-1})$ . Note that  $\bar{F}_n^{1,0}(0,0,c) = F(c) = P(\varepsilon_i \leq \sigma c)$ .

We embed these processes in the space  $D[0,1]$  of processes that are continuous from the right and with limits of left, where the space is endowed with the Skorokhod metric. We do this as follows. The indicator  $1_{(\varepsilon_i \leq c)}$  and the distribution function  $F(c)$  can be defined as 0 or 1 when  $c$  takes the values  $-\infty$  and  $\infty$  respectively. We can then define quantiles  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Correspondingly we can continuously extend the definition of the weighted and marked empirical distribution function and its compensator by choosing  $\widehat{F}_n^{g,p}(a,b,-\infty) = \bar{F}_n^{g,p}(a,b,-\infty) = 0$  while  $\widehat{F}_n^{g,p}(a,b,\infty) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p$  and  $\bar{F}_n^{g,p}(a,b,\infty) = \frac{1}{n} \sum_{i=1}^n g_{in} E_{i-1} \varepsilon_i^p$ .

We now define the empirical process, for  $0 \leq \psi \leq 1$ ,

$$\mathbf{F}_n^{g,p}(a,b,c_\psi) = n^{1/2} \{ \widehat{F}_n^{g,p}(a,b,c_\psi) - \bar{F}_n^{g,p}(a,b,c_\psi) \}. \quad (18)$$

We will show convergence that is uniform in  $a, b, c_\psi$  for the above process. This generalizes results in [22], which had no marks and no variation  $a$  in scale, in [20, Theorem 2.2.5], which had no marks, in [15, 17], which had marks, but no variation in quantile  $c$  and no variation  $a$  in scale respectively.

In the following we first present the new result concerning variation in the scale  $a$  and the quantile  $c$ . Subsequently, we combine this with existing results concerning variation in  $b, c$  in order to get a result that is uniform in all three arguments  $a, b, c$ .

#### 4.1 The case of estimated scale and known regression parameter

The main technical contribution of the paper is to analyze the empirical process in the case of estimated scale, but known regression parameter. Thus, we establish results for the empirical process that are uniform in  $a, c$ . Koul [20, Theorem 2.2.5] established a similar result for the case of unbounded weights  $g_{in}$  but no marks  $\varepsilon_i^p$ . His proof exploits that the function  $1_{(\varepsilon_i \leq \sigma c)}$  is monotone in  $c$  and bounded. These properties are not shared by  $\varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c)}$ , so we follow a different strategy for the

proof that exploits the iterated martingale inequality from [18] reported as Lemma 3 in the Appendix 1.

We first present the uniformity result for the empirical process and then a uniform linearization result for the compensator. The proof involves a chaining argument. For this, we apply an iterated martingale inequality, see Lemma 3, to explore the tail behaviour of the maximum of a family of martingales.

**Theorem 5.** *Let  $\mathcal{F}_i$  be an increasing sequence of  $\sigma$ -fields so  $\varepsilon_{i-1}$  and  $g_{in}$  are  $\mathcal{F}_{i-1}$  measurable and  $\varepsilon_i$  is independent of  $\mathcal{F}_{i-1}$ . Let  $\varepsilon_i/\sigma$  have a continuous density  $f$ . Let  $p$  and  $\eta$  be given so  $p \in \mathbf{N}_0$  and  $0 < \eta \leq 1/4$ . Suppose*

(i) *the density  $f$  satisfies*

(a) *moments:*  $\int_{-\infty}^{\infty} |u|^{4p} f(u) du < \infty$ ;

(b) *boundedness:*  $\sup_{u \in \mathbf{R}} |u|(1 + |u|^{4p})f(u) < \infty$ ;

(ii) *the weights  $g_{in}$  satisfy  $n^{-1} \mathbf{E} \sum_{i=1}^n |g_{in}|^4 = \mathbf{O}(1)$ .*

*Let  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for any  $B > 0$  and as  $n \rightarrow \infty$*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4 - \eta} B} |\mathbf{F}_n^{g,p}(a, 0, c_\psi) - \mathbf{F}_n^{g,p}(0, 0, c_\psi)| = \mathbf{o}_P(1).$$

The second result provides a linearization of the compensator.

**Theorem 6.** *Let  $\mathcal{F}_i$  be an increasing sequence of  $\sigma$ -fields so  $\varepsilon_{i-1}$  and  $g_{in}$  are  $\mathcal{F}_{i-1}$  measurable and  $\varepsilon_i$  is independent of  $\mathcal{F}_{i-1}$ . Let  $\varepsilon_i/\sigma$  have a differentiable density  $f$ . Let  $p$  and  $\eta$  be given so  $p \in \mathbf{N}_0$  and  $0 < \eta \leq 1/4$ . Suppose*

(i) *the density  $f$  satisfies*

(a) *moments:*  $\int_{-\infty}^{\infty} |u|^p f(u) du < \infty$ ;

(b) *boundedness:*  $\sup_{u \in \mathbf{R}} u^2 |u^{p-1} f(u) + u^p \dot{f}(u)| < \infty$ ;

(ii) *the weights  $g_{in}$  satisfy  $n^{-1} \sum_{i=1}^n |g_{in}| = \mathbf{O}_P(1)$ .*

*Let  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for any  $B > 0$  and as  $n \rightarrow \infty$*

$$\begin{aligned} \sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4 - \eta} B} & |n^{1/2} \{ \bar{\mathbf{F}}_n^{g,p}(a, 0, c_\psi) - \bar{\mathbf{F}}_n^{g,p}(0, 0, c_\psi) \} \\ & - \sigma^{p-1} c_\psi^p f(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} n^{-1/2} a c_\psi| = \mathbf{O}_P(n^{-2\eta}). \end{aligned}$$

## 4.2 The case of estimated scale and regression parameter

We now turn to the general one-sided empirical process with estimated scale and regression parameters. The case with known regression parameter was treated above while the case with known scale was treated in [18]. Through an argument reported in the appendix these results can be combined to prove the general result. For this we need the union of the various assumptions. This is listed below as Assumption 2. Note the density  $f$  is not necessarily symmetric in this section and Assumption 2 is weaker than Assumption 1.

**Assumption 2** Let  $\mathcal{F}_i$  be an increasing sequence of  $\sigma$ -fields so  $\varepsilon_{i-1}$ ,  $x_i$  and  $g_{in}$  are  $\mathcal{F}_{i-1}$  measurable and  $\varepsilon_i$  is independent of  $\mathcal{F}_{i-1}$ . Let  $\varepsilon_i/\sigma$  have a continuously differentiable density  $f$  which is positive on  $\mathbf{R}$ . Let  $p$ ,  $\eta$ ,  $\kappa$  be given so  $p \in \mathbf{N}_0$  and  $0 \leq \kappa < \eta \leq 1/4$ . Choose  $r \in \mathbf{N}_0$  so

$$2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim x). \quad (19)$$

Suppose

(i) the density  $f$  satisfies

(a) moments:  $\int_{-\infty}^{\infty} |u|^{2r} p f(u) du < \infty$ ;

(b) boundedness:  $\sup_{u \in \mathbf{R}} \{1 + |u|^{\max(4p+1, 2^r p-1)}\} f(u) + (1 + u^{2^r p+2}) |\dot{f}(u)| < \infty$ ;

(c) smoothness: a  $C_H > 0$  exists so that for all  $v > 0$

$$\frac{\sup_{u \geq v} (1 + u^{2^r p}) f(u)}{\inf_{0 \leq u \leq v} (1 + u^{2^r p}) f(u)} \leq C_H, \quad \frac{\sup_{u \leq -v} (1 + u^{2^r p}) f(u)}{\inf_{-v \leq u \leq 0} (1 + u^{2^r p}) f(u)} \leq C_H;$$

(ii) the regressors  $x_i$  satisfy  $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = O_P(1)$ ;

(iii) the weights  $g_{in}$  satisfy

(a)  $n^{-1} \mathbf{E} \sum_{i=1}^n |g_{in}|^{2^r} (1 + |n^{1/2} N' x_i|) = O(1)$ ;

(b)  $n^{-1} \sum_{i=1}^n |g_{in}| (1 + |n^{1/2} N' x_i|^2) = O_P(1)$ .

*Remark 1.* Assumption 1 (ia, iib, iic) implies Assumption 2 with  $r \geq 2$  satisfying (14) when  $g_{in}$  is either of 1,  $n^{1/2} N' x_i$ ,  $n N' x_i x_i' N$  and  $p$  is either of 0, 1, 2. Details are given in Lemma 4 in the appendix.  $\square$

We present two asymptotic results. The first theorem shows that the estimation error for the scale and regression parameter is negligible uniformly in the quantile.

**Theorem 7.** Suppose Assumption 2 is satisfied. Let  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for any  $B > 0$  and as  $n \rightarrow \infty$

$$\sup_{0 \leq \psi \leq 1} \sup_{|a|, |b| \leq n^{1/4-\eta} B} |\mathbf{F}_n^{g,P}(a, b, c_\psi) - \mathbf{F}_n^{g,P}(0, 0, c_\psi)| = o_P(1).$$

The proof has two parts. First, we keep  $a$  fixed and consider variation in  $b$ ,  $c$ . This has been done in [17, Theorem 4.1]. Secondly, we keep  $b$  fixed and consider variation in  $a$ ,  $c$  as done in Theorem 5.

The second result provides a linearization of the compensator.

**Theorem 8.** Suppose Assumption 2 (ia, ib, iiib) holds with  $r = 0$ . Let  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for any  $B > 0$  and as  $n \rightarrow \infty$

$$\begin{aligned} & \sup_{0 \leq \psi \leq 1} \sup_{|a|, |b| \leq n^{1/4-\eta} B} |n^{1/2} \{ \bar{\mathbf{F}}_n^{g,P}(a, b, c_\psi) - \bar{\mathbf{F}}_n^{g,P}(0, 0, c_\psi) \} \\ & - \sigma^{p-1} c_\psi^p f(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} (n^{-1/2} a c_\psi + x_{in}' b)| = O_P(n^{-2\eta}). \end{aligned}$$

Finally, the tightness of the empirical process  $\mathbf{F}_n^{g,p}(0, 0, c_\psi)$  was shown in [17, Theorem 4.4], see tightness in [3].

### 4.3 A result for the two-sided empirical process

The 1-step Huber-skip M-estimator involves indicators depending on the absolute value of the residuals. We therefore present some results for a class of two-sided weighted and marked empirical processes.

Define the weighted and marked absolute empirical distribution function

$$\widehat{\mathbf{G}}_n^{g,p}(a, b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p \mathbf{1}_{(|\varepsilon_i - x'_{in} b| \leq \sigma c + n^{-1/2} a c)}. \quad (20)$$

We suppose  $a$  so that  $\sigma + n^{-1/2} a > 0$ , in which case it suffices to consider  $c \geq 0$ . This restriction on  $a$  is satisfied when choosing  $a$  as  $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)$  such that  $\sigma + n^{-1/2} \tilde{a} = \tilde{\sigma} > 0$ . Introduce the compensator of  $\widehat{\mathbf{G}}_n^{g,p}(a, b, c)$

$$\overline{\mathbf{G}}_n^{g,p}(a, b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \varepsilon_i^p \mathbf{1}_{(|\varepsilon_i - x'_{in} b| \leq \sigma c + n^{-1/2} a c)}. \quad (21)$$

Note  $\overline{\mathbf{G}}_n^{1,0}(0, 0, c) = \mathbf{G}(c) = \mathbf{P}(|\varepsilon_i| \leq \sigma c)$ . Then the absolute empirical process is

$$\mathbf{G}_n^{g,p}(a, b, c) = n^{1/2} \{ \widehat{\mathbf{G}}_n^{g,p}(a, b, c) - \overline{\mathbf{G}}_n^{g,p}(a, b, c) \}. \quad (22)$$

We can now derive asymptotic theory for the absolute empirical process from Theorems 7, 8. These results are presented under more restrictive Assumption 1, where the innovation distribution is symmetric, see Remark 1 and Lemma 4. In this section, we only consider  $g_{in}$  chosen as  $1$ ,  $n^{1/2} N' x_i$ ,  $n N' x_i x'_i N$  and  $p$  as  $0, 1, 2$ .

**Theorem 9.** *Suppose Assumption 1 (ia, iib, iic) holds. Let  $c_\psi = \mathbf{G}^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for all  $B > 0$  and as  $n \rightarrow \infty$*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a|, |b| \leq n^{1/4 - \eta_B}} |\mathbf{G}_n^{g,p}(a, b, c_\psi) - \mathbf{G}_n^{g,p}(0, 0, c_\psi)| = o_{\mathbf{P}}(1).$$

**Theorem 10.** *Suppose Assumption 1 (ia, iic) holds. Let  $c_\psi = \mathbf{G}^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then for all  $B > 0$  and as  $n \rightarrow \infty$*

$$\begin{aligned} & \sup_{0 \leq \psi \leq 1} \sup_{|a|, |b| \leq n^{1/4 - \eta_B}} |n^{1/2} \{ \overline{\mathbf{G}}_n^{g,p}(a, b, c_\psi) - \overline{\mathbf{G}}_n^{g,p}(0, 0, c_\psi) \} \\ & - 2\sigma^{p-1} c_\psi^p f(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} \{ \mathbf{1}_{(p \text{ even})} n^{-1/2} a c_\psi + \mathbf{1}_{(p \text{ odd})} x'_{in} b \} | = O_{\mathbf{P}}(n^{-2\eta}). \end{aligned}$$

## 5 Discussion

This paper contributes to the asymptotic theory of iterated 1-step Huber-skip M-estimators. The results are derived under the null hypothesis that there are no outliers in the model. It is well known that the first order asymptotic approximation is fragile in some small finite sample situations. Therefore, it would be of interest to carry out simulation studies to evaluate the finite sample performance of the results in this paper. Likewise it would be of interest to extend the result to situations where outliers are actually present in the data generating process. Scenario possibly contain single outliers, clusters of outliers, level shifts, symmetric or non-symmetric outliers. In such situations, we would analyze the potency, which is the retention rate for relevant outliers. Moreover, it would be possible to compare the potency of two distinct outlier detection algorithms with the same gauge.

**Acknowledgements** The second author is grateful to the Programme of Economic Modelling, which is part of Institute for New Economic Thinking at the Oxford Martin School. We thank Jana Jurečková and two anonymous referees for many constructive suggestions for improvement of the manuscript.

## Appendix 1 A metric on $\mathbf{R}$ and some inequalities

The asymptotic theory uses a chaining argument. This involves a partitioning of the quantile axis using a metric, which is presented first. Then follows some preliminary inequalities including an iterated exponential martingale inequality.

Define the function

$$J_{i,p}(x,y) = \left(\frac{\varepsilon_i}{\sigma}\right)^p \{1_{(\varepsilon_i/\sigma \leq y)} - 1_{(\varepsilon_i/\sigma \leq x)}\}. \quad (23)$$

Our interest focus on  $J_{i,p}(x,y)$  of order  $2^r$  with  $r \in \mathbf{N}$ . Note that  $u^{2^r p}$  is non-negative since  $2^r p$  is even for  $p \in \mathbf{N}_0$  and  $r \in \mathbf{N}$ . Introduce a positive and increasing function

$$H_r(x) = \int_{-\infty}^x (1 + u^{2^r p}) f(u) du. \quad (24)$$

The derivative of this function is  $\dot{H}_r(x) = (1 + x^{2^r p}) f(x)$ . Then, denote the constant

$$H_r = H_r(\infty) = \int_{-\infty}^{\infty} (1 + u^{2^r p}) f(u) du, \quad (25)$$

which is finite by Assumption 2(*ia*). Selection of the specific  $r \in \mathbf{N}$  will be more clear in proofs of the empirical process results. The intuition of  $H_r(x)$  is obtained through setting  $p = 0$  so that  $H_r(x) = 2F(x)$ ,  $\dot{H}_r(x) = 2f(x)$  and  $H_r = 2$ . Therefore,  $H_r(x)$  is the generalization of the distribution  $F(x) \sim \varepsilon_i/\sigma$ . For  $x \leq y$  and  $0 \leq s \leq r$ ,

$$0 \leq |\mathbb{E}\{J_{i,p}(x,y)^{2^s}\}| \leq \mathbb{E}\{|J_{i,p}(x,y)|^{2^s}\} \leq H_r(y) - H_r(x), \quad (26)$$

as  $|u^p| < |u^q| + 1$  for  $q \geq p \geq 0$ . Let  $|H_r(x) - H_r(y)|$  be the  $H_r$ -distance for  $x, y \in \mathbf{R}$ .

In the context of chaining, partition the range of  $H_r(c)$  into  $K$  intervals of equal size  $H_r/K$ . In other words, partition the support into  $K$  intervals by endpoints

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \quad (27)$$

with  $c_{-k} = c_0$  for  $k \in \mathbf{N}$  so that for  $1 \leq k \leq K$

$$H_r(c_k) - H_r(c_{k-1}) = \frac{H_r}{K}. \quad (28)$$

We first present two preliminary inequalities.

**Lemma 1.** *If  $|\tilde{c} - c| \leq |Ac + B|$  and  $|A| \leq 1/2$ , then*

$$|c| \leq \frac{|\tilde{c}| + |B|}{1 - |A|}, \quad (Ac + B)^2 \leq 16(A^2\tilde{c}^2 + B^2).$$

*Proof (Lemma 1). First inequality.* Since  $|Ac + B| \leq |A||c| + |B|$ , the assumption implies  $c - |A||c| - |B| \leq \tilde{c} \leq c + |A||c| + |B|$ . Suppose  $c \geq 0$ , then the lower inequality gives  $c(1 - |A|) - |B| \leq \tilde{c}$  so that  $c \leq (\tilde{c} + |B|)/(1 - |A|)$ . Suppose  $c < 0$ , then the upper inequality gives  $\tilde{c} \leq c(1 - |A|) + |B|$  so that  $(\tilde{c} - |B|)/(1 - |A|) \leq c$ . Combine to get  $|c| \leq \max\{(\tilde{c} + |B|)/(1 - |A|), (\tilde{c} - |B|)/(1 - |A|)\} \leq (|\tilde{c}| + |B|)/(1 - |A|)$ .

*Second inequality.* The first inequality in the lemma,  $(x + y)^2 \leq 2(x^2 + y^2)$  and  $|A| \leq 1/2$  imply  $c^2 \leq 8(\tilde{c}^2 + B^2)$  and  $(Ac + B)^2 \leq 2(A^2c^2 + B^2)$ . Combine to get  $(Ac + B)^2 \leq 2(8A^2\tilde{c}^2 + 8A^2B^2 + B^2) \leq 16(A^2\tilde{c}^2 + B^2)$ .  $\square$

The following lemma concerns the  $H_r$ -distance of multiplicative shifts.

**Lemma 2.** *Let  $r \in \mathbf{N}_0$ . Suppose  $f$  is a continuous density satisfying*

(a) *moments:*  $\int_{-\infty}^{\infty} |u|^{2^r} p f(u) du < \infty$ ;

(b) *boundedness:*  $\sup_{c \in \mathbf{R}} |c|(1 + |c|^{2^r}) f(c) < \infty$ .

*Let  $c_\psi = F^{-1}(\psi)$  for  $0 \leq \psi \leq 1$ . Then, for any  $B > 0$ , there exists  $C > 0$  so*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4 - \eta} B} |H_r\{c_\psi(1 + n^{-1/2} \frac{a}{\sigma})\} - H_r(c_\psi)| \leq C n^{-1/4 - \eta}.$$

*Proof (Lemma 2).* Denote  $\mathcal{H} = |H_r\{c_\psi(1 + n^{-1/2} a/\sigma)\} - H_r(c_\psi)|$ . Apply the first order mean value theorem at the point  $c_\psi$  to get  $\mathcal{H} = |\sigma^{-1} n^{-1/2} a| |c_\psi| |\dot{H}_r(\tilde{c}_\psi)|$ , where  $|\tilde{c}_\psi - c_\psi| \leq |\sigma^{-1} n^{-1/2} a c_\psi|$  and  $\dot{H}_r(\tilde{c}_\psi) = (1 + \tilde{c}_\psi^{2^r}) f(\tilde{c}_\psi)$ .

There exists  $n_0$ , so for any  $n > n_0$  we have  $|\sigma^{-1} n^{-1/2} a| \leq 1/2$  uniformly in  $|a| \leq n^{1/4 - \eta} B$ . First, for  $n > n_0$ , we apply the first inequality in Lemma 1 to obtain  $|c_\psi| \leq |\tilde{c}_\psi|/(1 - |\sigma^{-1} n^{-1/2} a|) \leq 2|\tilde{c}_\psi|$ . It follows

$$\mathcal{H} \leq \sigma^{-1} n^{-1/2} n^{1/4 - \eta} B 2 |\tilde{c}_\psi| |\dot{H}_r(\tilde{c}_\psi)| \leq 2\sigma^{-1} B \sup_{c \in \mathbf{R}} |c| |\dot{H}_r(c)| n^{-1/4 - \eta}.$$

Thus  $\mathcal{H} \leq Cn^{-1/4-\eta}$  by condition (b) that  $|c\dot{H}_r(c)| = |c|(1+|c|^{2^r p})f(c)$  is bounded uniformly in  $c$ .

Second, consider  $n \leq n_0$ . Note  $H_r(x) \leq H_r(\infty) = H_r$  for any  $x$  so that the triangle inequality shows  $\mathcal{H} \leq 2H_r$ . With  $0 < \eta \leq 1/4$ , it follows

$$\mathcal{H} \leq 2H_r n^{1/4+\eta} n^{-1/4-\eta} \leq 2H_r n_0^{1/4+\eta} n^{-1/4-\eta} = Cn^{-1/4-\eta},$$

where  $C = 2H_r n_0^{1/4+\eta}$  is finite since  $H_r < \infty$  by condition (a).  $\square$

The chaining argument involves the tail behaviour of the maximum of a family of martingales which can be controlled using the following iterated martingale inequality taken from [17]. It builds on an exponential martingale inequality derived by Bercu and Touati [1, Theorem 2.1].

**Lemma 3.** (Johansen and Nielsen, 2016a, Theorem 5.2) For  $l$  so  $1 \leq l \leq L$ , let  $z_{l,i}$  be  $\mathcal{F}_i$  adapted satisfying  $Ez_{l,i}^{2^r} < \infty$  for some  $\bar{r} \in \mathbf{N}$ . Let  $D_r = \max_{1 \leq l \leq L} \sum_{i=1}^n E_{i-1} z_{l,i}^{2^r}$  for  $1 \leq r \leq \bar{r}$ . Suppose, for some  $\zeta \geq 0$ ,  $\lambda > 0$ , that  $L = O(n^\lambda)$  and  $ED_r = O(n^\zeta)$  for  $r \leq \bar{r}$ . If  $\nu > 0$  is chosen such that

- (i)  $\zeta < 2\nu$ ;
- (ii)  $\zeta + \lambda < \nu 2^{\bar{r}}$ ;

then, for all  $\kappa > 0$  and as  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{ \max_{1 \leq l \leq L} \left| \sum_{i=1}^n (z_{l,i} - E_{i-1} z_{l,i}) \right| > \kappa n^\nu \right\} = 0.$$

## Appendix 2 Proofs of empirical process results concerning scale

Here we prove the empirical process results concerning the variation in scale when the regression parameter is known. We use the distance function  $H_r$  with  $r = 2$ .

*Proof (Theorem 5).* Let  $c_{\psi^\dagger} = c_\psi(1 + n^{-1/2}a/\sigma)$  so  $\mathbf{F}_n^{g:P}(a, 0, c_\psi) = \mathbf{F}_n^{g:P}(0, 0, c_{\psi^\dagger})$ . Note  $c_{\psi^\dagger}$  can be greater or less than  $c_\psi$ , since  $a$  such that  $|a| \leq n^{1/4-\eta}B$  and  $c_\psi$  can be either positive or negative. Assume  $c_\psi < c_{\psi^\dagger}$  without loss of generality. Denote  $R(c_\psi, c_{\psi^\dagger}) = \mathbf{F}_n^{g:P}(0, 0, c_{\psi^\dagger}) - \mathbf{F}_n^{g:P}(0, 0, c_\psi)$ . The aim is to prove  $\mathcal{R}_n = o_P(1)$  for  $n \rightarrow \infty$  where  $\mathcal{R}_n = \sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4-\eta}B} |R(c_\psi, c_{\psi^\dagger})|$ .

1. *Partition the support.* For  $\delta, n > 0$  partition the range of quantiles  $c$  as laid out in (27) with  $K = \text{int}(H_r n^{1/2}/\delta)$  and  $r = 2$  since  $H_r < \infty$  by assumption (ia).

2. *Assign  $c_\psi$  and  $c_{\psi^\dagger}$  to the partitioned support.* For each  $\psi$  and  $\psi^\dagger$  there exist  $k \leq k^\dagger$  and grid points so that  $c_{k-1} < c_\psi \leq c_k$  and  $c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger}$ .

3. *Apply chaining.* Relate  $c_\psi$  to the nearest right grid point  $c_k$  and  $c_{\psi^\dagger}$  to the nearest left grid point  $c_{k^\dagger-1}$ . Add and subtract  $\mathbf{F}_n^{g:P}(0, 0, c_k)$  and  $\mathbf{F}_n^{g:P}(0, 0, c_{k^\dagger-1})$  to  $R(c_\psi, c_{\psi^\dagger})$ . The triangle inequality gives

$$|R(c_\psi, c_{\psi^\dagger})| \leq |R(c_\psi, c_k)| + |R(c_k, c_{k^\dagger-1})| + |R(c_{k^\dagger-1}, c_{\psi^\dagger})|.$$

Note that if  $c_\psi, c_{\psi^\dagger}$  are in the same interval, then  $|R(c_k, c_{k^\dagger-1})| = |R(c_{k-1}, c_k)|$ . If  $c_\psi, c_{\psi^\dagger}$  are in the neighbouring intervals, then  $|R(c_k, c_{k^\dagger-1})| = 0$ . Apply chaining to obtain  $\mathcal{R}_n \leq \mathcal{R}_{n,1} + \mathcal{R}_{n,2} + \mathcal{R}_{n,3} + \mathcal{R}_{n,4}$ , where

$$\begin{aligned} \mathcal{R}_{n,1} &= \max_{1 \leq k < k^\dagger - 1 < K} |R(c_k, c_{k^\dagger-1})|, \\ \mathcal{R}_{n,2} &= \max_{1 \leq k \leq K} |R(c_{k-1}, c_k)|, \\ \mathcal{R}_{n,3} &= \max_{1 \leq k \leq K} \sup_{c_{k-1} < c_\psi \leq c_k} |R(c_\psi, c_k)|, \\ \mathcal{R}_{n,4} &= \max_{1 \leq k^\dagger \leq K} \sup_{c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger}} |R(c_{k^\dagger-1}, c_{\psi^\dagger})|. \end{aligned}$$

Thus, it suffices to show  $\mathcal{R}_{n,j} = o_p(1)$  for  $j = 1, 2, 3, 4$  as  $n \rightarrow \infty$ .

4. *The term  $\mathcal{R}_{n,1}$  is  $o_p(1)$ .* Use Lemma 3 with  $v = 1/2$ . Let  $g_{in}$  have coordinates  $g_{in}^* = \sigma^p g_{in}$ . Recall the notation  $J_{i,p}(x, y)$  in (23). Write the coordinates of  $R(c_k, c_{k^\dagger-1})$  as  $n^{-1/2} \sum_{i=1}^n (z_{l,i} - E_{i-1} z_{l,i})$  with  $z_{l,i} = g_{in}^* J_{i,p}(c_k, c_{k^\dagger-1})$ , where  $l$  represents the indices  $k, k^\dagger$  with  $L \leq K^2$ . Two conditions of Lemma 3 need to be verified.

*The parameter  $\lambda$ .* The set of indices  $l$  has the size  $L = O(n^\lambda)$  where  $\lambda = 1$ , since  $L \leq K^2$  and  $K = O(n^{1/2})$ .

*The parameter  $\zeta$ .* Consider  $1 \leq s \leq r = 2$  (instead of  $1 \leq r \leq \bar{r} = 2$ ). By construction of partition and assignment in steps 1, 2, then  $c_\psi \leq c_k < c_{k^\dagger-1} < c_{\psi^\dagger}$ . Thus,

$$E_{i-1} J_{i,p}^{2^s}(c_k, c_{k^\dagger-1}) \leq H_r(c_{k^\dagger-1}) - H_r(c_k) \leq H_r(c_{\psi^\dagger}) - H_r(c_\psi) \leq Cn^{-1/4-\eta},$$

by Lemma 2 using assumption (i) for some finite  $C > 0$ . Since

$$D_s = \max_{1 \leq l \leq L} \sum_{i=1}^n E_{i-1} z_{l,i}^{2^s} = \max_{1 \leq k < k^\dagger - 1 < K} \sum_{i=1}^n g_{in}^{*2^s} E_{i-1} J_{i,p}^{2^s}(c_k, c_{k^\dagger-1}),$$

we then find  $D_s \leq Cn^{-1/4-\eta} \sum_{i=1}^n g_{in}^{*2^s}$ . Moreover, using assumption (ii) we find that  $E n^{-1} \sum_{i=1}^n g_{in}^{*2^s} = O(1)$ . Thus, with  $\zeta = 3/4 - \eta$ , we have  $ED_s = O(n^\zeta)$ .

*Condition (i)* is that  $\zeta < 2v$ . This holds since  $\eta > 0$  so  $\zeta = 3/4 - \eta < 1 = 2v$ .

*Condition (ii)* is that  $\zeta + \lambda < v2^r$  where  $r = 2$ . This is satisfied since  $\eta > 0$  so  $\zeta + \lambda = 7/4 - \eta < 2 = v2^r$ .

5. *The term  $\mathcal{R}_{n,2}$  is  $o_p(1)$ .* Use Lemma 3 with  $v = 1/2$  and  $z_{l,i} = g_{in}^* J_{i,p}(c_{k-1}, c_k)$ , where index  $l = k$  has the size  $L = K$ . Two conditions of Lemma 3 need to be shown.

*The parameter  $\lambda$ .* The size  $L = O(n^\lambda)$  where  $\lambda = 1/2$ , since  $L = K = O(n^{1/2})$ .

*The parameter  $\zeta$ .* Consider  $1 \leq s \leq r = 2$ . The equality (28) shows

$$E_{i-1} J_{i,p}^{2^s}(c_{k-1}, c_k) \leq H_r(c_k) - H_r(c_{k-1}) = \frac{H_r}{K} = O(n^{-1/2}).$$

Then, we find

$$D_s = \max_{1 \leq l \leq L} \sum_{i=1}^n \mathbb{E}_{i-1} z_{l,i}^{2s} = \max_{1 \leq k \leq K} \sum_{i=1}^n g_{in}^{*2s} \mathbb{E}_{i-1} J_{i,p}^{2s}(c_{k-1}, c_k) = O(n^{-1/2}) \sum_{i=1}^n g_{in}^{*2s}.$$

It follows that  $\mathbb{E}D_s = O(n^\zeta)$  where  $\zeta = 1/2$  by assumption (ii).

*Condition (i)* holds, since  $\zeta = 1/2 < 1 = 2\nu$ .

*Condition (ii)* holds, since  $\zeta + \lambda = 1 < 2 = \nu 2'$ .

6. *Decompose the term  $\mathcal{R}_{n,3}$ .* Apply the triangle and Jensen's inequality to obtain,

$$|R(c_\Psi, c_k)| \leq n^{-1/2} \sum_{i=1}^n |g_{in}^*| \{ |J_{i,p}(c_\Psi, c_k)| + \mathbb{E}_{i-1} |J_{i,p}(c_\Psi, c_k)| \}.$$

For  $c_{k-1} < c_\Psi \leq c_k$  where  $1 \leq k \leq K$ , we have  $|J_{i,p}(c_\Psi, c_k)| \leq |J_{i,p}(c_{k-1}, c_k)|$ . Then,

$$\mathcal{R}_{n,3} \leq \max_{1 \leq k \leq K} n^{-1/2} \sum_{i=1}^n |g_{in}^*| \{ |J_{i,p}(c_{k-1}, c_k)| + \mathbb{E}_{i-1} |J_{i,p}(c_{k-1}, c_k)| \}.$$

Therefore, it can be argued that  $\mathcal{R}_{n,3} \leq \tilde{\mathcal{R}}_{n,3} + 2\bar{\mathcal{R}}_{n,3}$ , where

$$\begin{aligned} \tilde{\mathcal{R}}_{n,3} &= \max_{1 \leq k \leq K} n^{-1/2} \sum_{i=1}^n |g_{in}^*| \{ |J_{i,p}(c_{k-1}, c_k)| - \mathbb{E}_{i-1} |J_{i,p}(c_{k-1}, c_k)| \}, \\ \bar{\mathcal{R}}_{n,3} &= \max_{1 \leq k \leq K} n^{-1/2} \sum_{i=1}^n |g_{in}^*| \mathbb{E}_{i-1} |J_{i,p}(c_{k-1}, c_k)|. \end{aligned}$$

Thus, it suffices to show  $\tilde{\mathcal{R}}_{n,3}$  and  $\bar{\mathcal{R}}_{n,3}$  are  $\text{o}_{\mathbb{P}}(1)$  as  $n \rightarrow \infty$ .

7. *The term  $\tilde{\mathcal{R}}_{n,3}$  is  $\text{o}_{\mathbb{P}}(1)$ .* Argue along the lines of step 5 to show  $\tilde{\mathcal{R}}_{n,3} = \text{o}_{\mathbb{P}}(1)$ .

8. *Bounding the term  $\bar{\mathcal{R}}_{n,3}$ .* Use the equality (28) and  $K = O(H_r n^{1/2}/\delta)$  to get

$$\mathbb{E}_{i-1} |J_{i,p}(c_{k-1}, c_k)| \leq H_r(c_k) - H_r(c_{k-1}) = \frac{H_r}{K} = O(n^{-1/2}\delta).$$

We then find  $\bar{\mathcal{R}}_{n,3} = O(n^{-1/2}\delta) n^{-1/2} \sum_{i=1}^n |g_{in}^*| = O_{\mathbb{P}}(\delta)$  by the Markov inequality and the assumption (ii) that  $n^{-1} \sum_{i=1}^n \mathbb{E} |g_{in}^*|^4 = O(1)$ . Thus, choose  $\delta$  sufficiently small so that  $\bar{\mathcal{R}}_{n,3} = \text{o}_{\mathbb{P}}(1)$ .

9. *The term  $\mathcal{R}_{n,4}$  is  $\text{o}_{\mathbb{P}}(1)$ .* This is similar as to show  $\mathcal{R}_{n,3} = \text{o}_{\mathbb{P}}(1)$ . Thus the same argument can be made through steps 6, 7, 8.  $\square$

*Proof (Theorem 6).* The term of interest is

$$\begin{aligned} D_n(a, c_\Psi) &= n^{1/2} \{ \bar{F}_n^{g,p}(a, 0, c_\Psi) - \bar{F}_n^{g,p}(0, 0, c_\Psi) \} \\ &\quad - \sigma^{p-1} c_\Psi^p f(c_\Psi) n^{-1/2} \sum_{i=1}^n g_{in} n^{-1/2} a c_\Psi, \end{aligned}$$

where  $\bar{F}_n^{g,p}$  is well-defined due to assumption (ia). Let  $w_i^{a, c_\Psi} = 1_{(\varepsilon_i \leq \sigma c_\Psi + n^{-1/2} a c_\Psi)} - 1_{(\varepsilon_i \leq \sigma c_\Psi)}$  and  $h_i(a, c_\Psi) = n^{-1/2} a c_\Psi / \sigma$  and denote  $s(c) = c^p f(c)$ . Define  $S_i(a, c_\Psi) =$

$E_{i-1} \varepsilon_i^p w_i^{a, c_\psi} - \sigma^p h_i(a, c_\psi) s(c_\psi)$  so that we have  $D_n(a, c_\psi) = n^{-1/2} \sum_{i=1}^n g_{in} S_i(a, c_\psi)$ . Write  $S_i(a, c_\psi)$  as an integral and apply the second order Taylor expansion at  $c_\psi$  to get

$$S_i(a, c_\psi) = \sigma^p \left\{ \int_{c_\psi}^{c_\psi + h_i(a, c_\psi)} s(u) du - h_i(a, c_\psi) s(c_\psi) \right\} = \sigma^p h_i^2(a, c_\psi) \dot{s}(\tilde{c}_\psi) / 2,$$

where  $|\tilde{c}_\psi - c_\psi| \leq |h_i(a, c_\psi)|$ . There exists  $n_0 > 0$  so for any  $n > n_0$  we have  $|\sigma^{-1} n^{-1/2} a| \leq 1/2$ . We then apply the second inequality in Lemma 1 to obtain  $h_i^2(a, c_\psi) \leq 16n^{-1} a^2 \tilde{c}_\psi^2 / \sigma^2$ . Exploit the bound  $|a| \leq n^{1/4-\eta} B$  to get

$$|S_i(a, c_\psi)| = O(n^{-1/2-2\eta}) \tilde{c}_\psi^2 |\dot{s}(\tilde{c}_\psi)| = O(n^{-1/2-2\eta})$$

uniformly in  $\psi, a$ , since  $\tilde{c}_\psi^2 |\dot{s}(\tilde{c}_\psi)| \leq \sup_{c \in \mathbf{R}} c^2 |\dot{s}(c)| < \infty$  by assumption (i) noting that  $\dot{s}(c) = c^{p-1} \dot{f}(c) + c^p \dot{f}'(c)$ . Then the triangle inequality gives

$$|D_n(a, c_\psi)| \leq n^{-1/2} \sum_{i=1}^n |g_{in}| |S_i(a, c_\psi)| = O(n^{-2\eta}) n^{-1} \sum_{i=1}^n |g_{in}|.$$

By assumption (ii), this term is of order  $O_P(n^{-2\eta})$  uniformly in  $\psi, a$ .  $\square$

### Appendix 3 Proofs of general empirical process results

*Proof (Theorem 7).* The term of interest is  $\mathcal{W} = \mathbf{F}_n^{g:P}(a, b, c_\psi) - \mathbf{F}_n^{g:P}(0, 0, c_\psi)$ . Denote  $c_{\psi^\dagger} = c_\psi(1 + n^{-1/2} a / \sigma)$ . Notice that  $\mathbf{F}_n^{g:P}(a, b, c_\psi) = \mathbf{F}_n^{g:P}(0, b, c_{\psi^\dagger})$  so that  $\mathcal{W} = \mathbf{F}_n^{g:P}(0, b, c_{\psi^\dagger}) - \mathbf{F}_n^{g:P}(0, 0, c_\psi)$ . Add and subtract  $\mathbf{F}_n^{g:P}(a, 0, c_\psi) = \mathbf{F}_n^{g:P}(0, 0, c_{\psi^\dagger})$  and apply the triangle inequality to get

$$|\mathcal{W}| \leq |\mathbf{F}_n^{g:P}(0, b, c_{\psi^\dagger}) - \mathbf{F}_n^{g:P}(0, 0, c_{\psi^\dagger})| + |\mathbf{F}_n^{g:P}(a, 0, c_\psi) - \mathbf{F}_n^{g:P}(0, 0, c_\psi)|.$$

Thus, the problem reduces to showing

$$\sup_{0 \leq \psi^\dagger \leq 1} \sup_{|b| \leq n^{1/4-\eta} B} |\mathbf{F}_n^{g:P}(0, b, c_{\psi^\dagger}) - \mathbf{F}_n^{g:P}(0, 0, c_{\psi^\dagger})| = o_P(1), \quad (29)$$

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4-\eta} B} |\mathbf{F}_n^{g:P}(a, 0, c_\psi) - \mathbf{F}_n^{g:P}(0, 0, c_\psi)| = o_P(1). \quad (30)$$

Then (29) is shown in [17, Theorem 4.1] by Assumption 2(i, ii, iii) with  $r \geq 2$  such that (14) holds. Further, (30) was considered in Theorem 5, which requires Assumption 2(ia, ib, iii) with  $r = 2$ .  $\square$

*Proof (Theorem 8).* We generalize the proof of Theorem 6. We note  $\bar{\mathbf{F}}_n^{g:P}$  is well-defined due to Assumption 2(ia). The term of interest is

$$D_n(a, b, c_\psi) = n^{1/2} \{ \bar{\mathbf{F}}_n^{g,p}(a, b, c_\psi) - \bar{\mathbf{F}}_n^{g,p}(0, 0, c_\psi) \} \\ - \sigma^{p-1} c_\psi^p \mathbf{f}(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} (n^{-1/2} a c_\psi + x'_{in} b).$$

Let  $w_i^{a,b,c_\psi} = \mathbf{1}_{(\varepsilon_i \leq \sigma c_\psi + n^{-1/2} a c_\psi + x'_{in} b)} - \mathbf{1}_{(\varepsilon_i \leq \sigma c_\psi)}$ ,  $h_i(a, b, c_\psi) = (n^{-1/2} a c_\psi + x'_{in} b) / \sigma$  and  $s(c) = c^p \mathbf{f}(c)$ . Define  $S_i(a, b, c_\psi) = \mathbf{E}_{i-1} \varepsilon_i^p w_i^{a,b,c_\psi} - \sigma^p h_i(a, b, c_\psi) s(c_\psi)$  so that  $D_n(a, b, c_\psi) = n^{-1/2} \sum_{i=1}^n g_{in} S_i(a, b, c_\psi)$ . Write  $S_i(a, b, c_\psi)$  as an integral

$$S_i(a, b, c_\psi) = \sigma^p \left\{ \int_{c_\psi}^{c_\psi + h_i(a,b,c_\psi)} s(u) du - h_i(a, b, c_\psi) s(c_\psi) \right\}.$$

Second order Taylor expansion at  $c_\psi$  shows  $S_i(a, b, c_\psi) = \sigma^p h_i^2(a, b, c_\psi) \dot{s}(\tilde{c}_\psi) / 2$ , where  $|\tilde{c}_\psi - c_\psi| \leq |h_i(a, b, c_\psi)|$ . There exists  $n_0 > 0$  so for any  $n > n_0$  we have  $|\sigma^{-1} n^{-1/2} a| \leq 1/2$ . We then apply the second inequality in Lemma 1 to obtain  $h_i^2(a, b, c_\psi) \leq 16 \{ n^{-1} a^2 \tilde{c}_\psi^2 + (x'_{in} b)^2 \} / \sigma^2$ . Exploit bounds  $|a|, |b| \leq n^{1/4-\eta} B$  and the inequality  $x^2 + y^2 \leq (1+x^2)(1+y^2)$  to get

$$|S_i(a, b, c_\psi)| = O(n^{-1/2-2\eta})(1 + |n^{1/2} x_{in}|^2)(1 + \tilde{c}_\psi^2) |\dot{s}(\tilde{c}_\psi)|.$$

Since  $(1 + \tilde{c}_\psi^2) |\dot{s}(\tilde{c}_\psi)| \leq \sup_{c \in \mathbf{R}} (1 + c^2) |\dot{s}(c)| < \infty$  by Assumption 2(ib) with  $r = 0$ , we have  $|S_i(a, b, c_\psi)| = O(n^{-1/2-2\eta})(1 + |n^{1/2} x_{in}|^2)$  uniformly in  $\psi, a, b$ . Then the triangle inequality gives

$$|D_n(a, b, c_\psi)| \leq n^{-1/2} \sum_{i=1}^n |g_{in}| |S_i(a, b, c_\psi)| = O(n^{-2\eta}) n^{-1} \sum_{i=1}^n |g_{in}| (1 + |n^{1/2} x_{in}|^2).$$

By Assumption 2(iii), this term is of order  $O_P(n^{-2\eta})$  uniformly in  $\psi, a, b$ .  $\square$

The absolute empirical process results are given under more restrictive Assumption 1, so the next lemma concerns the relationship between Assumption 1 and 2.

**Lemma 4.** *Suppose  $g_{in}$  is either of  $1, n^{1/2} N' x_i, n N' x_i x'_i N$  and  $p$  is either of  $0, 1, 2$ . Then Assumption 1(ia, iib, iic) implies Assumption 2 with  $r \geq 2$  satisfying (14).*

*Proof (Lemma 4).* Assumption 1(ia) shows Assumption 2(ia, ic), while Assumption 2(ib) further needs continuous differentiability of  $\mathbf{f}$ , see discussion in [17, Remark 4.1(c)]. Assumption 1(iib) is the same as Assumption 2(ii). Assumption 1(iic) implies Assumption 2(iia) and (iic) by Markov inequality.  $\square$

*Proof (Theorem 9).* The term of interest is  $\mathcal{G} = \mathbf{G}_n^{g,p}(a, b, c_\psi) - \mathbf{G}_n^{g,p}(0, 0, c_\psi)$ . Our focus is on the absolute quantile  $c_\psi = G^{-1}(\psi) > 0$  rather than the one-sided quantile  $c_{\psi^*} = F^{-1}(\psi^*) \in \mathbf{R}$ . Note  $|\varepsilon_i|/\sigma \sim G$  and  $\varepsilon_i/\sigma \sim F$ . Since

$$\mathbf{1}_{(|\varepsilon_i - x'_{in} b| \leq \sigma c + n^{-1/2} a c)} = \mathbf{1}_{(\varepsilon_i \leq \sigma c + n^{-1/2} a c + x'_{in} b)} - \mathbf{1}_{(\varepsilon_i \leq -\sigma c - n^{-1/2} a c + x'_{in} b)}$$

and by (18), (22), we have  $\mathbf{G}_n^{g,P}(a, b, c) = \mathbf{F}_n^{g,P}(a, b, c) - \lim_{c^\dagger \downarrow c} \mathbf{F}_n^{g,P}(a, b, -c^\dagger)$  for any  $c > 0$ . By this and the triangle inequality, then for any  $c_\psi = \mathbf{G}^{-1}(\psi) > 0$ ,

$$|\mathcal{G}| \leq |\mathbf{F}_n^{g,P}(a, b, c_\psi) - \mathbf{F}_n^{g,P}(0, 0, c_\psi)| + \lim_{c^\dagger \downarrow c_\psi} |\mathbf{F}_n^{g,P}(a, b, -c^\dagger) - \mathbf{F}_n^{g,P}(0, 0, -c^\dagger)|.$$

These vanish uniformly in  $\psi, a, b$  by Theorem 7 using Assumption 2 with  $r \geq 2$  such that (14) holds. Lemma 4 shows that Assumption 1 (ia, iib, iic) suffices.  $\square$

*Proof (Theorem 10).* Argue as in the proof of Theorem 9 but using Theorem 8 instead of Theorem 7. Due to the symmetry of  $f$ , the correction term is then

$$\sigma^{p-1} c_\psi^p f(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in}[\{1 + (-1)^p\} n^{-1/2} a c_\psi + \{1 - (-1)^p\} x'_m b].$$

This reduces as desired.  $\square$

## Appendix 4 Proofs of the main results

We first present an auxiliary result for asymptotic expansions of product moments. Then, the tightness and fixed point result are shown for the iterated estimators. At last, we provide the proof of the Poisson exceedence theory for the gauge.

The 1-step Huber-skip M-estimators are least squares estimators for selected observations. The following result describes the asymptotic behaviour of the corresponding product moments. For this purpose introduce the indicators

$$v_i^{a,b,c} = 1_{(|\varepsilon_i - x'_m b| \leq \sigma c + n^{-1/2} a c)}. \quad (31)$$

**Lemma 5.** *Suppose Assumption 1 (ia, ii) holds. Then we have expansions*

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n v_i^{a,b,c} &= n^{-1/2} \sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma c)} + 2f(c) \frac{ac}{\sigma} + R_v(a, b, c), \\ n^{-1/2} \sum_{i=1}^n \varepsilon_i^2 v_i^{a,b,c} &= n^{-1/2} \sum_{i=1}^n \varepsilon_i^2 1_{(|\varepsilon_i| \leq \sigma c)} + 2\sigma^2 c^2 f(c) \frac{ac}{\sigma} + R_{v\varepsilon\varepsilon}(a, b, c), \\ \sum_{i=1}^n N' x_i \varepsilon_i v_i^{a,b,c} &= \sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + 2cf(c) \Sigma b + R_{vx\varepsilon}(a, b, c), \\ n^{1/2} \sum_{i=1}^n N' x_i x'_i N v_i^{a,b,c} &= n^{1/2} \sum_{i=1}^n N' x_i x'_i N 1_{(|\varepsilon_i| \leq \sigma c)} + 2f(c) \Sigma \frac{ac}{\sigma} + R_{vxx}(a, b, c). \end{aligned}$$

Let  $R(a, b, c) = |R_v(a, b, c)| + |R_{v\varepsilon\varepsilon}(a, b, c)| + |R_{vx\varepsilon}(a, b, c)| + |R_{vxx}(a, b, c)|$ . Then for any  $B > 0$  and as  $n \rightarrow \infty$

$$\sup_{0 < c < \infty} \sup_{|a|, |b| \leq n^{1/4 - \eta} B} |R(a, b, c)| = o_{\mathbb{P}}(1).$$

*Remark 2.* The first and fourth item in Lemma 5 adjusted by  $n^{-1/2}$  have expansions

$$n^{-1} \sum_{i=1}^n v_i^{a,b,c} = \psi + R'_v(a, b, c), \quad \sum_{i=1}^n N' x_i x'_i N v_i^{a,b,c} = \psi \Sigma + R'_{vxx}(a, b, c),$$

where for any  $B > 0$  and as  $n \rightarrow \infty$

$$\sup_{0 < c < \infty} \sup_{|a|, |b| \leq n^{1/4 - \eta} B} |R'_v(a, b, c)| + |R'_{vxx}(a, b, c)| = o_{\mathbb{P}}(1).$$

Indeed, for the first expansion, we apply the law of large numbers to obtain  $n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma c)} = \psi + o_{\mathbb{P}}(1)$ , while  $\sup_{c \in \mathbf{R}} |c| f(c) < \infty$  by Assumption 1(*ia*) and  $n^{-1/2} a$  vanishes. For the second expansion, decompose

$$\sum_{i=1}^n N' x_i x'_i N 1_{(|\varepsilon_i| \leq \sigma c)} = \sum_{i=1}^n N' x_i x'_i N \{1_{(|\varepsilon_i| \leq \sigma c)} - \psi\} + \sum_{i=1}^n N' x_i x'_i N \psi.$$

The first item vanishes by the Chebyshev inequality and Assumption 1(*ia, ic*), while the second converges to  $\psi \Sigma$ .  $\square$

*Proof (Lemma 5).* The general class of empirical processes is

$$\mathcal{M}_n = n^{-1/2} \sum_{i=1}^n g_{in} \varepsilon_i^p v_i^{a,b,c}, \quad v_i^{a,b,c} = 1_{(|\varepsilon_i - x'_{in} b| \leq \sigma c + n^{-1/2} ac)}.$$

1. *Decompose  $\mathcal{M}_n$ .* Write  $\mathcal{M}_n = \mathcal{M}_{n,1} + \mathcal{M}_{n,2} + \mathcal{M}_{n,3}$ , where

$$\begin{aligned} \mathcal{M}_{n,1} &= n^{-1/2} \sum_{i=1}^n g_{in} \varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)}, \quad \mathcal{M}_{n,2} = n^{-1/2} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \varepsilon_i^p \{v_i^{a,b,c} - 1_{(|\varepsilon_i| \leq \sigma c)}\}, \\ \mathcal{M}_{n,3} &= n^{-1/2} \sum_{i=1}^n g_{in} \varepsilon_i^p \{v_i^{a,b,c} - 1_{(|\varepsilon_i| \leq \sigma c)}\} - n^{-1/2} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \varepsilon_i^p \{v_i^{a,b,c} - 1_{(|\varepsilon_i| \leq \sigma c)}\}. \end{aligned}$$

Therefore, the first term in stochastic expansion is  $\mathcal{M}_{n,1}$ . We will linearize  $\mathcal{M}_{n,2}$  to obtain the second term, and argue that  $\mathcal{M}_{n,3}$  is small in probability.

2. *Linearize  $\mathcal{M}_{n,2}$ .* Note  $\mathcal{M}_{n,2} = n^{1/2} \{\overline{\mathbf{G}}_n^{g,p}(a, b, c) - \overline{\mathbf{G}}_n^{g,p}(0, 0, c)\}$ , see (21). Theorem 10 by Assumption 1(*ia, ic*) shows  $\mathcal{M}_{n,2} = \overline{\mathcal{M}}_{n,2} + o_{\mathbb{P}}(n^{-2\eta})$ , where

$$\overline{\mathcal{M}}_{n,2} = 2\sigma^{p-1} c^p f(c) n^{-1/2} \sum_{i=1}^n g_{in} \{1_{(p \text{ even})} n^{-1/2} ac + 1_{(p \text{ odd})} x'_{in} b\}.$$

This reduces as desired by Assumption 1(*ia*). Note  $0 < \eta \leq 1/4$ . Thus, we have  $\mathcal{M}_{n,2} = \overline{\mathcal{M}}_{n,2} + o_{\mathbb{P}}(1)$  uniformly in  $0 < c < \infty$  and  $|a|, |b| \leq n^{1/4 - \eta} B$ .

3. *Bounding  $\mathcal{M}_{n,3}$ .* Note  $\mathcal{M}_{n,3} = \mathbf{G}_n^{g,P}(a, b, c) - \mathbf{G}_n^{g,P}(0, 0, c)$ , see (22). Due to Assumption 1 (ia, iib, iic), Theorem 9 shows  $\mathcal{M}_{n,3} = o_{\mathbb{P}}(1)$  uniformly in  $a, b, c$ .  $\square$

*Proof (Theorem 1).* The  $m+1$  step estimators for  $\beta, \sigma^2$  are defined in (10), (11). These are least squares estimators for the non-outlying observations and satisfy

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \left( \sum_{i=1}^n N' x_i x_i' N v_{i,c}^{(m)} \right)^{-1} \left( \sum_{i=1}^n N' x_i \varepsilon_i v_{i,c}^{(m)} \right), \quad (32)$$

$$\begin{aligned} n^{1/2} \{ (\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2 \} &= \zeta_c^{-2} (n^{-1} \sum_{i=1}^n v_{i,c}^{(m)})^{-1} n^{-1/2} \left\{ \sum_{i=1}^n (\varepsilon_i^2 - \zeta_c^2 \sigma^2) v_{i,c}^{(m)} \right. \\ &\quad \left. - \left( \sum_{i=1}^n \varepsilon_i x_i' N v_{i,c}^{(m)} \right) \left( \sum_{i=1}^n N' x_i x_i' N v_{i,c}^{(m)} \right)^{-1} \left( \sum_{i=1}^n N' x_i \varepsilon_i v_{i,c}^{(m)} \right) \right\}. \end{aligned} \quad (33)$$

We express the weight  $v_{i,c}^{(m)}$  in (9) as

$$v_{i,c}^{(m)} = \mathbf{1}_{(|y_i - x_i' \widehat{\beta}_c^{(m)}| \leq \widehat{\sigma}_c^{(m)} c)} = \mathbf{1}_{(|\varepsilon_i - x_i' \widehat{b}_c^{(m)}| \leq \sigma c + n^{-1/2} \widehat{a}_c^{(m)} c)} = v_i^{\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c},$$

where  $\widehat{b}_c^{(m)} = N^{-1}(\widehat{\beta}_c^{(m)} - \beta)$  and  $\widehat{a}_c^{(m)} = n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$  are the  $m$  step estimation errors for  $\beta$  and  $\sigma$ .

Since  $|\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = O_{\mathbb{P}}(1)$  and by Assumption 1 (ia, ii), then Lemma 5 and Remark 2 with  $\kappa = 0, \eta = 1/4$  show asymptotic expansions for product moments. Substitute these expansions into (32), (33) to first get

$$\widehat{b}_c^{(m+1)} = \frac{2\text{cf}(c)}{\psi} \widehat{b}_c^{(m)} + (\psi \Sigma)^{-1} \sum_{i=1}^n N' x_i \varepsilon_i \mathbf{1}_{(|\varepsilon_i| \leq \sigma c)} + R_{\beta}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where the remainder  $R_{\beta}(a, b, c)$  vanishes uniformly in  $c_0 \leq c < \infty$  and  $|a|, |b| \leq B$ . A key to this is that  $c$  is bounded away from zero and that  $\Sigma$  is positive definite by Assumption 1 (iia) so that the denominator  $\psi, \psi \Sigma$  is bounded away from zero.

Secondly, we get an expression for  $\widehat{\sigma}_c^{(m+1)}$ . By Taylor expansion, first note that

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \frac{1}{2\sigma} n^{1/2} \{ (\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2 \} + n^{-1/2} O[n \{ (\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2 \}^2].$$

Then apply arguments as above to get

$$\begin{aligned} \widehat{a}_c^{(m+1)} &= \frac{c(c^2 - \zeta_c^2) f(c)}{\tau_2^c} \widehat{a}_c^{(m)} + \frac{1}{2\sigma \tau_2^c} n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \zeta_c^2 \sigma^2) \mathbf{1}_{(|\varepsilon_i| \leq \sigma c)} \\ &\quad + R_{\sigma}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c), \end{aligned}$$

where the remainder  $R_{\sigma}(a, b, c)$  also vanishes uniformly.  $\square$

To prove the tightness and fixed point result, let  $|\cdot|$  refer to the usual Euclidean vector norm, while  $\|M\| = \max\{\text{eigen}(M'M)\}^{1/2}$  is the spectral norm for any matrix  $M$ . Note that the norms are compatible so that  $|Mx| \leq \|M\||x|$  for any vector  $x$ .

*Proof (Theorem 2).* Due to Assumption 1 (ia, ii), Theorem 1 shows

$$\widehat{u}_c^{(m+1)} = \Gamma_c \widehat{u}_c^{(m)} + K_c + R_u(\widehat{u}_c^{(m)}, c), \quad (34)$$

where the remainder term satisfies  $\sup_{c_0 \leq c < \infty} \sup_{|u| \leq B} |R_u(u, c)| = o_P(1)$  and

$$\widehat{u}_c^{(m)} = \begin{pmatrix} \widehat{b}_c^{(m)} \\ \widehat{a}_c^{(m)} \end{pmatrix} = \begin{Bmatrix} N^{-1}(\widehat{\beta}_c^{(m)} - \beta) \\ n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma) \end{Bmatrix}, \quad \Gamma_c = \begin{Bmatrix} \frac{2cf(c)}{\psi} I_{\dim x} & 0 \\ 0 & \frac{c(c^2 - \zeta_c^2)f(c)}{\tau_2^c} \end{Bmatrix}, \quad (35)$$

$$K_c = \begin{Bmatrix} (\psi\Sigma)^{-1} & 0 \\ 0 & (2\sigma\tau_2^c)^{-1} \end{Bmatrix} \sum_{i=1}^n \begin{Bmatrix} N'x_i\epsilon_i \\ n^{-1/2}(\epsilon_i^2 - \zeta_c^2\sigma^2) \end{Bmatrix} 1_{(|\epsilon_i| \leq \sigma c)}. \quad (36)$$

Apply the difference equation (34) recursively to obtain the representation

$$\widehat{u}_c^{(m+1)} = \Gamma_c^{m+1} \widehat{u}_c^{(0)} + \sum_{l=0}^m \Gamma_c^l \{K_c + R_u(\widehat{u}_c^{(m-l)}, c)\}. \quad (37)$$

Use the triangle inequality and  $|Mx| \leq \|M\||x|$  to get

$$|\widehat{u}_c^{(m+1)}| \leq \|\Gamma_c^{m+1}\| |\widehat{u}_c^{(0)}| + \{|K_c| + \max_{0 \leq l \leq m} |R_u(\widehat{u}_c^{(l)}, c)|\} \sum_{l=0}^m \|\Gamma_c^l\|.$$

Assumption 1 (ia) shows  $\sup_{c_0 \leq c < \infty} \max\{|2cf(c)/\psi|, |c(c^2 - \zeta_c^2)f(c)/\tau_2^c|\} < 1$ , see [16, Theorem 3.5], so  $\sup_{c_0 \leq c < \infty} \|\Gamma_c\| < 1$ . Gelfand's formula in [24, Theorem 3.4] gives  $\lim_{m \rightarrow \infty} \|M^m\|^{1/m} = \max\{\text{eigen}(M)\}$ . Therefore for some  $\omega$  satisfying that  $\sup_{c_0 \leq c < \infty} \|\Gamma_c\| < \omega < 1$  there exists  $m_0 > 0$  so for all  $m > m_0$

$$\sup_{c_0 \leq c < \infty} \|\Gamma_c^m\| < \omega^m < 1. \quad (38)$$

Also note  $(I_{\dim x+1} - \Gamma_c)^{-1} = \sum_{l=0}^{\infty} \Gamma_c^l$ . This in turn implies for some  $1 < B_0 < \infty$

$$\sup_{0 \leq m < \infty} \sup_{c_0 \leq c < \infty} \|\Gamma_c^m\| < B_0, \quad \sup_{c_0 \leq c < \infty} \|(I_{\dim x+1} - \Gamma_c)^{-1}\| \leq \sum_{l=0}^{\infty} \sup_{c_0 \leq c < \infty} \|\Gamma_c^l\| < B_0. \quad (39)$$

Therefore, (39) shows for all  $m \in [0, \infty)$

$$|\widehat{u}_c^{(m+1)}| < B_0 \{|\widehat{u}_c^{(0)}| + |K_c| + \max_{0 \leq l \leq m} |R_u(\widehat{u}_c^{(l)}, c)|\}. \quad (40)$$

For any  $c \in [c_0, \infty)$ , Assumption 1 (iii) with  $\eta = 1/4$  guarantees tightness of  $\widehat{u}_c^{(0)}$ , while the kernel  $K_c$  is tight by [17, Theorem 4.4] using Assumption 1 (ia, iib, iic). Thus, for all  $\varepsilon, \delta > 0$  there exist  $n_0, U_0 > 0$  so that the set

$$\mathcal{A}_n = \{B_0 \sup_{c_0 \leq c < \infty} (|\hat{u}_c^{(0)}| + |K_c|) \leq U_0/3, B_0 \sup_{c_0 \leq c < \infty} \sup_{|u| \leq U_0} |R_u(u, c)| < \delta/2\} \quad (41)$$

has probability larger than  $1 - \varepsilon$  for all  $n > n_0$ .

Mathematical induction over  $m$  is used to show  $\sup_{0 \leq m < \infty} \sup_{c_0 \leq c < \infty} |\hat{u}_c^{(m)}| \leq U_0$  on the set  $\mathcal{A}_n$ . For  $m = 0$  as induction starts,  $\sup_{c_0 \leq c < \infty} |\hat{u}_c^{(0)}| \leq B_0^{-1} U_0/3 < U_0$  holds since  $B_0 > 1$ . The induction assumption is that  $\sup_{0 \leq l \leq m} \sup_{c_0 \leq c < \infty} |\hat{u}_c^{(l)}| \leq U_0$ . This implies  $B_0 \max_{0 \leq l \leq m} |R_u(\hat{u}_c^{(l)}, c)| < \delta/2$ , and then the bound in (40) becomes  $\sup_{c_0 \leq c < \infty} |\hat{u}_c^{(m+1)}| < 2U_0/3 + \delta/2 < U_0$  so  $\sup_{0 \leq l \leq m+1} \sup_{c_0 \leq c < \infty} |\hat{u}_c^{(l)}| \leq U_0$ .  $\square$

*Proof (Theorem 3).* Due to Assumption 1 (ia, ii, iii), Theorem 1 provides the recursive equation (34). Then, Theorem 2 shows  $\sup_{0 \leq m < \infty} \sup_{c_0 \leq c < \infty} |\hat{u}_c^{(m)}| = O_P(1)$ , so the remainder term in (34) is  $o_P(1)$ . Thus, for  $m, n \rightarrow \infty$  the fixed point should satisfy  $\hat{u}_c^* = \Gamma_c \hat{u}_c^* + K_c$  so that

$$\hat{u}_c^* = (I_{\dim x+1} - \Gamma_c)^{-1} K_c. \quad (42)$$

Substitute (35), (36) of  $\hat{u}_c^*$ ,  $\Gamma_c$  and  $K_c$  into (42) to obtain

$$\begin{Bmatrix} N^{-1}(\hat{\beta}_c^* - \beta) \\ n^{1/2}(\hat{\sigma}_c^* - \sigma) \end{Bmatrix} = \begin{bmatrix} \frac{1}{\psi - 2cf(c)} \Sigma^{-1} \sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} \\ \frac{1}{2\sigma\{\tau_c^2 - c(c^2 - \zeta_c^2)f(c)\}} n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \zeta_c^2 \sigma^2) 1_{(|\varepsilon_i| \leq \sigma c)} \end{bmatrix}.$$

Replace (37) and (42) into the deviation  $\hat{\Delta}_c^{(m+1)} = \hat{u}_c^{(m+1)} - \hat{u}_c^*$ , and then apply  $\sum_{l=0}^m \Gamma_c^l = (I_{\dim x+1} - \Gamma_c^{m+1})(I_{\dim x+1} - \Gamma_c)^{-1}$  to attain

$$\hat{\Delta}_c^{(m+1)} = \Gamma_c^{m+1} \{\hat{u}_c^{(0)} - (I_{\dim x+1} - \Gamma_c)^{-1} K_c\} + \sum_{l=0}^m \Gamma_c^l R_u(\hat{u}_c^{(m-l)}, c).$$

To bound  $\hat{\Delta}_c^{(m+1)}$ , use the triangle inequality and  $|Mx| \leq \|M\| |x|$  to get

$$|\hat{\Delta}_c^{(m+1)}| \leq \|\Gamma_c^{m+1}\| \{|\hat{u}_c^{(0)}| + \|(I_{\dim x+1} - \Gamma_c)^{-1}\| |K_c|\} + \max_{0 \leq l \leq m} |R_u(\hat{u}_c^{(l)}, c)| \sum_{l=0}^m \|\Gamma_c^l\|.$$

By Assumption 1 (ia) and Gelfand's formula, (38) and the second inequality in (39) imply for  $m > m_0$

$$|\hat{\Delta}_c^{(m+1)}| < \omega^{m+1} (|\hat{u}_c^{(0)}| + B_0 |K_c|) + B_0 \max_{0 \leq l \leq m} |R_u(\hat{u}_c^{(l)}, c)|.$$

On the set  $\mathcal{A}_n$  as in (41), since  $\sup_{0 \leq m < \infty} \sup_{c_0 \leq c < \infty} |\hat{u}_c^{(m)}| \leq U_0$  by Theorem 2, we then have  $\sup_{c_0 \leq c < \infty} |\hat{\Delta}_c^{(m+1)}| < \omega^{m+1} (B_0^{-1} U_0/3 + U_0/3) + \delta/2 < \omega^{m+1} U_0 + \delta/2$ . As  $0 < \omega < 1$ ,  $\omega^{m+1}$  declines exponentially so  $m_0$  can be chosen sufficiently large that for all  $m > m_0$  then  $\omega^{m+1} U_0 < \delta/2$ . Thus  $P(\sup_{c_0 \leq c < \infty} |\hat{\Delta}_c^{(m+1)}| < \delta) > 1 - \varepsilon$  for all  $m > m_0$  and  $n > n_0$ .  $\square$

*Proof (Theorem 4).* Assumption 1(*ia*) implies  $E|\varepsilon_i/\sigma|^l < \infty$  for some  $l > 4$ . Apply (15) and the Chebyshev inequality to get  $\lambda/n = P(|\varepsilon_i| > \sigma c_n) \leq E|\varepsilon_i/\sigma|^l c_n^{-l}$ . Thus  $c_n \leq (E|\varepsilon_i/\sigma|^l)^{1/l} \lambda^{-1/l} n^{1/l}$  so that the divergence rate of  $c_n$  is  $O(n^{1/l}) = o(n^{1/4})$ .

1. *A bound on the sample space.* By Assumption 1(*ia, ii, iii*) with  $\eta = 1/4$ , Theorem 2 and 3 show that  $\widehat{\beta}_{c_n}^{(m)}$ ,  $(\widehat{\sigma}_{c_n}^{(m)})^2$  are tight. Assumption 1(*iib*) gives  $\max_{1 \leq i \leq n} |x_{in}| = O_P(n^{\kappa-1/2}) = o_P(n^{-1/4})$  for some  $0 \leq \kappa < 1/4$ . Thus, for all  $\varepsilon > 0$  there exists a large constant  $A_0$  so that the set

$$\mathcal{B}_n = \left\{ \sup_{0 \leq m < \infty} (|\widehat{b}_{c_n}^{(m)}| + |\widehat{a}_{c_n}^{(m)}|) + n^{1/4} \max_{1 \leq i \leq n} |x_{in}| \leq A_0 \right\}$$

has the probability larger than  $1 - \varepsilon$  for all  $n$ . Note that  $\widehat{b}_{c_n}^{(m)} = N^{-1}(\widehat{\beta}_{c_n}^{(m)} - \beta)$  and  $\widehat{a}_{c_n}^{(m)} = n^{1/2}(\widehat{\sigma}_{c_n}^{(m)} - \sigma)$ .

2. *Bound the indicator.* Define the random quantity,

$$s_{i,c_n}^{(m)} = \widehat{\sigma}_{c_n}^{(m)} c_n - y_i + x_i' \widehat{\beta}_{c_n}^{(m)} + \varepsilon_i = \sigma c_n + n^{-1/2} \widehat{a}_{c_n}^{(m)} c_n + x_i' \widehat{b}_{c_n}^{(m)}.$$

On the set  $\mathcal{B}_n$  and as  $c_n = o(n^{1/4})$ , we have for some  $A_1 > 0$

$$\begin{aligned} s_{i,c_n}^{(m)} &\leq \sigma c_n + n^{-1/2} A_0 c_n + n^{-1/4} A_0^2 \leq \sigma(c_n + n^{-1/4} A_1), \\ s_{i,c_n}^{(m)} &\geq \sigma c_n - n^{-1/2} A_0 c_n - n^{-1/4} A_0^2 \geq \sigma(c_n - n^{-1/4} A_1). \end{aligned}$$

Since the sets  $(y_i - x_i' \widehat{\beta}_{c_n}^{(m)} > \widehat{\sigma}_{c_n}^{(m)} c_n)$  and  $(\varepsilon_i > s_{i,c_n}^{(m)})$  are equal, we find

$$\mathbf{1}_{(\varepsilon_i/\sigma > c_n + n^{-1/4} A_1)} \leq \mathbf{1}_{(y_i - x_i' \widehat{\beta}_{c_n}^{(m)} > \widehat{\sigma}_{c_n}^{(m)} c_n)} \leq \mathbf{1}_{(\varepsilon_i/\sigma > c_n - n^{-1/4} A_1)}.$$

A similar argument shows

$$\mathbf{1}_{(\varepsilon_i/\sigma < -c_n - n^{-1/4} A_1)} \leq \mathbf{1}_{(y_i - x_i' \widehat{\beta}_{c_n}^{(m)} < -\widehat{\sigma}_{c_n}^{(m)} c_n)} \leq \mathbf{1}_{(\varepsilon_i/\sigma < -c_n + n^{-1/4} A_1)}.$$

Thus, we get the lower and upper bound for indicators uniformly in iteration  $m$  so

$$\mathbf{1}_{(|\varepsilon_i/\sigma| > c_n + n^{-1/4} A_1)} \leq \mathbf{1}_{(|y_i - x_i' \widehat{\beta}_{c_n}^{(m)}| > \widehat{\sigma}_{c_n}^{(m)} c_n)} \leq \mathbf{1}_{(|\varepsilon_i/\sigma| > c_n - n^{-1/4} A_1)}. \quad (43)$$

3. *Expectation of indicator bounds.* The aim is to prove

$$nE\mathbf{1}_{(|\varepsilon_i/\sigma| > c_n + n^{-1/4} A_1)} \rightarrow \lambda, \quad nE\mathbf{1}_{(|\varepsilon_i/\sigma| > c_n - n^{-1/4} A_1)} \rightarrow \lambda. \quad (44)$$

Since  $nE\mathbf{1}_{(|\varepsilon_i/\sigma| > c_n)} = \lambda$  by (15), it suffices to show

$$\mathcal{E}_n = nE\{\mathbf{1}_{(|\varepsilon_i/\sigma| > c_n - n^{-1/4} A_1)} - \mathbf{1}_{(|\varepsilon_i/\sigma| > c_n + n^{-1/4} A_1)}\} \rightarrow 0.$$

Note  $|\varepsilon_i/\sigma| \sim g, G$  and  $g = 2f, G = 2F - 1$ . By this and (15),  $2\{1 - F(c_n)\} = \lambda/n$ . Write  $\mathcal{E}_n$  as integral, apply the mean value theorem and use the above identity to get

$$\mathcal{E}_n = n \int_{c_n - n^{-1/4}A_1}^{c_n + n^{-1/4}A_1} 2f(u)du = 4nn^{-1/4}A_1 f(\tilde{c}) = \frac{4\lambda n^{-1/4}A_1 f(\tilde{c})}{2\{1 - F(c_n)\}},$$

where  $|\tilde{c} - c_n| \leq n^{-1/4}A_1$ . Then, we find

$$\mathcal{E}_n = 2\lambda A_1 \frac{f(\tilde{c})}{f(c_n - n^{-1/4}A_1)} \frac{f(c_n - n^{-1/4}A_1)}{f(c_n)} \frac{f(c_n)}{c_n\{1 - F(c_n)\}} n^{-1/4}c_n.$$

Since  $c_n - n^{-1/4}A_1 \leq \tilde{c}$  and  $f$  has the decreasing tail by Assumption 1(*ia*), the first ratio is bounded by 1. Since  $c_n = o(n^{1/4})$ , Assumption 1(*ib, ic*) shows the second and third ratio are bounded. Then use  $n^{-1/4}c_n = o(1)$  to get  $\mathcal{E}_n = o(1)$ .

4. *Poisson approximation.* On the set  $\mathcal{B}_n$ , apply (43) to obtain

$$\sum_{i=1}^n 1_{(|\varepsilon_i/\sigma| > c_n + n^{-1/4}A_1)} \leq \sum_{i=1}^n 1_{(|y_i - x_i \hat{\beta}_{c_n}^{(m)}| > \hat{\sigma}_{c_n}^{(m)} c_n)} \leq \sum_{i=1}^n 1_{(|\varepsilon_i/\sigma| > c_n - n^{-1/4}A_1)}.$$

Using (44), the Poisson limit theorem shows that the lower and upper bound have the Poisson limit with mean  $\lambda$ . By (16),  $n\hat{\gamma}_{c_n}^{(m)} \xrightarrow{D} \text{Poisson}(\lambda)$  for all  $0 \leq m < \infty$ .  $\square$

## References

1. Bercu, B. and Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848-1869.
2. Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428-434.
3. Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.
4. Cavaliere, G. and Georgiev, I. (2013). Exploiting infinite variance through dummy variables in nonstationary autoregression. *Econometric Theory* 29, 1162-1195.
5. Dollinger, M.B. and Staudte, R.G. (1991). Influence functions of iteratively re-weighted least squares estimators. *Journal of the American Statistical Association* 86, 709-716.
6. Doornik, J.A. (2009). Autometrics. In Castle, J.L. and Shephard, N. (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honor of David F. Hendry*, pp. 88-121. Oxford: Oxford University Press.
7. Freedman, D.A. (1975). On tail probabilities for martingales. *Annals of Probability* 3, 100-118.
8. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics*. New York: Wiley.
9. Hendry, D.F. (1999). An econometric analysis of US food expenditure, 1931-1989. In Magnus, J.R. and Morgan, M.S. (eds.) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, pp. 341-361. New York: Wiley.
10. Hendry, D.F. and Doornik, J.A. (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge MA: MIT Press.
11. Hendry, D.F., Johansen, S. and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23, 317-335. Erratum 337-339.
12. Hendry, D.F. and Santos, C. (2010). An automatic test of super exogeneity. In Bollerslev, T., Russell, J.R. and Watson, M.W. (eds.) *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pp. 164-193. Oxford: Oxford University Press.
13. Hoover, K.D. and Perez, S.J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal* 2, 167-191.

14. Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.
15. 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 Honor of David F. Hendry*, pp. 1-36. Oxford: Oxford University Press.
16. Johansen, S. and Nielsen, B. (2013). Asymptotic theory for iterated one-step Huber-skip estimators. *Econometrics* 1, 53-70.
17. Johansen, S. and Nielsen, B. (2016a). Analysis of the Forward Search using some new results for martingales and empirical processes. *Bernoulli* 22, 1131-1183.
18. 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.
19. Jurečková, J., Sen, P.K. and Picek, J. (2013). *Methodology in Robust and Nonparametric Statistics*. Boca Raton, FL: CRC Press.
20. Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edition. New York: Springer.
21. Koul, H.L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994-1012.
22. Koul, H.L. and Ossander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540-562.
23. Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828-838.
24. Varga, R.S. (2000). *Matrix Iterative Analysis*. Berlin: Springer.
25. 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.