Asymptotic Analysis of Outliers Detection Algorithms Using A New Empirical Process Result

Xiyu Jiao

Department of Economics, University of Oxford

Supervised by

Professor Bent Nielsen

Department of Economics, University of Oxford

Submitted for The Degree of

Master of Philosophy in Economics

May 11, 2015

Abstract

The Robustified Least Squares and the Impulse Indicator Saturation are iterative algorithms concerned with detecting outliers and other unsuspected structures in data. These two algorithms are constructed when the full or split sample least squares are chosen as the initial estimators for the iterated 1-step Huber-skip M-estimator. These methods classify observations as outliers or not. This paper establishes an asymptotic theory considering the role of the varying cut-off in such algorithms. In particular, we demonstrate the tightness property for estimates in each iterated step with the varying cut-off. Moreover, a Poisson exceedence theory to the gauge is built for all iterations. The argument involves a theory for a new class of the weighted and marked empirical process.

Keywords: Outliers robustness, the iterated 1-step Huber-skip M-estimator, Robustified Least Squares, Impulse Indicator Saturation, tightness, a fixed point, Poisson approximation to gauge, weighted and marked empirical processes.

Word count: (No. of words in a typical page, page 2) $\times$ (No. of pages) $= 384 \times 52 = 19968$.

Acknowledgement: The author would like to thank Bent Nielsen for his supervision, patient, suggestions and guidance. Without his time and effort devoted to my supervision, this paper cannot be written.

Contents

1 Introduction **1**

2 Model and outliers detection algorithms **3**

2.1 Model 4

2.2 Robust methods 5

2.3 Some statistical algorithms 8

3 A new result on iterated 1-step M-estimators **10**

3.1 Assumptions 10

3.2 A new tightness result 12

4 Gauge as a measure of false detection **14**

4.1 Background 14

4.2 Poisson approximation to gauge for 1-step M-estimators 15

4.3 A new Poisson approximation result 16

5 Absolute weighted and marked empirical process **18**

5.1 Background 18

5.2 A new empirical process result 20

5.3 Asymptotic expansion of empirical processes 21

6 Auxiliary weighted and marked empirical process **23**

6.1 Assumptions 24

6.2 Some existing empirical process results 25

6.3 A new empirical process result 26

7 Conclusion **27**

A Some martingale inequalities **28**

B A metric on $\mathbb{R}$ and some inequalities **30**

C Proofs of Theorem 6.3, 6.4 and 6.5 in §6 **34**

D Proofs in §3, §4 and §5 **41**

D.1 Proofs of Theorem 5.1, 5.2 and 5.3 in §5 42

D.2 Proofs of Theorem 3.1 and 3.2 in §3 44

D.3 Proofs of Theorem 4.2 and 4.3 in §4 49

References **50**

1 Introduction

Since the Lucas critique, researchers have recognized the structural instability due to outliers and other unsuspected structural changes in data. Thus, outliers are essential for building a robust econometric model for policy analysis. Moreover, we can use outliers as impulses or shocks to discover the causal structure in the economy. Due to an importance of detecting and treating outliers, many robust methods have been developed over past decades. These include the Huber-skip M-estimator, Least Trimmed Squares and 1-step versions. The iterated 1-step estimators formulate outliers detection algorithms such as the Impulse Indicator Saturation and the Forward Search. All these methods throw away observations detected as outliers according to binary decision rules. In this paper, we focus on the Robustified Least Squares and the Impulse Indicator Saturation, which are in fact iterated 1-step Huber-skip M-estimators with the full or split sample least squares as the initial estimators. The idea of the Impulse Indicator Saturation was introduced in the empirical work of US food expenditure by Hendry (1999). A formal version was then suggested and investigated by Hendry, Johansen and Santos (2008). Hendry and Doornik (2014) provided a comprehensive study for the Impulse Indicator Saturation as a special case of the model selection algorithm Autometrics.

With the purpose of evaluating outliers detection algorithms, the notion of a gauge is introduced as the expected frequency of falsely detected outliers when there is no data contamination in the model. The gauge is a similar but different concept with respect to the size of a statistical test due to estimation uncertainties. For hypothesis testing, we can choose the critical value indirectly from the size we are willing to endure. For outliers detection algorithms, the gauge can also help us select the cut-off value in the binary decision rule. The idea of a gauge originated in a simulation study testing properties of a general-to-specific model selection algorithm by Hoover and Perez (1999). They started to consider the gauge as one of various measures assessing the performance of such algorithms. Then, Hendry and Santos (2010) formally introduced the concept of a gauge. Together with Doornik (2009) and Hendry and Doornik (2014), they carried out a large number of simulation studies to analyze the gauge for irrelevant regressors and irrelevant outliers.

For the Impulse Indicator Saturation, Hendry, Johansen and Santos (2008) started to build an asymptotic distribution theory in the location model. Johansen and Nielsen (2009) extended an asymptotic theory to the multiple regression where regressors are allowed to be stationary, trend stationary and unit root processes. In contrast to the traditional approach of influence function in robust statistics, they constructed an empirical process theory to describe the least squares statistics in each step of the algorithm. Then, Johansen and Nielsen (2013) demonstrated a fixed point theory to investigate iterations involved in the algorithm. Based on the tightness result obtained, they showed the iterated 1-step M-estimator converges upon through infinite iterations. Johansen and Nielsen (2014c) established an asymptotic normal distribution to the gauge for various Huber-skip M-estimators. All these asymptotic analysis are restricted to the situation where the cut-off is held fixed. In addition, their derivations are conducted under the hypothesis that the statistical model contains no outliers.

All asymptotic theory reviewed was established when the cut-off is fixed and taken as given. 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 property for estimates in each iterated step with the varying cut-off. Moreover, this paper demonstrates an asymptotic Poisson distribution to the gauge in the situation where the varying cut-off is set to fix the number of wrongly discovered outliers regardless of the sample size. The proposal then is to control the cut-off from the gauge we are willing to tolerate. Briefly speaking, the argument involves a theory for a new class of the weighted and marked empirical process.

At first, this paper establishes an asymptotic theory to investigate the role of the varying cut-off in the iterated 1-step Huber-skip M-estimator. To be specific, we show iterated 1-step estimates are tight for all iterations as long as the cut-off value is selected rather large. The proof requires asymptotic expansions for a new class of empirical processes, a mathematical induction and a geometric sequence argument. Our new tightness result extends Johansen and Nielsen (2013) to a more general case where variation in the cut-off is considered. The tightness result is essential for proving the fixed point theory and Poisson approximations to the gauge.

Furthermore, a contribution of this paper is to prove a Poisson distribution theory to the gauge for any iteration of the iterated 1-step Huber-skip M-estimator using the new tightness result obtained. This arises in the situation where we fix the number of wrongly classified outliers by allowing variation in the cut-off as the sample size increases. Our result complements Johansen and Nielsen (2014c) who established normal approximations to the gauge when the proportion of falsely discovered outliers is controlled by fixing the cut-off as the sample size rises.

In order to carry out asymptotic analysis, we construct a theory for a new class of the weighted and marked empirical process. In the literature, weights represent functions of regressors and marks are functions of regression errors. The result in this paper generalizes those of Koul and Ossiander (1994) who did not allow the non-bounded marks and divergence of the normalized estimation errors. Moreover, we take into account both multiplicative and additive shifts due to all possible estimation uncertainties along with the varying quantile. This in turn extends the result in Johansen and Nielsen (2009, 2014a) who conducted derivations with the fixed quantile in 2009 and who did not consider the multiplicative estimation error for variance in 2014.

The outline of this paper is the following. We first review the model, robust methods and outliers detection algorithms in §2. Then, this paper presents assumptions and a new result on iterated 1-step Huber-skip M-estimators in §3. A Poisson exceedence theory to gauge is given in §4. §5 provides the result of the absolute weighted and marked empirical process. The one-sided empirical process result is given in §6 with proofs presented in the appendix §A, B, C. Proofs of main theorems in §3, §4 and §5 follow in the appendix §D.

2 Model and outliers detection algorithms

The regression model with some notations is described first. It is widely known that least squares estimators are sensitive to outliers. We review a number of robust methods which have a common property that outlying observations are dropped. Some statistical algorithms are listed including the Robustified Least Squares and the Impulse Indicator Saturation.

2.1 Model

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

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

This setting can represent both classical regression and time series models. Moreover, regressors x_i can be a deterministic or stochastic trend. The errors ε_i are independent with the filtration $\mathcal{F}_{i-1}$ generated by $(x_1, \ldots, x_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$. Furthermore, innovations ε_i are identically distributed with scale σ so that ε_i/σ has a known density f and a distribution function $F(c) = \mathrm{P}(\varepsilon_i/\sigma \leq c)$. In practice, the distribution f, F will often be standard normal or at least symmetric. When discuss the one-sided weighted and marked empirical process, a more general distribution is allowed.

In the statistics literature, the outliers are defined as pairs of observations (y_i, x_i) which do not conform with the regression model (2.1). An outlying observation is detected if the scaled error ε_i/σ does not conform with a reference density f. The different explanations of outliers arise for cross-sectional and time series data. As the nature of cross-sectional data that pairs of sample (y_i, x_i), $i = 1, 2, \ldots, n$ are unrelated, one observation (y_i, x_i) will be dropped if the scaled innovation ε_i/σ is classified as an outlier. This can be interpreted as an adjustment error ε_i not conforming with the density f. In other words, (y_i, x_i) are not correct with respect to the statistical model (2.1). A consequence of finding an outlier is distinct for time series, where lagged dependent variables are included in the regressors x_i. The simplest example is $x_i = y_{i-1}$ for a first order auto-regression model. Two distinguished concepts of outliers arise in this context. These are the so-called innovative outliers and additive outliers. An innovative outlier occurs when a single ε_i is classified as an outlying observation. In this situation, the evaluation of the dynamic model from y_{i-1} to y_i is discarded but observations y_{i-1} and y_i are kept. This is because y_{i-1} appears as the dependent variable at time $i-1$ and y_i as the independent variable at time $i+1$. Therefore, an innovative outlier means a pair of observations is correct but not generated by the model. An additive outlier happens if two errors ε_i and ε_{i+1} are captured as outliers. In this

case, y_i is wrongly measured and dropped by discarding evaluation of dynamics with observations (y_i, y_{i-1}) and (y_{i+1}, y_i).

Outliers detection algorithms use absolute residuals and then calculate robust least squares estimators from the non-outlying sample. This implicitly assumes a symmetric density. Thus, a non-symmetric density in practice will lead to biasness in the estimation procedure. It is necessary to correct the bias when applying robust methods.

Suppose the absolute error $|\varepsilon_i|/\sigma$ has a density g and a distribution function $G(c) = \mathrm{P}(|\varepsilon_i|/\sigma \leq c)$ for $c > 0$. With a symmetric assumption, $G(c) = F(c) - F(-c) = F(c) - (1 - F(c)) = 2F(c) - 1$ and $g(c) = 2f(c)$. Define $\psi = G(c_\psi)$ so that c_ψ is the ψ quantile such that $c_\psi = G^{-1}(\psi) = F^{-1}\{(\psi + 1)/2\}$. The probability of exceeding the cut-off is defined as $\gamma = 1 - \psi$. It is convenient to introduce moments and truncated moments

$$\tau_n = \int_{-\infty}^{\infty} u^n f(u)du, \qquad \tau_n^c = \int_{-c}^{c} u^n f(u)du. \tag{2.2}$$

Thus, $\tau_0^c = \psi$, $\tau_2 = 1$ and $\tau_n = \tau_n^c = 0$ for an odd n due to a symmetric density. Define also the conditional variance of ε_i/σ given $\{|\varepsilon_i|/\sigma \leq c\}$ as

$$\varsigma^2 = \frac{\tau_2^c}{\psi} = \frac{\int_{-c}^{c} u^2 f(u)du}{\mathrm{P}(|\varepsilon_i| \leq \sigma c)}. \tag{2.3}$$

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 in most cases, $\tau_2^c = \psi - 2cf(c)$, $\tau_4^c = 3\psi - 2c(c^2 + 3)f(c)$ and $\tau_4 = 3$.

2.2 Robust methods

The M-estimator was first proposed by Huber (1964) as a class of maximum likelihood estimators in an univariate setting. For the regression model (2.1), the M-estimator is the minimizer of

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \rho(y_i - x_i'\beta). \tag{2.4}$$

ρ is the absolutely continuous and non-negative criterion function. For instance, the least squares estimator can be obtained when $\rho(u) = u^2$. The median estimator is to set $\rho(u) = |u|$. The Huber-skip M-estimator is introduced through the non-convex criterion function $\rho(u) = u^2 1_{(|u| \leq \sigma c)} + c^2 1_{(|u| > \sigma c)}$ for some cut-off c and a known scale σ.

In the context of Huber-skip estimators, we have binary stochastic weights v_i which can be defined in many different ways for each observation. The objective function then is

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\{(y_i - x_i'\beta)^2 v_i + c^2(1 - v_i)\}. \tag{2.5}$$

The idea here is to pursue hard rejection of outliers. The Huber-skip M-estimator is to set a stochastic weight

$$v_i = 1_{(|y_i - x_i'\beta| \le \sigma c)}, \tag{2.6}$$

which depends on a cut-off point c and β. Note that σ is known for this estimate. Rousseeuw (1984) introduced the Least Trimmed Squares estimator where

$$v_i = 1_{(|y_i - x_i'\beta| \le \xi_{(k)})}. \tag{2.7}$$

$\xi_{(k)}$ is the k-th smallest order statistic of absolute residuals $\xi_i = |y_i - x_i'\beta|$ for $i = 1, 2, \ldots, n$. We can choose an integer $k \le n$ so that $\xi_{(k)} = \sigma G^{-1}(k/n)$ where σ is known. Notice that ψ, c and k are different ways of calibrating the robust methods, since, given an integer $k \le n$, ψ and c can be found such that $k/n = \psi = G(c)$. The regression estimator $\widehat{\beta}$ can be computed as a minimizer of $R_n(\beta)$ in (2.5) with the weight (2.6) or (2.7). Then the variance estimate can be calculated by

$$\widehat{\sigma}^2 = \varsigma^{-2}(\sum_{i=1}^{n} v_i)^{-1}\{\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta})^2 v_i\}. \tag{2.8}$$

Remember $\varsigma^2 = \tau_2^c/\psi$ is the bias correction factor defined in (2.3) for consistency.

For the Least Trimmed Squares estimator, $\sum_{i=1}^{n}(1 - v_i) = n - k$. The second term in the objective function (2.5) therefore does not depend on β. Thus, $\widehat{\beta}$ is the minimizer of the objective function

$$R_{n,LTS}(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 v_i. \tag{2.9}$$

In contrast to the Huber-skip M-indicator (2.6), the least trimmed weight (2.7) does not depend on scale σ. In practice, a binomial search algorithm is used to implement the LTS estimator. But Maronna, Martin and Yohai (2006) argued such an algorithm is uncomputable in most situations. Rousseeuw and Driessen (1998) studied and suggested several iterative approximations such as the Fast LTS algorithm in order to obtain the

estimate. Rousseeuw and Leroy (1987) discussed a breakdown point $\gamma = 1-\psi = 1-k/n$ for $\psi < 1/2$ for the LTS estimator.

The asymptotic theory of the robust methods in the literature has been developed in great detail in the situation where there are no outliers presented in the model. Huber (1964) established a theory of M-estimators for convex criterion functions ρ in a location model. Chen and Wu (1988) demonstrated strong consistency of M-estimators for general criterion functions ρ in an i.i.d. setting. Then, Johansen and Nielsen (2014b) extended the theory of M-estimators in the time series regression. The idea is to first prove the tightness of the estimator $\widehat{\beta}$ through a geometric sequence argument which requires the assumption to the frequency of small regressors. Then it is shown consistency and asymptotic expansion using the iterated martingale inequality. An asymptotic theory of the Least Trimmed Squares estimator is provided by Víšek (2006a, b, c).

Subsequently, consider some estimators which are easier to compute while preserve the robustness property. In the case where the stochastic weight does not depend on β, the least squares estimator can be written as

$$\widehat{\beta} = (\sum_{i=1}^{n} x_i x_i' v_i)^{-1} (\sum_{i=1}^{n} x_i y_i v_i). \tag{2.10}$$

Then the variance can be estimated as $\widehat{\sigma}^2$ in (2.8). The so-called 1-step Huber-skip M-estimator is based on the indicator

$$v_i = 1_{(|y_i - x_i'\widetilde{\beta}| \leq \widetilde{\sigma} c)}, \tag{2.11}$$

where $\widetilde{\beta}$ and $\widetilde{\sigma}$ are initial estimators. Similar, introduce the 1-step Huber-skip L-estimator depended on the weight

$$v_i = 1_{(|y_i - x_i'\widetilde{\beta}| \leq \widetilde{\xi}_{(k)})}, \tag{2.12}$$

where $\widetilde{\xi}_{(k)}$ is the k-th smallest order statistic of absolute residuals $\widetilde{\xi}_i = |y_i - x_i'\widetilde{\beta}|$. Notice that $k < n$ is the cut-off point. The corresponding iterated versions of these two estimators formulate outliers detection algorithms such as the Impulse Indicator Saturation and the Forward Search.

1-step Huber-skip estimators of M and L-types are computational attractive, however good starting initial estimators are required. As before, the 1-step M-estimate needs the

initial scale estimate in contrast to the L-type. Robustified least squares estimators can be obtained if $\widetilde{\beta}$ and $\widetilde{\sigma}$ are the full sample least squares estimators. Welsh and Ronchetti (2002) showed Robustified Least Squares can perform badly in high leverage situations.

Johansen and Nielsen (2009, 2014a) established an asymptotic theory of 1-step Huber-skip M and L-estimators. An asymptotic expansion involves the weighted and marked empirical process. The idea of the proof includes a chaining argument and the iterated martingale inequality. The M-type estimate requires a multiplicative shift inequality whereas an additive shift should be considered for the L-estimator.

2.3 Some statistical algorithms

1-step Huber-skip M-estimators can also be iterated. The iterated versions of 1-step Huber-skip estimates formulate some statistical outliers detection algorithms. These are the Iterated 1-step Huber-skip M-estimator, the Impulse Indicator Saturation. We first consider the Iterated 1-step Huber-skip M-estimator.

Algorithm 2.1. ***Iterated 1-step Huber-skip M-estimator****. Choose a cut-off* $c > 0$*.*
1. Choose initial estimators $\widehat{\beta}^{(0)}$*,* $(\widehat{\sigma}^{(0)})^2$ *and let* $m = 0$*.*
2. Define indicator variables $v_i^{(m)}$ *as in* (2.11)*, replacing* $\widetilde{\beta}$*,* $\widetilde{\sigma}^2$ *by* $\widehat{\beta}^{(m)}$*,* $(\widehat{\sigma}^{(m)})^2$*.*
3. Compute least squares estimators $\widehat{\beta}^{(m+1)}$*,* $(\widehat{\sigma}^{(m+1)})^2$ *as in* (2.10)*,* (2.8) *replacing* v_i *by* $v_i^{(m)}$*.*
4. Let $m = m + 1$ *and repeat 2 and 3.*

The idea of the Impulse Indicator Saturation Algorithm was initially applied by Hendry (1999) in his empirical work. Then the algorithm was studied comprehensively by Hendry, Johansen and Santos (2008). Also see Hendry and Doornik (2014). The IIS algorithm is to divide full sample into two sub-samples. It then use regression estimate calculated from each sub-sample to detect outliers in the other sub-sample. The IIS is still a version of the iterated 1-step M-estimators, however a different starting estimate is used comparing to the Robustified Least Squares.

Algorithm 2.2. ***Impulse Indicator Saturation****. Choose a cut-off* $c > 0$*.*
1.1. Split indices in sets $\mathcal{I}_j$ *for* $j = 1, 2$ *of* n_j *observations.*

1.2. Calculate the least squares estimators for (β, σ^2) *based upon sample* $\mathcal{I}_j$ *as*

$$\widehat{\beta}_j = (\sum_{i \in \mathcal{I}_j} x_i x_i')^{-1} (\sum_{i \in \mathcal{I}_j} x_i y_i), \qquad \widehat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i \in \mathcal{I}_j} (y_i - x_i' \widehat{\beta}_j)^2.$$

1.3. Define indicator variables for each observation

$$v_i^{(-1)} = 1_{(i \in \mathcal{I}_1)} 1_{(|y_i - x_i' \widehat{\beta}_2| \leq \widehat{\sigma}_2 c)} + 1_{(i \in \mathcal{I}_2)} 1_{(|y_i - x_i' \widehat{\beta}_1| \leq \widehat{\sigma}_1 c)}. \tag{2.13}$$

1.4. Compute least squares estimators $\widehat{\beta}^{(0)}$, $(\widehat{\sigma}^{(0)})^2$ *using* (2.10), (2.8), *replacing* v_i *by* $v_i^{(-1)}$ *and let* $m = 0$.

2. Define indicator variables $v_i^{(m)}$ *as in* (2.11), *replacing* $\widetilde{\beta}$, $\widetilde{\sigma}^2$ *by* $\widehat{\beta}^{(m)}$, $(\widehat{\sigma}^{(m)})^2$.

3. Compute least squares estimators $\widehat{\beta}^{(m+1)}$, $(\widehat{\sigma}^{(m+1)})^2$ *as in* (2.10), (2.8) *replacing* v_i *by* $v_i^{(m)}$.

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

The Impulse Indicator Saturation 2.2 is possibly more robust than the Robustified Least Squares 2.1 when we have prior knowledge that the outliers are located in a particular subset of whole sample. Doornik (2009) argued the choice of the initial sets $\mathcal{I}_1$ and $\mathcal{I}_2$ should be iterated since the location of contaminated observations is unknown in most practical situations.

Algorithm 2.1, 2.2 will iterate without a stopping rule. The effect of starting points will be lost through iterations. Thus, the essential issue is whether such algorithms converge to a fixed point which approximates the Huber-skip M-estimate when m and n increase. With the purpose of solving this problem, Johansen and Nielsen (2013) showed a fixed point result for infinite iterations. The significant step of their proof is to demonstrate the tightness of a sequence of the normalised estimators. The reason why it holds is that the asymptotic expansions are contract in each iterated step.

An interest in the literature investigates whether an initial estimator with a slow convergence rate can be refined upon through iteration. He and Portnoy (1992) showed a single step cannot improve the convergence rate for the Reweighted Least Squares estimators. However, the asymptotics of the Impulse Indicator Saturation and the Forward Search demonstrated a slow rate can be improved through iterations, see Johansen and Nielsen (2013, 2014a).

3 A new result on iterated 1-step M-estimators

At first, we list and discuss some assumptions made on the density f, regressors x_i, errors ε_i and the initial estimators $\widetilde{\beta}$, $\widetilde{\sigma}^2$. Then, this paper presents a new tightness result which permits variation in the cut-off c for the iterated 1-step Huber-skip M-estimator. This is an extension of Johansen and Nielsen (2013) who derived the result under the situation where the cut-off value c is fixed and taken as given.

3.1 Assumptions

We list the sufficient assumptions for an asymptotic theory of outliers detection algorithms. These assumptions are somewhat stronger than they need to be. In the section of the one-sided weighted and marked empirical processes, we will introduce some weaker and explicit assumptions, which are just required to the proof. Notice that here assumes the density f is symmetric, although it needs not necessarily to be in the analysis of the one-sided empirical processes.

The innovations ε_i and regressors x_i must satisfy some moment conditions so as to carry out asymptotic analysis. The regressors x_i can be temporally dependent and trended deterministically or stochastically. N is a normalisation matrix for x_i such that $N = n^{-1/2} I_{\dim(x)}$ for stationarity and $N = n^{-1} I_{\dim(x)}$ for unit roots. The minimal assumptions vary for different robust methods and their formulated outliers detection algorithms. We make the following assumptions in this paper.

Assumption 3.1. *Let $\mathcal{F}_i$ be an increasing sequence of σ-fields so ε_{i-1} and x_i are $\mathcal{F}_{i-1}$-measurable and ε_i is independent with $\mathcal{F}_{i-1}$. ε_i/σ has a symmetric, continuous, differentiable density f which is positive for $F^{-1}(0) < c < F^{-1}(1)$. For some $0 \leq \kappa < \eta \leq 1/4$ choose an $r \geq 2$ so $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$. Let $q_0 = 1 + 2^{r+1}$. Suppose*
(i) density satisfies

(a) $c^q f(c)$, $|c^{q-1} f'(c)|$ are decreasing for large c and some $q > q_0$;

(b) $f(c)/\{c[1 - F(c)]\} = \mathrm{O}(1)$ for $c \to \infty$;

(c) $f(c - n^{-1/4}A)/f(c) = \mathrm{O}(1)$ for all $A > 0$ and $c \to \infty$;

(d) $\sup_{c_0 \leq c < \infty} |2cf(c)/\psi| < 1$ and $\sup_{c_0 \leq c < \infty} |c(c^2 - \varsigma^2) f(c)/\tau_2^c| < 1$ for all $c_0 > 0$;

(*ii*) *regressors* x_i *are* $\mathcal{F}_{i-1}$*-measurable and satisfy*

(*a*) $V_n = N' \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{D} V$;

(*b*) $\Sigma_n = N' \sum_{i=1}^n x_i x_i' N \xrightarrow{D} \Sigma \overset{a.s.}{>} 0$;

(*c*) $\mu_n = n^{-1/2} N' \sum_{i=1}^n x_i \xrightarrow{D} \mu$;

(*d*) $\max_{1 \le i \le n} |n^{1/2-\kappa} N' x_i| = \mathrm{O_p}(1)$ *for some* $\kappa < \eta$;

(*e*) $n^{-1} \mathrm{E} \sum_{i=1}^n |n^{1/2} N' x_i|^{q_0} = \mathrm{O}(1)$;

(*iii*) *initial estimators satisfy*

(*a*) $N^{-1}(\widetilde{\beta} - \beta) = \mathrm{O_p}(n^{1/4-\eta})$ *for some* $\eta > 0$;

(*b*) $n^{1/2}(\widetilde{\sigma}^2 - \sigma^2) = \mathrm{O_p}(n^{1/4-\eta})$ *for some* $\eta > 0$.

At first, it is necessary to explain the meaning of κ, η, r and q_0. κ represents the order of magnitude of maximum of the normalized regressors, and η, such that $0 \le \kappa < \eta \le 1/4$, indicates the convergence rate of the initial estimators. $r \ge 2$ so $2^{r-1} \ge 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$ is defined to control the number of moments and the smoothness of the density f. Then, $q_0 = 1 + 2^{r+1}$ in condition (ia) can ensure the moments, boundedness and smoothness conditions required for f in order to prove the empirical process results.

Johansen and Nielsen (2013, 2014a, c) demonstrated density conditions (*i*) are satisfied in many situations, such as standard normal and student-t distributions. These general density assumptions allow us to choose a more flexible density with heavier tails, even the normal distribution is used in practice. Moreover, note that the tail monotonicity condition (*ia*) is more severe than that in the empirical process literature due to non-bounded marks ε_i^p. This is the main density condition which will be used to prove the empirical process results. Assumptions (ib, ic) are essential for building the Poisson approximation to the gauge in the situation where the number of outliers is fixed. The tightness and fixed point argument requires the density condition (*id*). (iia, iib, iic) are standard in regression analysis. Johansen and Nielsen (2014a) discussed (iid, iie) for stationary, deterministic and random walk regressors. Condition (*iii*) is a weaker assumption than normally seen, which allows the standardised estimation errors diverge at a rate of $n^{1/4-\eta}$ rather than being bounded. In particular, $\eta = 1/4$ can be chosen to guarantee the tightness assumption of initial estimators.

3.2 A new tightness result

Johansen and Nielsen (2009) started to derive an asymptotic theory for the 1-step Huber-skip M-estimator using the empirical process technique. Then, the fixed point result was established by Johansen and Nielsen (2013). This is to demonstrate the iterated 1-step Huber-skip M-estimator converges upon through infinite iterations. An essential step of their argument is to show tightness of the normalized estimates for all iterations. However, their research is restricted to the case where the cut-off value c is taken as given and fixed. In other words, Johansen and Nielsen (2009, 2013) did not build a formal theory exploring the role of the varying cut-off c in the algorithm. The purpose of this paper is to construct an asymptotic theory which can investigate how variation in the c affects the iterated 1-step M-estimator. In particular, we demonstrate a new tightness result considering the varying c for the Robustified Least Squares 2.1 and the Impulse Indicator Saturation 2.2.

For the iterated 1-step Huber-skip M-estimator, estimates $\widehat{\beta}^{(m)}$, $(\widehat{\sigma}^{(m)})^2$ and stochastic weights $v_i^{(m-1)}$ do depend on the cut-off c in each iterated step m. Thus, for any iteration m, we emphasize the subscript c in sequences of $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ and $v_{i,c}^{(m-1)}$. From (2.10), (2.8) and (2.11), we can define $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ and $v_{i,c}^{(m-1)}$ as

$$\widehat{\beta}_c^{(m)} = (\sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m-1)})^{-1} (\sum_{i=1}^{n} x_i y_i v_{i,c}^{(m-1)}), \tag{3.1}$$

$$(\widehat{\sigma}_c^{(m)})^2 = \varsigma_c^{-2} (\sum_{i=1}^{n} v_{i,c}^{(m-1)})^{-1} \{\sum_{i=1}^{n} (y_i - x_i' \widehat{\beta}_c^{(m)})^2 v_{i,c}^{(m-1)}\}, \tag{3.2}$$

$$v_{i,c}^{(m-1)} = 1_{(|y_i - x_i' \widehat{\beta}_c^{(m-1)}| \leq \widehat{\sigma}_c^{(m-1)} c)}. \tag{3.3}$$

Note $\widehat{\beta}_c^{(m-1)}$, $(\widehat{\sigma}_c^{(m-1)})^2$ are estimates in the previous step. $\varsigma_c^2 = \tau_2^c / \psi_c$ is the bias correction factor for the variance estimator as in (2.3).

Consider Algorithm 2.1, 2.2. This paper shows $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ are tight for all iterations m in a more general situation where the cut-off c can be varying in the interval $[c_0, \infty)$. $c_0 > 0$ is a finite positive number. This in turn extends Johansen and Nielsen (2013) who only proved the tightness result with the fixed c.

Theorem 3.1. *Consider the iterated Robustified Least Squares Algorithm 2.1. Suppose Assumption 3.1*(ia, id, ii) *holds. Then for* $n \to \infty$*, we have*

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

Theorem 3.2. *Consider the Impulse Indicator Saturation Algorithm 2.2. Suppose Assumption 3.1*(ia, id, ii) *holds for each sub-sample set* $\mathcal{I}_1$*,* $\mathcal{I}_2$*. Then for* $n \to \infty$*, we have*

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

For Algorithm 2.1, 2.2, the initial estimators are tight, since they are either the full sample or the split sample least squares which do not depend on the cut-off c. We will build asymptotic expansions for a new class of empirical processes which allow free variation in the quantile c and estimation uncertainties for β, σ. This can provide 1-step relationships between $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ and $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ which follow autoregressive structures. Then a mathematical induction argument can be used to show the tightness result for any finite iteration m and all $c \in [c_0, \infty)$. By Assumption 3.1(id) that $\sup_{c_0\leq c<\infty} |2cf(c)/\psi_c| < 1$, $\sup_{c_0\leq c<\infty} |c(c^2 - \varsigma_c^2)f(c)/\tau_2^c| < 1$, the stationarity property does hold for two autoregressive equations between $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ and $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$. Thus we can finish the proof for all iterations m and $c \in [c_0, \infty)$ by applying a geometric sequence argument.

For outliers detection algorithms, all asymptotic analysis is conducted under the hypothesis that there is no data contamination in the model. Thus, we consider the situation where the model contains no outliers. The iterated 1-step M-estimate will always throw away some observations wrongly detected as outliers, and thus has lower efficiency comparing to the full sample estimate. Although there is efficiency loss if apply Algorithm 2.1, 2.2, the tightness property of $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ can still be guaranteed for each step m as long as the cut-off value c is selected rather large. This is the main implication by Theorem 3.1, 3.2. In addition, this new tightness result is essential to construct a fixed point theory and Poisson approximations to the gauge for Algorithm 2.1, 2.2.

Proofs appear in the appendix. Before showing the formal argument, we need to first establish a theory for a new class of the empirical process.

4 Gauge as a measure of false detection

The background with some existing work is introduced initially. We then review some recent contributions in Johansen and Nielsen (2014c) who proved a Poisson exceedence theory to the gauge for the 1-step Huber-skip M-estimator. At last, based on the tightness result obtained, we can present a new Poisson approximation result which investigates iterations involved in the iterated 1-step Huber-skip M-estimator including the Robustified Least Squares and the Impulse Indicator Saturation.

4.1 Background

Outliers detection algorithms have a positive probability to find outliers even when, in fact, the data generation process has no outliers. False detection of outliers in general will lead to efficiency loss in the estimation procedure. Thus, the performance of such algorithms can be evaluated by the concept of a gauge, which is the expected retention rate of falsely discovered outliers. Moreover, the gauge is a related but different notion with respect to a size of a statistical test. For hypothesis testing, we choose the critical value indirectly from the size we are willing to tolerate. For outliers detection algorithms, the gauge can also help us to select the cut-off.

Hoover and Perez (1999) originally introduced the idea of a gauge when they studied properties of general-to-specific variable selection algorithms through simulation. Then the concept of a gauge was formally proposed by Hendry and Santos (2010). They defined the gauge as the expected retention rate of irrelevant regressors in the context of model selection algorithms. Hendry and Doornik (2014) presented a comprehensive study on the gauge for Autometrics and in particular the Impulse Indicator Saturation Algorithm.

The algorithms assign stochastic indicators v_i to all samples as in (2.11), (2.12). $v_i = 0$ represents the observation i is an outlier, otherwise $v_i = 1$. The sample gauge can be introduced as

$$\widehat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(1 - v_i). \tag{4.1}$$

The population gauge is then

$$\mathrm{E}\widehat{\gamma} = \mathrm{E}\{\frac{1}{n}\sum_{i=1}^{n}(1-v_i)\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(1-v_i). \tag{4.2}$$

These two concepts are defined in the situation where the model has no contamination. In general, the sample gauge is the retention rate of falsely discovered outlying observations. The population version arises when the expected value is taken.

For the Least Trimmed Squares and 1-step Huber-skip L-estimators, the gauge is trivially $\widehat{\gamma} = \gamma$. Recall $\gamma = 1 - \psi = \mathrm{P}(|\varepsilon_i|/\sigma > c)$ for $c > 0$. This is because L-type estimators are designed to keep the trimming proportion fixed. For the Forward Search, Johansen and Nielsen (2014c) derived a convergent distribution for the sample gauge, and then developed a stopping rule from this.

Johansen and Nielsen (2014c) established an asymptotic expansion for the sample gauge when they control the proportion of outliers by fixing the cut-off as the sample size increases. Then, a normal distribution theory follows for various Huber-skip M-estimators. Based on normal approximations, it suggests that the gauge should be selected rather small.

4.2 Poisson approximation to gauge for 1-step M-estimators

A Poisson exceedence theory arises in the scenario where the cut-off value c is set to allow the fixed number λ of outliers regardless of the sample size n.

For some $\lambda > 0$, the cut-off value c_n is set so as to let

$$n\mathrm{P}(|\varepsilon_i| > \sigma c_n) = \lambda. \tag{4.3}$$

We can notice immediately that $c_n \to \infty$ as $n \to \infty$. For the 1-step Huber-skip M-estimators, c_n appears in the stochastic weight, estimators and the sample gauge, thus we emphasize the subscript c_n in sequences of the stochastic weight v_{i,c_n}, estimators $\widehat{\beta}_{c_n}$, $(\widehat{\sigma}_{c_n})^2$ and the sample gauge $\widehat{\gamma}_{c_n}$ that may depend on c_n. The sample gauge in this case is

$$\widehat{\gamma}_{c_n} = \frac{1}{n}\sum_{i=1}^{n}(1-v_{i,c_n}) = \frac{1}{n}\sum_{i=1}^{n}1_{(|y_i - x_i'\widehat{\beta}_{c_n}| > \widehat{\sigma}_{c_n} c_n)}, \tag{4.4}$$

Where $v_{i,c_n} = 1_{(|y_i - x_i'\widehat{\beta}_{c_n}| \leq \widehat{\sigma}_{c_n} c_n)}$.

Assuming estimation errors $N^{-1}(\widehat{\beta}_{c_n}-\beta)$ and $n^{1/2}(\widehat{\sigma}_{c_n}-\sigma)$ are $\mathrm{O_p}(1)$, Johansen and Nielsen (2014c) proved a Poisson exceedence theory.

Theorem 4.1. *(Johansen and Nielsen, 2014c, Theorem 9.7) Consider the 1-step Huber-skip M-estimator. Let c_n be defined from (4.3). Suppose Assumption 3.1(ia, ib, ic, iid) holds, and that $N^{-1}(\widehat{\beta}_{c_n}-\beta)$ and $n^{1/2}(\widehat{\sigma}_{c_n}-\sigma)$ are $\mathrm{O_p}(1)$. Then for $n\to\infty$, the sample gauge $\widehat{\gamma}_{c_n}$ in (4.4) satisfies*

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

4.3 A new Poisson approximation result

Johansen and Nielsen (2014c) only proved a Poisson approximation around the 1-step Huber-skip M-estimator to the gauge based on initial estimates which do not depend on the varying cut-off. But the iterated results have not been established, since they did not have the theory of the weighted and marked empirical process allowing the varying quantile c with estimation errors for β and σ. A contribution of this paper is to show Poisson approximations for the iterated 1-step Huber-skip M-estimator including Algorithm 2.1, 2.2.

Consider the iterated 1-step Huber-skip M-estimator and let c_n be defined from (4.3). Notice $c_n\to\infty$ as $n\to\infty$. For each iterated step m, by replacing c by c_n in stochastic weights $v_{i,c}^{(m)}$, estimates $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ defined from (3.3), (3.1), (3.2), we can express $v_{i,c_n}^{(m)}$, $\widehat{\beta}_{c_n}^{(m)}$, $(\widehat{\sigma}_{c_n}^{(m)})^2$ as

$$v_{i,c_n}^{(m)} = 1_{(|y_i-x_i'\widehat{\beta}_{c_n}^{(m)}|\le\widehat{\sigma}_{c_n}^{(m)}c_n)},$$

$$\widehat{\beta}_{c_n}^{(m)} = (\sum_{i=1}^n x_ix_i'v_{i,c_n}^{(m-1)})^{-1}(\sum_{i=1}^n x_iy_iv_{i,c_n}^{(m-1)}),$$

$$(\widehat{\sigma}_{c_n}^{(m)})^2 = \varsigma_{c_n}^{-2}(\sum_{i=1}^n v_{i,c_n}^{(m-1)})^{-1}\{\sum_{i=1}^n(y_i-x_i'\widehat{\beta}_{c_n}^{(m)})^2v_{i,c_n}^{(m-1)}\}.$$

Note $v_{i,c_n}^{(m-1)}$ is the stochastic indicator in the previous step. $\varsigma_{c_n}^2=\tau_2^{c_n}/\psi_{c_n}$ is the bias correction factor for the variance estimator as in (2.3). Then, for any iteration m, we highlight the subscript c_n in the sample gauge $\widehat{\gamma}_{c_n}^{(m)}$ which can be defined as

$$\widehat{\gamma}_{c_n}^{(m)} = \frac{1}{n}\sum_{i=1}^n(1-v_{i,c_n}^{(m)}) = \frac{1}{n}\sum_{i=1}^n 1_{(|y_i-x_i'\widehat{\beta}_{c_n}^{(m)}|>\widehat{\sigma}_{c_n}^{(m)}c_n)}. \tag{4.5}$$

For the iterated Robustified Least Squares Algorithm 2.1, $\widehat{\beta}_{c_n}^{(0)} = \widetilde{\beta}$, $(\widehat{\sigma}_{c_n}^{(0)})^2 = \widetilde{\sigma}^2$ are tight since they are the full sample least squares which do not depend on c_n. For the Impulse Indicator Saturation Algorithm 2.2, the initial estimators $\widehat{\beta}_j$ and $\widehat{\sigma}_j^2$ are tight in the sub-sample set $\mathcal{I}_j$ for $j = 1, 2$ due to the same reason that they are the split sample least squares which are irrelevant of c_n. Thus, Theorem 4.1 trivially applies so that an asymptotic Poisson distribution can be demonstrated for $\widehat{\gamma}_{c_n}^{(0)}$ in the Algorithm 2.1 and for $\widehat{\gamma}_{c_n}^{(-1)}$ in the Algorithm 2.2. Note

$$v_{i,c_n}^{(-1)} = 1_{(i\in\mathcal{I}_1)}1_{(|y_i - x_i'\widehat{\beta}_2|\leq\widehat{\sigma}_2 c_n)} + 1_{(i\in\mathcal{I}_2)}1_{(|y_i - x_i'\widehat{\beta}_1|\leq\widehat{\sigma}_1 c_n)}$$

as in (2.13) in the Algorithm 2.2. However, the updated estimators $\widehat{\beta}_{c_n}^{(m)}$ and $(\widehat{\sigma}_{c_n}^{(m)})^2$ do depend on the cut-off c_n. An extra argument is required to show $\widehat{\beta}_{c_n}^{(m)}$ and $(\widehat{\sigma}_{c_n}^{(m)})^2$ are tight in order to prove a Poisson exceedence theory for any iteration m of Algorithm 2.1, 2.2. Thus, a new Poisson approximation result can be shown by applying a special case of Theorem 3.1, 3.2 where $c = c_n \to \infty$ together with Theorem 4.1.

Theorem 4.2. *Consider the iterated Robustified Least Squares Algorithm 2.1. Let c_n be defined from (4.3). Suppose Assumption 3.1(i, ii) holds. Then for all $0 \leq m < \infty$ and $n \to \infty$, the sample gauge $\widehat{\gamma}_{c_n}^{(m)}$ in (4.5) satisfies*

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

Theorem 4.3. *Consider the Impulse Indicator Saturation Algorithm 2.2. Let c_n be defined from (4.3). Suppose Assumption 3.1(i, ii) holds for each sub-sample set $\mathcal{I}_1$, $\mathcal{I}_2$. Then for all $-1 \leq m < \infty$ and $n \to \infty$, the sample gauge $\widehat{\gamma}_{c_n}^{(m)}$ in (4.5) satisfies*

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

Table 1 assumes ε_i/σ follows a standard normal distribution. For a given λ and $n = 100, 200$, the cut-off can be found from (4.3) so that

$$c_n = G^{-1}(\psi) = F^{-1}(\frac{1+\psi}{2}) = \Phi^{-1}\{(1 + \frac{n-\lambda}{n})/2\} = \Phi^{-1}(1 - \frac{\lambda}{2n}).$$

Then, Poisson approximations give 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 λ

			x					
λ	$c_{n=100}$	$c_{n=200}$	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 λ, and the cut-off $c_n = \Phi^{-1}\{1 - \lambda/(2n)\}$ for $n = 100, 200$.

declines from 5 to 0.1. The reason is due to the left skewness of the Poisson distribution. In particular, we can 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. In summary, Poisson approximations is possibly more accurate than normal approximations for the small λ and n.

Proofs appear in the appendix. We establish a theory for a new class of the weighted and marked empirical process before showing the tightness result for Algorithm 2.1, 2.2. Subsequently, a Poisson exceedence theory can be demonstrated.

5 Absolute weighted and marked empirical process

In order to analyze outliers detection algorithms, we need a new class of the weighted and marked empirical process. Thus, this paper presents a new empirical process result which generalizes Johansen and Nielsen (2009, 2014a).

Some concepts and existing results are introduced initially. Then, we can provide ingredients of the empirical process results for asymptotic analysis.

5.1 Background

Estimators and regressors are normalized during the paper. Define

$$\widetilde{b} = N^{-1}(\widetilde{\beta} - \beta), \quad x_{in} = N' x_i, \tag{5.1}$$

where N is a normalisation matrix based on the stochastic properties of the regressors x_i. Choose N so that $\sum_{i=1}^{n} x_{in}x_{in}'$ converges, and $n^{-1/2}\sum_{i=1}^{n}|x_{in}|$ is bounded. For instance, $N = n^{-1/2}I_{\dim(x)}$ if (y_i, x_i) is stationary for $i = 1, 2, \ldots, n$. If $\{x_i\}_{i=1}^{n}$ is a unit root process, $N = n^{-1}I_{\dim(x)}$. Further more, define

$$\widetilde{a} = n^{1/2}(\widetilde{\sigma} - \sigma), \quad \widetilde{d} = n^{1/2}(\frac{\widetilde{\xi}_{(k)}}{\sigma} - c). \tag{5.2}$$

$\widetilde{a}$, $\widetilde{b}$ and $\widetilde{d}$ measure estimation errors for σ, β and c respectively. Then introduce matrix-valued weights g_{in} of the form 1, $n^{1/2}N'x_i$ or $nN'x_ix_i'N$. The weights should satisfy the condition $n^{-1}\sum_{i=1}^{n}|g_{in}|$ is bounded.

The empirical distribution function of $|\varepsilon_i|/\sigma \sim G(c)$ is $\widehat{G}_n(c) = n^{-1}\sum_{i=1}^{n} 1_{(|\widehat{\varepsilon}_i|/\sigma \le c)}$. Then, after some arrangements, obtain

$$\widehat{G}_n(c) = \frac{1}{n}\sum_{i=1}^{n} 1_{(|y_i - x_i'\widehat{\beta}| \le \sigma c)} = \frac{1}{n}\sum_{i=1}^{n} 1_{(|\varepsilon_i - x_i'(\widehat{\beta}-\beta)| \le \sigma c)} = \frac{1}{n}\sum_{i=1}^{n} 1_{(|\varepsilon_i - x_{in}'\widehat{b}| \le \sigma c)}.$$

Based on this, it is quite natural to define the weighted and marked absolute empirical distribution function

$$\widehat{G}_n^{g,p}(b, c) = \frac{1}{n}\sum_{i=1}^{n} g_{in}\varepsilon_i^p 1_{(|\varepsilon_i - x_{in}'b| \le \sigma c)}. \tag{5.3}$$

Since the $\widehat{G}_n^{g,p}(b, c)$ is analyzed through martingale arguments, we introduce its compensator

$$\overline{G}_n^{g,p}(b, c) = \mathrm{E}_{i-1}\{\frac{1}{n}\sum_{i=1}^{n} g_{in}\varepsilon_i^p 1_{(|\varepsilon_i - x_{in}'b| \le \sigma c)}\} = \frac{1}{n}\sum_{i=1}^{n} g_{in}\mathrm{E}_{i-1}\{\varepsilon_i^p 1_{(|\varepsilon_i - x_{in}'b| \le \sigma c)}\}. \tag{5.4}$$

In the above, $b \in \mathbb{R}^{\dim(x)}$ and $c \in \mathbb{R}$. Let $\mathcal{F}_i$ be an increasing sequence of σ-fields so that ε_{i-1}, x_i and g_{in} are $\mathcal{F}_{i-1}$ measurable and ε_i is independent with $\mathcal{F}_{i-1}$. Note that $\mathrm{E}_{i-1}\{\varepsilon_i^p 1_{(|\varepsilon_i| \le \sigma c)}\} = \mathrm{E}\{\varepsilon_i^p 1_{(|\varepsilon_i| \le \sigma c)}\} = 0$ for an odd p and $b = 0$ if f is symmetric. Define the weighted and marked empirical process

$$\mathbb{G}_n^{g,p}(b, c) = n^{1/2}\{\widehat{G}_n^{g,p}(b, c) - \overline{G}_n^{g,p}(b, c)\}. \tag{5.5}$$

We focus on four combinations of weights and marks when carrying out asymptotic analysis of outliers detection algorithms. $g_{in} = n^{1/2}N'x_i$, $p = 1$ and $g_{in} = nN'x_ix_i'N$, $p = 0$ are involved in the least squares estimator. The residuals involve $g_{in} = 1$, $p = 0$.

The mentioned terms above as well as $g_{in} = 1$, $p = 2$ are sufficient statistics for the variance estimator.

This class of empirical processes generalizes that of Koul (2002) who did not allow non-bounded marks ε_i^p. Therefore, these are independent of interest. Moreover, the weighted and marked empirical processes are essential for analyzing the least squares statistics and the gauge in the 1-step Huber-skip estimators and the corresponding iterated versions. Thus it is a key to understand outliers detection algorithms, such as the Impulse Indicator Saturation and the Forward Search.

5.2 A new empirical process result

The Impulse Indicator Saturation Algorithm 2.2 is the version of the iterated 1-step Huber-skip M-estimator with the indicator (2.11)

$$1_{(|\widetilde{\varepsilon}_i|/\widetilde{\sigma}\leq c)} = 1_{(|\varepsilon_i - x_{in}'\widetilde{b}|\leq \sigma c(1+n^{-1/2}\widetilde{a}/\sigma))},$$

whereas the Forward Search is the iterated 1-step Huber-skip L-estimator with the indicator (2.12)

$$1_{(|\widetilde{\varepsilon}_i|\leq \widetilde{\xi}_{(k)})} = 1_{(|\varepsilon_i - x_{in}'\widetilde{b}|\leq \sigma(c+n^{-1/2}\widetilde{d}))}.$$

Thus, the main interest of the empirical process concentrates on $\mathbb{G}_n^{g,p}(b, c(1+n^{-1/2}a/\sigma))$ and $\mathbb{G}_n^{g,p}(b, c+n^{-1/2}d)$. $\mathbb{G}_n^{g,p}(b, c(1+n^{-1/2}a/\sigma))$ includes estimation errors for the parameter β and variance σ, while $\mathbb{G}_n^{g,p}(b, c+n^{-1/2}d)$ contains estimation errors for the parameter β and quantile c. Johansen and Nielsen (2014a) proved asymptotic results of $\mathbb{G}_n^{g,p}(b, c+n^{-1/2}d)$ and the tightness property of $\mathbb{G}_n^{g,p}(b, c)$. See the tightness definition and argument in Billingsley (1968). In this paper, we demonstrate an asymptotic result for another essential class of the empirical process $\mathbb{G}_n^{g,p}(b, c(1+n^{-1/2}a/\sigma))$ which takes into account of the multiplicative estimation error a.

Theorem 5.1. *Let $c_\psi = G^{-1}(\psi)$. Suppose Assumption 3.1(ia, iid, iie) holds for some $0 < \eta \leq 1/4$. Then, for all $B > 0$ and $n \to \infty$, it holds*

$$\sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B} |\mathbb{G}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{G}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

Theorem 5.1 extends Johansen and Nielsen (2009) who established the result for the fixed quantile c_ψ and with $\eta > 1/4$. An idea of chaining is applied to prove Theorem 5.1. This is to cover the domain of estimation uncertainties and the varying quantile with a finite number of chaining intervals through the partitioned points. Then solving the supremum over the large compact set can be transferred to finding the maximum value over the partitioned points and the maximum of the variation within the chaining interval. Thus, an iterated inequality is required to explore the tail behaviour of the maximum of a family of martingales. At last, the argument will be a consideration of additive and multiplicative shifts.

Furthermore, we can combine all additive and multiplicative shifts due to estimation errors a, b and d with the varying quantile c for the process $\mathbb{G}_n^{g,p}(b, c(1 + n^{-1/2}a/\sigma) + n^{-1/2}d)$.

Theorem 5.2. *Let $c_\psi = G^{-1}(\psi)$. Suppose Assumption 3.1(ia, iid, iie) holds for some $0 \leq \kappa < \eta \leq 1/4$. Then, for all $B > 0$ and $n \to \infty$, it holds*

$$\sup_{0\leq\psi\leq 1}\sup_{|a|,|b|,|d|\leq n^{1/4-\eta}B} |\mathbb{G}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma}) + n^{\kappa-1/2}d) - \mathbb{G}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

Theorem 5.2 is proved by associating the result of the empirical process $\mathbb{G}_n^{g,p}(b, c + n^{\kappa-1/2}d)$ with $\mathbb{G}_n^{g,p}(b, c(1 + n^{-1/2}a/\sigma))$. This theorem extends Johansen and Nielsen (2014a) where the estimation error a for σ is fixed.

5.3 Asymptotic expansion of empirical processes

The empirical process theory is significant for analyzing Huber-skip estimators and their related outliers detection algorithms. This is because the 1-step estimators for β and σ^2 can be expressed as the following forms,

$$n^{-1}\sum_{i=1}^{n} v_i, \quad n^{-1/2}\sum_{i=1}^{n} \varepsilon_i^2 v_i, \quad N'\sum_{i=1}^{n} x_i\varepsilon_i v_i, \quad N'\sum_{i=1}^{n} x_i x_i' v_i N. \tag{5.6}$$

v_i represents stochastic weights (2.11), (2.12) for residuals for M, L-estimates. The unbounded $\mathcal{F}_i$-measurable ε_i and ε_i^2 are defined as marks in Johansen and Nielsen (2009). Koul (2002) called the $\mathcal{F}_{i-1}$-adapted x_i and $x_i x_i'$ as weights. For 1-step M-estimators,

the indicator functions are

$$v_{M,i} = v_i^{\widetilde{a},\widetilde{b},c} = 1_{(|\varepsilon_i - x_{in}'\widetilde{b}| \le \sigma c(1+n^{-1/2}\widetilde{a}/\sigma))}, \tag{5.7}$$

where $\widetilde{b} = N^{-1}(\widetilde{\beta} - \beta)$ and $\widetilde{a} = n^{1/2}(\widetilde{\sigma} - \sigma)$ are estimation uncertainties for β and σ. For 1-step L-estimators, the indicator has the form

$$v_{L,i} = v_i^{\widetilde{b},c,\widetilde{d}} = 1_{(|\varepsilon_i - x_{in}'\widetilde{b}| \le \sigma(c+n^{-1/2}\widetilde{d}))}, \tag{5.8}$$

where estimation errors $\widetilde{b} = N^{-1}(\widetilde{\beta} - \beta)$ and $\widetilde{d} = n^{1/2}(\widetilde{\xi}_{(k)}/\sigma - c)$ are allowed for β and quantiles c. See (5.1) and (5.2).

The M-type Algorithm 2.1, 2.2 and the Forward Search are governed by 1-step Huber-skip estimates in each iterated step with statistics (5.6) and v_i given by (5.7), (5.8). Thus, we need an asymptotic expansion of the product moment (5.6) to control each step of such algorithms. The general indicator can be considered by

$$v_i^{a,b,c,d} = 1_{(|\varepsilon_i - x_{in}'b| \le \sigma[c(1+n^{-1/2}a/\sigma)+n^{-1/2}d])}. \tag{5.9}$$

Estimation uncertainties in (5.7), (5.8) are changed by the bounded and deterministic terms a, b, d. An interest in the literature is thus to replace a, b, d by $\widetilde{a}, \widetilde{b}, 0$ and by $0, \widetilde{b}, \widetilde{d}$. The following linear approximation is a generalization of Johansen and Nielsen (2009, 2014a).

Theorem 5.3. *Suppose Assumption 3.1*(ia, iid, iie) *holds. Consider the product moments (5.6) with indicators* $v_i^{a,b,c,d}$ *give by (5.9). Then we have expansions*

$$n^{-1/2}\sum_{i=1}^n v_i^{a,b,c,d} = n^{-1/2}\sum_{i=1}^n 1_{(|\varepsilon_i| \le \sigma c)} + 2f(c)(\frac{ac}{\sigma} + d) + R_v(a,b,c,d),$$

$$n^{-1/2}\sum_{i=1}^n \varepsilon_i^2 v_i^{a,b,c,d} = n^{-1/2}\sum_{i=1}^n \varepsilon_i^2 1_{(|\varepsilon_i| \le \sigma c)} + 2\sigma^2 c^2 f(c)(\frac{ac}{\sigma} + d) + R_{v\varepsilon\varepsilon}(a,b,c,d),$$

$$N'\sum_{i=1}^n x_i\varepsilon_i v_i^{a,b,c,d} = N'\sum_{i=1}^n x_i\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + 2cf(c)N'\sum_{i=1}^n x_i x_i' Nb + R_{vx\varepsilon}(a,b,c,d),$$

$$n^{1/2}N'\sum_{i=1}^n x_i x_i' v_i^{a,b,c,d} N = n^{1/2}N'\sum_{i=1}^n x_i x_i' 1_{(|\varepsilon_i| \le \sigma c)} N + 2f(c)N'\sum_{i=1}^n x_i x_i' N(\frac{ac}{\sigma} + d)$$
$$+ R_{vxx}(a,b,c,d).$$

Let

$$R(a,b,c,d) = |R_v(a,b,c,d)| + |R_{v\varepsilon\varepsilon}(a,b,c,d)| + |R_{vx\varepsilon}(a,b,c,d)| + |R_{vxx}(a,b,c,d)|.$$

Then it holds for all large $B > 0$*, all small* $0 < \eta \leq 1/4$ *and* $n \to \infty$ *that*

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

An idea here is to subtract compensators of the product moments (5.6). Subsequently, applying the second order Taylor expansion to compensators can achieve the second term in the linearization. The empirical process theory shows the rest terms are $\mathrm{o_p}(1)$.

Proofs of Theorem 5.1, 5.2 and 5.3 are given in the appendix. Since the asymptotic result of the absolute empirical process can be reduced to the one-sided case, we will then present some general results for the one-sided empirical process before getting into the proof.

6 Auxiliary weighted and marked empirical process

The one-sided empirical processes are considered for errors ε_i as opposed to absolute errors $|\varepsilon_i|$. An asymptotic theory in this section is more general, and thus can be transferred to the absolute case. The analysis generalizes the result of Koul and Ossiander (1994), who did not allow non-bounded marks ε_i^p. Furthermore, they also assumed estimation errors must be bounded. In contrast, marks ε_i^p are included in our more general class of the empirical process, and estimation errors are permitted to be diverge at a rate of $n^{1/4-\eta}$. In addition, we also generalize the empirical process results of Johansen and Nielsen (2009) in the way that the quantile $c_\psi = F^{-1}(\psi)$ can be varying in its domain rather than being fixed. Based on these, a more general empirical process result is extended for all additive and multiplicative shifts due to estimation uncertainties a, b and d with the varying quantile c_ψ. This work is an extension of Johansen and Nielsen (2014a) since they only considered additive estimation errors b, d in the process.

The explicit assumptions are listed first. Then, we discuss the relationship between the conditions assumed here and previously. After reviewing some existing results, a new asymptotic theory is presented for the one-sided empirical process.

6.1 Assumptions

During this section, the density f is allowed to be more general. Especially, f is not necessarily symmetric. The conditions listed below are more explicit and weaker than Assumption 3.1. We will keep track of these assumptions in the proof of the one-sided empirical process result.

Assumption 6.1. *Let $\mathcal{F}_i$ be an increasing sequence of σ-fields so ε_{i-1}, x_i and g_{in} are $\mathcal{F}_{i-1}$-measurable and ε_i is independent with $\mathcal{F}_{i-1}$. ε_i/σ has a continuous, differentiable density f which is positive for $F^{-1}(0) < c < F^{-1}(1)$. Let p, r, η, κ, ν be given so $p, r \in \mathbb{N}_0$, $0 \leq \kappa < \eta \leq 1/4$ and $\nu \leq 1$. Suppose*

(i) density satisfies

(a) moments: $\int_{-\infty}^{\infty} |\epsilon|^{2^r p/\nu} f(\epsilon) d\epsilon < \infty$;

(b) boundedness: $\sup_{c\in\mathbb{R}}\{(1+|c|^{2^r p-1})f(c)+c(1+|c|^{2^r p})f(c)+(1+|c|^{2^r p})|f'(c)|\} < \infty$;

(c) smoothness: a $C_{\mathrm{H}} \in \mathbb{N}$ exist so that for all $c > 0$

$$\frac{\sup_{\epsilon\geq c}(1+\epsilon^{2^r p})f(\epsilon)}{\inf_{0\leq\epsilon\leq c}(1+\epsilon^{2^r p})f(\epsilon)} \leq C_{\mathrm{H}}, \quad \frac{\sup_{\epsilon\leq -c}(1+\epsilon^{2^r p})f(\epsilon)}{\inf_{-c\leq\epsilon\leq 0}(1+\epsilon^{2^r p})f(\epsilon)} \leq C_{\mathrm{H}};$$

(ii) regressors x_i are $\mathcal{F}_{i-1}$-measurable and satisfy $\max_{1\leq i\leq n} |n^{1/2-\kappa} N' x_i| = \mathrm{O_p}(1)$ for some $\kappa < \eta$;

(iii) weights g_{in} are matrix valued, $\mathcal{F}_{i-1}$-measurable and satisfy

(a) $n^{-1}\mathrm{E}\sum_{i=1}^n |g_{in}|^{2^r}(1+|n^{1/2}N'x_i|) = \mathrm{O}(1)$ for some $r \in \mathbb{N}_0$;

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

Johansen and Nielsen (2014a) demonstrated Assumption 3.1(ia) is sufficient for Assumption 6.1(i) if f is symmetric and differentiable. That is to say, the moments, boundedness and smoothness of a symmetric density f can be controlled by its tail monotonicity. $\nu < 1$ in Assumption 6.1(ia) is only used to deal with the tightness property of the empirical process, otherwise $\nu = 1$ is enough to our proof. Condition (iii) provides more accurate assumptions on weights g_{in}, which can be showed by condition (ii) made on regressors x_i. Weights g_{in} can be 1, $n^{1/2}N'x_i$ and $nN'x_ix_i'N$, where $N = n^{-1/2}I_{\dim(x)}$ for stationary regressors and $N = n^{-1}I_{\dim(x)}$ for random walks.

6.2 Some existing empirical process results

The one-sided weighted and marked empirical distribution function is defined by

$$\widehat{F}_n^{g,p}(b,c) = \frac{1}{n}\sum_{i=1}^{n} g_{in}\varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + x_{in}' b)}, \tag{6.1}$$

for $b \in \mathbb{R}^{\dim(x)}$ and $c \in \mathbb{R}$ with weights g_{in} and marks ε_i^p. Then, introduce its compensator

$$\overline{F}_n^{g,p}(b,c) = \frac{1}{n}\sum_{i=1}^{n} g_{in}\mathrm{E}_{i-1}\{\varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + x_{in}' b)}\}, \tag{6.2}$$

and the one-sided empirical process

$$\mathbb{F}_n^{g,p}(b,c) = n^{1/2}\{\widehat{F}_n^{g,p}(b,c) - \overline{F}_n^{g,p}(b,c)\}. \tag{6.3}$$

Recall $x_{in} = N'x_i$ are normalized regressors, where N is the corresponding normalisation matrix depending on the stochastic properties of regressors. For instance, $N = n^{-1/2}I_{\dim(x)}$ if (y_i, x_i) is stationary for $i = 1, 2, \ldots, n$. If $\{x_i\}_{i=1}^n$ is a unit root process, $N = n^{-1}I_{\dim(x)}$. The main interest of the empirical process focuses on combinations between $p = 0, 1, 2$ and $g_{in} = 1, n^{1/2}N'x_i, nN'x_ix_i'N$.

Johansen and Nielsen (2014a) built an asymptotic theory for the one-sided empirical process $\mathbb{F}_n^{g,p}(b, c_\psi + n^{-1/2}d)$ in order to analyze the Forward Search. The first result shows that the dependence on the estimation error b is negligible for $\mathbb{F}_n^{g,p}$. This is proved using a chaining argument with an additive shift and iterated martingale inequlities.

Theorem 6.1. *(Johansen and Nielsen, 2014a, Theorem 4.1) Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 6.1$(i, ii, iiia)$ holds with $\nu = 1$, some $0 \leq \kappa < \eta \leq 1/4$, and an $r \geq 2$ such that $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$. Then, for any $B > 0$ and $n \to \infty$, it holds that*

$$\sup_{0\leq\psi\leq 1}\sup_{|b|\leq n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(b, c_\psi) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

The second result considers an additive shift due to both estimation errors b and d. It proves the dependence on b and d is negligible for $\mathbb{F}_n^{g,p}$. The argument for this theorem is similar to the proof of Theorem 6.1.

Theorem 6.2. *(Johansen and Nielsen, 2014a, Theorem 4.2) Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 6.1$(i, ii, iiia)$ holds with $\nu = 1$, some $0 \leq \kappa < \eta \leq 1/4$, and an $r \geq 2$ such*

that $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$. *Then, for any* $B > 0$ *and* $n \to \infty$, *it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4-\eta} B} |\mathbb{F}_n^{g,p}(b, c_\psi + n^{\kappa - 1/2} d) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

6.3 A new empirical process result

We establish an asymptotic result for another essential class of the process $\mathbb{F}_n^{g,p}(0, c_\psi(1 + n^{-1/2} a/\sigma))$ so as to study the Impulse Indicator Saturation. Denote $c_{\psi\dagger} = c_\psi(1 + n^{-1/2} a/\sigma)$. The first result demonstrates that the dependence on the estimation error a is negligible for $\mathbb{F}_n^{g,p}$ when $b = 0$. In other words, $\mathbb{F}_n^{g,p}(0, c_{\psi\dagger})$ is equivalent to $\mathbb{F}_n^{g,p}(0, c_\psi)$ asymptotically as $n \to \infty$. This is proved by applying a chaining argument with a multiplicative shift inequality.

Theorem 6.3. *Let* $c_\psi = F^{-1}(\psi)$. *Suppose Assumption 6.1*(ia, ib, ii, iii) *holds with* $\nu = 1$, *some* $0 < \eta \leq 1/4$, *and an* $r \geq 2$. *Then, for any* $B > 0$ *and* $n \to \infty$, *it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4-\eta} B} |\mathbb{F}_n^{g,p}(0, c_\psi(1 + n^{-1/2} \frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

The next result conbines estimation errors b and a, which takes into account both additive and multiplicative shifts. This is shown by associating Theorem 6.1 with Theorem 6.3. The theorem below generalizes Johansen and Nielsen (2009) who set the quantile c_ψ fixed.

Theorem 6.4. *Let* $c_\psi = F^{-1}(\psi)$. *Suppose Assumption 6.1*(i, ii, iii) *holds with* $\nu = 1$, *some* $0 \leq \kappa < \eta \leq 1/4$, *and an* $r \geq 2$ *such that* $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$. *Then, for any* $B > 0$ *and* $n \to \infty$, *it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a|, |b| \leq n^{1/4-\eta} B} |\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2} \frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

Furthermore, we extend the asymptotic result of the empirical process $\mathbb{F}_n^{g,p}(b, c_\psi + n^{\kappa - 1/2} d)$ and $\mathbb{F}_n^{g,p}(0, c_\psi(1 + n^{-1/2} a/\sigma))$ to a more general class $\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2} a/\sigma) + n^{\kappa - 1/2} d)$ where all possible additive and multiplicative shifts are included. This result demonstrates the dependence on estimation errors a, b and d for $\mathbb{F}_n^{g,p}$ is negligible uniformly in the quantile c_ψ when $n \to \infty$. We can prove this by combining Theorem 6.2

and Theorem 6.3. This in turn is an extension of Johansen and Nielsen (2014a) where the estimation error a is selected as 0.

Theorem 6.5. *Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 6.1(i, ii, iii) holds with $\nu = 1$, some $0 \leq \kappa < \eta \leq 1/4$, and an $r \geq 2$ such that $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$. Then, for any $B > 0$ and $n \to \infty$, it holds that*

$$\sup_{0\leq\psi\leq 1} \sup_{|a|,|b|,|d|\leq n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma}) + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).$$

Proofs appear in the appendix. Since the proof involves a chaining argument, a metric is set up with a multiplicative shift inequality. Then, introduce some martingale results exploring the tail behaviour of the maximum of a family of martingales. Finally, proofs of Theorem 6.3, 6.4 and 6.5 are provided.

7 Conclusion

This paper builds an asymptotic theory exploring how variation in the cut-off influences the iterated 1-step Huber-skip M-estimator. Especially, we establish the tightness result for estimates in each iterated step with the varying cut-off. Moreover, a Poisson exceedence theory is demonstrated for calibrating the cut-off value of Algorithm 2.1, 2.2. This happens if the cut-off is set through the way that the number of falsely detected outliers is held fixed regardless of the sample size. The argument involves a theory for a new class of the weighted and marked empirical process. These results are derived under the hypothesis that there are no outliers in the model. In addition, this paper assumes the density is symmetric in the main sections, even it is allowed to be non-symmetric in the one-sided empirical process section.

For the Robustified Least Squares Algorithm 2.1 and the Impulse Indicator Saturation Algorithm 2.2, we define the same cut-off $c > 0$ in each iterated step. In practice, however, the cut-off value c can be chosen differently through iterations. From this perspective, future work would carry out asymptotic analysis on such general RLS and IIS algorithms.

In contrast to the traditional approach of influence function in robust statistics, the results presented are based on the empirical process techniques. These are two different

approaches for analysing robust methods asymptotically. Thus, we intend to compare our approach with the traditional idea in the future work.

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 given in this paper.

An asymptotic analysis does not provide any result on the behaviour of outliers detection algorithms when outlying observations are actually presented in the model. Leading on from that, the further research would investigate situations with outliers to check how well detection algorithms capture these. An interesting scenario possibly contains single outliers, clusters of outliers, level shifts, symmetric or non-symmetric outliers. To be specific, the future work could compare the potency of two distinct detection algorithms which have the same gauge. The potency is the retention rate of relevant outliers or relevant regressors in general.

A Some martingale inequalities

The asymptotic theory of the empirical processes can be handled by applying a chaining argument. This reduces the problem to finding the tail probability of the maximum of a certain family of martingales. Thus, we review some useful martingale inequalities proposed to deal with this concern. An iterated martingale inequality was constructed by Johansen and Nielsen (2014a) with the purpose of exploring the tail behaviour of the maximum of a family of martingales. It builds on an exponential martingale inequality derived by Bercu and Touati (2008).

Theorem A.1. *(Bercu and Touati, 2008, Theorem 2.1) For $i = 1, ..., n$, let m_i be a locally square integrable difference sequence with respect to a filtration $\mathcal{F}_i$. Then, for all $x, y > 0$,*

$$\mathrm{P}[|\sum_{i=1}^{n} m_i| \geq x, \sum_{i=1}^{n} \{m_i^2 + \mathrm{E}_{i-1}(m_i^2)\} \leq y] \leq 2\exp(-\frac{x^2}{2y}).$$

Theorem A.2. *(Johansen and Nielsen, 2014a, Theorem 5.1) For l so $1 \leq l \leq L$, let $z_{l,i}$ be $\mathcal{F}_i$ adapted satisfying $\mathrm{E}z_{l,i}^{2^{\bar{r}}} < \infty$ for some $\bar{r} \in \mathbb{N}$. Let $D_r = \max_{1 \leq l \leq L} \sum_{i=1}^{n} \mathrm{E}_{i-1} z_{l,i}^{2^r}$*

for $1 \leq r \leq \bar{r}$. *Then, for all* $\kappa_0, \kappa_1, ..., \kappa_{\bar{r}}$, *it holds*

$$\mathrm{P}\{\max_{1\leq l\leq L} |\sum_{i=1}^{n} (z_{l,i} - \mathrm{E}_{i-1} z_{l,i})| \geq \kappa_0\} \leq L\frac{\mathrm{E}D_{\bar{r}}}{\kappa_{\bar{r}}} + \sum_{r=1}^{\bar{r}} \frac{\mathrm{E}D_r}{\kappa_r} + 2L\sum_{r=0}^{\bar{r}-1} \exp(-\frac{\kappa_r^2}{14\kappa_{r+1}}).$$

Parameters $\kappa_0, \kappa_1, ..., \kappa_{\bar{r}}$ in an iterated martingale inequality can be chosen in various ways. Two special cases were analyzed by Johansen and Nielsen (2014a).

The first type considers a family of size $L = \mathrm{O}(n^{\lambda})$ where $\lambda > 0$, and chooses $\kappa_q = (\kappa n^{1/2})^{2^q} (28\lambda \log n)^{1-2^q}$, so that $\kappa_q^2/\kappa_{q+1} = 28\lambda \log n$ and $\kappa_0 = \kappa n^{1/2}$. It seeks to prove that the maximum of a family of martingales is of order $\mathrm{o_p}(n^{1/2})$. We can obtain the above statement by setting $\upsilon = 1/2$ through the following theorem. This will be used to prove the asymptotic result of the empirical process.

In the second type, a family of size L is fixed. Parameters are selected as $\kappa_q = \kappa n^{2^{q-1}} \theta^{1-2^q}$ so that $\kappa_q^2/\kappa_{q+1} = \kappa\theta$ and $\kappa_0 = \kappa n^{1/2}$. This type of the iterated martingale inequality can be used to demonstrate the tightness result of the empirical process.

Theorem A.3. *(Johansen and Nielsen, 2014a, Theorem 5.2) For* l *so* $1 \leq l \leq L$, *let* $z_{l,i}$ *be* $\mathcal{F}_i$ *adapted satisfying* $\mathrm{E}z_{l,i}^{2^{\bar{r}}} < \infty$ *for some* $\bar{r} \in \mathbb{N}$. *Let* $D_r = \max_{1\leq l\leq L} \sum_{i=1}^{n} \mathrm{E}_{i-1} z_{l,i}^{2^r}$ *for* $1 \leq r \leq \bar{r}$. *Suppose, for some* $\varsigma \geq 0$, $\lambda > 0$, *that* $L = \mathrm{O}(n^{\lambda})$ *and* $\mathrm{E}D_r = \mathrm{O}(n^{\varsigma})$ *for* $r \leq \bar{r}$. *Then, if* $\upsilon > 0$ *is chosen such that*

(*i*) $\varsigma < 2\upsilon$;

(*ii*) $\varsigma + \lambda < \upsilon 2^{\bar{r}}$;

it holds that for all $\kappa > 0$ *and* $n \to \infty$ *that*

$$\lim_{n\to\infty} \mathrm{P}\{\max_{1\leq l\leq L} |\sum_{i=1}^{n} (z_{l,i} - \mathrm{E}_{i-1} z_{l,i})| \geq \kappa n^{\upsilon}\} = 0.$$

Theorem A.4. *(Johansen and Nielsen, 2014a, Theorem 5.3) For* l *so* $1 \leq l \leq L$, *let* $z_{l,i}$ *be* $\mathcal{F}_i$ *adapted satisfying* $\mathrm{E}z_{l,i}^4 < \infty$. *Suppose* $\mathrm{E}\max_{1\leq l\leq L} \sum_{i=1}^{n} \mathrm{E}_{i-1} z_{l,i}^{2q} \leq Cn$ *for* $q = 1, 2$ *and some* $C > 0$. *Then it holds, for all* $\theta > 0$,

$$\mathrm{P}\{\max_{1\leq l\leq L} |\sum_{i=1}^{n} (z_{l,i} - \mathrm{E}_{i-1} z_{l,i})| \geq \kappa n^{1/2}\} \leq \frac{(L+1)\theta^3 C}{\kappa n} + \frac{\theta C}{\kappa} + 4L\exp(-\frac{\kappa\theta}{14}).$$

B A metric on $\mathbb{R}$ and some inequalities

A chaining argument involves the partition of the axis so that a metric is set up in this section. Then one inequality about the multiplicative shift related to this metric is shown in order to prove the empirical process results.

Before getting into the lemma, it is necessary to first introduce the background and some notations. Define the function

$$J_{i,p}(x,y) = (\frac{\varepsilon_i}{\sigma})^p\{1_{(\varepsilon_i/\sigma\leq y)} - 1_{(\varepsilon_i/\sigma\leq x)}\}, \tag{B.1}$$

where $p \in \mathbb{N}_0$ and ε_i/σ has probability density $f(x)$ and cumulative distribution $F(x)$. In this context, $J_{i,p}(x,y)$ of order 2^r is our main interest, where $r \in \mathbb{N}$. Note that $\epsilon^{2^r p}$ is non-negative since $2^r p$ is even for $p \in \mathbb{N}_0$ and $r \in \mathbb{N}$. Thus, define the increasing function

$$\mathrm{H}_r(x) = \int_{-\infty}^{x}(1+\epsilon^{2^r p})f(\epsilon)d\epsilon. \tag{B.2}$$

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

$$H_r = \mathrm{H}_r(\infty) = \int_{-\infty}^{\infty}(1+\epsilon^{2^r p})f(\epsilon)d\epsilon. \tag{B.3}$$

How to choose $r \in \mathbb{N}$ and $p \in \mathbb{N}_0$ will be more clear in the rest of the paper. The intuition of the $\mathrm{H}_r(x)$ function can be obtained through setting $p = 0$. By doing this, $\mathrm{H}_r(x) = 2F(x)$ with the derivative $\mathrm{H}_r'(x) = 2f(x)$ along with the constant $H_r = 2$. Therefore, $\mathrm{H}_r(x)$ is the generalization of the cumulative distribution function of ε_i/σ. Based on these notations, we introduce the following inequality which will be heavily used during the whole paper. For $x \leq y$ and $0 \leq s \leq r$, then

$$0 \leq |\mathrm{E}\{J_{i,p}(x,y)^{2^s}\}| \leq \mathrm{E}\{|J_{i,p}(x,y)|^{2^s}\} < \mathrm{H}_r(y) - \mathrm{H}_r(x), \tag{B.4}$$

as, for $q \geq p \geq 0$ and $\epsilon \in \mathbb{R}$, $|\epsilon^p| \leq |\epsilon^q| + 1$.

Suppose an assumption that $\int_{-\infty}^{\infty}|\epsilon|^{2^r p}f(\epsilon)d\epsilon < \infty$, so $H_r = \int_{-\infty}^{\infty}(1+\epsilon^{2^r p})f(\epsilon)d\epsilon = 1 + \int_{-\infty}^{\infty}\epsilon^{2^r p}f(\epsilon)d\epsilon < \infty$. In the context of chaining, partition the range of $\mathrm{H}_r(c)$ into K intervals of equal size. In other words, partition the support c into K intervals by the endpoints

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \tag{B.5}$$

and for $1 \leq k \leq K$,

$$\mathrm{H}_r(c_k) - \mathrm{H}_r(c_{k-1}) = \frac{H_r}{K}.$$

Let $c_{-k} = c_0$ for $k \in \mathbb{N}$. In the proof of the main theorem, the number of intervals is chosen as $K = \mathrm{int}(H_r n^{1/2}/\delta)$, where $\delta > 0$ is a very small number.

The first inequality concerns the H_r-distance between $c_{\psi\dagger} = c_\psi(1+n^{-1/2}a/\sigma)$ and c_ψ. $c_\psi = F^{-1}(\psi)$ is the quantile of the distribution function $F(c)$. c_ψ is varying in its domain, $-\infty < c_\psi < \infty$, rather than being fixed in Johansen and Nielsen (2009). $\widetilde{a} = n^{1/2}(\widetilde{\sigma} - \sigma)$ measures the estimation error for σ. Instead of using an assumption $|\widetilde{a}| \leq B$, the weaker assumption $|\widetilde{a}| \leq n^{1/4-\eta}B$ is allowed. $B > 0$ is the finite number and $0 < \eta \leq 1/4$. The lemma is to control the multiplicative shift of c_ψ especially when c_ψ is large. This will be used to prove the asymptotic results of the weighted and marked empirical processes by applying a chaining argument with the iterated martingale inequality.

Lemma B.1. *Suppose* $\sup_{c\in\mathbb{R}} c(1 + c^{2r_p})f(c) < \infty$, *and* $\int_{-\infty}^{\infty} |\epsilon|^{2r_p} f(\epsilon)d\epsilon < \infty$ *so that* $H_r < \infty$. *If we choose* $K = \mathrm{int}(H_r n^{1/2}/\delta)$, *then, for any* $B > 0$ *and some* $0 < \eta \leq 1/4$, *there exists* $C > 0$ *and it holds that*

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

Proof of Lemma B.1. In order to demonstrate the above inequality, we should show it in each of four situations where c_ψ and a are greater or less than zero, since $-\infty < c_\psi < \infty$ can be any real number for $0 \leq \psi \leq 1$ and $|a| \leq n^{1/4-\eta}B$. The proof of the case $c_\psi \leq 0$ is similar to that of $c_\psi \geq 0$. Therefore, without loss of generality, we focus on the proof of $c_\psi \geq 0$. Note, by the assumption, $H_r < \infty$ and choose $K = \mathrm{int}(H_r n^{1/2}/\delta)$.

1. *Let* $c_\psi \geq 0$ *and* $0 \leq a \leq n^{1/4-\eta}B$. In this case, $c_\psi \leq c_\psi(1 + n^{-1/2}a/\sigma)$, and since $\mathrm{H}_r(x)$ is a positive and increasing function for any x, then we have

$$|\mathrm{H}_r(c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| = \mathrm{H}_r(c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi).$$

By applying the first order Taylor expansion at the point c_ψ,

$$\mathrm{H}_r(c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) = \mathrm{H}_r(c_\psi) + (c_\psi n^{-1/2}\frac{a}{\sigma})\mathrm{H}'_r(c^*_\psi),$$

for $c_\psi \le c_\psi^* \le c_\psi(1+n^{-1/2}a/\sigma)$, along with $\mathrm{H}_r'(c_\psi^*) = (1+c_\psi^{*2^r p})f(c_\psi^*)$. Notice that c_ψ^* depends on both c_ψ and a. Based on these relationships,

$$\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi) = (c_\psi n^{-1/2}\frac{a}{\sigma})\mathrm{H}_r'(c_\psi^*) = c_\psi n^{-1/2}\frac{a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*).$$

As $c_\psi \le c_\psi^* \le c_\psi(1+n^{-1/2}a/\sigma)$, the following inequalities can be shown,

$$c_\psi n^{-1/2}\frac{a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*) \le c_\psi^* n^{-1/2}\frac{a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*) \le n^{-1/2}\frac{a}{\sigma}\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c).$$

Then, with $0 \le a \le n^{1/4-\eta}B$, it holds

$$n^{-1/2}\frac{a}{\sigma}\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c) \le n^{-1/2}\frac{1}{\sigma}n^{1/4-\eta}B\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c).$$

After some arrangements, we can conclude

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \le n^{-1/2}\frac{1}{\sigma}n^{1/4-\eta}B\{\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c)\}\frac{K}{H_r}\frac{H_r}{K}.$$

As $n^{-1/2}K/H_r \le \delta^{-1}$, the final inequality is

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \le Cn^{1/4-\eta}\frac{H_r}{K},$$

where $C = (\delta\sigma)^{-1}B\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c)$ is the positive finite number by the assumption $\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c) < \infty$.

2. *Let $c_\psi \ge 0$ and $-n^{1/4-\eta}B \le a \le 0$.* In this case, $c_\psi(1+n^{-1/2}a/\sigma) \le c_\psi$ and $\mathrm{H}_r(x)$ is a positive and increasing function for any x, so

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| = \mathrm{H}_r(c_\psi) - \mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})).$$

The first order Taylor expansion at the point c_ψ shows

$$\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) = \mathrm{H}_r(c_\psi) + (c_\psi n^{-1/2}\frac{a}{\sigma})\mathrm{H}_r'(c_\psi^*),$$

along with $\mathrm{H}_r'(c_\psi^*) = (1+c_\psi^{*2^r p})f(c_\psi^*)$ for $c_\psi(1+n^{-1/2}a/\sigma) \le c_\psi^* \le c_\psi$. It can be shown here,

$$\mathrm{H}_r(c_\psi) - \mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) = (c_\psi n^{-1/2}\frac{-a}{\sigma})\mathrm{H}_r'(c_\psi^*) = c_\psi n^{-1/2}\frac{-a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*).$$

Then there must exist n_0, and for any $n > n_0$ we have $1 + n^{-1/2}a/\sigma \geq 1/2$. We first investigate the situation where $n > n_0$ and subsequently $n \leq n_0$. For $n > n_0$, since $c_\psi(1 + n^{-1/2}a/\sigma) \leq c_\psi^*$, $c_\psi \leq c_\psi^*/(1 + n^{-1/2}a/\sigma) \leq 2c_\psi^*$. We can obtain

$$c_\psi n^{-1/2}\frac{-a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*) \leq 2c_\psi^* n^{-1/2}\frac{-a}{\sigma}(1+c_\psi^{*2^r p})f(c_\psi^*) \leq 2n^{-1/2}\frac{-a}{\sigma}\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c).$$

Since $-n^{1/4-\eta}B \leq a \leq 0$, the inequality can be derived further,

$$2n^{-1/2}\frac{-a}{\sigma}\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c) \leq 2n^{-1/2}\frac{1}{\sigma}n^{1/4-\eta}B\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c).$$

Thus, the following can be attained,

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \leq n^{-1/2}\frac{1}{\sigma}n^{1/4-\eta}2B\{\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c)\}\frac{K}{H_r}\frac{H_r}{K}.$$

As $n^{-1/2}K/H_r \leq \delta^{-1}$, the final inequality is

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \leq Cn^{1/4-\eta}\frac{H_r}{K},$$

where $C = (\delta\sigma)^{-1}2B\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c)$ is the positive finite number by the assumption $\sup_{c\in\mathbb{R}} c(1+c^{2^r p})f(c) < \infty$.

Secondly, prove the lemma in the case where $n \leq n_0$. Since $\mathrm{H}_r(x)$ is a positive and increasing function, then $\mathrm{H}_r(x) \leq H_r$ for any x. By this and the triangle inequality, it holds

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \leq |\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma}))| + |\mathrm{H}_r(c_\psi)| \leq 2H_r.$$

Since $K = \mathrm{int}(H_r n^{1/2}/\delta)$ and $n \leq n_0$, then we have

$$2H_r = 2\frac{K}{n^{1/4-\eta}}n^{1/4-\eta}\frac{H_r}{K} \leq 2\frac{H_r n^{1/4+\eta}}{\delta}n^{1/4-\eta}\frac{H_r}{K} \leq 2\frac{H_r n_0^{1/4+\eta}}{\delta}n^{1/4-\eta}\frac{H_r}{K}.$$

Note $0 < \eta \leq 1/4$ in the above. Therefore the following holds,

$$|\mathrm{H}_r(c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathrm{H}_r(c_\psi)| \leq 2\frac{H_r n_0^{1/4+\eta}}{\delta}n^{1/4-\eta}\frac{H_r}{K} \leq Cn^{1/4-\eta}\frac{H_r}{K},$$

where $C = 2H_r n_0^{1/4+\eta}/\delta$ is the positive finite number by the assumption $H_r < \infty$.

3. *Let* $c_\psi \leq 0$ *and* $0 \leq a \leq n^{1/4-\eta}B$. A quite similar argument can be applied in this situation as in 2.

4. *Let* $c_\psi \leq 0$ *and* $-n^{1/4-\eta}B \leq a \leq 0$. A quite similar argument can be applied in this situation as in 1. ■

C Proofs of Theorem 6.3, 6.4 and 6.5 in §6

Proof of Theorem 6.3. Note $c_{\psi^\dagger} = c_\psi(1 + n^{-1/2}a/\sigma)$ can be greater or less than c_ψ since a such that $|a| \leq n^{1/4-\eta}B$ is either positive or negative. Without loss of generality, assume $c_{\psi^\dagger} < c_\psi$. The situation where $c_{\psi^\dagger} > c_\psi$ can be demonstrated using the same argument. Then, the aim is to prove $\sup_{0\leq\psi\leq 1}\sup_{|a|\leq n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1)$.

1. *Partition the support.* For $\delta, n > 0$ partition the range of quantiles c as laid out in (B.5) with $K = \mathrm{int}(H_r n^{1/2}/\delta)$ using Assumption 6.1(ia) $\int_{-\infty}^{\infty}|\epsilon|^{2^r p/\nu}f(\epsilon)d\epsilon < \infty$ with $\nu = 1$ so that $H_r < \infty$.

2. *Assign $c_{\psi^\dagger} = c_\psi(1 + n^{-1/2}a/\sigma)$ and c_ψ to the partitioned support.* Consider the first situation where $c_{\psi^\dagger}$ and c_ψ are in the same partitioned interval. There must exist $c_{k^\dagger-1} = c_{k-1}$ and $c_{k^\dagger} = c_k$ such that $c_{k-1} < c_{\psi^\dagger} < c_\psi \leq c_k$. For $k < K$ so $c_{\psi^\dagger} < c_\psi \leq c_{K-1}$, then relate $c_{\psi^\dagger}$ and c_ψ to the nearest right grid pont c_k. Whereas for $k = K$ so $c_{K-1} < c_{\psi^\dagger} < c_\psi < c_K = \infty$, assign $c_{\psi^\dagger}$ and c_ψ to the nearest left grid point c_{K-1}.

The second case happens when $c_{\psi^\dagger}$ and c_ψ are located into the different partitioned intervals. There exist $c_{k^\dagger-1}$, $c_{k^\dagger}$, c_{k-1} and c_k, so $c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger} \leq c_{k-1} < c_\psi \leq c_k$. For $1 \leq k^\dagger < k \leq K$, relate $c_{\psi^\dagger}$ to its nearest right grid point $c_{k^\dagger}$ and assign c_ψ to its nearest left grid point c_{k-1}. Notice that some terms in the first and second situation will collapse after applying the chaining.

3. *Apply chaining.* In order to apply a chaining argument, we need to consider two situations discussed in the step 2 assignment. In the first case where $c_{\psi^\dagger}$ and c_ψ are in the same partitioned interval, we have $\mathbb{F}_n^{g,p}(0, c_{k^\dagger}) = \mathbb{F}_n^{g,p}(0, c_k)$ as $c_{k^\dagger} = c_k$. For $k < K$, add and subtract $\mathbb{F}_n^{g,p}(0, c_{k^\dagger})$ and $\mathbb{F}_n^{g,p}(0, c_k)$,

$$\begin{aligned}|\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)| &= |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{k^\dagger}) + \mathbb{F}_n^{g,p}(0, c_k) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\ &\leq |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{k^\dagger})| + |\mathbb{F}_n^{g,p}(0, c_k) - \mathbb{F}_n^{g,p}(0, c_\psi)|.\end{aligned}$$

For $k = K$, add and subtract $\mathbb{F}_n^{g,p}(0, c_{K-1})$,

$$\begin{aligned}|\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)| &= |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{K-1}) + \mathbb{F}_n^{g,p}(0, c_{K-1}) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\ &\leq |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{K-1})| + |\mathbb{F}_n^{g,p}(0, c_{K-1}) - \mathbb{F}_n^{g,p}(0, c_\psi)|.\end{aligned}$$

Then, it can be shown

$$\sup_{0\le\psi\le 1}\sup_{|a|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_\psi)| \tag{C.1}$$

$$\le \max_{1\le k^\dagger<K}\sup_{c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k^\dagger})|+\sup_{c_{K-1}<c_{\psi^\dagger}}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{K-1})|$$

$$+\max_{1\le k<K}\sup_{c_{k-1}<c_\psi\le c_k}|\mathbb{F}_n^{g,p}(0,c_\psi)-\mathbb{F}_n^{g,p}(0,c_k)|+\sup_{c_{K-1}<c_\psi}|\mathbb{F}_n^{g,p}(0,c_\psi)-\mathbb{F}_n^{g,p}(0,c_{K-1})|.$$

It suffices to show each term on the right hand side of (C.1) is $\mathrm{o_p}(1)$.

Furthermore, consider the second situation where $c_{\psi^\dagger}$ and c_ψ are in the different partitioned intervals, add and subtract $\mathbb{F}_n^{g,p}(0,c_{k^\dagger})$ and $\mathbb{F}_n^{g,p}(0,c_{k-1})$,

$$\begin{aligned}&|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_\psi)|\\ &=|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k^\dagger})+\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})+\mathbb{F}_n^{g,p}(0,c_{k-1})-\mathbb{F}_n^{g,p}(0,c_\psi)|\\ &\le|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k^\dagger})|+|\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})|+|\mathbb{F}_n^{g,p}(0,c_{k-1})-\mathbb{F}_n^{g,p}(0,c_\psi)|.\end{aligned}$$

Then, it can be shown

$$\sup_{0\le\psi\le 1}\sup_{|a|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_\psi)|\le\max_{1\le k^\dagger<k\le K}|\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})| \tag{C.2}$$

$$+\max_{1\le k^\dagger<K}\sup_{c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k^\dagger})| \tag{C.3}$$

$$+\max_{1<k\le K}\sup_{c_{k-1}<c_\psi\le c_k}|\mathbb{F}_n^{g,p}(0,c_\psi)-\mathbb{F}_n^{g,p}(0,c_{k-1})|. \tag{C.4}$$

Then it suffices to demonstrate terms (C.2), (C.3) and (C.4) are $\mathrm{o_p}(1)$ for $c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}\le c_{k-1}<c_\psi\le c_k$.

In summary, notice the fact that two terms on the right hand side of (C.1) are $\mathrm{o_p}(1)$ in the first case implies (C.3) and (C.4) are $\mathrm{o_p}(1)$ in the second case. By combining two situations mentioned above, the following three non-redundant terms are essential to prove the problem.

$$\max_{1\le k^\dagger<k\le K}|\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})|\quad for\quad c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}\le c_{k-1}<c_\psi\le c_k, \tag{C.5}$$

$$\max_{1\le k^\dagger<K}\sup_{c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k^\dagger})|+\sup_{c_{K-1}<c_{\psi^\dagger}}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_{K-1})|, \tag{C.6}$$

and

$$\max_{1\le k<K}\sup_{c_{k-1}<c_\psi\le c_k}|\mathbb{F}_n^{g,p}(0,c_\psi)-\mathbb{F}_n^{g,p}(0,c_k)|+\sup_{c_{K-1}<c_\psi}|\mathbb{F}_n^{g,p}(0,c_\psi)-\mathbb{F}_n^{g,p}(0,c_{K-1})|. \quad \text{(C.7)}$$

Therefore, it suffices to demonstrate (C.5), (C.6) and (C.7) are $\mathrm{o_p}(1)$.

4. *The term (C.5) is* $\mathrm{o_p}(1)$. Firstly notice that if $k^\dagger = k-1$ and $c_{k^\dagger} = c_{k-1}$, the problem in this step is immediately solved since the term (C.5) is 0 by its definition. In order to analyze the problem, it is better to rewrite $\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})$ into the martingale form using the notation $J_{i,p}(c_{k-1},c_{k^\dagger})$. See the definition (B.1).

$$\begin{aligned}
&\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})\\
&=n^{1/2}\{\widehat{F}_n^{g,p}(0,c_{k^\dagger})-\overline{F}_n^{g,p}(0,c_{k^\dagger})\}-n^{1/2}\{\widehat{F}_n^{g,p}(0,c_{k-1})-\overline{F}_n^{g,p}(0,c_{k-1})\}\\
&=n^{-1/2}\{\sum_{i=1}^n g_{in}\varepsilon_i^p 1_{(\varepsilon_i\le\sigma c_{k^\dagger})}-\sum_{i=1}^n g_{in}\mathrm{E}_{i-1}[\varepsilon_i^p 1_{(\varepsilon_i\le\sigma c_{k^\dagger})}]\}\\
&\quad-n^{-1/2}\{\sum_{i=1}^n g_{in}\varepsilon_i^p 1_{(\varepsilon_i\le\sigma c_{k-1})}-\sum_{i=1}^n g_{in}\mathrm{E}_{i-1}[\varepsilon_i^p 1_{(\varepsilon_i\le\sigma c_{k-1})}]\}\\
&=n^{-1/2}\sum_{i=1}^n g_{in}\varepsilon_i^p\{1_{(\varepsilon_i\le\sigma c_{k^\dagger})}-1_{(\varepsilon_i\le\sigma c_{k-1})}\}\\
&\quad-n^{-1/2}\sum_{i=1}^n g_{in}\mathrm{E}_{i-1}[\varepsilon_i^p\{1_{(\varepsilon_i\le\sigma c_{k^\dagger})}-1_{(\varepsilon_i\le\sigma c_{k-1})}\}]\\
&=n^{-1/2}\sum_{i=1}^n \sigma^p g_{in}J_{i,p}(c_{k-1},c_{k^\dagger})-n^{-1/2}\sum_{i=1}^n \sigma^p g_{in}\mathrm{E}_{i-1}J_{i,p}(c_{k-1},c_{k^\dagger})\\
&=n^{-1/2}\sum_{i=1}^n \sigma^p g_{in}\{J_{i,p}(c_{k-1},c_{k^\dagger})-\mathrm{E}_{i-1}J_{i,p}(c_{k-1},c_{k^\dagger})\}.
\end{aligned}$$

Since the term (C.5) is the maximum of a family of martingales, Theorem A.3 can be directly applied to show it converges to 0 in probability if two technical conditions are satisfied in the theorem. Theorem A.3 aims to show for any $\kappa>0$ and $n\to\infty$

$$\lim_{n\to\infty}\mathrm{P}\{\max_{1\le l\le L}|\sum_{i=1}^n(z_{l,i}-\mathrm{E}_{i-1}z_{l,i})|\ge\kappa n^\upsilon\}=0.$$

Let $\upsilon=1/2$ and let g_{in} have coordinates $g_{in}^*=\sigma^p g_{in}$. Then $z_{l,i}=g_{in}^*J_{i,p}(c_{k-1},c_{k^\dagger})$, where l represents the indices $k^\dagger,k$ with $L\le K^2$. We can conclude the term (C.5) $\max_{1\le k^\dagger<k\le K}|\mathbb{F}_n^{g,p}(0,c_{k^\dagger})-\mathbb{F}_n^{g,p}(0,c_{k-1})|=\mathrm{o_p}(1)$ for $c_{k^\dagger-1}<c_{\psi^\dagger}\le c_{k^\dagger}\le c_{k-1}<c_\psi\le c_k$ by demonstrating two conditions of Theorem A.3.

The parameter λ. The set of indices l has size $L = \mathrm{O}(n^{\lambda})$. $L \leq K^2 = \mathrm{O}(n)$ as $K = \mathrm{O}(H_r n^{1/2}/\delta)$. Thus $\lambda = 1$.

The parameter ς. Note $\mathrm{E}D_r = \mathrm{O}(n^{\varsigma})$ where $D_r = \max_{1\leq l\leq L}\sum_{i=1}^{n}\mathrm{E}_{i-1}z_{l,i}^{2^r}$ for $1 \leq r \leq \bar{r}$ and $\bar{r} \in \mathbb{N}$. $\bar{r} = 2$ can be chosen in this proof. By construction of partition and assignment in the step 1 and 2 and Lemma B.1 using Assumption 6.1(ia, ib), for some $0 < \eta \leq 1/4$ and $C > 0$,

$$|\mathrm{H}_r(c_{k^\dagger}) - \mathrm{H}_r(c_{k-1})| = \mathrm{H}_r(c_{k-1}) - \mathrm{H}_r(c_{k^\dagger}) \leq \mathrm{H}_r(c_{\psi}) - \mathrm{H}_r(c_{\psi^\dagger}) \leq Cn^{1/4-\eta}\frac{H_r}{K}.$$

Note the above relationship holds since $c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger} \leq c_{k-1} < c_{\psi} \leq c_k$ and $\mathrm{H}_r(x)$ is an increasing positive function. By the inequality (B.4),

$$\mathrm{E}_{i-1}J_{i,p}^{2^r}(c_{k-1}, c_{k^\dagger}) = \mathrm{E}J_{i,p}^{2^r}(c_{k-1}, c_{k^\dagger}) \leq |\mathrm{H}_r(c_{k^\dagger}) - \mathrm{H}_r(c_{k-1})| \leq Cn^{1/4-\eta}\frac{H_r}{K}.$$

Notice that the first equality follows from the independence of ε_i with $\mathcal{F}_{i-1}$. The inequality shown above is uniform in i and indices $k^\dagger, k$ such that $c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger} \leq c_{k-1} < c_{\psi} \leq c_k$. For $1 \leq r \leq \bar{r} = 2$, and indices $k^\dagger, k$ so that $c_{k^\dagger-1} < c_{\psi^\dagger} \leq c_{k^\dagger} \leq c_{k-1} < c_{\psi} \leq c_k$,

$$D_r = \max_{1\leq l\leq L}\sum_{i=1}^{n}\mathrm{E}_{i-1}z_{l,i}^{2^r} = \max_{1\leq k^\dagger<k\leq K}\sum_{i=1}^{n} g_{in}^{*2^r}\mathrm{E}_{i-1}J_{i,p}^{2^r}(c_{k-1}, c_{k^\dagger}).$$

Then, we can get

$$D_r \leq \max_{1\leq k^\dagger<k\leq K}\sum_{i=1}^{n} g_{in}^{*2^r} Cn^{1/4-\eta}\frac{H_r}{K} \leq Cn^{1/4-\eta}\frac{H_r}{K}\sum_{i=1}^{n} g_{in}^{*2^r}.$$

Therefore

$$\mathrm{E}D_r \leq Cn^{1/4-\eta}\frac{H_r}{K}n\mathrm{E}\{n^{-1}\sum_{i=1}^{n} g_{in}^{*2^r}\} = n^{1/4-\eta}n^{-1/2}n\cdot\mathrm{O}(1) = \mathrm{O}(n^{3/4-\eta}),$$

since by Assumption 6.1$(iiia)$ $\mathrm{E}\{n^{-1}\sum_{i=1}^{n} g_{in}^{*2^r}\} = \mathrm{O}(1)$ and $K = \mathrm{O}(H_r n^{1/2}/\delta)$. Finally, $\varsigma = 3/4 - \eta$ as $\mathrm{E}D_r = \mathrm{O}(n^{\varsigma})$ for $r \leq \bar{r} = 2$.

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

Condition (ii) is that $\varsigma + \lambda < \upsilon 2^{\bar{r}}$ where $\bar{r} = 2$. This also holds since $\eta > 0$ so $\varsigma + \lambda = 7/4 - \eta < 2 = \upsilon 2^{\bar{r}}$.

5. *Decompose the term (C.6).* Again, it is necessary to write the first term in (C.6) into the martingale form as in the step 4. By applying the triangle and Jensen's

inequalities, we can obtain

$$\begin{aligned}
|\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{k^\dagger})| &= |n^{-1/2} \sum_{i=1}^{n} g_{in}^* \{J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger}) - \mathrm{E}_{i-1} J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})\}| \\
&\leq n^{-1/2} \sum_{i=1}^{n} |g_{in}^* \{J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger}) - \mathrm{E}_{i-1} J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})\}| \\
&\leq n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})| + |\mathrm{E}_{i-1} J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})|\} \\
&\leq n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})| + \mathrm{E}_{i-1} |J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})|\}.
\end{aligned}$$

Since it is easy to see $|J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})| \leq |J_{i,p}(c_{k^\dagger - 1}, c_{k^\dagger})|$ for $c_{k^\dagger - 1} < c_{\psi^\dagger} \leq c_{k^\dagger}$ and $1 \leq k^\dagger < K$,

$$\begin{aligned}
&\max_{1 \leq k^\dagger < K} \sup_{c_{k^\dagger - 1} < c_{\psi^\dagger} \leq c_{k^\dagger}} |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{k^\dagger})| \\
\leq &\max_{1 \leq k^\dagger < K} \sup_{c_{k^\dagger - 1} < c_{\psi^\dagger} \leq c_{k^\dagger}} n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})| + \mathrm{E}_{i-1} |J_{i,p}(c_{\psi^\dagger}, c_{k^\dagger})|\} \\
\leq &\max_{1 \leq k^\dagger < K} n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{k^\dagger - 1}, c_{k^\dagger})| + \mathrm{E}_{i-1} |J_{i,p}(c_{k^\dagger - 1}, c_{k^\dagger})|\}.
\end{aligned}$$

The similar argument can be carried out for analyzing the second term in (C.6). Again, after using the triangle and Jensen's inequalities,

$$|\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{K-1})| \leq n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{\psi^\dagger}, c_{K-1})| + \mathrm{E}_{i-1} |J_{i,p}(c_{\psi^\dagger}, c_{K-1})|\}.$$

Since $|J_{i,p}(c_{\psi^\dagger}, c_{K-1})| \leq |J_{i,p}(c_{K-1}, c_K)|$ for $c_{K-1} < c_{\psi^\dagger}$,

$$\begin{aligned}
&\sup_{c_{K-1} < c_{\psi^\dagger}} |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_{K-1})| \\
\leq &\sup_{c_{K-1} < c_{\psi^\dagger}} n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{\psi^\dagger}, c_{K-1})| + \mathrm{E}_{i-1} |J_{i,p}(c_{\psi^\dagger}, c_{K-1})|\} \\
\leq &\, n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{K-1}, c_K)| + \mathrm{E}_{i-1} |J_{i,p}(c_{K-1}, c_K)|\}.
\end{aligned}$$

Therefore, the term (C.6) can be decomposed into two terms. It suffices to demonstrate

$$\max_{1 \leq k^\dagger \leq K} n^{-1/2} \sum_{i=1}^{n} |g_{in}^*| \{|J_{i,p}(c_{k^\dagger - 1}, c_{k^\dagger})| - \mathrm{E}_{i-1} |J_{i,p}(c_{k^\dagger - 1}, c_{k^\dagger})|\} = \mathrm{o_p}(1), \tag{C.8}$$

and

$$2 \max_{1\le k^\dagger\le K} n^{-1/2} \sum_{i=1}^{n} |g^*_{in}| \mathrm{E}_{i-1} |J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})| = \mathrm{o_p}(1). \tag{C.9}$$

6. *The term (C.8) is* $\mathrm{o_p}(1)$. This can be proved by applying Theorem A.3. To see this, note that the term (C.8) is the maximum of a family of martingales of the required form with $z_{l,i} = |g^*_{in}||J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})|$. In this case, $\upsilon = 1/2$ and the index $l = k^\dagger$ has the size $L = K$. It suffices to set $\bar{r} = 2$ as in the step 4. Two conditions of Theorem A.3 need to be shown.

The parameter λ. $\lambda = 1/2$ since $L = K = \mathrm{O}(H_r n^{1/2}/\delta)$ and $L = \mathrm{O}(n^\lambda)$.

The parameter ς. By the inequality (B.4) and since that the ε_i is independent with $\mathcal{F}_{i-1}$,

$$\mathrm{E}_{i-1}|J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})|^{2^r} = \mathrm{E}J^{2^r}_{i,p}(c_{k^\dagger-1}, c_{k^\dagger}) \le \mathrm{H}_r(c_{k^\dagger}) - \mathrm{H}_r(c_{k^\dagger-1}) = \frac{H_r}{K}.$$

Notice that the above inequality is uniform in i and the index $k^\dagger$. For $1 \le r \le \bar{r}$,

$$D_r = \max_{1\le l\le L} \sum_{i=1}^{n} \mathrm{E}_{i-1} z^{2^r}_{l,i} = \max_{1\le k^\dagger\le K} \sum_{i=1}^{n} |g^*_{in}|^{2^r} \mathrm{E}_{i-1}|J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})|^{2^r}.$$

Then, we can obtain

$$D_r \le \max_{1\le k^\dagger\le K} \sum_{i=1}^{n} |g^*_{in}|^{2^r} \frac{H_r}{K} \le \frac{H_r}{K} \sum_{i=1}^{n} |g^*_{in}|^{2^r}.$$

Therefore

$$\mathrm{E}D_r \le \frac{H_r}{K} n \mathrm{E}\{n^{-1} \sum_{i=1}^{n} |g^*_{in}|^{2^r}\} = n^{-1/2} n \cdot \mathrm{O}(1) = \mathrm{O}(n^{1/2}),$$

since by Assumption 6.1(*iiia*) $\mathrm{E}\{n^{-1}\sum_{i=1}^{n} |g^*_{in}|^{2^r}\} = \mathrm{O}(1)$ and $K = \mathrm{O}(H_r n^{1/2}/\delta)$. Finally, $\varsigma = 1/2$ as $\mathrm{E}D_r = \mathrm{O}(n^\varsigma)$ for $r \le \bar{r} = 2$.

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

Condition (*ii*) *holds*, since $\varsigma + \lambda = 1 < 2 = \upsilon 2^{\bar{r}}$.

7. *Bounding the term (C.9).* Again by the inequality (B.4) and since that the ε_i is independent with $\mathcal{F}_{i-1}$,

$$\mathrm{E}_{i-1}|J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})| = \mathrm{E}|J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})| \le \mathrm{H}_r(c_{k^\dagger}) - \mathrm{H}_r(c_{k^\dagger-1}) = \frac{H_r}{K}.$$

Notice that the above inequality is uniform in i and the index $k^\dagger$. Thus,

$$\max_{1\le k^\dagger\le K} n^{-1/2} \sum_{i=1}^{n} |g^*_{in}| \mathrm{E}_{i-1}|J_{i,p}(c_{k^\dagger-1}, c_{k^\dagger})| \le \max_{1\le k^\dagger\le K} n^{-1/2} \sum_{i=1}^{n} |g^*_{in}| \frac{H_r}{K} = \frac{H_r}{K} n^{-1/2} \sum_{i=1}^{n} |g^*_{in}|.$$

The term (C.9) is bounded by the rate $\mathrm{O_p}(\delta)$ for $\delta > 0$ by Assumption 6.1(*iiib*) that $n^{-1}\sum_{i=1}^{n}|g_{in}^{*}| = \mathrm{O_p}(1)$ and $K = \mathrm{O}(H_r n^{1/2}/\delta)$. Choose δ sufficiently small, we can conclude that

$$\max_{1\le k^{\dagger}\le K} n^{-1/2}\sum_{i=1}^{n}|g_{in}^{*}|\mathrm{E}_{i-1}|J_{i,p}(c_{k^{\dagger}-1}, c_{k^{\dagger}})| = \mathrm{o_p}(1).$$

8. *The term (C.7) is* $\mathrm{o_p}(1)$. This can be shown by the same argument as in the steps $5, 6, 7$.

9. *Consider the situation where* $c_{\psi^{\dagger}} > c_{\psi}$. This can be demonstrated by the same argument as in the case $c_{\psi^{\dagger}} < c_{\psi}$ from the step 2 to 8. ■

Proof of Theorem 6.4. Decompose the term $|\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}a/\sigma)) - \mathbb{F}_n^{g,p}(0, c_\psi)|$ into two terms by adding and subtracting $\mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}a/\sigma))$,

$$\begin{aligned}
&|\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\
&= |\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) \\
&\quad + \mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\
&\le |\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma}))| \\
&\quad + |\mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)|.
\end{aligned}$$

Then, it suffices to prove

$$\sup_{0\le\psi\le 1}\sup_{|a|,|b|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma}))| = \mathrm{o_p}(1), \quad \text{(C.10)}$$

and

$$\sup_{0\le\psi\le 1}\sup_{|a|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1). \quad \text{(C.11)}$$

Since $|a| \le n^{1/4-\eta}B$ and c_ψ is varying in its domain, $c_{\psi^{\dagger}} = c_\psi(1+n^{-1/2}a/\sigma)$ performs very similar to c_ψ and also can be varying in the interval $[-\infty, +\infty]$. Based on this, (C.10) can be rewritten as

$$\begin{aligned}
&\sup_{0\le\psi\le 1}\sup_{|a|,|b|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(b, c_\psi(1+n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi(1+n^{-1/2}\frac{a}{\sigma}))| \\
&= \sup_{0\le\psi^{\dagger}\le 1}\sup_{|b|\le n^{1/4-\eta}B}|\mathbb{F}_n^{g,p}(b, c_{\psi^{\dagger}}) - \mathbb{F}_n^{g,p}(0, c_{\psi^{\dagger}})|.
\end{aligned}$$

Then (C.10) can be proved using Theorem 6.1 by Assumption 6.1$(i, ii, iiia)$ with $\nu = 1$ and an $r \geq 2$ such that $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$ for some $0 \leq \kappa < \eta \leq 1/4$. With Assumption 6.1$(ia, ib, ii, iii)$ with $\nu = 1$ and an $r \geq 2$, Theorem 6.3 shows (C.11). ■

Proof of Theorem 6.5. Note $c_{\psi^\dagger} = c_\psi(1+n^{-1/2}a/\sigma)$. The main term can be rewritten as the following,

$$|\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma}) + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_\psi)| = |\mathbb{F}_n^{g,p}(b, c_{\psi^\dagger} + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_\psi)|.$$

Then it can be decomposed into two terms by adding and subtracting $\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger})$,

$$\begin{aligned}
&|\mathbb{F}_n^{g,p}(b, c_{\psi^\dagger} + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\
&= |\mathbb{F}_n^{g,p}(b, c_{\psi^\dagger} + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) + \mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)| \\
&\leq |\mathbb{F}_n^{g,p}(b, c_{\psi^\dagger} + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_{\psi^\dagger})| + |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)|.
\end{aligned}$$

Thus, the problem reduces to

$$\sup_{0\leq\psi^\dagger\leq 1} \sup_{|b|,|d|\leq n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(b, c_{\psi^\dagger} + n^{\kappa-1/2}d) - \mathbb{F}_n^{g,p}(0, c_{\psi^\dagger})| = \mathrm{o_p}(1), \tag{C.12}$$

and

$$\sup_{0\leq\psi\leq 1} \sup_{|a|\leq n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(0, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1). \tag{C.13}$$

Then (C.12) can be proved using Theorem 6.2 by Assumption 6.1$(i, ii, iiia)$ with $\nu = 1$ and an $r \geq 2$ such that $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim(x))$ for some $0 \leq \kappa < \eta \leq 1/4$. With Assumption 6.1$(ia, ib, ii, iii)$ with $\nu = 1$ and an $r \geq 2$, Theorem 6.3 shows (C.13). ■

D Proofs in §3, §4 and §5

Asymptotic results of the one-sided empirical process can be transferred to the absolute case. Thus, we first prove the result of the absolute empirical process in §5. Then, based on asymptotic expansions of empirical processes, the tightness result in §3 can be shown. At last, we provide the proof of the Poisson exceedence theory in §4.

D.1 Proofs of Theorem 5.1, 5.2 and 5.3 in §5

Proof of Theorem 5.1. For the absolute process, the asymptotic arguments are made on the probability scale $\psi = G(c_\psi)$. The absolute quantile is $c_\psi = G^{-1}(\psi) > 0$ instead of the quantile in the one-sided process $c_{\psi^*} = F^{-1}(\psi^*)$ defined in the interval $[-\infty, +\infty]$. $G(c)$ is the distribution function of $|\varepsilon_i|/\sigma$, whereas ε_i/σ has the distribution $F(c)$. After investigating the relationship between the one-sided and absolute empirical processes, asymptotic results of the one-sided process can be easily transferred to the absolute case. Since $1_{(|\varepsilon_i - x'_{in}b| \le \sigma c)} = 1_{(-\sigma c + x'_{in}b \le \varepsilon_i \le \sigma c + x'_{in}b)} = 1_{(\varepsilon_i \le \sigma c + x'_{in}b)} - 1_{(\varepsilon_i \le -\sigma c + x'_{in}b)}$, the process $\mathbb{G}_n^{g,p}(b,c)$ can be expressed in terms of $\mathbb{F}_n^{g,p}(b,c)$ through their definitions (5.5), (6.3) quite easily by $\mathbb{G}_n^{g,p}(b,c) = \mathbb{F}_n^{g,p}(b,c) - \mathbb{F}_n^{g,p}(b,-c)$ for any $c > 0$. For $c_\psi = G^{-1}(\psi) > 0$,

$$
\begin{aligned}
&|\mathbb{G}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{G}_n^{g,p}(0, c_\psi)| \\
&= |\{\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(b, -c_\psi(1 + n^{-1/2}\frac{a}{\sigma}))\} - \{\mathbb{F}_n^{g,p}(0, c_\psi) - \mathbb{F}_n^{g,p}(0, -c_\psi)\}| \\
&\le |\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| + |\mathbb{F}_n^{g,p}(b, -c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, -c_\psi)|.
\end{aligned}
$$

By Assumption 3.1(ia, iid, iie), Theorem 6.4 holds, for any $-\infty < c_\psi < \infty$,

$$
\sup_{0 \le \psi \le 1} \sup_{|a|, |b| \le n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{F}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1).
$$

Then we can demonstrate

$$
\sup_{0 \le \psi \le 1} \sup_{|a|, |b| \le n^{1/4-\eta}B} |\mathbb{G}_n^{g,p}(b, c_\psi(1 + n^{-1/2}\frac{a}{\sigma})) - \mathbb{G}_n^{g,p}(0, c_\psi)| = \mathrm{o_p}(1). \qquad \blacksquare
$$

Proof of Theorem 5.2. Asymptotic results of the one-sided empirical process can be transferred to the absolute process. We can simply repeat the idea here. Thus, as a similar argument in the proof of Theorem 5.1, Theorem 5.2 can be demonstrated using Theorem 6.5 by Assumption 3.1(ia, iid, iie). $\blacksquare$

Proof of Theorem 5.3. We can investigate the general class of empirical processes

$$
n^{-1/2} \sum_{i=1}^{n} g_{in} \varepsilon_i^p v_i^{a,b,c,d}, \quad v_i^{a,b,c,d} = 1_{(|\varepsilon_i - x'_{in}b| \le \sigma[c(1 + n^{-1/2}a/\sigma) + n^{-1/2}d])}. \tag{D.1}
$$

1. *Decompose the term (D.1).* We can decompose the term (D.1) as a sum of the following three terms,

$$
n^{-1/2} \sum_{i=1}^{n} g_{in} \varepsilon_i^p 1_{(|\varepsilon_i| \le \sigma c)}, \tag{D.2}
$$

$$n^{-1/2}\sum_{i=1}^{n} g_{in}\mathrm{E}_{i-1}\{\varepsilon_i^p[v_i^{a,b,c,d}-1_{(|\varepsilon_i|\le\sigma c)}]\}, \tag{D.3}$$

$$n^{-1/2}\sum_{i=1}^{n} g_{in}\varepsilon_i^p[v_i^{a,b,c,d}-1_{(|\varepsilon_i|\le\sigma c)}]-n^{-1/2}\sum_{i=1}^{n} g_{in}\mathrm{E}_{i-1}\{\varepsilon_i^p[v_i^{a,b,c,d}-1_{(|\varepsilon_i|\le\sigma c)}]\}. \tag{D.4}$$

Then, the first term in expansions can be immediately obtained as (D.2).

2. *Approximate the term (D.3).* The indicators $v_i^{a,b,c,d}$ can be rearranged as

$$1_{(-c(1+n^{-1/2}a/\sigma)-n^{-1/2}d+x_{in}'b/\sigma\le\varepsilon_i/\sigma\le c(1+n^{-1/2}a/\sigma)+n^{-1/2}d+x_{in}'b/\sigma)}.$$

Then, we can get

$$\begin{aligned}
&\mathrm{E}_{i-1}\{\varepsilon_i^p[v_i^{a,b,c,d}-1_{(|\varepsilon_i|\le\sigma c)}]\}\\
&=\sigma^p[\int_{-c(1+n^{-1/2}a/\sigma)-n^{-1/2}d+x_{in}'b/\sigma}^{c(1+n^{-1/2}a/\sigma)+n^{-1/2}d+x_{in}'b/\sigma}\epsilon^p f(\epsilon)d\epsilon-\int_{-c}^{c}\epsilon^p f(\epsilon)d\epsilon]\\
&=\sigma^p[\int_{c}^{c(1+n^{-1/2}a/\sigma)+n^{-1/2}d+x_{in}'b/\sigma}\epsilon^p f(\epsilon)d\epsilon-\int_{-c}^{-c(1+n^{-1/2}a/\sigma)-n^{-1/2}d+x_{in}'b/\sigma}\epsilon^p f(\epsilon)d\epsilon].
\end{aligned}$$

Apply the second order Taylor expansion to each of the above two integrals at points c and $-c$ respectively,

$$\int_{c}^{c(1+n^{-1/2}a/\sigma)+n^{-1/2}d+x_{in}'b/\sigma}\epsilon^p f(\epsilon)d\epsilon=(n^{-1/2}\frac{ac}{\sigma}+n^{-1/2}d+\frac{x_{in}'b}{\sigma})c^p f(c)+\mathrm{O}_p(n^{-1}),$$

$$\int_{-c}^{-c(1+n^{-1/2}a/\sigma)-n^{-1/2}d+x_{in}'b/\sigma}\epsilon^p f(\epsilon)d\epsilon=(-n^{-1/2}\frac{ac}{\sigma}-n^{-1/2}d+\frac{x_{in}'b}{\sigma})(-c)^p f(c)+\mathrm{O}_p(n^{-1}).$$

Note in the above, $f(c)=f(-c)$ as f is symmetric. And the last term is $\mathrm{O}_p(n^{-1})$ in the Taylor expansion due to the boundedness condition of the density f by Assumption 3.1(ia). Thus, it holds

$$\begin{aligned}
&\mathrm{E}_{i-1}\{\varepsilon_i^p[v_i^{a,b,c,d}-1_{(|\varepsilon_i|\le\sigma c)}]\}\\
&=\sigma^p c^p f(c)[1+(-1)^p]n^{-1/2}(\frac{ac}{\sigma}+d)+\sigma^p c^p f(c)[1-(-1)^p]\frac{x_{in}'b}{\sigma}+\mathrm{O}_p(n^{-1})\\
&=\sigma^p c^p f(c)\{[1+(-1)^p]n^{-1/2}(\frac{ac}{\sigma}+d)+[1-(-1)^p]\frac{x_{in}'b}{\sigma}\}+\mathrm{O}_p(n^{-1}).
\end{aligned}$$

At last, the term (D.3) can be approximated by

$$n^{-1/2}\sum_{i=1}^{n} g_{in}\sigma^p c^p f(c)\{[1+(-1)^p]n^{-1/2}(\frac{ac}{\sigma}+d)+[1-(-1)^p]\frac{x_{in}'b}{\sigma}\}, \tag{D.5}$$

along with the small term $\mathrm{O}_p(n^{-1/2})$. This is due to $n^{-1}\sum_{i=1}^n g_{in} = \mathrm{O}_p(1)$ by Assumption 3.1(iid), then $n^{-1/2}\sum_{i=1}^n g_{in}\mathrm{O}_p(n^{-1}) = n^{-1}\sum_{i=1}^n g_{in}\mathrm{O}_p(n^{-1/2}) = \mathrm{O}_p(n^{-1/2})$. Hence a second term in expansions have the form (D.5) which is a linearization of (D.3).

3. *Bounding the term (D.4).* The term (D.4) can be expressed as

$$\mathbb{G}_n^{g,p}(b, c(1+n^{-1/2}\frac{a}{\sigma}) + n^{-1/2}d) - \mathbb{G}_n^{g,p}(0,c).$$

See definitions of absolute empirical processes (5.3), (5.4) and (5.5). Then, by Assumption 3.1(ia, iid, iie), we can apply Theorem 5.2 for $\kappa = 0$. Thus we have

$$\sup_{0\le c<\infty}\sup_{|a|,|b|,|d|\le n^{1/4-\eta}B} |\mathbb{G}_n^{g,p}(b, c(1+n^{-1/2}\frac{a}{\sigma}) + n^{-1/2}d) - \mathbb{G}_n^{g,p}(0,c)| = \mathrm{o_p}(1),$$

where $B > 0$ is a large bound and $0 < \eta \le 1/4$ is small. It proves (D.4) is $\mathrm{o_p}(1)$.

4. *Set the particular g_{in} and p.* Expansions of Theorem 5.3 arise if set ($g_{in} = 1, p = 0$), ($g_{in} = 1, p = 2$), ($g_{in} = n^{1/2}N'x_i, p = 0$) and ($g_{in} = nN'x_ix_i'N, p = 0$) with $x_{in} = N'x_i$. Based on steps 1, 2 and 3, (D.2) and (D.5) are the first term and second term in expansions respectively. Moreover the term left can be argued

$$\sup_{0<c<\infty}\sup_{|a|,|b|,|d|\le n^{1/4-\eta}B} |R(a,b,c,d)| = \mathrm{o_p}(1). \qquad \blacksquare$$

D.2 Proofs of Theorem 3.1 and 3.2 in §3

Proof of Theorem 3.1. We first use a mathematical induction argument to establish tightness of $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ for any finite $m = 0, 1, 2, \cdots$, and all $c \in [c_0, \infty)$. This involves asymptotic expansions of empirical processes which give a 1-step relationship between $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ and $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$. Moreover, it can be argued $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ are tight and then have a fixed point for infinite iterations $m \to \infty$ and $c \in [c_0, \infty)$ by a geometric sequence argument. Thus, we can conclude tightness of $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ for all $0 \le m < \infty$ and $c_0 \le c < \infty$. Note c_0 is a finite positive number.

1. *Tightness of the initial estimators.* For the Algorithm 2.1, $\widehat{\beta}_c^{(0)} = \widetilde{\beta}$ and $(\widehat{\sigma}_c^{(0)})^2 = \widetilde{\sigma}^2$ are the full sample least squares estimators, which do not depend on the cut-off c. Thus by Assumption 3.1(iia, iib), $\widetilde{\beta}$ and $\widetilde{\sigma}^2$ are tight. Then, for $n \to \infty$, we have

$$\sup_{c_0\le c<\infty} |N^{-1}(\widehat{\beta}_c^{(0)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(0)} - \sigma)| = \mathrm{O_p}(1).$$

2. *Assume the m step estimates* $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ *are tight for all* $c \in [c_0, \infty)$. $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ can be expressed as in (3.1), (3.2),

$$\widehat{\beta}_c^{(m)} = (\sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m-1)})^{-1} (\sum_{i=1}^{n} x_i y_i v_{i,c}^{(m-1)}),$$

$$(\widehat{\sigma}_c^{(m)})^2 = \varsigma_c^{-2} (\sum_{i=1}^{n} v_{i,c}^{(m-1)})^{-1} \{\sum_{i=1}^{n} (y_i - x_i' \widehat{\beta}_c^{(m)})^2 v_{i,c}^{(m-1)}\}.$$

$v_{i,c}^{(m-1)} = 1_{(|y_i - x_i' \widehat{\beta}_c^{(m-1)}| \le \widehat{\sigma}_c^{(m-1)} c)}$ is a stochastic indicator in (3.3). $\widehat{\beta}_c^{(m-1)}$, $(\widehat{\sigma}_c^{(m-1)})^2$ are estimates in the $m-1$ step. $\varsigma_c^2 = \tau_2^c / \psi_c$ is the bias correction factor in (2.3). For $n \to \infty$, we suppose the m step estimates $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ satisfy

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O_p}(1).$$

3. *Comparison with the least squares.* The estimates $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ in the $m+1$ step are computed through,

$$\widehat{\beta}_c^{(m+1)} = (\sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m)})^{-1} (\sum_{i=1}^{n} x_i y_i v_{i,c}^{(m)}),$$

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

$v_{i,c}^{(m)} = 1_{(|y_i - x_i' \widehat{\beta}_c^{(m)}| \le \widehat{\sigma}_c^{(m)} c)}$ is a stochastic weight. After some manipulations, it holds

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = (N' \sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m)} N)^{-1} (N' \sum_{i=1}^{n} x_i \varepsilon_i v_{i,c}^{(m)}), \tag{D.6}$$

$$n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} = \varsigma_c^{-2} (n^{-1} \sum_{i=1}^{n} v_{i,c}^{(m)})^{-1} \{n^{-1/2} \sum_{i=1}^{n} (\varepsilon_i^2 - \varsigma_c^2 \sigma^2) v_{i,c}^{(m)}\} + \mathrm{O_p}(n^{-1/2}). \tag{D.7}$$

Thus, it is sufficient to analyze the following statistics for all $c \in [c_0, \infty)$

$$n^{-1} \sum_{i=1}^{n} v_{i,c}^{(m)}, \quad n^{-1/2} \sum_{i=1}^{n} \varepsilon_i^2 v_{i,c}^{(m)}, \quad N' \sum_{i=1}^{n} x_i \varepsilon_i v_{i,c}^{(m)}, \quad N' \sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m)} N. \tag{D.8}$$

4. *Asymptotic expansions of (D.8).* As in (5.7), we can express

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

See (5.1) and (5.2), $\widehat{b}_c^{(m)} = N^{-1}(\widehat{\beta}_c^{(m)} - \beta)$ and $\widehat{a}_c^{(m)} = n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are estimation errors for β and σ in the m step of the algorithm. Since we suppose the m step estimates $\widehat{\beta}_c^{(m)}$

and $(\widehat{\sigma}_c^{(m)})^2$ are tight for $c \in [c_0, \infty)$, then $\sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1)$. Thus, asymptotic expansions of (D.8) can be obtained using Theorem 5.3 with $\eta = 1/4$ and $d = 0$ by Assumption 3.1(ia, iid, iie). Replacing a, b, c, d by $\widehat{a}_c^{(m)}$, $\widehat{b}_c^{(m)}$, c, 0 can achieve

$$n^{-1}\sum_{i=1}^{n} v_{i,c}^{(m)} = n^{-1}\sum_{i=1}^{n} 1_{(|\varepsilon_i| \le \sigma c)} + 2n^{-1/2} f(c)\frac{\widehat{a}_c^{(m)} c}{\sigma} + R_v(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

$$n^{-1/2}\sum_{i=1}^{n} \varepsilon_i^2 v_{i,c}^{(m)} = n^{-1/2}\sum_{i=1}^{n} \varepsilon_i^2 1_{(|\varepsilon_i| \le \sigma c)} + 2\sigma^2 c^2 f(c)\frac{\widehat{a}_c^{(m)} c}{\sigma} + R_{v\varepsilon\varepsilon}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

$$N'\sum_{i=1}^{n} x_i \varepsilon_i v_{i,c}^{(m)} = N'\sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + 2cf(c)N'\sum_{i=1}^{n} x_i x_i' N\widehat{b}_c^{(m)} + R_{vx\varepsilon}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

$$N'\sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m)} N = N'\sum_{i=1}^{n} x_i x_i' 1_{(|\varepsilon_i| \le \sigma c)} N + 2n^{-1/2} f(c) N'\sum_{i=1}^{n} x_i x_i' N\frac{\widehat{a}_c^{(m)} c}{\sigma} + R_{vxx}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c).$$

Then for $n \to \infty$ and a finite positive number B, we have

$$\sup_{c_0 \le c < \infty} \sup_{|\widehat{a}_c^{(m)}|, |\widehat{b}_c^{(m)}| \le B} |R(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| = \mathrm{o_p}(1).$$

5. *Tightness of the $m+1$ step estimates $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ for all $c \in [c_0, \infty)$.* We analyze $\widehat{\beta}_c^{(m+1)}$ first and then $(\widehat{\sigma}_c^{(m+1)})^2$. Using asymptotic expansions of (D.8), (D.6) can be arranged for any $c \in [c_0, \infty)$ as

$$\widehat{b}_c^{(m+1)} = \{\psi_c \Sigma_n + \mathrm{O_p}(n^{-1/2})\}^{-1}\{N'\sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + 2cf(c)\Sigma_n \widehat{b}_c^{(m)} + \mathrm{o_p}(1)\}.$$

To obtain this, we apply Assumption 3.1(ia, iib) that $\sup_{c_0 \le c < \infty} cf(c) = \mathrm{O_p}(1)$ and $N'\sum_{i=1}^{n} x_i x_i' N = \Sigma_n$ along with $\sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1)$. Thus, we can get

$$\widehat{b}_c^{(m+1)} = \frac{2cf(c)}{\psi_c}\widehat{b}_c^{(m)} + (\psi_c \Sigma_n)^{-1} N'\sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c), \tag{D.9}$$

where $\sup_{c_0 \le c < \infty} \sup_{|\widehat{a}_c^{(m)}|, |\widehat{b}_c^{(m)}| \le B} |R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| = \mathrm{o_p}(1)$. Furthermore, by Assumption 3.1(ia, iib, iic) and tightness of the m step estimates for $c \in [c_0, \infty)$, we can argue from (D.9)

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta)| = \mathrm{O_p}(1), \quad \sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m+1)}| = \mathrm{O_p}(1).$$

After obtaining the tightness result for $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ can be investigated subsequently. Replace asymptotic expansions of (D.8) into (D.7), it holds for any $c \in [c_0, \infty)$

$$n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} = \varsigma_c^{-2}\{\psi_c + \mathrm{O_p}(n^{-1/2})\}^{-1}$$
$$\{n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i^2 - \varsigma_c^2\sigma^2)1_{(|\varepsilon_i| \le \sigma c)} + 2\sigma c f(c)(c^2 - \varsigma_c^2)\widehat{a}_c^{(m)} + \mathrm{o_p}(1)\} + \mathrm{O_p}(n^{-1/2}).$$

Assumption 3.1(ia) and $\sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1)$ are required to get this. Therefore, we have

$$n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} = 2\sigma\frac{c(c^2 - \varsigma_c^2)f(c)}{\tau_2^c}\widehat{a}_c^{(m)} + \frac{1}{\tau_2^c}n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i^2 - \varsigma_c^2\sigma^2)1_{(|\varepsilon_i| \le \sigma c)}$$
$$+ R_{\sigma^2}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c), \quad \text{(D.10)}$$

where $\sup_{c_0 \le c < \infty} \sup_{|\widehat{a}_c^{(m)}|, |\widehat{b}_c^{(m)}| \le B} |R_{\sigma^2}(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| = \mathrm{o_p}(1)$. By Assumption 3.1($ia$) that $\sup_{c_0 \le c < \infty} c^3 f(c) = \mathrm{O_p}(1)$ and $\sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1)$, it can be obtained from (D.10)

$$\sup_{c_0 \le c < \infty} |n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\}| = \mathrm{O_p}(1).$$

It can also be written as

$$\sup_{c_0 \le c < \infty} |n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma)| = \mathrm{O_p}(1), \quad \sup_{c_0 \le c < \infty} |\widehat{a}_c^{(m+1)}| = \mathrm{O_p}(1).$$

Based on this, (D.10) can be rearranged as

$$\widehat{a}_c^{(m+1)} = \frac{c(c^2 - \varsigma_c^2)f(c)}{\tau_2^c}\widehat{a}_c^{(m)} + \frac{1}{2\sigma\tau_2^c}n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i^2 - \varsigma_c^2\sigma^2)1_{(|\varepsilon_i| \le \sigma c)} + R_\sigma(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c), \quad \text{(D.11)}$$

where $\sup_{c_0 \le c < \infty} \sup_{|\widehat{a}_c^{(m)}|, |\widehat{b}_c^{(m)}| \le B} |R_\sigma(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| = \mathrm{o_p}(1)$. In summary, we have proved the $m+1$ step estimates $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ satisfy

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma)| = \mathrm{O_p}(1).$$

6. *Tightness of* $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ *for any finite* $m = 0, 1, 2, \cdots$, *and all* $c \in [c_0, \infty)$. By a mathematical induction argument, it can be concluded for $n \to \infty$ and any finite $m = 0, 1, 2, \cdots$,

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O_p}(1).$$

7. *Tightness of* $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ *for infinite iterations* $m \to \infty$, *and all* $c \in [c_0, \infty)$. For any finite m, it holds

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O_p}(1), \sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1).$$

Thus, the remainder terms in (D.9), (D.11) satisfy

$$\sup_{c_0 \le c < \infty} \sup_{|\widehat{a}_c^{(m)}|, |\widehat{b}_c^{(m)}| \le B} |R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| + |R_\sigma(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c)| = \mathrm{o_p}(1).$$

Then for any finite m and $c \in [c_0, \infty)$, (D.9), (D.11) can be expressed as

$$\widehat{b}_c^{(m+1)} = \frac{2cf(c)}{\psi_c} \widehat{b}_c^{(m)} + (\psi_c \Sigma_n)^{-1} N' \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + \mathrm{o_p}(1), \tag{D.12}$$

$$\widehat{a}_c^{(m+1)} = \frac{c(c^2 - \varsigma_c^2) f(c)}{\tau_2^c} \widehat{a}_c^{(m)} + \frac{1}{2\sigma\tau_2^c} n^{-1/2} \sum_{i=1}^{n} (\varepsilon_i^2 - \varsigma_c^2 \sigma^2) 1_{(|\varepsilon_i| \le \sigma c)} + \mathrm{o_p}(1). \tag{D.13}$$

We can immediately find (D.12), (D.13) have autoregressive structures between estimation errors $\widehat{b}_c^{(m+1)}$, $\widehat{a}_c^{(m+1)}$ and $\widehat{b}_c^{(m)}$, $\widehat{a}_c^{(m)}$. By Assumption 3.1(id), it holds

$$\sup_{c_0 \le c < \infty} |\frac{2cf(c)}{\psi_c}| < 1, \quad \sup_{c_0 \le c < \infty} |\frac{c(c^2 - \varsigma_c^2) f(c)}{\tau_2^c}| < 1.$$

Thus, (D.12), (D.13) are stationary autoregressive processes. We can apply a geometric sequence argument to stationary processes (D.12), (D.13), see the similar analysis made by Johansen and Nielsen (2013) or Theorem D.13 in Johansen and Nielsen (2014a). It can then be argued even for $m \to \infty$

$$\sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O_p}(1), \sup_{c_0 \le c < \infty} |\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O_p}(1).$$

8. *Tightness of* $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ *for all* $0 \le m < \infty$ *and* $c_0 \le c < \infty$. Collecting these evaluations, we can conclude for $n \to \infty$,

$$\sup_{0 \le m < \infty} \sup_{c_0 \le c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O_p}(1). \qquad \blacksquare$$

Proof of Theorem 3.2. Suppose Assumption 3.1(ia, id, ii) holds for each sub-sample set $\mathcal{I}_1$ and $\mathcal{I}_2$. $\widehat{\beta}_j$, $\widehat{\sigma}_j^2$ for $j = 1, 2$ are the split sample least squares estimators. These do not depend on c. $\widehat{\beta}_c^{(-1)}$, $(\widehat{\sigma}_c^{(-1)})^2$ are tight for any $c \in [c_0, \infty)$ since the Impulse Indicator Saturation Algorithm 2.2 uses the initial estimates as $\widehat{\beta}_j$, $\widehat{\sigma}_j^2$ for $j = 1, 2$. Thus, Theorem 3.2 can be demonstrated by a similar argument as in the proof of Theorem 3.1. $\blacksquare$

D.3 Proofs of Theorem 4.2 and 4.3 in §4

Proof of Theorem 4.2. For the Robustified Least Squares Algorithm 2.1, by Assumption 3.1(ia, id, ii), Theorem 3.1 demonstrates $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ are tight for all $m \in [0, \infty)$ and $c \in [c_0, \infty)$. Let c_n be defined from (4.3). We know $c_n \to \infty$ as $n \to \infty$. Then, it also holds that $\widehat{\beta}_{c_n}^{(m)}$, $(\widehat{\sigma}_{c_n}^{(m)})^2$ are tight for all $m \in [0, \infty)$. Therefore, by Assumption 3.1(ia, ib, ic, iid), Theorem 4.1 trivially applies so as to finish the proof. ■

Proof of Theorem 4.3. For the Impulse Indicator Saturation Algorithm 2.2, by Assumption 3.1(ia, id, ii), Theorem 3.2 demonstrates $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ are tight for all $m \in [-1, \infty)$ and $c \in [c_0, \infty)$. Let c_n be defined from (4.3) so that $c_n \to \infty$ as $n \to \infty$. Then, it also holds that $\widehat{\beta}_{c_n}^{(m)}$, $(\widehat{\sigma}_{c_n}^{(m)})^2$ are tight for all $m \in [-1, \infty)$. Therefore, by Assumption 3.1(ia, ib, ic, iid), Theorem 4.1 trivially applies so as to finish the proof. ■

References

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

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

[3] Chen, X.R. and Wu, Y.H. (1988). Strong consistency of M-estimates in linear models. *Journal of Multivariate Analysis* 27, 116-130.

[4] 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.

[5] He, X. and Portnoy, S. (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *The Annals of Statistics* 20, 2161-2167.

[6] 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.

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

[8] Hendry, D.F., Johansen, S. and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23(2), 317-335.

[9] 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.

[10] Hoover, K.D. and Perez, S.J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The Econometrics Journal* 2, 167-191.

[11] Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35, 73-101.

[12] 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.

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

[14] Johansen, S. and Nielsen, B. (2014a). Analysis of the Forward Search using some new results for martingales and empirical processes. Forthcoming *Bernoulli*.

[15] Johansen, S. and Nielsen, B. (2014b). Asymptotic theory of M-estimators for multiple regression. Discussion Paper.

[16] Johansen, S. and Nielsen, B. (2014c). Outlier detection algorithms for least squares time series regression. Discussion Paper.

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

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

[19] Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Chicester: Wiley.

[20] Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association* 79, 871-880.

[21] Rousseeuw, P.J. and van Driessen, K. (1998). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212-223.

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

[23] Víšek, J.Á. (2006a). The least trimmed squares. Part I: Consistency. *Kybernetika* 42, 1-36.

[24] Víšek, J.Á. (2006b). The least trimmed squares. Part II: $\sqrt{n}$-consistency. *Kybernetika* 42, 181-202.

[25] Víšek, J.Á. (2006c). The least trimmed squares. Part III: Asymptotic normality. *Kybernetika* 42, 203-224.

[26] 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.