Department of Economics
Discussion Paper Series

Marked and Weighted Empirical Processes of Residuals with Applications to Robust Regressions

Vanessa Berenguer-Rico and Bent Nielsen

Number 841
December, 2017

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

Marked and Weighted Empirical Processes of Residuals with Applications to Robust Regressions

Vanessa Berenguer-Rico* and Bent Nielsen†

16 December 2017

Abstract

A new class of marked and weighted empirical processes of residuals is introduced. The framework is general enough to accommodate both stationary and non-stationary regressions as well as a wide class of estimation procedures with applications in misspecification testing and robust statistics. Two applications are presented.

First, we analyze the relationship between truncated moments and linear statistical functionals of residuals. In particular, we show that the asymptotic behaviour of these functionals, expressed as integrals with respect to their empirical distribution functions, can be easily analyzed given the main theorems of the paper. In our context the integrands can be unbounded provided that the underlying distribution meets certain moment conditions. A general first order asymptotic approximation of the statistical functionals is derived and then applied to some cases of interest.

Second, the consequences of using the standard cumulant based normality test for robust regressions are analyzed. We show that the rescaling of the moment based statistic is case dependent, i.e., it depends on the truncation and the estimation method being used. Hence, using the standard least squares normalizing constants in robust regressions will lead to incorrect inferences. However, if appropriate normalizations, which we derive, are used then the test statistic is asymptotically chi-square.

1 Introduction

Weighted and marked empirical processes have many statistical applications. Two related types of empirical distribution functions have been analyzed previously. On the one hand, empirical distribution functions of residuals weighted by some function of the regressors have been studied by, for instance, [11, 13, 14]. Applications of this approach include asymptotic theory of robust estimators and goodness of fit tests. On the other hand, the empirical distribution functions of regressors marked by the residuals have been analyzed by, for instance, [6, 15, 20]. Applications of this approach include model specification checks, which rely on moment conditions between the errors and the regressors.

We consider the regression $y_i = x_i'\beta + \varepsilon_i$, where the regressors can be i.i.d., stationary or non-stationary while the error term is i.i.d. with an unknown scale σ; see §2.2 for details. Our concern is the exceedingly common research strategy of robustifying the regression, where the investigator first estimates the parameters by some consistent and, preferably, robust estimators $\tilde{\beta}, \tilde{\sigma}$; then deselects observations with large residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$; and finally re-estimates the parameters by estimators $\hat{\beta}, \hat{\sigma}$ with residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$ for the selected observations. Properties of the updated estimators have been analyzed by [18, 22] and more recently [9, 11, 12]. It is common to apply standard misspecification tests at the end of the above mentioned robust regression procedure.

*Department of Economics, University of Oxford, Mansfield College, and Programme for Economic Modelling.
†Department of Economics, University of Oxford, Nuffield College, and Programme for Economic Modelling.

The properties of such tests are unknown. To analyze these properties we generalize the class of weighted and marked empirical distribution functions further and consider

$$\hat{\mathsf{F}}_n^{w,m}(c) = n^{-1}\textstyle\sum_{i=1}^n w(x_{in})m(\hat{\varepsilon}_i/\hat{\sigma})1_{(\tilde{\varepsilon}_i\le\tilde{\sigma}c)}, \tag{1.1}$$

where $w(x_{in})$ is a weight function and $m(\hat{\varepsilon}_i/\hat{\sigma})$ is the mark, for some smooth function m. [11] allowed polynomial marks $m(\varepsilon_i/\sigma) = (\varepsilon_i/\sigma)^p$, without estimation error, which form the basis for analyzing estimators. Here, we allow for marks with estimation error, which are the basis for analyzing misspecification statistics. In doing so we generalize and improve some of the results in [11], simplify the proofs and relax the regularity assumptions.

We note that the marked and weighted distribution function $\hat{\mathsf{F}}_n^{w,m}(c)$ defines a very general class that includes a number of special cases which are relevant in theory and in applications. In particular, we consider two applications. First, we study linear statistical functionals of truncated residuals

$$\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n) = \textstyle\int_{-\infty}^c m(u)d\hat{\mathsf{F}}_n^{1,1}(u) = n^{-1}\textstyle\sum_{i=1}^n m(\tilde{\varepsilon}_i/\tilde{\sigma})1_{(\tilde{\varepsilon}_i\le\tilde{\sigma}c)}.$$

Here, $\hat{\mathsf{F}}_n(c) = n^{-1}\sum_{i=1}^n 1_{(\tilde{\varepsilon}_i\le\tilde{\sigma}c)} = \hat{\mathsf{F}}_n^{1,1}(c)$ is the empirical distribution of the scaled residuals. When m is unbounded the asymptotic theory of $\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n)$ cannot be derived from an asymptotic theory for $\hat{\mathsf{F}}_n$ in an obvious way. However, since $\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n) = \hat{\mathsf{F}}_n^{1,m}(c)$, this is a marked empirical process and an asymptotic theory follows from our results. Specifically, $\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n)$ is expanded in terms of $\mathsf{T}^{m,c}(\mathsf{F}_n)$.

In our second application, we consider testing for normality in robust regressions. It is common to check normality after having eliminated outlying observations using a moment based test statistic. The asymptotic properties of the moment based test after having implemented this robustifying procedure are not known. In order to derive those properties, we express the test statistics in terms of the weighted and marked empirical distribution function noting that the estimation errors in the indicators and marks will now be different. In particular, we look at statistics of the form

$$\hat{\mu}_{k,c} = \frac{\sum_{i=1}^n (\hat{\varepsilon}_i/\hat{\sigma})^k 1_{(|\tilde{\varepsilon}_i|\le\tilde{\sigma}c)}}{\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i|\le\tilde{\sigma}c)}},$$

for $k = 3, 4$. [4] suggest a related robust test using an information matrix approach. We show that the normalizations in the moment based test for normality depend on the estimation method being used and the outlier detection procedure. In particular, using the standard least squares normalization constants, that is 3, 6, 24, can lead to misleading inferences. We derive the correct normalizations for some robust estimators of interest.

The paper is organized as follows. §2 contains the main asymptotic results for marked and weighted empirical processes of residuals. Two applications of these general results follow: linear statistical functionals of truncated residuals are analyzed in §3 and normality tests for robust regressions are studied in §4. All proofs are collected in the Appendix.

2 Marked and Weighted Empirical Processes

2.1 Model and Notation

We consider marked and weighted empirical processes of residuals estimated from the model

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

which includes an intercept and where $\beta = (\mu, \alpha')'$ and $x_i = (1, z_i')'$ are k-vectors. In the empirical process theory we will introduce weights w_{in} that are typically derived from the regressors. The innovations are i.i.d. with distribution function $\mathsf{F}(c) = \mathsf{P}(\varepsilon_i \le \sigma c)$ with unknown scale $\sigma > 0$. Overall the model satisfies the following martingale structure.

Assumption 2.1. *Let $\mathcal{F}_i$ be a filtration so that ε_{i-1}, x_i, and w_{in} are $\mathcal{F}_{i-1}$-measurable and ε_i/σ is independent of $\mathcal{F}_{i-1}$ with distribution function* F *and positive density* f *on* $\mathbb{R}$ *with derivative* $\dot{\mathsf{f}}$.

We normalize the regressors as $x_{in} = N'x_i$ where the normalization matrix N is chosen so that the normalized information $\sum_{i=1}^n x_{in}x_{in}'$ has a positive limit. For example, $N = n^{-1/2}I_k$ for stationary regressors, $N = n^{-1}I_k$ for random walk regressors, while $N = \mathrm{diag}(n^{-1/2}, n^{-3/2})$ if $z_i = i$. The weights are typically chosen so that $n^{-1}\sum_{i=1}^n w_{in}$ has a limit and standard examples include 1, $n^{1/2}N'x_i$ or $nN'x_ix_i'N$.

The error of the initial regression estimator is scaled as $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)$, while the rescaled scale estimation error is $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)$. In standard situtations $\tilde{a}, \tilde{b}$ converge in distribution, but we allow $\tilde{a}, \tilde{b}$ to diverge slightly in the following theory. The standarized residuals can be rewritten using the model equation (2.1) as

$$\frac{\tilde{\varepsilon}_i}{\tilde{\sigma}} = \frac{y_i - x_i'\tilde{\beta}}{\tilde{\sigma}} = \frac{\varepsilon_i - x_i'NN^{-1}(\tilde{\beta} - \beta)}{\sigma + n^{-1/2}n^{1/2}(\tilde{\sigma} - \sigma)} = \frac{\varepsilon_i - x_{in}'\tilde{b}}{\sigma + n^{-1/2}\tilde{a}}.$$

Likewise we introduce notation $\hat{b} = N^{-1}(\hat{\beta} - \beta)$ and $\hat{a} = n^{1/2}(\hat{\sigma} - \sigma)$ for the updated estimators with $\hat{\varepsilon}_i = \varepsilon_i - x_{in}'\hat{b}$ for the selected observations.

In the asymptotic theory, Lemma A.1 in §A.1 allows us to replace the estimation errors $\hat{\theta} = (\tilde{a}, \tilde{b}, \hat{a}, \hat{b})$ with deterministic values $\theta = (a_1, b_1, a_m, b_m)$ varying in some set. Here the subscripts indicate whether parameters appear in the indicator or the mark. The marked and weighted empirical distribution of interest is then

$$\mathsf{F}_n^{w,m}(\theta,c) = n^{-1}\textstyle\sum_{i=1}^n w_{in} m\left(\dfrac{\varepsilon_i - x_{in}'b_m}{\sigma + n^{-1/2}a_m}\right) 1_{(\varepsilon_i \le \sigma c + n^{-1/2}a_1 c + x_{in}'b_1)}, \tag{2.2}$$

so that $\mathsf{F}_n^{w,m}(\hat{\theta},c) = \hat{\mathsf{F}}_n^{w,m}(c)$ is the empirical distribution function in (1.1). For later reference we also introduce $\theta_1 = (a_1, b_1)$ and $\theta_m = (a_m, b_m)$ where θ_1 and θ_m collect the estimation errors in the indicator function and the marks, respectively.

The statistical analysis of $\mathsf{F}_n^{w,m}$ uses martingale theory, hence, we define the compensator as the weighted sum of conditional expectations

$$\overline{\mathsf{F}}_n^{w,m}(\theta,c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\mathsf{E}_{i-1}\{m\left(\dfrac{\varepsilon_i - x_{in}'b_m}{\sigma + n^{-1/2}a_m}\right) 1_{(\varepsilon_i \le \sigma c + n^{-1/2}a_1 c + x_{in}'b_1)}\}, \tag{2.3}$$

and define the empirical process as the martingale

$$\mathbb{F}_n^{w,m}(\theta,c) = n^{1/2}\{\mathsf{F}_n^{w,m}(\theta,c) - \overline{\mathsf{F}}_n^{w,m}(\theta,c)\}. \tag{2.4}$$

2.2 Asymptotic expansions

Theorem 2.1 below shows that the empirical process $\mathbb{F}_n^{w,m}(\theta,c)$ is asymptotically equivalent to $\mathbb{F}_n^{w,m}(0,c)$ uniformly in θ and c. As a consequence we have that $\mathbb{F}_n^{w,m}(\hat{\theta},c)$ and $\mathbb{F}_n^{w,m}(0,c)$ are asymptotically equivalent; see Lemma A.1. The theorem also gives an asymptotic uniform linearization of the compensator $\overline{\mathsf{F}}_n^{w,m}(\theta,c)$.

Assumption 2.2. *Set $0 \le \kappa < \eta \le 1/4$. Let m be a differentiable mark function and let $\tilde{m}(u)$ represent each of $m(u)$, $u\dot{m}(u)$, $\dot{m}(u)$. Suppose:*
(i) density and marks satisfy
(a) moments: $\int_{-\infty}^{\infty} \tilde{m}^4(u)\mathsf{f}(u)du < \infty$;
(b1) boundedness: $\sup_{u\in\mathbb{R}} |u|\{1 + m^4(u)\}\mathsf{f}(u) < \infty$
(b2) boundedness: $\sup_{u\in\mathbb{R}} |\frac{\partial}{\partial u}[\{1 + \tilde{m}^4(u)\}\mathsf{f}(u)]| < \infty$
(b3) boundedness: $\sup_{u\in\mathbb{R}}(1 + u^2)|\dot{m}(u)\mathsf{f}(u)| < \infty$;
(c) smoothness: Let $\tilde{h}(u) = \{1 + \tilde{m}^4(u)\}\mathsf{f}(u)$ so that

$$\sup_{c>0} \frac{\sup_{u\ge c}\tilde{h}(u)}{\inf_{0\le u\le c}\tilde{h}(u)} < \infty, \quad \sup_{c>0} \frac{\sup_{u\le -c}\tilde{h}(u)}{\inf_{-c\le u\le 0}\tilde{h}(u)} < \infty.$$

(*d*) *local Lipschitz:* $\exists \omega > 0$ *and a function* $\mathring{m}(u) \geq 0$ *so* $\forall u^*, u$ *so that* $|u^* - u| \leq (1 + |u|)\omega$ *then* $|\dot{m}(u^*) - \dot{m}(u)| \leq |u^* - u|\mathring{m}(u)$ *and where* $\int_{-\infty}^{\infty}(1 + |u|^2)\mathring{m}(u)\mathsf{f}(u)du < \infty$*;*

(*ii*) *regressors:* $\max_{1\leq i\leq n} |n^{1/2-\kappa}N'x_i| = \mathrm{O}_{\mathsf{P}}(1)$ *for some non-stochastic normalization matrix* N *;*

(*iii*) *weights:* $\mathsf{E}\sum_{i=1}^n |w_{in}|^{2+\omega} = \mathrm{O}(n)$ *for some* $\omega > 0$*.*

Theorem 2.1. *Let Assumptions 2.1, 2.2 hold. Then,* $\forall B > 0$*,* $n \to \infty$*,*

$$\sup_{|\theta|\leq Bn^{1/4-\eta}} \sup_{c\in\mathbb{R}} |\mathbb{F}_n^{w,m}(\theta,c) - \mathbb{F}_n^{w,m}(0,c)| = \mathrm{o}_{\mathsf{P}}(1),$$

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

where $\mathcal{B}_n^{w,m}(\theta,c) = \mathcal{B}_{1n}^{w,m}(\theta_1,c) - \mathcal{B}_{mn}^{w,m}(\theta_m,c)$ *and*

$$\mathcal{B}_{1n}^{w,m}(\theta_1,c) = \sigma^{-1}m(c)\mathsf{f}(c)n^{-1/2}\textstyle\sum_{i=1}^n w_{in}(n^{-1/2}a_1c + x_{in}'b_1), \tag{2.5}$$

$$\mathcal{B}_{mn}^{w,m}(\theta_m,c) = \sigma^{-1}n^{-1/2}\textstyle\sum_{i=1}^n w_{in}\{n^{-1/2}a_m\mathsf{E}(\varepsilon_i/\sigma)\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c)} + x_{in}'b_m\mathsf{E}\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c)}\}. \tag{2.6}$$

Remark 1. *Theorem 2.1 generalizes [14] who had marks* $m(u) = 1$*, known scale* σ*, and bounded normalized estimators so that* $\eta = 1/4$*. Their result essentially required 2nd moments for regressors and weights, which is slightly weaker than for the present result. Both proofs evolve around chaining and martingale inequalities combined with truncation of regressors and weights. The basic setup is a locally quadratic* $\mathcal{F}_i$*-martingale* $V_n = \sum_{i=1}^n v_i$ *with predictable quadratic variation* $\langle V\rangle_n = \sum_{i=1}^n \mathsf{E}(v_i^2|\mathcal{F}_{i-1})$ *and total quadratic variation* $[V]_n = \sum_{i=1}^n v_i^2$*. Since [14] had bounded marks they could use the [7] inequality: for* $|v_i| \leq c$*, then*

$$\mathsf{P}(V_n \geq x, \langle V\rangle_n \leq y) \leq \exp[-x^2/\{2(y + cx)\}].$$

With unbounded marks we will instead use the [3] inequality, refined by [5], see also [2] and embedded in the iterated martingale inequalities in Theorems A.2, A.3: for any $|v_i|$*, then* $\forall x, y > 0$

$$\mathsf{P}(V_n \geq x, [V]_n + 2\langle V\rangle_n \leq y) \leq \exp\{-3x^2/(2y)\}. \tag{2.7}$$

The total quadratic variation $[V]_n$ *is harder to control than the predictable quadratic variation* $\langle V\rangle_n$*, so that slightly stronger conditions are needed here. Theorem 2.1 also generalizes results in [11], where the marks are* $m(u) = u^p$ *without estimation errors. That result required that the number of moments grows with the dimension of the regressors, which is relaxed here.*

Remark 2. *Assumption 2.2(i, c) is a smoothness condition that is satisfied if* $\{1 + \tilde{m}^4(u)\}\mathsf{f}(u)$ *is monotone for large* $|u|$*; see [11, Remark 4.1].*

Remark 3. *The local Lipschitz Assumption 2.2(id) is satisfied for polynomials* $m(u) = u^p$*. Note* $\dot{m}(u) = pu^{p-1}$ *and* $\dot{m}(u^\dagger) - \dot{m}(u) = p(u^\dagger - u)\sum_{j=0}^{p-2} u^j(u^\dagger)^{k-2-j}$*. For* $|u^\dagger - u| \leq (1+|u|)\omega$ *and* $\omega = 1$ *we can choose* $\mathring{m}(u) = p\sum_{j=0}^{p-2} |u|^j(1 + |u|)^{k-2-j}$ *with integrability condition* $\mathsf{E}|u|^p < \infty$*.*

Remark 4. *Assumption 2.2(i) is satisfied for the normal distribution with* $m(u) = u^p$ *for any* $p \in \mathbb{N}$ *since the derivatives of the normal density are bounded, tail monotone and locally Lipschitz.*

Remark 5. *Assumption 2.2(ii) allows a general class of regressors. For stationary regressors the assumption is satisfied if there exists a* κ_0 *so* $1/4 > \kappa > \kappa_0 > 0$ *and* $\mathsf{E}|x_i|^{1/\kappa_0} < \infty$*, since by the Boole and Markov inequalities*

$$\mathsf{P}(\max_{1\leq i\leq n} |x_i| > y) = \mathsf{P}\textstyle\bigcup_{i=1}^n(|x_i| > y) \leq \sum_{i=1}^n \mathsf{P}(|x_i| > y) \leq \sum_{i=1}^n \mathsf{E}|x_i/y|^{1/\kappa_0}$$

vanishes for $y = n^{1/2-\kappa}$*. For deterministic regressors and random walk regressors we can choose* $\kappa = 0$*. See [11, Example 3.2] for details.*

Combining Theorem 2.1 and Lemma A.1 in Appendix A.1 we expand

$$n^{1/2}\{\mathsf{F}_n^{w,m}(\hat{\theta},c) - \overline{\mathsf{F}}_n^{w,m}(0,c)\} = \mathbb{F}_n^{w,m}(0,c) + \mathcal{B}_n^{w,m}(\hat{\theta},c) + \mathrm{o_P}(1), \tag{2.8}$$

uniformly in c. The next step is to find an asymptotic theory for the process $\mathbb{F}_n^{w,m} + \mathcal{B}_n^{w,m}$. Often the process $\mathbb{F}_n^{w,m} + \mathcal{B}_n^{w,m}$ is asymptotically Gaussian by the Central Limit Theorem even in the presence of random walk regressors with examples following in the subsequent sections. But, the asymptotic distribution could also involve stochastic integrals, see §1.5.4 of [9]. The tightness of the process $\mathbb{F}_n^{w,m}$ is analyzed by [14] with unit marks, $m(u) = 1$, while the following result applies for polynomial marks.

Assumption 2.3. *Let $m(u) = u^p$ for some $p \in \mathbb{N}_0$. Suppose*
(i) density satisfies: $\mathsf{E}|\varepsilon_i|^{4p+\nu} < \infty$ for some $\nu > 0$;
(ii) weights satisfies: $\mathsf{E}\sum_{i=1}^n |w_{in}|^4(1 + |n^{1/2}N'x_i|) = \mathrm{O}(n)$.

Theorem 2.2. *([11, Th. 4.2]) Let Assumptions 2.1, 2.3 hold. Then, $\forall\epsilon > 0$,*

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

3 Truncated Moments & Linear Statistical Functionals

Many test statistics can be expressed as statistical functionals of empirial processes of residuals, say $\mathsf{T}(\hat{\mathsf{F}}_n)$ where $\hat{\mathsf{F}}_n(u) = n^{-1}\sum_{i=1}^n 1_{(\tilde{\varepsilon}_i\le\tilde{\sigma}c)}$. Conditions ensuring the weak convergence of $\hat{\mathsf{F}}_n$ are insufficient to describe the asymptotic theory of $\mathsf{T}(\hat{\mathsf{F}}_n)$ in genereal, especially when considering functionals of the form

$$\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n) = \int_{-\infty}^c m(u)d\hat{\mathsf{F}}_n(u) = n^{-1}\sum_{i=1}^n m(\tilde{\varepsilon}_i/\tilde{\sigma})1_{(\tilde{\varepsilon}_i\le\tilde{\sigma}c)}, \tag{3.1}$$

and the integrand is unbounded. However, this statistical functional is a weighted and marked empirical distribution function, that is, $\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n) = \mathsf{F}_n^{1,m}(\hat{\theta},c)$. Theorem 2.1 expands such statistical functionals.

Corollary 3.1. *Suppose $n^{1/2}(\tilde{\sigma}-\sigma)$ and $N^{-1}(\tilde{\beta}-\beta)$ are $\mathrm{O_P}(n^{1/4-\eta})$ for some $\eta > 0$. Under Assumptions 2.1, 2.2 with that η, mark m defined from $\mathsf{T}^{m,c}$ and weights $w_{in} = 1$, then*

$$n^{1/2}\{\mathsf{T}^{m,c}(\hat{\mathsf{F}}_n) - \mathsf{T}^{m,c}(\mathsf{F})\} = n^{1/2}\{\mathsf{T}^{m,c}(\mathsf{F}_n) - \mathsf{T}^{m,c}(\mathsf{F})\} + \mathcal{B}_n^{1,m}(\hat{\theta},c) + \mathrm{o_P}(1),$$

uniformly in $c \in \mathbb{R}$, and where $\mathcal{B}_n^{1,m}(\hat{\theta},c) = \mathcal{B}_{1n}^{1,m}(\hat{\theta},c) - \mathcal{B}_{mn}^{1,m}(\hat{\theta},c)$ with

$$\begin{aligned}\mathcal{B}_{1n}^{1,m}(\hat{\theta},c) =& cm(c)\mathsf{f}(c)n^{1/2}(\tilde{\sigma}/\sigma-1) + m(c)\mathsf{f}(c)\sum_{i=1}^n x_{in}'(\tilde{\beta}-\beta),\\ \mathcal{B}_{mn}^{1,m}(\hat{\theta},c) =& n^{1/2}(\tilde{\sigma}/\sigma-1)\mathsf{E}\{(\varepsilon_1/\sigma)\dot{m}(\varepsilon_1/\sigma)1_{(\varepsilon_i\le\sigma c)}\}\\ &+ \sigma^{-1}\sum_{i=1}^n x_{in}'(\tilde{\beta}-\beta)\mathsf{E}\{\dot{m}(\varepsilon_1/\sigma)1_{(\varepsilon_i\le\sigma c)}\}.\end{aligned}$$

We note that the bias term $\mathcal{B}_n^{1,1}(\hat{\theta},c)$ only depends indirectly on the regressors. In particular, for least squares estimators we have

$$n^{1/2}(\tilde{\sigma}/\sigma-1) = 2^{-1}n^{-1/2}\sum_{i=1}^n\{(\varepsilon_i/\sigma)^2-1\} + \mathrm{o_P}(1), \tag{3.2}$$

$$\sigma^{-1}\sum_{i=1}^n x_{in}'(\tilde{\beta}-\beta) = n^{-1/2}\sum_{i=1}^n(\varepsilon_i/\sigma) + \mathrm{o_P}(1), \tag{3.3}$$

see Lemmas B.5, B.6 in Appendix B.

We now consider the special case $m(u) = u^p$ in some detail. We focus on the results without further attention to the regularity conditions set out in Corollary 3.1. Denote $\mathsf{T}^{m,c}$ by $\mathsf{T}^{p,c}$, let $\mathsf{T}^{p,\infty} = \mathsf{T}^p$ and consider a symmetric density f for simplicity. The idea is to highlight some subtle differences that arise when applying the statistical functional $\mathsf{T}^{p,c}$ to F_n and $\hat{\mathsf{F}}_n$, respectively.

Denote one-sided truncated moments by $\mathsf{T}^{p,c}(\mathsf{F}) = \mathsf{E}(\varepsilon_i/\sigma)^p 1_{(\varepsilon_i \le \sigma c)}$ for $c \in \mathbb{R}$, while two-sided truncated moments are given by

$$\tau_p^c = \mathsf{E}(\varepsilon_i/\sigma)^p 1_{(|\varepsilon_i| \le \sigma c)} = \mathsf{T}^{p,c}(\mathsf{F}) - \mathsf{T}^{p,-c}(\mathsf{F}) \qquad \text{for } c > 0. \tag{3.4}$$

In particular, $\mathsf{T}^{0,c}(\mathsf{F}) = \mathsf{P}(\varepsilon_i \le \sigma c)$ for $c \in \mathbb{R}$ and $\tau_0^c = \mathsf{P}(|\varepsilon_i| \le \sigma c)$ for $c \in \mathbb{R}$. Denote $\tau_p^\infty = \tau_p$. We note that when F is the standard normal distribution function then, for $p \in \mathbb{N}_0$,

$$\tau_{2p+1}^c = 0, \qquad \tau_{2p}^c = \{(2p-1)!!\}\mathsf{P}(\chi^2_{2p+1} \le c^2), \tag{3.5}$$

where the odd factorial $(2p-1)!!$ is one for $p = 0$ and $\prod_{\ell=1}^p (2\ell - 1)$ for $p \in \mathbb{N}$. This is proved by first integrating u^p with respect to the standard normal density $\varphi(u)$ and substituting $u^2 = v$ and then noting $\Gamma\{(p+1)/2\} = \Gamma(1/2)\prod_{\ell=1}^{p/2}\{(2\ell-1)/2\}$ by the functional equation for the gamma function. [1] have similar formulas for τ_1^c, τ_2^c. Now, inserting $c = \infty$ in the above formula gives the moments of the standard normal distribution: $\tau_0 = \tau_2 = 1$, $\tau_4 = 3$, $\tau_6 = 15$, $\tau_8 = 105$. Exploiting that the normal density satisfies $(\partial/\partial u)\{-u\mathsf{f}(u)\} = (u^2-1)\mathsf{f}(u)$ we also get

$$\tau_2^c = \textstyle\int_{-c}^c u^2 \mathsf{f}(u)du = \tau_0^c - 2c\mathsf{f}(c). \tag{3.6}$$

Example 1. ***The sample central moments*** *of* F_n *are*

$$\tilde{\mathsf{T}}^p(\mathsf{F}_n) = \textstyle\int_{-\infty}^\infty \{u - \int_{-\infty}^\infty v d\mathsf{F}_n(v)\}^p d\mathsf{F}_n(u).$$

This is analyzed as a non-linear statistical functional of F_n *by [19, p. 232f]. However, we can also analyze this as a linear statistical functional of* $\tilde{\mathsf{F}}_n(c) = n^{-1}\sum_{i=1}^n 1_{(\tilde\varepsilon_i \le \sigma c)}$ *where* $\tilde\varepsilon_i = \varepsilon_i - \bar\varepsilon$. *To do so let* $x_i = 1$ *in (2.1), that is* $y_i = \mu + \varepsilon_i$. *We then get that* $\tilde{\mathsf{T}}^p(\mathsf{F}_n) = \mathsf{T}^p(\tilde{\mathsf{F}}_n)$, *which is a linear statistical functional in* $\tilde{\mathsf{F}}_n$. *Recall the properties of least squares estimators in (3.3) and apply Corollary 3.1 with* $\tilde\sigma = \sigma$ *to get*

$$n^{1/2}\{\mathsf{T}^p(\tilde{\mathsf{F}}_n) - \mathsf{T}^p(\mathsf{F})\} = n^{-1/2}\textstyle\sum_{i=1}^n(\varepsilon_i^p/\sigma^p - \tau_p) - p\tau_{p-1}n^{-1/2}\sum_{i=1}^n \varepsilon_i + \mathsf{o}_\mathsf{P}(1).$$

Assuming $\tau_1 = 0$ *the asymptotic variance is found to be*

$$\mathsf{var} = \tau_{2p} - (\tau_p)^2 - 2p\tau_{p-1}\tau_{p+1} + p^2\tau_{p-1}^2\tau_2,$$

in agreement with [19, p. 233]. Serfling leaves it to the reader to check whether his condition A_1 *applies to his remainder term* R_{1n}. *Here, this is done through Corollary 3.1 with its more primitive conditions, which are satisfied for instance for a normal distribution.*

Example 2. ***Standardized sample moments*** *of* $\hat{\mathsf{F}}_n$. *Let* $\tilde\beta, \tilde\sigma$ *be least squares estimators in model (2.1). This includes an intercept so* $\sum_{i=1}^n \tilde\varepsilon_i = 0$. *We get the standardized moments* $\mathsf{T}^p(\hat{\mathsf{F}}_n) = n^{-1}\sum_{i=1}^n(\tilde\varepsilon_i/\tilde\sigma)^p$, *so that* $\mathsf{T}^1(\hat{\mathsf{F}}_n) = 0$ *and* $\mathsf{T}^2(\hat{\mathsf{F}}_n) = 1$ *while* $\mathsf{T}^3(\hat{\mathsf{F}}_n)$ *and* $\mathsf{T}^4(\hat{\mathsf{F}}_n)$ *are sample skewness and kurtosis. Corollary 3.1 combined with (3.2), (3.3) shows that*

$$\begin{aligned} n^{1/2}\{\mathsf{T}^p(\hat{\mathsf{F}}_n) - \mathsf{T}^p(\mathsf{F})\} &= n^{-1/2}\textstyle\sum_{i=1}^n\{(\varepsilon_i/\sigma)^p - \tau_p\} \\ &\quad - p\tau_{p-1}n^{-1/2}\textstyle\sum_{i=1}^n(\varepsilon_i/\sigma) - p\tau_p\frac{1}{2}n^{-1/2}\sum_{i=1}^n\{(\varepsilon_i/\sigma)^2 - 1\} + \mathsf{o}_\mathsf{P}(1). \end{aligned}$$

The asymptotic variance, assuming $\tau_1 = 0$, *is*

$$\begin{aligned} \mathsf{var}_p = \tau_{2p} - \tau_p^2 - 2p\tau_{p-1}\tau_{p+1} + p^2\tau_{p-1}^2\tau_2 - p\tau_p(\tau_{p+2} - \tau_p\tau_2) \\ + p^2\tau_p\tau_{p-1}\tau_3 + p^2\tau_p^2(\tau_4 - \tau_2^2)/4. \end{aligned} \tag{3.7}$$

In particular, in the normal case this reduces to $\mathsf{var}_3 = 6$ *and* $\mathsf{var}_4 = 24$.

4 Testing for Normality in Robust Regressions

We now apply the general theory of §2 to normality testing in robust regressions. We consider the widely used "data-analytic strategy" described in [22]; that is, outliers are first detected using an initial estimator and then, after eliminating them, the model is estimated by least squares on the retained observations. The properties of such statistical procedure are unknown but can be studied using the above general theory.

4.1 Estimators, test statistics, assumptions and notation

Suppose we have initial estimators $\tilde{\beta}, \tilde{\sigma}$. We then select observations where $|\tilde{\varepsilon}_i| \leq \widetilde{\sigma c}$ with $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ and run a regression on those observations giving

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

This results in updated residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$ and a residual variance estimator of the form

$$\hat{\sigma}^2 = \varsigma_c^{-2} \{\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i| \leq \widetilde{\sigma c})}\}^{-1} \textstyle\sum_{i=1}^n \hat{\varepsilon}_i^2 1_{(|\tilde{\varepsilon}_i| \leq \widetilde{\sigma c})}, \tag{4.1}$$

where the consistency factor is $\varsigma_c^2 = \tau_2^c / \tau_0^c$. Table 1 gives numerical values for ς_c^2 under the hypothesis of normal errors without outliers. The above estimators are refered to as 1-step Huber-skip estimators and are analyzed in [9, 10, 11, 12]. Examples include:

Example 3. *The least squares estimator where* $\widetilde{\sigma c} = \infty$.

Example 4. *The robustified least squares estimator where* $\tilde{\beta}, \tilde{\sigma}$ *are full-sample least squares estimators and* c *is a user-specified cut-off so that* $\widetilde{\sigma c} = \tilde{\sigma} c$. *We will write* $\hat{\beta}_{RLS}$ *for* $\hat{\beta}$.

Example 5. *The least trimmed squares estimator of [17]. Let* $\xi_i(\beta) = |y_i - x_i'\beta|$ *with order statistics* $\xi_{(i)}(\beta)$ *in increasing order. Let* $n - h$ *be the number of trimmed observations corresponding to a trimming proportion of* $(n-h)/n = \mathsf{P}(\varepsilon_1^2 \geq \sigma^2 c^2)$. *Then the least trimmed squares estimator is* $\tilde{\beta}_{LTS} = \arg\min_\beta \sum_{i=1}^h \xi_{(i)}^2(\beta)$. *Let* $\tilde{\xi}_i = \xi_i(\tilde{\beta}_{LTS})$. *This estimator selects* h *observations with smallest residuals* $\tilde{\xi}_{(i)}$ *so that the cut-off value is* $\widetilde{\sigma c} = \tilde{\xi}_{(h)}$. *In our setup with* $\tilde{\beta} = \hat{\beta} = \tilde{\beta}_{LTS}$ *giving residuals* $\tilde{\varepsilon}_i = \hat{\varepsilon}_i = y_i - x_i'\tilde{\beta}_{LTS}$. *Inserting this in (4.1) gives* $\tilde{\sigma}_{LTS}^2 = \hat{\sigma}^2$.

We consider the moment based normality test on the robustified (truncated) sub-sample of second stage residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$. Let s denote the estimation procedure being used and define the conditional sample moments

$$\hat{\mu}_{p,c}^s = \frac{\sum_{i=1}^n (\hat{\varepsilon}_i/\hat{\sigma})^p 1_{(|\tilde{\varepsilon}_i| \leq \widetilde{\sigma c})}}{\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i| \leq \widetilde{\sigma c})}} \qquad \text{for } p \in \mathbb{N}. \tag{4.2}$$

We then study the following truncated normality test statistics

$$\hat{T}_{3,c}^s = \frac{n^{1/2} \hat{\mu}_{3,c}^s}{(\lambda_{6,c}^s)^{1/2}}, \qquad \hat{T}_{4,c}^s = \frac{n^{1/2}(\hat{\mu}_{4,c}^s - \lambda_{3,c}^s)}{(\lambda_{24,c}^s)^{1/2}}, \tag{4.3}$$

where $\lambda_{3,c}^s, \lambda_{6,c}^s, \lambda_{24,c}^s$ are normalizing constants that depend on the selection stage through c and the estimation method. We note that when $\widetilde{\sigma c} = \infty$ there is no selection over observations and the statistics reduce to the standard cumulant based test statistics for normality of residuals with $\lambda_{3,\infty}^{OLS} = 3$, $\lambda_{6,\infty}^{OLS} = 6$ and $\lambda_{24,\infty}^{OLS} = 24$, see Example 2. When there is selection, the normalizing factors $\lambda_{3,c}^s$, $\lambda_{6,c}^s$, $\lambda_{24,c}^s$ depend on the truncated moments and certain constants entering the first order asymptotic expansions of the estimation method being used.

We analyze the normality test when there is no contamination and normal errors. Assumption 2.2 reduces as follows, see Remark 4.

Assumption 4.1. *Let* $0 \le \kappa < 1/4$. *Suppose*
(i) ε_i/σ *is* $\mathsf{N}(0,\sigma^2)$ *distributed;*
(ii) regressors: $\max_{1\le i\le n} |n^{1/2-\kappa}N'x_i| = \mathrm{O}_{\mathsf{P}}(1)$ *for some non-stochastic normalization matrix* N, *where* $N^{-1} = \mathrm{O}(n^\ell)$ *for some* $\ell > 0$ *and where* $\inf\{n : \sum_{i=1}^n x_i x_i'$ *invertible*$\} < \infty$ *a.s.*

For the least trimmed squares estimator further assumptions are needed.

Assumption 4.2. *The regressors are nonrandom so that* $\sum_{i=1}^n |x_i|^4 = \mathrm{O}(n)$ *and* $n^{-1}\sum_{i=1}^n x_i x_i'$ *has a positive definite limit.*

4.2 The robustified least squares case in Example 4

The normality test is based on the truncated empirical moments in (4.2) where $\tilde\beta, \tilde\sigma$ are full sample least squares estimators and $\hat\beta, \hat\sigma$ are the corresponding 1-step Huber skip estimators with residuals $\tilde\varepsilon_i = y_i - x_i'\tilde\beta$ and $\hat\varepsilon_i = y_i - x_i'\hat\beta$, respectively, while the cut-off is $\widetilde{\sigma c} = \tilde\sigma c$.

The truncated normality test statistics $\hat T^{RLS}_{3,c}$, $\hat T^{RLS}_{4,c}$ in (4.3) are computed as follows. The asymptotic expansions will involve the vectors

$$z^c_{3,i} = \left\{\begin{array}{c} (\varepsilon_i/\sigma)^3 1_{(|\varepsilon_i|\le\sigma c)} \\ (\varepsilon_i/\sigma)1_{(|\varepsilon_i|\le\sigma c)} \\ (\varepsilon_i/\sigma) \end{array}\right\}, \qquad z^c_{4,i} = \left\{\begin{array}{c} (\varepsilon_i/\sigma)^4 1_{(|\varepsilon_i|\le\sigma c)} - \tau^c_4 \\ (\varepsilon_i/\sigma)^2 1_{(|\varepsilon_i|\le\sigma c)} - \tau^c_2 \\ 1_{(|\varepsilon_i|\le\sigma c)} - \tau^c_0 \\ (\varepsilon_i/\sigma)^2 - 1 \end{array}\right\}. \tag{4.4}$$

For normal and hence symmetric ε_i these vectors are uncorrelated. The Central Limit Theorem then shows that $z^c_{3,i}, z^c_{4,i}$ are asymptotically normal and independent with variances

$$\Omega^c_3 = \begin{pmatrix} \tau^c_6 & \tau^c_4 & \tau^c_4 \\ \tau^c_4 & \tau^c_2 & \tau^c_2 \\ \tau^c_4 & \tau^c_2 & 1 \end{pmatrix}, \tag{4.5}$$

$$\Omega^c_4 = \left\{\begin{array}{cccc} \tau^c_8 - \tau^c_4\tau^c_4 & \tau^c_6 - \tau^c_2\tau^c_4 & \tau^c_4(1-\tau^c_0) & \tau^c_6 - \tau^c_4 \\ \tau^c_6 - \tau^c_2\tau^c_4 & \tau^c_4 - \tau^c_2\tau^c_2 & \tau^c_2(1-\tau^c_0) & \tau^c_4 - \tau^c_2 \\ \tau^c_4(1-\tau^c_0) & \tau^c_2(1-\tau^c_0) & \tau^c_0(1-\tau^c_0) & \tau^c_2 - \tau^c_0 \\ \tau^c_6 - \tau^c_4 & \tau^c_4 - \tau^c_2 & \tau^c_2 - \tau^c_0 & 2 \end{array}\right\}. \tag{4.6}$$

We compute the vectors

$$\zeta^{RLS}_{3,c} = \{1, -3\tau^c_2/\tau^c_0, 2(c^2 - 3\tau^c_2/\tau^c_0)c\mathsf{f}(c)\}', \tag{4.7}$$
$$\zeta^{RLS}_{4,c} = \{1, -2\tau^c_4/\tau^c_2, \tau^c_4/\tau^c_0, (c^4 - c^2 2\tau^c_4/\tau^c_2 + \tau^c_4/\tau^c_0)c\mathsf{f}(c)\}', \tag{4.8}$$

and define the normalizations, for $s = RLS$,

$$\lambda^s_{3,c} = \tau^c_4/\tau^c_0, \quad \lambda^s_{6,c} = \zeta^{s\prime}_{3,c}\Omega^c_3\zeta^s_{3,c}/(\tau^c_0)^2, \quad \lambda^s_{24,c} = \zeta^{s\prime}_{4,c}\Omega^c_4\zeta^s_{4,c}/(\tau^c_0)^2. \tag{4.9}$$

Table 1 gives numerical values for $\lambda^{RLS}_{3,c}$, $\lambda^{RLS}_{6,c}$, $\lambda^{RLS}_{24,c}$. We note that these normalizations depend substantially on the choice of c.

We get the following asymptotic result.

Theorem 4.1. *Let Assumptions 2.1, 4.1 hold and* $c_0 > 0$. *Then, uniformly in* $c \ge c_0$, *for* $p = 3, 4$, *we get* $\hat T^{RLS}_{p,c} = T^{RLS}_{p,c,n} + \mathrm{o}_{\mathsf{P}}(1)$ *where*

$$T^{RLS}_{p,c,n} = \{(\zeta^{RLS}_{p,c})'\Omega^c_p(\zeta^{RLS}_{p,c})\}^{-1/2}(\zeta^{RLS}_{p,c})' n^{-1/2}\textstyle\sum_{i=1}^n z^c_{p,i}.$$

For $c \ge c_0$ *then* $T^{RLS}_{3,c,n}$ *and* $T^{RLS}_{4,c,n}$ *converge to independent Gaussian processes with zero mean and unit variance. In particular, for fixed* $c \ge c_0$ *then* $(\hat T^{RLS}_{3,c})^2 + (\hat T^{RLS}_{4,c})^2$ *is asymptotically* χ^2_2.

Table 1: Normality test for robust regressions. Normalization factors under normality.

$\tau_0^c = \mathsf{P}(\lvert\varepsilon_1\rvert < \sigma c)$	0.5	0.95	0.99	0.999	0.9999	0.99999	1
c	0.67	1.96	2.58	3.29	3.89	4.42	∞
ς_c^{-1}	2.6477	1.1480	1.0399	1.0059	1.0008	1.0001	1
$\lambda_{3,c}^{RLS} = \lambda_{3,c}^{LTS}$	0.0379	1.3501	2.2750	2.8381	2.9709	2.9954	3
$\lambda_{6,c}^{RLS}$	0.0111	0.8865	2.4986	4.6725	5.6472	5.9250	6
$\lambda_{6,c}^{LTS}$	0.0041	0.8313	2.4908	4.6724	5.6472	5.9250	6
$\lambda_{24,c}^{RLS}$	0.0012	1.1211	4.5439	12.9758	19.7877	22.7983	24
$\lambda_{24,c}^{LTS}$	0.0029	2.0489	7.6335	16.8966	21.9157	23.5276	24

Example 6. *The normalizations $\lambda_{3,c}^{RLS}$, $\lambda_{6,c}^{RLS}$, $\lambda_{24,c}^{RLS}$ found in Theorem 4.1 are substantially different from the traditional values 3, 6, 24. Those incorrect values are commonly applied in practice after outlier detection. This leads to severe size distortions as we are comparing $\hat{\mu}_{3,c}^{RLS}$ and $\hat{\mu}_{4,c}^{RLS}$ with $\mathsf{N}(0, 6/n)$ and $\mathsf{N}(3, 24/n)$ distributions rather than $\mathsf{N}(0, \lambda_{6,c}^{RLS}/n)$ and $\mathsf{N}(\lambda_{3,c}^{RLS}, \lambda_{24,c}^{RLS}/n)$ distributions. Then the 3^{rd} moment test is under-sized while the 4^{th} moment test has asymptotic size of unity. Indeed, suppose we set $c = 2.58$ corresponding to a 1% trimming and let $n = 100$. The incorrect normalizations give 95% sampling regions of $[-0.48, 0.48]$ and $[2.04, 3.96]$, respectively, instead of the correct $[-0.30, 0.30]$ and $[1.86, 2.69]$, leading to sizes of 0.24% and 13.5%, respectively. For $n = 200$ and $n = 400$ the fourth moment test has sizes increasing to 62.0% and 98.9%, respectively.*

4.3 The least trimmed squares case in Example 5

The result in this case is similar to the previous one. The main difference is technical: since order statistics are used the proof involves empirical and quantile processes.

The truncated empirical moments in (4.2) have $\tilde{\beta} = \hat{\beta}$ as the least trimmed squares estimator with residuals $\tilde{\varepsilon}_i = \hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$. The cut-off is $\widetilde{\sigma c} = \tilde{\xi}_{(h)}$ which is the hth smallest order statistic of $\tilde{\xi}_i = |\tilde{\varepsilon}_i|$. The least trimmed estimators $\tilde{\beta}$ and $\tilde{\sigma}$ were analyzed by [21] and [12], respectively.

The truncated normality test statistics $\hat{T}_{3,c}^{LTS}$ and $\hat{T}_{4,c}^{LTS}$ are computed from (4.3) where the normalizations are expressed as follows. Recall the covariances Ω_3^c, Ω_4^c in (4.6) and compute vectors

$$\zeta_{3,c}^{LTS} = \{1, 2c^3\mathsf{f}(c)/\tau_2^c - 3, 0\}', \tag{4.10}$$

$$\zeta_{4,c}^{LTS} = \{1 - \tau_4^c/\tau_0^c, -2\tau_4^c/\tau_2^c, \tau_4^c/\tau_0^c + 2c^2\tau_4^c/\tau_2^c - c^4, 0\}', \tag{4.11}$$

and the normalizations $\lambda_{3,c}^{LTS}$, $\lambda_{6,c}^{LTS}$, $\lambda_{24,c}^{LTS}$ from (4.9) with $s = LTS$. The normalizations are tabulated in Table 1. We have the following result.

Theorem 4.2. *Let Assumptions 2.1, 4.1, 4.2 hold and choose a fixed $h/n \in (0,1)$. Then $\hat{T}_{3,c}^{LTS}$ and $\hat{T}_{4,c}^{LTS}$ are asymptotically independent χ_1^2.*

A Empirical Processes Results

We prove Theorem 2.1. The weights w_{in} may be arrays, but to show that the resulting array of empirical processes vanishes it suffices to show this for each element. Thus, we proceed in this appendix as if w_{in} is scalar.

A.1 The Chaining Setup

We consider processes $M_n(\hat{\theta}, c)$ depending on estimation errors $\hat{\theta}$. If $\hat{\theta}$ is bounded in probability, then $M_n(\hat{\theta}, c)$ can be analyzed by studying the behaviour of $M_n(\theta, c)$ uniformly in $\theta \in \Theta$ for a compact Θ. This is due to the following result.

Lemma A.1. *If* $\forall\epsilon > 0$ *a compact set* Θ *exists so* $\lim_{n\to\infty}\mathsf{P}(\hat\theta\in\Theta^c) < \epsilon$ *then* $\mathsf{P}\{|M_n(\hat\theta,c)| > \epsilon\} \leq \mathsf{P}\{\sup_{\theta\in\Theta}|M_n(\theta,c)| > \epsilon\} + \epsilon$.

Proof of Lemma A.1: Boole's inequality shows $\mathsf{P}(\mathcal{A}) \leq \mathsf{P}(\mathcal{A}\cap\mathcal{B}) + \mathsf{P}(\mathcal{B}^c)$ for events $\mathcal{A} = \{|M_n(\hat\theta,c)| > \epsilon\}$, $\mathcal{B} = (\hat\theta\in\Theta)$. The probability $\mathsf{P}(\mathcal{A}\cap\mathcal{B})$ is bounded by considering the largest possible outcome of $|M_n(\theta,c)|$ for $\theta\in\Theta$. The probability $\mathsf{P}(\mathcal{B}^c)$ vanishes by assumption. $\square$

We generalize the norm on $\mathbb{R}$ developed in [11] in order to cover $\mathbb{R}$ with a finite number of chaining points c_k. The norm evolves around *fourth power of* the variables $m(\varepsilon_i/\sigma)1_{(\varepsilon_i/\sigma\leq c)}$ for $c\in\mathbb{R}$, where ε_i/σ has density f and m is a (mark) function that will be chosen in various ways throughout the proof of Theorem 2.1. We define

$$\mathsf{H}^m(c) = \int_{-\infty}^{c}\{1+|m(u)|^4\}\mathsf{f}(u)du = \mathsf{E}\{1+m^4(\varepsilon_i/\sigma)\}1_{(\varepsilon_i\leq\sigma c)}, \tag{A.1}$$

with derivative $\dot{\mathsf{H}}^m(c) = \{1+|m(c)|^4\}\mathsf{f}(c)$. The function H^m is increasing by construction and bounded by Assumption 2.2(ia). Let

$$H^m = \mathsf{H}^m(\infty) = \mathsf{E}[\{1+m^4(\varepsilon_i/\sigma)\}] = \int_{-\infty}^{\infty}\{1+m^4(u)\}\mathsf{f}(u)du < \infty.$$

The inequality $m^{2^q} \leq 1+m^4$ for $0\leq q\leq 2$ implies that, for $c\leq c^\dagger$,

$$\mathsf{E}\{|m(\varepsilon_i/\sigma)|1_{(c<\varepsilon_i/\sigma\leq c^\dagger)}\}^{2^q} \leq \mathsf{E}\{1+m^4(\varepsilon_i/\sigma)\}1_{(c<\varepsilon_i/\sigma\leq c^\dagger)} = \mathsf{H}^m(c^\dagger) - \mathsf{H}^m(c). \tag{A.2}$$

We denote $\mathsf{H}^m(c^\dagger) - \mathsf{H}^m(c)$ the H^m-distance between c and $c^\dagger$.

For the chaining, partition the range of $\mathsf{H}^m(c)$ into K intervals of equal size H^m/K. We choose $K = \mathrm{int}(n^{1/2}/\delta)$ for some $\delta > 0$, and, accordingly, partition the support into K intervals defined by the grid points

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \tag{A.3}$$

so that $\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1}) = H^m/K = \mathrm{O}(\delta n^{-1/2})$.

A chaining argument is used to show $\sup_{c\in\mathbb{R}}|M_n(\theta,c)|$ is small. That is

$$\sup_{c\in\mathbb{R}}|M_n(\theta,c)| \leq \max_{1\leq k\leq K}|M_n(\theta,c_k)| + \max_{1\leq k\leq K}\sup_{c:c_{k-1}<c\leq c_k}|M_n(\theta,c) - M_n(\theta,c_k)|. \tag{A.4}$$

We refer to these two terms as the *discrete points* and the *oscillation* terms.

A.2 Iterated exponential martingale inequalities

In the chaining arguments we investigate the tail probability for the maximum of a certain family of martingales. We now modify the iterated martingale inequality in [11, Th. 5.1]. The new inequality is sharper as the proof uses the Delyon inequality (2.7) instead of the Bercu and Touati inequality. Only a single iteration is presented as this suffices with the subsequent proofs. However, the main difference is the intersection with the set bounding the weights inspired by the [14, Lemma 2.3] version of the Freedman [7] inequality.

Theorem A.2. *For* $1\leq\ell\leq L$, *let* $w_{\ell,i}$ *and* $y_{\ell,i+1}$ *be* $\mathcal{F}_i$*-adapted with* $\mathsf{E}z_{\ell i}^4 < \infty$. *Let* $D_r = \max_{1\leq\ell\leq L}\sum_{i=1}^n\mathsf{E}_{i-1}z_{\ell i}^{2^r}$ *for* $r = 1,2$. *Then, for all* $\kappa_w,\kappa_0,\kappa_1,\kappa_2 > 0$ *and for* $\mathcal{D}_n = (\max_{1\leq\ell\leq L}\max_{1\leq i\leq n}|w_{\ell i}|\leq\kappa_w)$,

$$\mathsf{P}\{\max_{1\leq\ell\leq L}|\sum_{i=1}^n w_{\ell i}(z_{\ell i} - \mathsf{E}_{i-1}z_{\ell i})| > \kappa_0\cap\mathcal{D}_n)\}$$
$$\leq \frac{\kappa_w^2}{\kappa_1}\mathsf{E}D_1 + \frac{\kappa_w^4 L}{3\kappa_2}\mathsf{E}D_2 + 2L\{\exp(-\frac{\kappa_0^2}{6\kappa_1}) + \exp(-\frac{\kappa_1^2}{6\kappa_2})\}.$$

Proof of Theorem A.2: Let $\mathcal{P}_n$ be the probability of interest.
1. *Truncation.* Define $\bar{w}_{\ell i} = w_{\ell i} 1_{(|w_{\ell i}| \leq \kappa_w)}$ so that $w_{\ell i} = \bar{w}_{\ell i}$ on $\mathcal{D}_n$. Let $A_\ell = \sum_{i=1}^n \bar{w}_{\ell i}(z_{\ell i} - \mathsf{E}_{i-1} z_{\ell i})$ and $\mathcal{A} = (\max_{1\leq\ell\leq L} |A_\ell| > \kappa_0)$ so that we get $\mathcal{P}_n = \mathsf{P}(\mathcal{A} \cap \mathcal{D}_n)$ which is bounded by $\mathcal{P}_n \leq \mathsf{P}(\mathcal{A})$.
2. *Martingale A_ℓ and quadratic variation.* The weight $\bar{w}_{\ell i}$ is $\mathcal{F}_{i-1}$-adapted and bounded, so that A_ℓ is a martingale with bounded weights and its sum of predictable and total quadratic variation is $B_\ell = \sum_{i=1}^n \bar{w}_{\ell i}^2 B_{\ell i}$ where $B_{\ell i} = (z_{\ell i} - \mathsf{E}_{i-1} z_{\ell i})^2 + 2\mathsf{E}_{i-1}(z_{\ell i} - \mathsf{E}_{i-1} z_{\ell i})^2$. This requires that $\mathsf{E} z_{\ell i}^2 < \infty$, which is assumed. Let $\mathcal{B} = (\max_{1\leq\ell\leq L} B_\ell \leq 9\kappa_1)$ and note the inequality

$$\mathsf{P}(\mathcal{A}) = \mathsf{P}(\mathcal{A} \cap \mathcal{B}) + \mathsf{P}(\mathcal{A} \cap \mathcal{B}^c) \leq \mathsf{P}(\mathcal{A} \cap \mathcal{B}) + \mathsf{P}(\mathcal{B}^c). \tag{A.5}$$

3. *Bounding* $\mathsf{P}(\mathcal{A} \cap \mathcal{B})$. Let $\mathcal{A}_\ell = (|A_\ell| > \kappa_0)$ and $\mathcal{B}_\ell = (B_\ell \leq 9\kappa_1)$. Apply Boole's inequality noting that $\mathcal{A} = \cup_{\ell=1}^L \mathcal{A}_\ell$ and then that $\mathcal{B} \subset \mathcal{B}_\ell$ to get

$$\mathsf{P}(\mathcal{A} \cap \mathcal{B}) \leq \textstyle\sum_{\ell=1}^L \mathsf{P}(\mathcal{A}_\ell \cap \mathcal{B}) \leq \sum_{\ell=1}^L \mathsf{P}(\mathcal{A}_\ell \cap \mathcal{B}_\ell).$$

The martingale exponential inequality in (2.7) shows

$$\mathsf{P}(\mathcal{A} \cap \mathcal{B}) \leq \textstyle\sum_{\ell=1}^L 2\exp\{-3\kappa_0^2/(18\kappa_1)\} = \sum_{\ell=1}^L 2\exp\{-\kappa_0^2/(6\kappa_1)\}.$$

4. *Martingale decomposition of B_ℓ.* Ignore the indices on $B_{\ell i}$, E_{i-1} and $z_{\ell i}$, and apply the inequality $(z - \mathsf{E}z)^2 \leq 2(z^2 + \mathsf{E}^2 z)$ along with $\mathsf{E}^2 z \leq \mathsf{E}z^2$ and $\mathsf{E}(z - \mathsf{E}z)^2 \leq \mathsf{E}z^2$ to get that $B = (z - \mathsf{E}z)^2 + 2\mathsf{E}(z - \mathsf{E}z)^2$ satisfies the inequality $B \leq 2z^2 + 5\mathsf{E}z^2 = 2(z^2 - \mathsf{E}z^2) + 7\mathsf{E}z^2$. Thus,

$$\mathsf{P}(\mathcal{B}^c) \leq \mathsf{P}[\max_{1\leq\ell\leq L} \textstyle\sum_{i=1}^n \kappa_w^2 \{2(z_{\ell i}^2 - \mathsf{E}_{i-1} z_{\ell i}^2) + 7\mathsf{E}_{i-1} z_{\ell i}^2\} > 9\kappa_1].$$

Let $\tilde{A}_\ell = \kappa_w^2 \sum_{i=1}^n (z_{\ell i}^2 - \mathsf{E}_{i-1} z_{\ell i}^2)$ and $\tilde{\mathcal{A}} = (\max_{1\leq\ell\leq L} |\tilde{A}_\ell| > \kappa_1)$.
Further, let $\tilde{\mathcal{C}} = (\kappa_w^2 \max_{1\leq\ell\leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell i}^2 > \kappa_1)$. Noting that $\mathsf{P}(2x + 7y > 9\kappa) \leq \mathsf{P}\{(2x > 2\kappa) \cup (7y > 7\kappa)\}$ we get the further bound $\mathsf{P}(\mathcal{B}^c) \leq \mathsf{P}(\tilde{\mathcal{A}}) + \mathsf{P}(\tilde{\mathcal{C}})$.
5. *Bounding* $\mathsf{P}(\tilde{\mathcal{C}})$. Note $\bar{w}_{\ell i}^2 \leq \kappa_w$ and apply the Markov inequality to get

$$\mathsf{P}(\tilde{\mathcal{C}}) \leq \kappa_1^{-1} \kappa_w^2 \mathsf{E} \max_{1\leq\ell\leq L} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell,i}^2 = \kappa_1^{-1} \kappa_w^2 \mathsf{E} D_1.$$

6. *Martingale $\tilde{A}_\ell$ and quadratic variation.* The martingale $\tilde{A}_\ell$ has quadratic variation $\tilde{B}_\ell = \kappa_w^4 \sum_{i=1}^n \tilde{B}_{\ell i}$ where $\tilde{B}_{\ell i} = (z_{\ell i}^2 - \mathsf{E}_{i-1} z_{\ell i}^2)^2 + 2\mathsf{E}_{i-1}(z_{\ell i}^2 - \mathsf{E}_{i-1} z_{\ell i}^2)^2$, requiring $\mathsf{E} z_{\ell i}^4 < \infty$. Thus, the triangle inequality and (A.5) with $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}} = (\max_{1\leq\ell\leq L} \tilde{B}_\ell \leq 9\kappa_2)$ give $\mathsf{P}(\tilde{\mathcal{A}}) \leq \mathsf{P}(\tilde{\mathcal{A}} \cap \tilde{\mathcal{B}}) + \mathsf{P}(\tilde{\mathcal{B}}^c)$.
7. *Bounding* $\mathsf{P}(\tilde{\mathcal{A}} \cap \tilde{\mathcal{B}})$. Proceed as in item 3 to get the bound

$$\mathsf{P}(\tilde{\mathcal{A}} \cap \tilde{\mathcal{B}}) \leq 2L \exp\{-(\kappa_1/\kappa_w^2)^2/(6\kappa_2/\kappa_w^4)\} = 2L \exp\{-\kappa_1^2/(6\kappa_2)\}.$$

8. *Bounding* $\mathsf{P}(\tilde{\mathcal{B}}^c)$. By Boole's inequality

$$\mathsf{P}(\tilde{\mathcal{B}}^c) = \mathsf{P}\textstyle\bigcup_{\ell=1}^L (\kappa_w^4 \tilde{B}_\ell > 9\kappa_2) \leq L \max_{1\leq\ell\leq L} \mathsf{P}(\kappa_w^4 \tilde{B}_\ell > 9\kappa_2).$$

The Markov inequality and $\mathsf{E}\tilde{B}_{\ell i} = 3\mathsf{E}(z_{\ell i}^2 - \mathsf{E}_{i-1} z_{\ell i}^2)^2 \leq 3\mathsf{E} z_{\ell i}^4$ give

$$\mathsf{P}(\tilde{\mathcal{B}}^c) \leq \frac{\kappa_w^4 L}{9\kappa_2} \max_{1\leq\ell\leq L} \mathsf{E}\tilde{B}_\ell \leq \frac{3\kappa_w^4 L}{9\kappa_2} \max_{1\leq\ell\leq L} \mathsf{E} \textstyle\sum_{i=1}^n z_{\ell i}^4.$$

Use iterated expectations and $\max_{1\leq\ell\leq L} \mathsf{E} x_\ell \leq \mathsf{E} \max_{1\leq\ell\leq L} x_\ell$ to get

$$\mathsf{P}(\tilde{\mathcal{B}}^c) \leq \frac{\kappa_w^4 L}{3\kappa_2} \mathsf{E} \max_{1\leq\ell\leq L} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell,i}^4 = \frac{\kappa_w^4 L}{3\kappa_2} \mathsf{E} D_2.$$

9. *Combine* the bounds $\mathsf{P}(\mathcal{A} \cap \mathcal{B})$, $\mathsf{P}(\tilde{\mathcal{C}})$, $\mathsf{P}(\tilde{\mathcal{A}} \cap \tilde{\mathcal{B}})$, $\mathsf{P}(\tilde{\mathcal{B}}^c)$ in items 3,5,7,8. $\square$

The next result is a corollary to Theorem A.2 and modifies [11, Th. 5.2].

Theorem A.3. *For $1 \leq \ell \leq L$, let $\tilde{z}_{\ell i}$, $\tilde{w}_{i+1,n}$ be $\mathcal{F}_i$-adapted so $\mathsf{E}\tilde{z}^4_{\ell,i} < \infty$. Let $D_r = \max_{1\leq \ell\leq L}\sum_{i=1}^n \mathsf{E}_{i-1}\tilde{z}^{2^r}_{\ell i}$ for $r = 1, 2$. Suppose, $\exists\varsigma \geq 0$, $\lambda > 0$ so that $L = \mathrm{O}(n^\lambda)$ and $\mathsf{E}D_r = \mathrm{O}(n^\varsigma)$ for $r = 1, 2$. Then, $\forall\upsilon > 0$ so that*

(i) $\varsigma < 2\upsilon$, (ii) $\varsigma + \lambda < 4\upsilon$

it holds, $\forall\tilde{\kappa} \geq 0$, $\forall\gamma > 0$ and $\tilde{\mathcal{D}}_n = (\max_{1\leq i\leq n} |\tilde{w}_{in}| \leq n^{\tilde{\kappa}})$, that

$$\lim_{n\to\infty} \mathsf{P}\{\max_{1\leq \ell\leq L} |\textstyle\sum_{i=1}^n \tilde{w}_{in}(\tilde{z}_{\ell i} - \mathsf{E}_{i-1}\tilde{z}_{\ell i})| > \gamma n^{\upsilon+\tilde{\kappa}} \cap \tilde{\mathcal{D}}_n\} = 0.$$

Proof of Theorem A.3: Apply Theorem A.2, for a fixed n, while

$$\kappa_0 = \gamma n^{\upsilon+\tilde{\kappa}}, \qquad \kappa_1 = \frac{(\gamma n^{\upsilon+\tilde{\kappa}})^2}{7\lambda \log n}, \qquad \kappa_2 = \frac{(\gamma n^{\upsilon+\tilde{\kappa}})^4}{(7\lambda \log n)^3}, \qquad \kappa_w = n^{\tilde{\kappa}}$$

for any $\gamma > 0$ and $\tilde{\kappa} \geq 0$ so that $\kappa_0^2/\kappa_1 = \kappa_1^2/\kappa_2 = 7\lambda \log n$ and exploit conditions (i, ii) to see that the probability of interest for the particular coordinate of $n^{1/2}N'x_i$ satisfies

$$\mathcal{P}_n = \mathrm{O}(\frac{n^{2\tilde{\kappa}}}{n^{2\upsilon+2\tilde{\kappa}}/\log n} n^\varsigma + \frac{n^{4\tilde{\kappa}}n^\lambda}{n^{4\upsilon+4\tilde{\kappa}}/\log^3 n} n^\varsigma + n^\lambda n^{-7\lambda/6}) = \mathrm{o}(1),$$

as desired since $\varsigma < 2\upsilon$ and $\varsigma + \lambda < 4\upsilon$ while $\lambda > 0$. □

A.3 Preliminary Lemmas

Lemma A.4. *(Jiao, Nielsen, [8]): If $|\tilde{c} - c| \leq |Ac + B|$ and $|A| \leq 1/2$ then $|c| \leq 2(|\tilde{c}| + |B|)$ and $(Ac + B)^2 \leq 16(A^2\tilde{c}^2 + B^2)$.*

The next result generalises [8, Lemma 2].

Lemma A.5. *Suppose Assumption 2.2(ib3) holds. Define the function $\mathcal{H}(a, b, c) = \mathsf{E}_{i-1}\{(1 + |\varepsilon_i/\sigma|)|\dot{m}(\varepsilon_i/\sigma)|\{1_{(\varepsilon_i\leq\sigma c+n^{-1/2}ac/\sigma+x'_{in}b)} - 1_{(\varepsilon_i\leq\sigma c)}\}|\}$. Then, $\forall B > 0$, $\exists n_0, C > 0$, $\forall n > n_0$ so that*

$$\sup_{|a|,|b|\leq n^{1/4-\eta}B}\ \sup_{c\in\mathbb{R}} \mathcal{H}(a, b, c) \leq Cn^{-1/4-\eta}(1 + n^{1/2}|x_{in}|).$$

Proof of Lemma A.5. Apply the mean value theorem at the point c to get $\mathcal{H} = |cn^{-1/2}a/\sigma + x'_{in}b|\mathsf{J}(\tilde{c})$ where $\mathsf{J}(c) = (1 + |c|)|\dot{m}(c)|\mathsf{f}(c)$ while $|\tilde{c} - c| \leq |\sigma^{-1}n^{-1/2}ac + x'_{in}b|$. Bound $\mathcal{H} \leq (|\sigma^{-1}n^{-1/2}a||c| + |x'_{in}b|)\mathsf{J}(c)$ using the triangle inequality. Notice that there exists an n_0, so that for any $n > n_0$ then $|n^{-1/2}a/\sigma| \leq 1/2$ uniformly in $|a| \leq n^{1/4-\eta}B$. Hence, for $n > n_0$, the first inequality in Lemma A.4 shows $|c| \leq 2(|\tilde{c}| + |x'_{in}b|)$. Combine the bounds to get $\mathcal{H} \leq \{2|n^{-1/2}a/\sigma||\tilde{c}| + (1 + 2|n^{-1/2}a/\sigma|)|x'_{in}b|\}\mathsf{J}(\tilde{c})$. Now, note that $|a|, |b| \leq Bn^{1/4-\eta}$ while $\sup_{c\in\mathbb{R}}(1 + |c|)\mathsf{J}(c) < \infty$ by assumption. □

We now bound differences of $\dot{\mathsf{H}}^m$ over grid points. The result generalizes Johansen and Nielsen (2016a, Lemma B.1) with a simplied proof.

Lemma A.6. *Apply Assumption 2.2$(ia, ib2, ic)$ with $\tilde{m} = m$ only. Then*

$$\max_{1\leq k\leq K} |\dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1})| = \mathrm{O}(K^{-1/2}).$$

Proof of Lemma A.6: 1. *Definitions.* Introduce the bounding functions

$$\underline{\dot{\mathsf{H}}}(c) = \begin{cases} \inf_{0\leq d\leq c} \dot{\mathsf{H}}^m(d) \text{ for } c \geq 0, \\ \inf_{c\leq d\leq 0} \dot{\mathsf{H}}^m(d) \text{ for } c \leq 0, \end{cases} \quad \overline{\dot{\mathsf{H}}}(c) = \begin{cases} \sup_{0\leq c\leq d} \dot{\mathsf{H}}^m(d) \text{ for } c \geq 0, \\ \sup_{d\leq c\leq 0} \dot{\mathsf{H}}^m(d) \text{ for } c \leq 0. \end{cases}$$

The functions $\underline{\dot{\mathsf{H}}}(c)$ and $\overline{\dot{\mathsf{H}}}(c)$ are monotonic on $\mathbb{R}_+$ and $\mathbb{R}_-$. Assumption 2.2(ic), implies $\exists C_H > 0$, $\forall c \in \mathbb{R}$ then $\overline{\dot{\mathsf{H}}}(c) \leq C_{\mathsf{H}}\underline{\dot{\mathsf{H}}}(c)$ so that

$$C_{\mathsf{H}}^{-1}\overline{\dot{\mathsf{H}}}(c) \leq \underline{\dot{\mathsf{H}}}(c) \leq \dot{\mathsf{H}}^m(c) \leq \overline{\dot{\mathsf{H}}}(c) \leq C_{\mathsf{H}}\underline{\dot{\mathsf{H}}}(c). \tag{A.6}$$

2. *Bounding* $\dot{\mathcal{H}}_k = \dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1})$. We prove $\dot{\mathcal{H}}_k = \mathrm{O}(K^{-1/2})$ uniformly in k. Condition (i) shows $\dot{\mathsf{H}}$ is continuous and integrable. Thus, $\dot{\mathsf{H}}$ vanishes for large $|c|$ and for large K there exist $c_- \leq 0 \leq c_+$ so that

$$\dot{\mathsf{H}}^m(c_-) = \dot{\mathsf{H}}^m(c_+) = H^m/K^{1/2}. \tag{A.7}$$

We consider 5 cases depending on the location of c_k, c_{k-1} relative to c_+, c_-.
2.1. *When* $c_- \leq c_{k-1} \leq c_k \leq c_+$. The mean value theorem gives, for an intermediate point $\tilde{c}_k$, $\dot{\mathcal{H}}_k = \dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1}) = (c_k - c_{k-1})\ddot{\mathsf{H}}^m(\tilde{c}_k)$. Since $|\ddot{\mathsf{H}}|$ is uniformly bounded by Assumption 2.2$(ib2)$, then $|\dot{\mathcal{H}}_k| \leq C(c_k - c_{k-1})$ for some constant C, and it suffices to show $(c_k - c_{k-1}) = \mathrm{O}(K^{-1/2})$ uniformly in k. Note that $H^m < \infty$ by Assumption 2.2(ia), while $\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1}) = H^m/K$ by construction. The mean value theorem gives, for an intermediate point c^* that $\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1}) = (c_k - c_{k-1})\dot{\mathsf{H}}^m(c^*)$. The ordering (A.6) shows $\dot{\mathsf{H}}^m(\cdot) \geq \underline{\dot{\mathsf{H}}}(\cdot)$ so that $c_k - c_{k-1} \leq (H^m/K)/\underline{\dot{\mathsf{H}}}(c^*)$.

We now argue that $\underline{\dot{\mathsf{H}}}(c^*) \geq C_{\mathsf{H}}^{-1} H^m/K^{1/2}$. First, for $c^* \geq 0$ and noting $c^* \leq c_k \leq c_+$, the monotonicity of $\underline{\dot{\mathsf{H}}}$ gives $\underline{\dot{\mathsf{H}}}(c^*) \geq \underline{\dot{\mathsf{H}}}(c_+)$. The ordering in (A.6) gives $\underline{\dot{\mathsf{H}}}(c_+) \geq C_{\mathsf{H}}^{-1}\dot{\mathsf{H}}^m(c_+)$ while $\dot{\mathsf{H}}^m(c_+) = H^m/K^{1/2}$ by the construction (A.7). Similarly, for $c^* \leq 0$ and noting $c_- \leq c_{k-1} \leq c^*$ we get $\underline{\dot{\mathsf{H}}}(c^*) \geq \underline{\dot{\mathsf{H}}}(c_-)$ where $\underline{\dot{\mathsf{H}}}(c_-) \geq C_{\mathsf{H}}^{-1}\dot{\mathsf{H}}^m(c_-)$ while $\dot{\mathsf{H}}^m(c_-) = H^m/K^{1/2}$.

Combining the inequalities $c_k - c_{k-1} \leq (H^m/K)/\{\underline{\dot{\mathsf{H}}}(c^*)\}$ and $\underline{\dot{\mathsf{H}}}(c^*) \geq C_{\mathsf{H}}^{-1} H^m/K^{1/2}$ gives that $c_k - c_{k-1} \leq (H^m/K)/\{C_{\mathsf{H}}^{-1} H^m/K^{1/2}\} = C_H/K^{1/2}$ uniformly in k. Thus, $|\dot{\mathcal{H}}_k| \leq CC_H/K^{1/2}$.
2.2. *When* $c_+ \leq c_{k-1} \leq c_k$ use the triangle inequality and then the bound (A.6) to get $|\dot{\mathcal{H}}_k| \leq \dot{\mathsf{H}}^m(c_k) + \dot{\mathsf{H}}^m(c_{k-1}) \leq \overline{\dot{\mathsf{H}}}(c_k) + \overline{\dot{\mathsf{H}}}(c_{k-1})$. Noting $c_+ \leq c_{k-1} \leq c_k$, the monotonicity of $\overline{\dot{\mathsf{H}}}$, the ordering (A.6) and the construction (A.7) give $|\dot{\mathcal{H}}_k| \leq 2\overline{\dot{\mathsf{H}}}(c_+) \leq 2C_{\mathsf{H}}\dot{\mathsf{H}}^m(c_+) = 2C_{\mathsf{H}} H^m/K^{1/2}$.
2.3. *When* $c_{k-1} \leq c_k \leq c_-$ follow item 2.2 using c_- instead of c_+.
2.4. *When* $c_{k-1} \leq c^+ \leq c_k$. Recall $\dot{\mathcal{H}}_k = \dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1})$. Add and subtract $\dot{\mathsf{H}}^m(c_+)$ to $\dot{\mathcal{H}}_k$ and apply the triangle inequality to bound $|\dot{\mathcal{H}}_k| \leq |\dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_+)| + |\dot{\mathsf{H}}^m(c_+) - \dot{\mathsf{H}}^m(c_{k-1})|$. The first term involves the points $c_k \geq c_+$ while the second term involves the points $c_{k-1} \leq c_+$. Thus, modifying the arguments in items 2.2, 2.1, respectively, gives the further bound $|\dot{\mathcal{H}}_k| \leq 2C_{\mathsf{H}} H^m/K^{1/2} + CC_H/K^{1/2} = \mathrm{O}(K^{-1/2})$.
2.5. *When* $c_{k-1} \leq c_- \leq c_k$ follow item 2.4, using at c_- instead of c_+. ☐

A.4 Chaining Lemmas without estimation error

We present a maximal inequality for sums of $z_i(c) = w_{in} m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)}$ without estimation error. The first two lemmas analyze the discrete points term and the oscillation term. The third lemma combines the two results.

Lemma A.7. ***Discrete points term.*** *Apply Assumptions 2.1, 2.2(ia, ii, iii) with* $\tilde{m} = m$, $\omega = 0$ *only. Apply the chaining setup in §A.1 for some* $\delta > 0$. *Let* $d = 0, 1$ *and* $\kappa \geq 0$. *Let* $z_{ki} = w_{in}(n^{1/2}x_{in})^d m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c_k)}$. *Then,* $\forall \psi > 0$, *we get* $\max_{1 \leq k \leq K} |\sum_{i=1}^n (z_{ki} - \mathsf{E}_{i-1} z_{ki})| = \mathrm{o}_{\mathsf{P}}(n^{3/4 + d\kappa + \psi})$.

Proof of Lemma A.7: 1. *Truncation.* For some $C_x > 0$, $\psi > 0$, let

$$\mathcal{C}_n = (\max_{1 \leq i \leq n} |n^{1/2}x_{in}| \leq C_x n^\kappa), \qquad \mathcal{D}_n = (\max_{1 \leq i \leq n} |w_{in}| \leq n^{1/2+\psi}). \tag{A.8}$$

By Assumption 2.2(ii) $\forall \epsilon > 0$ $\exists C_x, n_0 > 0$: $\mathsf{P}(\mathcal{C}_n^c) < \epsilon$ for $n > n_0$ and $\mathsf{P}(\mathcal{D}_n^c)$ vanishes since Assumption 2.2(iii) and Boole and Markov inequalities imply

$$\mathsf{P}(\mathcal{D}_n^c) = \mathsf{P}\textstyle\bigcup_{i=1}^n (|w_{in}| > n^\alpha) \leq n^{-\alpha/\alpha_0} \sum_{i=1}^n \mathsf{E}|w_{in}|^\alpha = \mathrm{o}(1). \tag{A.9}$$

Thus, it suffices to show the result on $\mathcal{C}_n \cap \mathcal{D}_n$. We note that on $\mathcal{C}_n \cap \mathcal{D}_n$ and for $d = 0, 1$ then $\max_{1 \leq i \leq n} |n^{1/2}x_{in}|^d |w_{in}|^{1/2} \leq n^{\tilde{\kappa}}$ with $\tilde{\kappa} = d\kappa + 1/4 + \psi/2$.
2. *Apply Theorem* A.3 with $\ell = k$ and $L = K = \mathrm{O}(n^\lambda)$ so $\lambda = 1/2$; let $\upsilon = 1/2 + \psi/2$ and $\tilde{\kappa}$ as above so that $\upsilon + \tilde{\kappa} = 3/4 + d\kappa + \psi$; choose $\tilde{z}_{\ell i} = |w_{in}|^{-1/2} w_{in} m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c_k)}$ and $\tilde{w}_{in} = |w_{in}|^{1/2}(n^{1/2}x_{in})^d$

so that $\tilde{w}_{in}\tilde{z}_{\ell i} = z_{ki}$ and $\tilde{\mathcal{D}}_n = (\max_{1\leq i\leq n}|\tilde{w}_{in}| \leq n^{\tilde{\kappa}})$ with $\tilde{\mathcal{D}}_n^c \subset \mathcal{C}_n^c \cup \mathcal{D}_n^c$; while $\varsigma = 1$. We check the conditions of Theorem A.3.
2.1. *Condition* $\mathsf{E}\tilde{z}_{\ell i}^4 < \infty$ holds since $\mathsf{E}(\tilde{z}_{\ell i}^4) = \mathsf{E}|w_{in}|^2\mathsf{E}|m(\varepsilon_i/\sigma)|^4 < \infty$ by independence and Assumption 2.2(ia, ii).
2.2. *Condition* $\mathsf{E}D_q = \mathrm{O}(n^\varsigma)$ for $1 \leq q \leq 2$. The bound (A.2) and Assumption 2.2(ia) gives, for $q = 1, 2$, that, $\mathsf{E}_{i-1}\{|m(\varepsilon_i/\sigma)|^{2^q}1_{(\varepsilon_i\leq\sigma c_k)}\} \leq \mathsf{H}^m(c_k) \leq H^m < \infty$. As a consequence $\mathsf{E}_{i-1}\tilde{z}_{\ell i}^{2^q} = |w_{in}|^{2^{q-1}}\mathsf{E}_{i-1}\{m^{2^q}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c_k)}\}$ is bounded by $\mathsf{E}_{i-1}\tilde{z}_{\ell i}^{2^q} \leq |w_{in}|^{2^{q-1}}H^m$, uniformly in k, so that $\mathsf{E}D_q = \mathsf{E}\max_{1\leq k\leq K}\sum_{i=1}^n\mathsf{E}_{i-1}\tilde{z}_{ki}^{2^q}$ is bounded by $\mathsf{E}D_q \leq H^m\mathsf{E}\sum_{i=1}^n|w_{in}|^{2^{q-1}}$, which is $\mathrm{O}(n) = \mathrm{O}(n^\varsigma)$ by Assumption 2.2(ii).
2.3. *Conditions* $(i), (ii)$: $\varsigma = 1 < 2\upsilon = 1 + \psi$ and $\varsigma + \lambda = 1 + 1/2 < 4\upsilon$. □

Lemma A.8. *Oscillation term.* *Apply Assumptions 2.1, 2.2(ia, ii, iii) with $\tilde{m} = m$ only. Apply the chaining setup in §A.1 for some $\delta > 0$. Let $d = 0, 1$ while $\kappa \geq 0$. Define $z_i(c, c_k) = w_{in}(n^{1/2}x_{in})^d m(\varepsilon_i/\sigma)1_{(c<\varepsilon_i/\sigma\leq c_k)}$ for $c_{k-1} < c \leq c_k$. Then*

$$\max_{1\leq k\leq K}\sup_{c:c_{k-1}<c\leq c_k}|\textstyle\sum_{i=1}^n\{z_i(c, c_k) - \mathsf{E}_{i-1}z_i(c, c_k)\}| = \mathrm{o}_{\mathsf{P}}(n^{1/2+d\kappa}).$$

Proof of Lemma A.8: 1. *Truncation.* By Assumption 2.2(ii, iii) we can choose $\mathcal{C}_n$ as in (A.8) and $\mathcal{D}_n = (\max_{1\leq i\leq n}|w_{in}| \leq n^{1/2-\omega/5})$ for $\omega > 0$ given in Assumption 2.2(iii), so that $\mathsf{P}(\mathcal{C}_n^c \cup \mathcal{D}_n^c)$ vanishes. Thus, it suffices to show the result on $\mathcal{C}_n \cap \mathcal{D}_n$.
2. *A first bound.* Let $\mathcal{M}_{nkc} = \sum_{i=1}^n\{z_i(c, c_k) - \mathsf{E}_{i-1}z_i(c, c_k)\}$. Note

$$|z_i(c, c_k)| \leq z_{ki} = |w_{in}||n^{1/2}x_{in}|^d|m(\varepsilon_i/\sigma)|1_{(c_{k-1}<\varepsilon_i/\sigma\leq c_k)}, \tag{A.10}$$

for $c_{k-1} \leq c \leq c_k$. Thus, $|\mathcal{M}_{nkc}| \leq \mathcal{M}_{nk} = \sum_{i=1}^n(z_{ki} + \mathsf{E}_{i-1}z_{ki})$ uniformly in c. Decompose $\mathcal{M}_{nk} = \tilde{\mathcal{M}}_{nk} + 2\bar{\mathcal{M}}_{nk}$ where $\tilde{\mathcal{M}}_{nk} = \sum_{i=1}^n(z_{ki} - \mathsf{E}_{i-1}z_{ki})$ and $\bar{\mathcal{M}}_{nk} = \sum_{i=1}^n\mathsf{E}_{i-1}z_{ki}$. Thus, it suffices to show that $\max_{1\leq k\leq K}|\tilde{\mathcal{M}}_{nk}|$ and $\max_{1\leq k\leq K}\bar{\mathcal{M}}_{nk}$ are of the desired order.
3. *The compensator* is $\max_k\bar{\mathcal{M}}_{nk} = \mathrm{o}_{\mathsf{P}}(n^{1/2+d\kappa})$. Since $\mathsf{P}(\mathcal{C}_n^c) \to 0$ it suffices that $1_{\mathcal{C}_n}\max_k\bar{\mathcal{M}}_{nk} = \mathrm{o}_{\mathsf{P}}(n^{1/2+d\kappa})$. Apply the bound (A.2) and the z_{ki} expression in (A.10) to bound
$\mathsf{E}_{i-1}z_{ki} \leq |w_{in}||n^{1/2}x_{in}|^d\{\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1})\}$. The chaining setup in §A.1 and Assumption 2.2(i) give $\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1}) = H^m/K = \delta\mathrm{O}(n^{-1/2})$ so that $\mathsf{E}_{i-1}z_{ki} = |w_{in}||n^{1/2}x_{in}|^d\delta\mathrm{O}(n^{-1/2})$. By Assumption 2.2($iii$) then $\mathsf{E}\sum_{i=1}^n|w_{in}|1_{\mathcal{C}_n} \leq \mathsf{E}\sum_{i=1}^n|w_{in}| = \mathrm{O}(n)$, so that

$$\mathsf{E}1_{\mathcal{C}_n}\max_k\bar{\mathcal{M}}_{nk} = \mathsf{E}1_{\mathcal{C}_n}\max_{1\leq k\leq K}\textstyle\sum_{i=1}^n\mathsf{E}_{i-1}z_{ki} \leq \delta\mathrm{O}(n^{d\kappa-1/2})\mathsf{E}\sum_{i=1}^n|w_{in}| = \delta\mathrm{O}(n^{1/2+d\kappa}). \tag{A.11}$$

The Markov inequality gives $\mathsf{P}(1_{\mathcal{C}_n}\max_k n^{-1/2-d\kappa}\bar{\mathcal{M}}_{nk} > \epsilon) \leq \mathrm{O}(1)\delta/\epsilon$ for all $\delta, \epsilon > 0$. For any $\epsilon > 0$ we can choose δ small. The desired bound follows.
4. *The martingale* is $\max_{1\leq k\leq K}|\tilde{\mathcal{M}}_{nk}| = \mathrm{o}_{\mathsf{P}}(n^{1/2})$. Use Theorem A.3 while truncating to $\mathcal{C}_n\cap\mathcal{D}_n$. In Theorem A.3 let $\ell = k$ and $L = K = \mathrm{O}(n^\lambda)$ with $\lambda = 1/2$; with $\upsilon = 1/4 + \omega/10$ and $\tilde{\kappa} = 1/4 + d\kappa - \omega/10$ so that $\upsilon + \tilde{\kappa} = 1/2 + d\kappa$; with $\tilde{w}_{in} = |w_{in}|^{1/2}|n^{1/2}x_{in}|^d$ and $\tilde{z}_{\ell i} = |w_{in}|^{-1/2}w_{in}m(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c_k)}$ so that $\tilde{w}_{in}\tilde{z}_{\ell i} = z_{ki}$; and with $\varsigma = 1/2$. We check the Lemma A.3 conditions .
4.1. *Condition* $\mathsf{E}(\tilde{z}_{\ell i}^4) < \infty$ holds by Assumption 2.2(i), (iii), see also proof of Lemma A.7, item 2.1.
4.2. *Condition* $\mathsf{E}D_q = \mathrm{O}(n^\varsigma)$ for $q = 1, 2$. Note $D_q = \max_{1\leq k\leq K}\sum_{i=1}^n\mathsf{E}_{i-1}\tilde{z}_{ki}^{2^q}$ so that $D_q = \max_{1\leq k\leq K}\sum_{i=1}^n\mathsf{E}_{i-1}|w_{in}|^{2^{q-1}}m^{2^q}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c_k)} = \delta\mathrm{O}(n^{1/2})$ as in (A.11).
4.3. *Conditions* $(i), (ii)$: $\varsigma = 1/2 < 2\upsilon = 1/2 + \omega/5$ and $\varsigma + \lambda = 1/2 + 1/2 = 1 < 4\upsilon = 1 + 2\omega/5$. □

Lemma A.9. *Maximal inequality.* *Let $d = 0, 1$ while $\kappa \geq 0$.*
Define $z_i(c) = w_{in}(n^{1/2}x_{in})^d m(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c)}$. Apply Assumptions 2.1, 2.2(ia, ii, iii) with $\tilde{m} = m$ only. Then $\sup_{c\in\mathbb{R}}|\sum_{i=1}^n\{z_i(c) - \mathsf{E}_{i-1}z_i(c)\}| = \mathrm{o}_{\mathsf{P}}(n^{3/4+d\kappa+\omega})$.

Proof of Lemma A.9: We prove this result by chaining over c. For any $\delta > 0$ consider the distance function H^m defined in (A.1) and $K = \mathrm{int}(n^{1/2}/\delta)$ discrete points grid points c_k chosen in §A.1. Apply then the chaining inequality (A.4). Lemmas A.7 and A.8 analyze the discrete points term and the oscillation term, respectively. □

A.5 Chaining Lemmas with estimation error

Two maximal inequalities are presented. The first result is concerned with additive estimation error. It replaces the polynomial marks $(\varepsilon_i/\sigma)^p$ in [11, Theorem 4.1] by a general mark function $m(\varepsilon_i/\sigma)$ while improving the proof of that result.

Lemma A.10. *Apply Assumptions 2.1, 2.2(ia, ib2, ic, ii, iii) with $\tilde{m} = m$ only. Let $0 \leq \kappa < \eta \leq 1/4$. Consider the distance function $\mathsf{H}^m(c)$ in (A.1) and its derivative $\dot{\mathsf{H}}^m(c) = \{1 + m^4(c)\}\mathsf{f}(c)$. Let $d = 0, 1$ and $z_i(b,c) = w_{in}(n^{1/2}N'x_i)^d m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \leq \sigma c + x_{in}'b)} - 1_{(\varepsilon_i \leq \sigma c)}\}$. Then*

$$\sup_{|b| \leq Bn^{1/4-\eta}} \sup_{c \in \mathbb{R}} |\textstyle\sum_{i=1}^n \{z_i(b,c) - \mathsf{E}_{i-1} z_i(b,c)\}| = \mathsf{o}_\mathsf{P}(n^{1/2+d\kappa}).$$

Proof of Lemma A.10: 1. *Notation.* Let $M_n(b,c) = \sum_{i=1}^n \{z_i(b,c) - \mathsf{E}_{i-1} z_i(b,c)\}$. We want to show $M_n(b,c) = \mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in b, c. Let c_k be the nearest right grid point to c. We rewrite $z_i(b,c)$ by adding and subtracting $1_{(\varepsilon_i \leq \sigma c_k)}$ to get $z_i(b,c) = z_i^\dagger(b,c,c_k) - z_i^\dagger(0,c,c_k)$, where

$$z_i^\dagger(b,c,c_k) = w_{in}(n^{1/2}N'x_i)^d m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \leq \sigma c + x_{in}'b)} - 1_{(\varepsilon_i \leq \sigma c_k)}\}.$$

Hence, we have $M_n(b,c) = M_n^\dagger(b,c,c_k) - M_n^\dagger(0,c,c_k)$ with

$$M_n^\dagger(b,c,c_k) = \textstyle\sum_{i=1}^n \{z_i^\dagger(b,c,c_k) - \mathsf{E}_{i-1} z_i^\dagger(b,c,c_k)\}.$$

Thus, $M_n(b,c) = \mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in b, c if

$$\sup_{|b| \leq Bn^{1/4-\eta}} \max_{1 \leq k \leq K} \sup_{c_{k-1} \leq c \leq c_k} |M_n^\dagger(b,c,c_k)|, \qquad \max_{1 \leq k \leq K} \sup_{c_{k-1} \leq c \leq c_k} |M_n^\dagger(0,c,c_k)|.$$

are both $\mathsf{o}_\mathsf{P}(n^{1/2+d\kappa})$. The second term was analyzed in Lemma A.8. It is also bounded by the first term, so it suffices to show that $M_n^\dagger(b,c,c_k) = \mathsf{o}_\mathsf{P}(n^{1/2})$.

2. *Truncating regressors and martingale decomposition.* Following (A.8) then Assumption 2.2(ii) shows $\forall \epsilon > 0\ \exists C_x, n_0 > 0$ so that the sets

$$\mathcal{C}_n = (\max_{1 \leq i \leq n} |n^{1/2}x_{in}| \leq C_x n^\kappa), \qquad \mathcal{C}_{in} = (|n^{1/2}x_{in}| \leq C_x n^\kappa), \tag{A.12}$$

satisfy $\mathsf{P}(\mathcal{C}_n^c) < \epsilon$ for $n > n_0$, while $\mathcal{C}_n \subseteq \mathcal{C}_{in}$ and $\mathcal{C}_{in}$ is $\mathcal{F}_{i-1}$-adapted. Thus, $n^{-1/2}|M_n^\dagger(b,c,c_k)|$ vanishes if $n^{-1/2}|M_n^\dagger(b,c,c_k)|1_{\mathcal{C}_n}$ vanishes. By the triangle inequality and $\mathcal{C}_n \subseteq \mathcal{C}_{in}$ we get

$$|M_n^\dagger(b,c,c_k)|1_{\mathcal{C}_n} \leq \textstyle\sum_{i=1}^n \{|z_i^\dagger(b,c,c_k)1_{\mathcal{C}_{in}}| + \mathsf{E}_{i-1}|z_i^\dagger(b,c,c_k)1_{\mathcal{C}_{in}}|\}.$$

We bound $z_i^\dagger(b,c,c_k)1_{\mathcal{C}_{in}}$. First, recalling the bound to b, we get, on $\mathcal{C}_{in}$,

$$|x_{in}'b| \leq |b||x_{in}| \leq Bn^{1/4-\eta}C_x n^{\kappa-1/2} = BC_x n^{\kappa-\eta-1/4} \leq K^{-1/2},$$

where the last inequality holds for large n since $\eta > \kappa$ while $K = \text{int}(n^{1/2}/\delta)$ for fixed δ. Since $c_{k-1} < c \leq c_k$ we can now bound the indicator functions in the summands $z_i^\dagger(b,c,c_k)$, on $\mathcal{C}_{in}$,

$$|1_{(\varepsilon_i \leq \sigma c + x_{in}'b)} - 1_{(\varepsilon_i \leq \sigma c_k)}| \leq 1_{(\varepsilon_i \leq \sigma c_k + K^{-1/2})} - 1_{(\varepsilon_i \leq \sigma c_{k-1} - K^{-1/2})}.$$

Exploiting the truncation on $\mathcal{C}_{in}$ and the above bounds we get, for $d = 0, 1$ that
$0 \leq |z_i^\dagger(b,c,c_k)|1_{\mathcal{C}_{in}} \leq z_i^\ddagger(c_k, c_{k-1})$ uniformly in b, c, where

$$z_i^\ddagger(c_k, c_{k-1}) = C_x^d n^{d\kappa}|w_{in}||m(\varepsilon_i/\sigma)|\{1_{(\varepsilon_i \leq \sigma c_k + K^{-1/2})} - 1_{(\varepsilon_i \leq \sigma c_{k-1} - K^{-1/2})}\}.$$

Thus, we can bound

$$|M_n^\dagger(b,c,c_k)|1_{\mathcal{C}_n} \leq M_n^\ddagger(c_k, c_{k-1}) = \textstyle\sum_{i=1}^n \{z_i^\ddagger(c_k, c_{k-1}) + \mathsf{E}_{i-1} z_i^\ddagger(c_k, c_{k-1})\}.$$

Now, $M_n^\ddagger$ has martingale decomposition $M_n^\ddagger = \tilde{M}_n^\ddagger + 2\bar{M}_n^\ddagger$ where

$$\begin{aligned}\tilde{M}_n^\ddagger(c_k, c_{k-1}) &= \textstyle\sum_{i=1}^n \{z_i^\ddagger(c_k, c_{k-1}) - \mathsf{E}_{i-1} z_i^\ddagger(c_k, c_{k-1})\},\\ \bar{M}_n^\ddagger(c_k, c_{k-1}) &= \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_i^\ddagger(c_k, c_{k-1}).\end{aligned}$$

3. *The compensator* is $\max_k \bar{M}_n^\ddagger(c_k, c_{k-1}) = \mathrm{o}_\mathsf{P}(n^{1/2})$. To see this define $\mathcal{H}_k = \mathsf{H}^m(c_k + K^{-1/2}) - \mathsf{H}^m(c_{k-1} - K^{-1/2})$, so that the bound (A.2) implies
$\max_{1\le k\le K} \mathsf{E}_{i-1} z_i^\ddagger(c_k, c_{k-1}) \le C_x^d n^{d\kappa}|w_{in}| \max_{1\le k\le K} \mathcal{H}_k$. The mean value theorem gives, for c_k^*, c_{k-1}^* so $|c_k^* - c_k|, |c_{k-1}^* - c_{k-1}|) \le 2K^{-1/2}$ that

$$\begin{aligned}\mathsf{H}^m(c_k + K^{-1/2}) &= \mathsf{H}^m(c_k) + K^{-1/2}\dot{\mathsf{H}}^m(c_k) + (K^{-1}/2)\ddot{\mathsf{H}}^m(c_k^*),\\ \mathsf{H}^m(c_{k-1} - K^{-1/2}) &= \mathsf{H}^m(c_{k-1}) + K^{-1/2}\dot{\mathsf{H}}^m(c_{k-1}) + (K^{-1}/2)\ddot{\mathsf{H}}^m(c_{k-1}^*).\end{aligned}$$

Taking difference and using the triangle inequality

$$|\mathcal{H}_k| = |\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1})| + K^{-1/2}|\dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1})| + (K^{-1}/2)|\ddot{\mathsf{H}}^m(c_k^*) - \ddot{\mathsf{H}}^m(c_{k-1}^*)|.$$

The first term is $H/K = \mathrm{O}(K^{-1})$ by construction. The second term is $\mathrm{O}(K^{-1})$ since $|\dot{\mathsf{H}}^m(c_k) - \dot{\mathsf{H}}^m(c_{k-1})| = \mathrm{O}(K^{-1/2})$ by Lemma A.6 and its assumptions are satisfied by Assumption 2.2 $(ia, ib2, ic)$. The third term is $\mathrm{O}(K^{-1})$ since $\ddot{\mathsf{H}}^m(.)$ is uniformly bounded by Assumption 2.2$(ib2)$. Hence, $|\mathcal{H}_k| = \mathrm{O}(K^{-1})$ so that

$$\max_{1\le k\le K} \mathsf{E}_{i-1} z_i^\ddagger(c_k, c_{k-1}) \le n^{d\kappa}|w_{in}|\mathrm{O}(K^{-1}) = |w_{in}|\delta\mathrm{O}(n^{d\kappa-1/2}). \tag{A.13}$$

Thus, we get uniformly in k that $\bar{M}_n^\ddagger(c_k, c_{k-1}) = \sum_{i=1}^n |w_{in}|\delta\mathrm{O}(n^{d\kappa-1/2})$. Since $\sum_{i=1}^n |w_{in}| = \mathrm{O}_\mathsf{P}(n)$ by Assumption 2.2(iii) and the Markov inequality we get $\bar{M}_n^\ddagger(c_k, c_{k-1}) = \delta\mathrm{O}_\mathsf{P}(n^{1/2+d\kappa})$. Since $\delta > 0$ can be chosen arbitrarily small, then $\bar{M}_n^\ddagger = \mathrm{o}_\mathsf{P}(n^{1/2+d\kappa})$.
4. *The martingale* is $\max_k \tilde{M}_n^\ddagger(c_k, c_{k-1}) = \mathrm{o}_\mathsf{P}(n^{1/2+d\kappa})$. It suffices to show the result on $\mathcal{D}_n = (\max_{1\le i\le n} |w_{in}| \le n^{1/2-\omega/5})$ since $\mathsf{P}(\mathcal{D}_n^c)$ vanishes by Assumption 2.2(ii), see (A.9). Apply Lemma A.3 with $\ell = k$ and $L = K = \mathrm{O}(n^{1/2}/\delta)$ so $\lambda = 1/2$; let $\upsilon = 1/4 + \omega/10$, $\kappa = 1/4 - \omega/10$ and $\varsigma = 1$. Choose $\tilde{w}_{in} = C_x^d n^{d\kappa}|w_{in}|^{1/2}$ and $\tilde{z}_{\ell i} = |w_{in}|^{-1/2} z_i^\ddagger(c_k, c_{k-1})$ so that

$$\tilde{z}_{\ell i} = |w_{in}|^{-1/2}|w_{in}||m(\varepsilon_i/\sigma)|\{1_{(\varepsilon_i \le \sigma c_k + K^{-1/2})} - 1_{(\varepsilon_i \le \sigma c_{k-1} - K^{-1/2})}\},$$

so that $\tilde{w}_{in}\tilde{z}_{\ell i} = z_i^\ddagger(c_k, c_{k-1})$. We check the conditions of Lemma A.3.
Condition $\mathsf{E}\tilde{z}_{\ell i}^4 < \infty$ holds by Assumption 2.2(ia, ii), see also proof of Lemma A.7, item 2.1.
Condition $\mathsf{E}D_q = \mathrm{O}(n^\varsigma)$ for $q = 1, 2$. Analyze $D_q = \max_{1\le k\le K} \sum_{i=1}^n \mathsf{E}_{i-1}\tilde{z}_{\ell i}^{2^q}$. Proceed as in (A.13) to get $D_q = \sum_{i=1}^n |w_{in}|^2 \delta\mathrm{O}(n^{-1/2})$ for $q = 1, 2$ so that $\mathsf{E}D_q = \delta\mathrm{O}_\mathsf{P}(n^{1/2})$ by Assumption 2.2(ii).
Conditions $(i), (ii)$: $\varsigma = 1/2 < 2\upsilon = 1/2 + \omega/5$ and $\varsigma + \lambda = 1/2 + 1/2 = 1 < 4\upsilon = 1 + 2\omega/5$. □

The next result concerns the scale estimation error. It generalizes [8, Theorem 5] and uses a bivariate chaining argument in the proof.

Lemma A.11. *Apply Assumptions 2.1, 2.2$(ia, ib1, iii)$ with $\tilde{m} = m$ only.*
Let $z_i(a, c) = w_{in} m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \le \sigma c + n^{-1/2} ac)} - 1_{(\varepsilon_i \le \sigma c)}\}$. Then, $\forall \eta > 0$,

$$\sup_{|a|\le Bn^{1/4-\eta}} \sup_{c\in\mathbb{R}} |\textstyle\sum_{i=1}^n \{z_i(a, c) - \mathsf{E}_{i-1} z_i(a, c)\}| = \mathrm{o}_\mathsf{P}(n^{1/2}).$$

Proof of Lemma A.11: Let $z_i(c^*, c) = z_i(a, c)$ and $c^* = c + n^{-1/2}ac/\sigma$ so

$$z_i(c^*, c) = w_{in} m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \le \sigma c^*)} - 1_{(\varepsilon_i \le \sigma c)}\}, \tag{A.14}$$

and the object of interest is the martingale $\mathcal{M}_n(c^*, c) = \sum_{i=1}^n \{z_i(c^*, c) - \mathsf{E}_{i-1} z_i(c^*, c)\}$. Let c_{k^*} and c_k be the nearest right grid points to c^* and c, respectively, and decompose

$$\mathcal{M}_n(c^*, c) = \mathcal{M}_n(c^*, c_{k^*}) + \mathcal{M}_n(c_{k^*}, c_k) + \mathcal{M}_n(c_k, c).$$

We show that each term is $\mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in a, c.
1. *The oscillation terms* $\mathcal{M}_n(c^*, c_{k^*})$ *and* $\mathcal{M}_n(c_k, c)$ are $\mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in a, c by Lemma A.8 with $d = 0$ using Assumption 2.2 (ia, iii).
2. *The discrete points term* $\mathcal{M}_n(c_{k^*}, c_k)$ is $\mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in a, c.
2.1. *Distance between* c_{k^*}, c_k. Let $\mathcal{H} = |\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k^*})|$. Note first that

$$\mathcal{H} \leq |\mathsf{H}^m(c) - \mathsf{H}^m(c^*)| + \{\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k-1})\} + \{\mathsf{H}^m(c_{k^*}) - \mathsf{H}^m(c_{k^*-1})\}.$$

Apply the mean value theorem to the first term while the last two terms equal H^m/K to get that $\mathcal{H} \leq |c - c^*| \dot{\mathsf{H}}^m(\tilde{c}) + 2H^m K^{-1}$, for an intermediate point $\tilde{c}$ so $|\tilde{c} - c| \leq |c - c^*|$. Now, $c - c^* = n^{-1/2} ac/\sigma$ where $|n^{-1/2} a/\sigma| \leq n^{-1/4-\eta} B/\sigma$. Lemma A.4 with $A = n^{-1/2} a/\sigma$ and $B = 0$ shows that $|c| \leq 2|\tilde{c}|$. Further, $|u| \dot{\mathsf{H}}^m(u) = |u|\{1 + m^4(u)\}\mathsf{f}(u)$ is bounded uniformly in $u \in \mathbb{R}$ by Assumption 2.2$(ib1)$, while $K = \text{int}(n^{1/2}/\delta)$. Hence,

$$\mathcal{H} \leq C|n^{-1/2} a/\sigma| + 2H^m K^{-1} = \{\mathrm{O}(n^{1/4-\eta}) + 2H^m\}K^{-1} = \mathrm{O}(n^{1/4-\eta})(H^m/K). \tag{A.15}$$

Thus, the number of grid points between c_{k^*}, c_k is $\mathrm{O}(n^{1/4-\eta})$.
2.2. *Cover of* c_{k^*}, c_k *for all* k^*, k. Since c_k takes $K = \mathrm{O}(n^{1/2})$ values and there are $\mathrm{O}(n^{1/4-\eta})$ grid points between c_{k^*}, c_k, uniformly in k, we have $L = \mathrm{O}(n^{1/2})\mathrm{O}(n^{1/4-\eta})$ combinations of k^*, k.
2.3. *Apply Theorem A.3* with $\ell = (k^*, k)$ and $L = \mathrm{O}(n^\lambda)$ so $\lambda = 3/4 - \eta$. Let $\upsilon = 3/8 - \eta/2 + \omega$ and $\tilde{\kappa} = 1/8 + \eta/2 - \omega$ so that $\upsilon + \tilde{\kappa} = 1/2$. Choose $\tilde{w}_{in} = |w_{in}|^{1/2}$ so that, using (A.14),

$$\tilde{z}_{\ell i} = |w_{in}|^{-1/2} z_i(c^*, c) = |w_{in}|^{-1/2} w_{in} m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \leq \sigma c^*)} - 1_{(\varepsilon_i \leq \sigma c)}\}.$$

It suffices to show the result on $\mathcal{D}_n = (\max_{1 \leq i \leq n} |w_{in}| \leq n^{1/2-\omega/5})$ since $\mathsf{P}(\mathcal{D}_n^c)$ vanishes for large n by Assumption 2.2$(ib1)$, see (A.9). We check the conditions of Theorem A.3.
Condition $\mathsf{E}(\tilde{z}_{\ell i}^4) < \infty$ holds by Assumption 2.2(i, iii), see also proof of Lemma A.7, item 2.1.
Condition $\mathsf{E}D_q = \mathrm{O}(n^\varsigma)$ for $q = 1, 2$. Let $D_q = \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} \tilde{z}_{\ell i}^{2^q}$. By the bound (A.2) then

$$\mathsf{E}_{i-1} \tilde{z}_{\ell i}^{2^q} = |w_{in}|^{2^{q-1}} \mathsf{E}_{i-1} m^{2^q}(\varepsilon_i/\sigma)|1_{(\varepsilon_i \leq \sigma c^*)} - 1_{(\varepsilon_i \leq \sigma c)}| \leq |w_{in}|^{2^{q-1}} |\mathsf{H}^m(c_k) - \mathsf{H}^m(c_{k^*})|. \tag{A.16}$$

By (A.15) and $K^{-1} = \mathrm{O}(n^{-1/2})$ then $\mathsf{E}_{i-1} \tilde{z}_{\ell i}^{2^q} = |w_{in}|^{2^{q-1}} \mathrm{O}(n^{-1/4-\eta})$ uniformly in ℓ. By Assumption 2.2(iii) then

$$\mathsf{E}D_q = \mathrm{O}(n^{-1/4-\eta}) \mathsf{E} \sum_{i=1}^n |w_{in}|^{2^{q-1}} = \mathrm{O}(n^{3/4-\eta}) = \mathrm{O}(n^\varsigma).$$

Conditions $(i), (ii)$: $\varsigma = 3/4 - \eta < 3/4 - \eta + 2\omega = 2\upsilon$ and $\varsigma + \lambda = 3/2 - 2\eta \leq 3/2 - 2\eta + 4\omega = 4\upsilon$.□

A.6 A Lipschitz result

This is the last ingredient to the main result.

Lemma A.12. *Apply Assumption 2.1, 2.2$(ia, ib2, ic, id, ii, iii)$ where $\tilde{m}$ is either $u\dot{m}(u)$ or $\dot{m}(u)$ only. Let $s_{mi} = n^{-1/2} a_m \varepsilon_i/\sigma + n^{-1/2} b_m'(n^{1/2} x_{in})$ and $v_i(\theta, c) = w_{in}\sigma^{-1}[m\{(\varepsilon_i - s_{mi})/\sigma\} - m(\varepsilon_i/\sigma)]1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)}$ where $\theta = (0, b_1, a_m, b_m)$. Then, $\forall B > 0$, it holds*

$$\sup_{|\theta| \leq Bn^{1/4-\eta}} \sup_{c \in \mathbb{R}} |n^{-1/2} \sum_{i=1}^n \{v_i(\theta, c) - \mathsf{E}_{i-1} v_i(\theta, c)\}| = \mathsf{o}_\mathsf{P}(1).$$

Proof of Lemma A.12: *Decomposition.* By the mean value theorem,

$$m(\frac{\varepsilon_i - s_{mi}}{\sigma}) = m(\frac{\varepsilon_i}{\sigma}) - \frac{s_{mi}}{\sigma}\dot{m}(\frac{\varepsilon_i}{\sigma}) - \frac{s_{mi}}{\sigma}\{\dot{m}(\frac{\varepsilon_i^*}{\sigma}) - \dot{m}(\frac{\varepsilon_i}{\sigma})\},$$

for intermediate points ε_i^* so $|\varepsilon_i^* - \varepsilon_i| \leq |s_{mi}|$. Insert this in the expression for $v_i(\theta, c)$, add and subtract $w_{in}s_{mi}\sigma^{-1}\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c)}$ to get $v_i(\theta, c) = \sum_{s=1}^3 v_{si}(\theta, c)$ where

$$v_{1i}(\theta, c) = -w_{in}s_{mi}\sigma^{-1}\dot{m}(\varepsilon_i/\sigma)\{1_{(\varepsilon_i\leq\sigma c+x_{in}'b_1)} - 1_{(\varepsilon_i\leq\sigma c)}\},$$
$$v_{2i}(\theta, c) = -w_{in}s_{mi}\sigma^{-1}\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i\leq\sigma c)},$$
$$v_{3i}(\theta, c) = -w_{in}s_{mi}\sigma^{-1}\{\dot{m}(\varepsilon_i^*/\sigma) - \dot{m}(\varepsilon_i/\sigma)\}1_{(\varepsilon_i\leq\sigma c+x_{in}'b_1)}.$$

Likewise, let $\mathcal{V}_{sn}(\theta, c) = n^{-1/2}\sum_{i=1}^n\{v_{si}(\theta, c) - \mathsf{E}_{i-1}v_{si}(\theta, c)\}$. By the triangle inequality it suffices to show that each $\mathcal{V}_{sn}$ is $\mathrm{o_P}(1)$ uniformly in θ, c.

1. *The term* $\mathcal{V}_{1n}(\theta, c)$ is $\mathrm{o_P}(1)$ uniformly in θ, c. Insert s_{mi} to get

$$v_{1i}(\theta, c) = w_{in}\sigma^{-1}\{(n^{-1/2}a_m)(\varepsilon_i/\sigma) + (n^{-1/2}b_m')(n^{1/2}x_{in})\}\dot{m}(\varepsilon_i/\sigma)\}\{1_{(\varepsilon_i\leq\sigma c+x_{in}'b_1)} - 1_{(\varepsilon_i\leq\sigma c)}\}, \quad (A.17)$$

where $n^{-1/2}a_m$, $n^{-1/2}b_m$ are $\mathrm{O}(n^{-1/4-\eta})$. Apply Lemma A.10 coordinate-wise to the sums involving $w_{in}(\varepsilon_i/\sigma)\dot{m}(\varepsilon_i/\sigma)$ and $w_{in}n^{1/2}x_{in}\dot{m}(\varepsilon_i/\sigma)$ to see that $\mathcal{V}_{1n}(\theta, c)$ is $\mathrm{O}(n^{-1/4-\eta})\mathrm{o_P}(n^{\kappa}) = \mathrm{o_P}(n^{-1/4})$, recalling $\kappa < \eta$. Assumptions in Lemma A.10 are met for weights and marks $w_{in}, u\dot{m}(u)$ and $w_{in}n^{1/2}x_{in}, \dot{m}(u)$ by Assumption 2.2($ia, ib2, ic, ii, iii$).

2. *The term* $\mathcal{V}_{2n}(\theta, c)$ is $\mathrm{o_P}(1)$ uniformly in θ, c. Insert s_{mi} to get

$$v_{2i}(\theta, c) = w_{in}\sigma^{-1}\{(n^{-1/2}a_m)(\varepsilon_i/\sigma) + (n^{-1/2}b_m')(n^{1/2}x_{in})\}\dot{m}(\varepsilon_i/\sigma)\}1_{(\varepsilon_i\leq\sigma c)}.$$

Apply Lemma A.9 coordinate-wise to the sums involving $w_{in}(\varepsilon_i/\sigma)\dot{m}(\varepsilon_i/\sigma)$ and $w_{in}n^{1/2}x_{in}\dot{m}(\varepsilon_i/\sigma)$ with $d = 0, 1$ to see that $\forall\omega > 0$ then $\mathcal{V}_{2n}(\theta, c)$ is $\mathrm{O}(n^{-1/4-\eta})\mathrm{o_P}(n^{1/4+\kappa+\omega})$. In particular, for $\omega < \eta - \kappa$ the product of remainder terms is $\mathrm{o_P}(1)$. The assumptions in Lemma A.9 are met for weights and marks $w_{in}, u\dot{m}(u)$ and $n^{1/2}w_{in}x_{in}, \dot{m}(u)$ by Assumption 2.2(ia, ii, iii).

3. *The term* $\mathcal{V}_{3n}(\theta, c)$ is $\mathrm{o_P}(1)$ uniformly in θ, c. We use a Lipschitz argument.

3.1. *Truncate* x_{in} using the sets $\mathcal{C}_n \subset \mathcal{C}_{in}$ outlined in (A.12). Thus, $|\mathcal{V}_{3n}(\theta, c)|$ vanishes if $|\mathcal{V}_{3n}(\theta, c)|1_{\mathcal{C}_n}$ vanishes. The triangle inequality and $\mathcal{C}_n \subset \mathcal{C}_{in}$ show
$|\mathcal{V}_{3n}(\theta, c)|1_{\mathcal{C}_n} \leq \sum_{i=1}^n\{|v_{3i}(\theta, c)1_{\mathcal{C}_{in}}| + \mathsf{E}_{i-1}|v_{3i}(\theta, c)1_{\mathcal{C}_{in}}|\}$.

3.2. *Bound* v_{3i} *on* $\mathcal{C}_{in}$: Recalling $s_{mi} = \varepsilon_i n^{-1/2}a_m/\sigma + x_{ni}'b_m$, $|n^{1/2}x_{ni}| \leq C_x n^{\kappa}$ on $\mathcal{C}_{in}$, $\kappa < \eta$, while $n^{-1/2}a_m, n^{-1/2}b_m = \mathrm{O}(n^{-1/4-\eta})$ shows

$$|s_{mi}|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/4})(1 + |\varepsilon_i/\sigma|)1_{\mathcal{C}_{in}}. \quad (A.18)$$

We note that $|\varepsilon_i^* - \varepsilon_i| \leq |s_{mi}|$. Thus, the local Lipschitz condition (id) shows $|\dot{m}(\varepsilon_i^*/\sigma) - \dot{m}(\varepsilon_i/\sigma)|1_{\mathcal{C}_{in}} \leq \sigma^{-1}|\varepsilon_i^* - \varepsilon_i|\mathring{m}(\varepsilon_i/\sigma)1_{\mathcal{C}_{in}}$, which can be bounded further by
$\mathrm{o}(n^{-1/4})(1 + |\varepsilon_i/\sigma|)\mathring{m}(\varepsilon_i/\sigma)1_{\mathcal{C}_{in}}$. Insert this in v_{3i}, apply (A.18) and $(1 + |\varepsilon|)^2 \leq 2(1 + |\varepsilon|^2)$ to get $|v_{3i}(\theta, c)|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/2})\tilde{v}_i$ where $\tilde{v}_i = |w_{in}|(1 + |\varepsilon_i/\sigma|^2)\mathring{m}(\varepsilon_i/\sigma)$. Here $\mathsf{E}_{i-1}\tilde{v}_i = |w_{in}|\mathrm{o}(1)$ by Assumption 2.2(id) and the $\mathrm{o}(1)$ term is uniform in i.

3.3. *Bound* $\mathcal{V}_{3n}$ *on* $\mathcal{C}_{in}$: Insert the $|v_{3i}|1_{\mathcal{C}_{in}}$ bound in that of $|\mathcal{V}_{3n}|1_{\mathcal{C}_n}$ to get

$$\mathcal{V}_{3n}^{\sup} = \sup_{|\theta|\leq Bn^{-\eta}}\sup_{c\in\mathbb{R}}|\mathcal{V}_{3n}(\theta, c)|1_{\mathcal{C}_n} = \mathrm{o}(n^{-1/2})n^{-1/2}\sum_{i=1}^n(\tilde{v}_i + \mathsf{E}_{i-1}\tilde{v}_i).$$

Taking expectations and using iterated expectations shows that $\mathsf{E}\mathcal{V}_{3n}^{\sup} = \mathrm{o}(n^{-1})\mathsf{E}\sum_{i=1}^n 2\mathsf{E}_{i-1}\tilde{v}_i$. The bound $\mathsf{E}_{i-1}\tilde{v}_i = |w_{in}|\mathrm{o}(1)$ from item 3.2 gives $\mathsf{E}\mathcal{V}_{3n}^{\sup} = \mathrm{o}(n^{-1})\mathsf{E}\sum_{i=1}^n|w_{in}|$, which then vanishes by Jensen's inequality and Assumption 2.2(iii). Then, the Markov inequality shows $\mathcal{V}_{3n}^{\sup}$ vanishes. □

A.7 Proof of Theorem 2.1

Part A: The Empirical Process: Define $\mathcal{V}_n(\theta, c) = \mathbb{F}_n^{w,m}(\theta, c) - \mathbb{F}_n^{w,m}(0, c)$ where $\theta = (a_1, b_1, a_m, b_m)$. It has to be shown that $\mathcal{V}_n(\theta, c)$ vanishes uniformly in θ, c. Add and subtract $\mathbb{F}_n^{w,m}(\theta_a, c)$ with $\theta_{a_1} = (a_1, 0, 0, 0)$ to decompose $\mathcal{V}_n = \mathcal{V}_{1n} + \mathcal{V}_{2n}$ where

$$\mathcal{V}_{1n}(\theta, c) = \mathbb{F}_n^{w,m}(\theta, c) - \mathbb{F}_n^{w,m}(\theta_{a_1}, c), \quad \mathcal{V}_{2n}(\theta, c) = \mathbb{F}_n^{w,m}(\theta_{a_1}, c) - \mathbb{F}_n^{w,m}(0, c).$$

Here, $\mathcal{V}_{1n}$ is concerned with estimation error in the location b_1 as well as in the marks, a_m, b_m, while $\mathcal{V}_{2n}$ is concerned with estimation error in scale a_1.

1. *The term* $\mathcal{V}_{1n}(\theta, c)$ equals $\mathcal{V}_{1n}\{0, b_1, a_m, b_m, c(1 + n^{-1/2}a_1/\sigma)\}$ so that

$$\sup_{|a_1| \leq Bn^{1/4-\eta}} \sup_{c \in \mathbb{R}} |\mathcal{V}_{1n}(a_1, b_1, a_m, b_m, c)| = \sup_{c \in \mathbb{R}} |\mathcal{V}_{1n}(0, b_1, a_m, b_m, c)|.$$

Thus, it suffices, in this part of the proof, to let $a_1 = 0$ and consider $\theta = (0, b_1, a_m, b_m)$, so that $\mathcal{V}_{1n}(\theta, c) = n^{-1/2}\sum_{i=1}^n \{v_{1i}(\theta, c) - \mathsf{E}_{i-1} v_{1i}(\theta, c)\}$ where

$$v_{1i}(\theta, c) = w_{in}\{m(\varepsilon_i^{a,b}/\sigma)1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)} - m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)}\},$$

with $\varepsilon_i^{a,b}/\sigma = (\varepsilon_i - x_{in}' b_m)/(\sigma + n^{-1/2}a_m)$. Linearize

$$\varepsilon_i^{a,b} = \varepsilon_i(1 - n^{-1/2}\tilde{a}_m/\sigma) - x_{in}'\tilde{b}_m = \varepsilon_i - \tilde{s}_{mi}, \tag{A.19}$$

where $\tilde{a}_m = a_m/(1 + n^{-1/2}a_m/\sigma)$ and $\tilde{b}_m = b_m/(1 + n^{-1/2}a_m/\sigma)$ and $\tilde{s}_{mi} = n^{-1/2}\tilde{a}_m \varepsilon_i/\sigma - x_{in}'\tilde{b}_m$ Given the bounds to a_m, b_m there exist $n_0, \tilde{B} > 0$ so that $|\tilde{a}_m|, |\tilde{b}_m| \leq \tilde{B}n^{1/4-\eta}$ for $n \geq n_0$. It suffices to show the uniform result over this larger region. Henceforth, we work with the linearized estimation error and, for the remainder of part A, ignore the tildes so that

$$v_{1i}(\theta, c) = w_{in}\{m(\frac{\varepsilon_i - s_{mi}}{\sigma})1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)} - m(\frac{\varepsilon_i}{\sigma})1_{(\varepsilon_i \leq \sigma c)}\}.$$

Add and subtract $w_{in} m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)}$ to get $v_{1i} = \sum_{s=1}^2 v_{1si}$ and $\mathcal{V}_{1n} = \sum_{s=1}^2 \mathcal{V}_{1ns}$ where

$$v_{11i}(\theta, c) = w_{in} m(\varepsilon_i/\sigma)\{1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)} - 1_{(\varepsilon_i \leq \sigma c)}\},$$
$$v_{12i}(\theta, c) = w_{in}\{m(\frac{\varepsilon_i - s_{mi}}{\sigma}) - m(\varepsilon_i/\sigma)\}1_{(\varepsilon_i \leq \sigma c + x_{in}' b_1)}.$$

Due to the triangle inequality it suffices to show that each of $\mathcal{V}_{11n}(\theta, c)$ and $\mathcal{V}_{12n}(\theta, c)$ are $\mathsf{o_P}(1)$ uniformly in θ, c by applying Lemma A.10 with $d = 0$ and Lemma A.12, respectively. All assumptions are satisfied for weights and marks w_{in}, m by Assumption 2.2($ia, ib1, ib2, ic, id, ii, iii$).

2. *The term* $\mathcal{V}_{2n}(a_1, c)$ is $\mathsf{o_P}(1)$ uniformly in θ, c. To see this apply Lemma A.11 noting that its assumptions are satisfied for weights and marks w_{in}, m by Assumption 2.2($ia, ib1, iii$).

Part B: The Compensator: We let $\bar{\mathcal{V}}_n(\theta, c) = n^{1/2}\{\overline{\mathsf{F}}_n^{w,m}(\theta, c) - \overline{\mathsf{F}}_n^{w,m}(0, c) - \mathcal{B}_n^{w,m}(\theta, c)\}$ for a parameter $\theta = (a_m, a_1, b_m, b_1)$ and show that $\bar{\mathcal{V}}_n$ vanishes uniformly in θ, c. Use (A.19) and write

$$\overline{\mathsf{F}}_n^{w,m}(\theta, c) = n^{-1}\sum_{i=1}^n w_{in} \mathsf{E}_{i-1} m(\frac{\varepsilon_i - \tilde{s}_{mi}}{\sigma})1_{(\varepsilon_i \leq \sigma c + s_{1i})},$$

where $\tilde{s}_{mi} = n^{-1/2}\tilde{a}_m(\varepsilon_i/\sigma) + x_{in}'\tilde{b}_m$ and $s_{1i} = cn^{-1/2}a_1 + x_{in}' b_1$, where $\tilde{a}_m, \tilde{b}_m$ are defined above. The bias term $\mathcal{B}_n^{w,m}(\theta, c)$ can be expressed in terms of $s_{mi} = n^{-1/2}a_m(\varepsilon_i/\sigma) + x_{in}' b_m$ as

$$\mathcal{B}_n^{w,m}(\theta, c) = \sigma^{-1}n^{-1/2}\sum_{i=1}^n w_{in}\{s_{1i} m(c)\mathsf{f}(c) - \mathsf{E}_{i-1} s_{mi}\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)}\}.$$

Thus, we can write $\bar{\mathcal{V}}_n(\theta, c) = n^{-1/2}\sum_{i=1}^n w_{in}\bar{v}_i(\theta, c)$ where

$$\bar{v}_i(\theta, c) = w_{in}\mathsf{E}_{i-1}\{m\{(\varepsilon_i - \tilde{s}_{mi})/\sigma\}1_{(\varepsilon_i \leq \sigma c + s_{1i})} - m(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)} - s_{1i} m(c)\mathsf{f}(c) + s_{mi}\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)}\}. \tag{A.20}$$

Add and subtract the terms $\{m(\varepsilon_i/\sigma) + (\tilde{s}_{mi}/\sigma)\dot{m}(\varepsilon_i/\sigma)\}1_{(\varepsilon_i \leq \sigma c + s_{1i})}$ and $(\tilde{s}_{mi}/\sigma)\dot{m}(\varepsilon_i/\sigma)1_{(\varepsilon_i \leq \sigma c)}$ to get $\bar{v}_i = \sum_{s=1}^4 \bar{v}_{si}$ and $\bar{\mathcal{V}}_n = \sum_{s=1}^4 \bar{\mathcal{V}}_{sn}$ where

$$\bar{v}_{1i}(\theta, c) = w_{in}\mathsf{E}_{i-1}\{m(\frac{\varepsilon_i - \tilde{s}_{mi}}{\sigma}) - m(\frac{\varepsilon_i}{\sigma}) + \frac{\tilde{s}_{mi}}{\sigma}\dot{m}(\frac{\varepsilon_i}{\sigma})\}1_{(\varepsilon_i \leq \sigma c + s_{1i})},$$
$$\bar{v}_{2i}(\theta, c) = w_{in}[\mathsf{E}_{i-1} m(\frac{\varepsilon_i}{\sigma})\{1_{(\varepsilon_i \leq \sigma c + s_{1i})} - 1_{(\varepsilon_i \leq \sigma c)}\} - \frac{s_{1i}}{\sigma} m(c)\mathsf{f}(c)],$$
$$\bar{v}_{3i}(\theta, c) = w_{in}\mathsf{E}_{i-1}\frac{\tilde{s}_{mi}}{\sigma}\dot{m}(\frac{\varepsilon_i}{\sigma})\{1_{(\varepsilon_i \leq \sigma c + s_{1i})} - 1_{(\varepsilon_i \leq \sigma c)}\},$$
$$\bar{v}_{4i}(\theta, c) = w_{in}\mathsf{E}_{i-1}\frac{s_{mi} - \tilde{s}_{mi}}{\sigma}\dot{m}(\frac{\varepsilon_i}{\sigma})1_{(\varepsilon_i \leq \sigma c)}.$$

We truncate x_{in} using the sets $\mathcal{C}_n \subset \mathcal{C}_{in}$ outlined in (A.12) and using Assumption 2.2(ii). Thus, $\bar{\mathcal{V}}_{sn}(\theta, c)$ vanishes if $\bar{\mathcal{V}}_{sn}(\theta, c)1_{\mathcal{C}_n}$ vanishes. In turn, $\bar{\mathcal{V}}_n(\theta, c)$ will vanish by the triangle inequality. Since $\mathcal{C}_n \subset \mathcal{C}_{in}$ and $\mathcal{C}_{in}$ is $\mathcal{F}_{i-1}$-adapted then $|\bar{\mathcal{V}}_{sn}(\theta, c)|1_{\mathcal{C}_n} \leq \sum_{i=1}^n \mathsf{E}_{i-1}|\bar{v}_{si}(\theta, c)1_{\mathcal{C}_{in}}|$. We show that each summand satisfies, for $s = 1, \ldots, 4$ and uniformly in θ, c,

$$|\bar{v}_{si}(\theta, c)|1_{\mathcal{C}_{in}} = \mathrm{O}(n^{-1/2})|w_{in}|1_{\mathcal{C}_{in}}. \tag{A.21}$$

Then, Assumption 2.2(iii) and the Markov inequality gives $|\bar{\mathcal{V}}_{sn}|1_{\mathcal{C}_n} = \mathrm{o}_\mathsf{P}(1)$.
1. $\bar{\mathcal{V}}_{1n}(\theta, c)$. Consider $\breve{m}_{1i} = m\{(\varepsilon_i - \tilde{s}_{mi})/\sigma\} - m(\varepsilon_i/\sigma) - \sigma^{-1}\tilde{s}_{mi}\dot{m}(\varepsilon_i/\sigma)$. We get $|\tilde{s}_{mi}|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/4})(1 + |\varepsilon_i/\sigma|)1_{\mathcal{C}_{in}}$ as in (A.18). The mean value theorem gives, for an intermediate point ε_i^* so $|\varepsilon_i^* - \varepsilon_i| \leq |\tilde{s}_{mi}|$ the expansion $\breve{m}_{1i} = -\sigma^{-1}\tilde{s}_{mi}\{\dot{m}(\varepsilon_i^*/\sigma) - \dot{m}(\varepsilon_i/\sigma)\}$. Thus, $|\breve{m}_{1i}|1_{\mathcal{C}_{in}} \leq \sigma^{-1}\tilde{s}_{mi}^2\mathring{m}(\varepsilon_i)1_{\mathcal{C}_{in}}$ by the local Lipschitz condition in Assumption 2.2(id). Further, $|\breve{m}_{1i}|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/2})(1 + |\varepsilon_i/\sigma|^2)\mathring{m}(\varepsilon_i)1_{\mathcal{C}_{in}}$ by the bound to $|\tilde{s}_{mi}|1_{\mathcal{C}_{in}}$. The integrability of $\mathring{m}$ then shows $\mathsf{E}_{i-1}|\breve{m}_{1i}|1_{\mathcal{C}_{in}} = \mathrm{O}(n^{-1/2-2\eta})1_{\mathcal{C}_{in}}$, uniformly in θ, c. In turn, $|\bar{v}_{1i}(\theta, c)|1_{\mathcal{C}_{in}}$ has the form (A.21).
2. $\bar{\mathcal{V}}_{2n}(\theta, c)$. Write $\bar{v}_{2i}(\theta, c) = w_{in}\{\int_c^{c+s_{1i}/\sigma} m(u)\mathsf{f}(u)du - \frac{s_{1i}}{\sigma}m(c)\mathsf{f}(c)\}$. Taylor expand $K(s) = \int_c^{c+s} k(u)du$ as $K(s) = sk(c) + 2^{-1}\dot{k}(c^*)$ for some c^* so $|c^* - c| \leq s$.
Hence, $\bar{v}_{2i}(\theta, c) = (w_{in}/2)(s_{1i}/\sigma)^2\{\dot{m}(c^*)\mathsf{f}(c^*) + m(c^*)\dot{\mathsf{f}}(c^*)\}$ for c^* so that $|c^* - c| \leq |s_{1i}|$ where $s_{1i} = n^{-1/2}a_1c + x_{in}'b_1$. Since $|a_1| \leq Bn^{1/4-\eta}$ then, for large n, we can use the second inequality in Lemma A.4 with $A = n^{-1/2}a_1$, $B = x_{in}'b_1$ and $\tilde{c} = c^*$ to get

$$|\bar{v}_{2i}(\theta, c)| \leq |w_{in}|8\sigma^{-2}[(c^*)^2(n^{-1/2}a_1)^2 + (x_{in}'b_1)^2]|\dot{m}(c^*)\mathsf{f}(c^*) + m(c^*)\dot{\mathsf{f}}(c^*)|.$$

Recall $|a_1|, |b_1| \leq Bn^{1/4-\eta}$ while $n^{1/2}x_{ni} = \mathrm{O}(n^\kappa) = \mathrm{o}(n^\eta)$ on $\mathcal{C}_{in}$ to get

$$|\bar{v}_{2i}(\theta, c)|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/2})|w_{in}|\sup_{u\in\mathbb{R}}(1 + u^2)|\dot{m}(u)\mathsf{f}(u) + m(u)\dot{\mathsf{f}}(u)|1_{\mathcal{C}_{in}}.$$

By Assumption 2.2$(ib3)$ we get $|\bar{v}_{2i}(\theta, c)|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/2})|w_{in}|1_{\mathcal{C}_{in}}$, which is of the form (A.21).
3. $\bar{\mathcal{V}}_{3n}(\theta, c)$. Note $|\tilde{s}_{mi}|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/4})(1 + |\varepsilon_i/\sigma|)1_{\mathcal{C}_{in}}$ as in (A.18) and let $\mathcal{H} = \mathsf{E}_{i-1}(1 + |\varepsilon_i/\sigma|)|\dot{m}(\varepsilon_i/\sigma)||1_{(\varepsilon_i \leq \sigma c + s_{1i})} - 1_{(\varepsilon_i \leq \sigma c)}|$ so that $|\bar{v}_{3i}(\theta, c)|1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/4})|w_{in}|\mathcal{H}1_{\mathcal{C}_{in}}$. Apply Lemma A.5 using Assumption 2.2 $(ia, ib3)$ to get $\mathcal{H} = \mathrm{O}(n^{-1/4-\eta})(1 + |n^{1/2}x_{in}|)$. The truncation of x_{ni} gives $\mathcal{H}1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/4})1_{\mathcal{C}_{ni}}$. Thus $\bar{v}_{3i}(\theta, c)$ satisfies (A.21).
4. $\bar{\mathcal{V}}_{4n}(\theta, c)$. Expand $s_{mi} - \tilde{s}_{mi} = n^{-1}\{\varepsilon_i(a_m^\sigma)^2 + n^{1/2}x_{in}'b_ma_m^\sigma\}/(1 + n^{-1/2}a_m^\sigma)$ where $a_m^\sigma = a_m/\sigma$. Use the bounds $a_m, b_m, a_m^\sigma = \mathrm{O}(n^{1/4-\eta})$ to get that $s_{mi} - \tilde{s}_{mi} = \mathrm{O}(n^{-1/2-2\eta})(1 + |n^{-1/2}x_{in}|)(1 + |\varepsilon_i/\sigma|)$. The x_{ni} truncation gives $(s_{mi} - \tilde{s}_{mi})1_{\mathcal{C}_{in}} = \mathrm{o}(n^{-1/2})(1 + |\varepsilon_i/\sigma|)1_{\mathcal{C}_{in}}$. Since $\mathsf{E}_{i-1}(1 + |\varepsilon_i/\sigma|)|\dot{m}(\varepsilon_i/\sigma)| < \infty$ by Assumption 2.2(ia) then $\bar{v}_{4i}(\theta, c)1_{\mathcal{C}_{in}}$ satisfies (A.21). □

B Proof of robust normality results

B.1 Two sided empirical processes

Introduce

$$\mathsf{G}_n^{w,m}(\theta,c) = \frac{1}{n}\sum_{i=1}^n m\left(\frac{\varepsilon_i - x_{in}'b_m}{\sigma + n^{-1/2}a_m}\right)1_{(|\varepsilon_i - x_{in}'b| \leq \sigma c + n^{-1/2}a_1c)}, \tag{B.1}$$

so that $\mathsf{G}_n^{w,m}(\theta,c) = \mathsf{F}_n^{w,m}(\theta,c) - \lim_{h\downarrow 0}\mathsf{F}_n^{w,m}(\theta, -c-h)$. The corresponding compensator $\bar{\mathsf{G}}_n^{w,m}$, bias term $\mathcal{G}_n^{w,m} = \mathcal{G}_{1n}^{w,m} - \mathcal{G}_{mn}^{w,m}$ and empirical process $\mathbb{G}_n^{w,m}$ are defined similarly. The following three results are immediate consequences of Theorems 2.1, 2.2.

Corollary B.1. *Suppose Assumptions 2.1, 2.2 hold. Then,* $\forall B > 0$,

$$\sup_{|\theta| \leq Bn^{1/4-\eta}}\sup_{c>0}|\mathbb{G}_n^{w,m}(\theta,c) - \mathbb{G}_n^{w,m}(0,c)| = \mathrm{o}_\mathsf{P}(1),$$

$$\sup_{|\theta| \leq n^{1/4-\eta}B}\sup_{c>0}|n^{1/2}\{\bar{\mathsf{G}}_n^{w,m}(\theta,c) - \bar{\mathsf{G}}_n^{w,m}(0,c)\} - \mathcal{G}_n^{w,m}(\theta,c)| = \mathrm{o}_\mathsf{P}(1).$$

Corollary B.2. *Suppose Assumptions 2.1, 2.3 hold. Then,* $\forall \epsilon > 0$,

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

B.2 Preliminary Lemmas

Lemma B.3. *If $w_{in} = 1$, $m(c) = c^k$ and f is symmetric, then the two sided bias term of Corollary B.1 is $\mathcal{G}_n^k(\theta, c) = \mathcal{G}_{1n}^k(\theta_1, c) - \mathcal{G}_{mn}^k(\theta_m, c)$, where*

$$\mathcal{G}_{1n}^k(\theta_1, c) = 1_{(k\ even)} 2c^{k+1}\mathsf{f}(c)\sigma^{-1}a_1 + 1_{(k\ odd)}2c^k\mathsf{f}(c)\sigma^{-1}n^{-1/2}\textstyle\sum_{i=1}^n x'_{in}b_1,$$
$$\mathcal{G}_{mn}^k(\theta_m, c) = 1_{(k\ even)}k\tau_k^c\sigma^{-1}a_m + 1_{(k\ odd)}k\tau_{k-1}^c\sigma^{-1}n^{-1/2}\textstyle\sum_{i=1}^n x'_{in}b_m.$$

Proof of Lemma B.3. By definition $\mathcal{G}_{jn}^k(\cdot, c) = \mathcal{B}_{jn}^{1,k}(\cdot, c) - \lim_{\tilde{c}\downarrow c}\mathcal{B}_{jn}^{1,k}(\cdot, -\tilde{c})$ for $j = 1, m$. Set $w_{in} = 1$, $m(c) = c^k$, note f is symmetric and compute $\mathcal{B}_{jn}^{1,k}(\theta_j, c)$ from the general formulas (2.5), (2.6) to get the result. □

Lemma B.4. *Let Assumption 2.2 hold with $w_{in} = 1$ and $m(u) = u^k$ and symmetric density f. Let $\theta_1 = (a_1, b_1)$, $\theta_m = (a_m, b_m)$ be estimation errors for indicator and mark. Let $c^d = c + n^{-1/2}d_1$ be an additively shifted quantile and let $\theta_1^d = (\sigma d_1 c^{-1}, 0)$, while $\theta = (\theta_1, \theta_m)$ and $\theta^d = (\theta_1 + \theta_1^d, \theta_m)$. Then, $\forall B, \eta > 0$ and uniformly in c and $|\theta|, |d_1| \in n^{1/4-\eta}B$ we get:*
(a) $\mathcal{G}_{jn}^k(\theta_j, c^d) = \mathcal{G}_{jn}^k(\theta_j, c) + \mathrm{o_P}(1)$ for $j = 1, m$;
(b) $n^{1/2}\{\bar{\mathsf{G}}_n^k(0,c^d) - \bar{\mathsf{G}}_n^k(0,c)\} = \mathcal{G}_{1n}^k(\theta_1^d, c) + \mathrm{o_P}(1)$
where $\mathcal{G}_{1n}^k(\theta_1^d, c) = 1_{(k\ even)}2c^k\mathsf{f}(c)\sigma^{-1}d_1$;
(c) $n^{1/2}\{\mathsf{G}_n^k(\theta,c^d) - \bar{\mathsf{G}}_n^k(0, c)\} = n^{1/2}\{\mathsf{G}_n^k(0, c) - \bar{\mathsf{G}}_n^k(0, c)\} + \mathcal{G}_n^k(\theta^d, c) + \mathrm{o_P}(1)$;
(d) $\mathsf{G}_n^k(\theta,c^d) = \bar{\mathsf{G}}_n^k(0, c) + \mathrm{o_P}(1)$.

Proof of Lemma B.4. (a) Evaluate the biases in Lemma B.3 at c^d to get

$$\begin{aligned}
\mathcal{G}_{1n}^k(\theta_1, c^d) &= 1_{(k\text{ even})}2(c^d)^{k+1}\mathsf{f}(c^d)\sigma^{-1}a_1 \\
&\quad +1_{(k\text{ odd})}2(c^d)^k\mathsf{f}(c^d)\sigma^{-1}n^{-1/2}\textstyle\sum_{i=1}^n x'_{in}b_1, \\
\mathcal{G}_{mn}^k(\theta_m, c^d) &= 1_{(k\text{ even})}k\tau_k^{c^d}\sigma^{-1}a_h + 1_{(k\text{ odd})}k\tau_{k-1}^{c^d}\sigma^{-1}n^{-1/2}\textstyle\sum_{i=1}^n x'_{in}b_h.
\end{aligned}$$

For $\mathcal{G}_{1n}^k$ we note that $v_p(c) = c^p\mathsf{f}(c)$ for $p = k, k+1$ has bounded derivatives by Assumption 2.2$(ib3)$. Then the mean value theorem shows $v_p(c^d) = v_p(c) + (c^d - c)\dot{v}_p(c^*)$ for an intermediate point c^* so that $|c^* - c| \leq |c^d - c|$. Since $c^d - c = n^{-1/2}d_1 = \mathrm{O}(n^{-1/4-\eta})$ we get $v_p(c^d) = v_p(c) + \mathrm{O}(n^{-1/4-\eta})$. Since $\theta_1 = \mathrm{O}(n^{1/4-\eta})$ and $n^{-1/2}\sum_{i=1}^n |x_{in}| \leq n^{1/2}\max_{1\leq i\leq n}|x_{in}| = \mathrm{O_P}(n^\kappa)$ with $\kappa < \eta$ by Assumption 2.2(ii) then $n^{-1/2}\sum_{i=1}^n x'_{in}b_1 = \mathrm{o_P}(n^{1/4})$ so that $\mathcal{G}_{1n}^k(\theta_1, c^d) = \mathcal{G}_{1n}^k(\theta_1, c) + \mathrm{o_P}(n^{-\eta})$.

For $\mathcal{G}_{mn}^k$ the only difference in the argument is that we replace the function $v(c)$ by $w(c) = \tau_p^c$ defined in (3.4) for $p = k-1, k$. This function has derivate $\dot{w}(c) = c^p\mathsf{f}(c)$ which is also bounded by assumption 2.2$(ib3)$. Hence, $\mathcal{G}_{mn}^k(\theta_m, c^d) = \mathcal{G}_{mn}^k(\theta_m, c) + \mathrm{O_P}(n^{-2\eta})$.
(b) The term of interest is $\bar{\mathcal{S}}\sum_{i=1}^n \mathsf{E}_{i-1}(\varepsilon_i/\sigma)^k\{1_{(|\varepsilon_i|\leq\sigma c^d)} - 1_{(|\varepsilon_i|\leq\sigma c)}\}$. Write $c^d = c + n^{-1/2}d_1$ and let $\tilde{x}_{in} = n^{-1/2}$ and $b^d = \sigma d_1$ so that $1_{(|\varepsilon_i|\leq\sigma c^d)} = 1_{(|\varepsilon_i|\leq\sigma c+\tilde{x}_{in}b^d)}$. Thus
$\bar{\mathcal{S}} = n^{-1/2}\sum_{i=1}^n \mathsf{E}_{i-1}(\varepsilon_i/\sigma)^k\{1_{(|\varepsilon_i|\leq\sigma c+\tilde{x}_{in}b^d)} - 1_{(|\varepsilon_i|\leq\sigma c)}\}$, so that with $\tilde{\theta} = (0, 0, 0, b^d)$ we can write

$$\bar{\mathcal{S}} = n^{1/2}\{\bar{\mathsf{F}}_n^k(\tilde{\theta}, c) - \bar{\mathsf{F}}_n^k(0, c)\} - n^{1/2}\{\lim_{c\downarrow c}\bar{\mathsf{F}}_n^k(-\tilde{\theta}, -c) - \lim_{c\downarrow c}\bar{\mathsf{F}}_n^k(0, -c)\}$$

Now we apply Theorem 2.1 to each term on the right hand side. For both terms the regressors are $\tilde{x}_{in} = n^{-1/2}$ while θ, c are $\tilde{\theta}, c$ and $-\tilde{\theta}, -c$, respectively, and where $b^d = \sigma d_1$ is of order $n^{1/4-\eta}$. Hence, the assumptions of Theorem 2.1 are met in this situation by assumption 2.2, so that

$$\bar{\mathcal{S}} = \{\mathcal{B}_n^{1,k}(\tilde{\theta}, c) + \mathrm{o_P}(1)\} - \{\mathcal{B}_n^{1,k}(-\tilde{\theta}, -c) + \mathrm{o_P}(1)\}.$$

Recall the expression $\mathcal{B}_n^{1,k}$ in (2.5) and note that $\tilde{x}_{in} = n^{-1/2}$ so that $n^{-1/2}\sum_{i=1}^n \tilde{x}_{in}b^d = b^d$ to get $\bar{\mathcal{S}} = \sigma^{-1}c^k\mathsf{f}(c)b^d - \sigma^{-1}(-c)^k\mathsf{f}(-c)(-b^d) + \mathrm{o_P}(1)$. Given the symmetry of the density f and $b^d = \sigma d_1$ while $\theta_1^d = (\sigma d_1 c^{-1}, 0)$ we get as desired $\bar{\mathcal{S}} = 1_{(k\text{ even})}2\sigma^{-1}c^{k+1}\mathsf{f}(c)\sigma d_1 c^{-1} + \mathrm{o_P}(1) = \mathcal{G}_{1n}^k(\theta_1^d, c) + \mathrm{o_P}(1)$.

(c) The bias $\mathcal{G}$ in Lemma B.3 shows that for $\theta^d = (\theta_1 + \theta_1^d, \theta_m)$ we have $\mathcal{G}_n^k(\theta^d, c) = \mathcal{G}_n^k(\theta, c) + \mathcal{G}_{1n}^k(\theta_1^d, c)$. Thus, we can write

$$n^{1/2}\{\mathsf{G}_n^k(\theta,c^d) - \bar{\mathsf{G}}_n^k(0,c)\} = n^{1/2}\{\mathsf{G}_n^k(0,c) - \bar{\mathsf{G}}_n^k(0,c)\} + \mathcal{G}_n^k(\theta^d, c) + \textstyle\sum_{j=1}^5 \mathcal{R}_j,$$

with remainder terms

$$\begin{aligned}
\mathcal{R}_1 &= n^{1/2}\{\mathsf{G}_n^k(\theta,c^d) - \bar{\mathsf{G}}_n^k(\theta,c^d) - \mathsf{G}_n^k(0,c^d) + \bar{\mathsf{G}}_n^k(0,c^d)\},\\
\mathcal{R}_2 &= n^{1/2}\{\mathsf{G}_n^k(0,c^d) - \bar{\mathsf{G}}_n^k(0,c^d) - \mathsf{G}_n^k(0,c) + \bar{\mathsf{G}}_n^k(0,c)\},\\
\mathcal{R}_3 &= n^{1/2}\{\bar{\mathsf{G}}_n^k(\theta,c^d) - \bar{\mathsf{G}}_n^k(0,c^d)\} - \mathcal{G}_n^k(\theta, c^d),\\
\mathcal{R}_4 &= n^{1/2}\{\bar{\mathsf{G}}_n^k(0,c^d) - \bar{\mathsf{G}}_n^k(0,c)\} - \mathcal{G}_{1n}^k(\theta_1^d, c),\\
\mathcal{R}_5 &= \mathcal{G}_n^k(\theta, c^d) - \mathcal{G}_n^k(\theta, c).
\end{aligned}$$

We get $\mathcal{R}_1, \mathcal{R}_3 = \mathrm{o}_{\mathsf{P}}(1)$ by Corollary B.1, $\mathcal{R}_2 = \mathrm{o}_{\mathsf{P}}(1)$ by Corollary B.2, $\mathcal{R}_4 = \mathrm{o}_{\mathsf{P}}(1)$ by part (b) and $\mathcal{R}_5 = \mathrm{o}_{\mathsf{P}}(1)$ by part (a) noting $\mathcal{G}_n^k = \mathcal{G}_{1n}^k - \mathcal{G}_{mn}^k$.
(d) Apply part (c) multiplied by $n^{-1/2}$. The first term $\mathsf{G}_n^k(0,c) - \bar{\mathsf{G}}_n^k(0,c)$ vanishes uniformly in c since the finite dimensional distributions vanish by the Law of Large Numbers and the process is tight by Corollary B.2. Finally, for the second term, note that $\mathcal{G}_n^k(\theta^d, c) = \mathcal{G}_{1n}^k(\theta_1, c) + \mathcal{G}_{mn}^k(\theta_m, c) + \mathcal{G}_{1n}^k(\theta_1^d, c)$. Recall the expressions of these terms in Lemma B.3 and part (c). Apply the triangle inequality to get

$$\begin{aligned}
|\mathcal{G}_{1n}^k(\theta_1, c)| &\le 1_{(k \text{ even})} 2|c^{k+1}|\mathsf{f}(c)\sigma^{-1}|a_1|\\
&\quad + 1_{(k \text{ odd})} 2|c^k|\mathsf{f}(c)\sigma^{-1} n^{-1/2} \textstyle\sum_{i=1}^n |x_{in}||b_1|,\\
|\mathcal{G}_{mn}^k(\theta_m, c)| &\le 1_{(k \text{ even})} k\tau_k^c \sigma^{-1}|a_m| + 1_{(k \text{ odd})} k\tau_{k-1}^c \sigma^{-1} n^{-1/2} \textstyle\sum_{i=1}^n |x_{in}||b_m|,\\
|\mathcal{G}_{1n}^k(\theta_1^d, c)| &\le 1_{(k \text{ even})} 2|c^k|\mathsf{f}(c)|d_1|
\end{aligned}$$

Note that $|c^{k+1}|\mathsf{f}(c)$ and $|c^k|\mathsf{f}(c)$ are bounded by Assumption 2.2$(ib3)$, the estimation errors, a_1, a_m, d_1 are assumed $\mathrm{O}(n^{1/4-\eta})$ while $n^{-1/2}\sum_{i=1}^n x_{in}' b_j = \mathrm{o}_{\mathsf{P}}(n^{1/4})$ for $j = 1, m$ as in part (a). Hence, $\mathcal{G}_n^k(\theta^d, c) = \mathrm{O}(n^{1/4})$ uniformly in c, θ, d_1 so that $n^{-1/2}\mathcal{G}_n^k(\theta^d, c)$ vanishes. □

B.3 Preliminary Results on Estimators

The estimators we consider have an expansion with a leading term that is of least squares form. For such estimators we can exploit the following result for the sum of predictors.

Lemma B.5. *Let $x_i = (1, z_i')'$ while $(m_i)_{i\in\mathbb{N}}$ is a random sequence, and*

$$N^{-1}(\hat{\beta} - \beta) = (N' \textstyle\sum_{i=1}^n x_i x_i' N)^{-1} N' \sum_{i=1}^n x_i m_i + \mathrm{o}_{\mathsf{P}}(1). \tag{B.2}$$

Assume $\sum_{i=1}^n N' x_i = \mathrm{O}_{\mathsf{P}}(n^{1/2})$ and $\sum_{i=1}^n x_i x_i'$ is invertible. Then,

$$\textstyle\sum_{i=1}^n x_i'(\hat{\beta} - \beta) = \sum_{i=1}^n m_i + \mathrm{o}_{\mathsf{P}}(n^{-1/2}).$$

Proof of Lemma B.5. The sum of predictors satisfy $\sum_{i=1}^n x_i'(\hat{\beta} - \beta) = \sum_{i=1}^n x_i' N N^{-1}(\hat{\beta} - \beta)$. Given expansion (B.2) we can write

$$\textstyle\sum_{i=1}^n x_i'(\hat{\beta} - \beta) = \sum_{i=1}^n x_i' N (N' \sum_{i=1}^n x_i x_i' N)^{-1} N' \sum_{i=1}^n x_i m_i + \mathrm{o}_{\mathsf{P}}(1) \sum_{i=1}^n x_i' N.$$

The normalizations cancel in the first term. Thus, we can normalize by any invertible matrix A. For the second term note $\sum_{i=1}^n x_i' N = \mathrm{O}_{\mathsf{P}}(n^{1/2})$ by assumption. Hence, we can write

$$\textstyle\sum_{i=1}^n x_i'(\hat{\beta} - \beta) = \sum_{i=1}^n x_i' A (A' \sum_{i=1}^n x_i x_i' A)^{-1} A' \sum_{i=1}^n x_i m_i + \mathrm{o}_{\mathsf{P}}(n^{1/2}).$$

Noting $x_i = (1, z_i')'$ define the sample average $\bar{z} = n^{-1}\sum_{i=1}^n z_i$ and choose

$$A' = \begin{pmatrix} 1 & -\bar{z}' \\ 0 & I_{\dim z} \end{pmatrix} \qquad \text{so that} \qquad A'x_i = \begin{pmatrix} 1 \\ z_i - \bar{z} \end{pmatrix}.$$

Since $\sum_{i=1}^n (z_i - \bar{z}) = 0$ we get

$$\sum_{i=1}^n x_i' A (A' \sum_{i=1}^n x_i x_i' A)^{-1} A' \sum_{i=1}^n x_i m_i$$
$$= \begin{pmatrix} n \\ 0 \end{pmatrix}' \left\{ \begin{matrix} n & 0 \\ 0 & \sum_{i=1}^n (z_i - \bar{z})(z_i - \bar{z})' \end{matrix} \right\}^{-1} \sum_{i=1}^n \begin{pmatrix} 1 \\ z_i - \bar{z} \end{pmatrix} m_i,$$

which equals $\sum_{i=1}^n m_i$. Insert in the above expression for $\sum_{i=1}^n x_i'(\hat{\beta} - \beta)$. □

Lemma B.6. *Let $\hat{\beta}$, $\hat{\sigma}$ be full sample least squares estimators of β, σ. Suppose Assumptions 2.1 4.1(ii) and $\mathsf{E}(\varepsilon_i|\mathcal{F}_{i-1}) = 0$, $\sup_n \mathsf{E}(\varepsilon_n^2|\mathcal{F}_{n-1}) < \infty$ a.s. Then*

$$N^{-1}(\hat{\beta} - \beta) = (N'\sum_{i=1}^n x_i x_i' N)^{-1} N \sum_{i=1}^n x_i \varepsilon_i, \tag{B.3}$$
$$n^{1/2}(\hat{\sigma} - \sigma) = (\sigma/2) n^{-1/2} \sum_{i=1}^n \{(\varepsilon_i/\sigma)^2 - 1\} + \mathrm{o}_{\mathsf{P}}(1). \tag{B.4}$$

Proof of Lemma B.6. (B.3) follows by the definition of the least squares estimator and linearity of the model. For (B.4) note that

$$n^{1/2}(\hat{\sigma}^2 - \sigma^2) = n^{-1/2} \sum_{i=1}^n (\hat{\varepsilon}_i^2 - \sigma^2)$$
$$= n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) - n^{-1/2} \sum_{i=1}^n \varepsilon_i x_i' (\sum_{i=1}^n x_i x_i')^{-1} \sum_{i=1}^n x_i \varepsilon_i. \tag{B.5}$$

The second term is of order $\mathrm{o}\{n^{-1/2}(\log \lambda_{\max})^2\}$ *a.s.*, see [16, Lemma 1]. This vanishes since $\log \lambda_{\max} = \mathrm{O}_{\mathsf{P}}(\log n)$ by assumption. Further, write

$$\hat{\sigma} - \sigma = (\sigma^2 + \hat{\sigma}^2 - \sigma^2)^{1/2} - \sigma = \sigma\{(1 + \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2})^{1/2} - 1\}.$$

Expand $(1 + x)^{1/2} = 1 + x/2 + (1 + x^*)^{-3/2} x^2/8$ for some x^* so $|x^*| \leq |x|$. For small x then $(1 + x)^{1/2} = 1 + x/2 + \mathrm{O}(x^2)$. Insert $x = (\hat{\sigma}^2 - \sigma^2)/\sigma^2$. □

Lemma B.7 (Jiao, Nielsen [8])**.** *Consider the robustified least squares estimator. Suppose Assumption 2.2 holds. Then, uniformly in $c \in [c_0, \infty)$,*

$$N^{-1}(\hat{\beta}^{RLS} - \beta) = (\tau_0^c N' \sum_{i=1}^n x_i x_i' N)^{-1} N' \sum_{i=1}^n x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)}$$
$$+ \{2c\mathsf{f}(c)/\tau_0^c\} N^{-1}(\hat{\beta}^{(0)} - \beta) + \mathrm{o}_{\mathsf{P}}(1), \tag{B.6}$$

$$n^{1/2}(\hat{\sigma}^{RLS} - \sigma) = \{\sigma/(2\tau_2^c)\} n^{-1/2} \sum_{i=1}^n \{(\varepsilon_i/\sigma)^2 1_{(|\varepsilon_i| \leq \sigma c)} - \tau_2^c\}$$
$$- \{\sigma/(2\tau_0^c)\} n^{-1/2} \sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}$$
$$+ \{2c(c^2 - \tau_2^c/\tau_0^c)\mathsf{f}(c)\}/(2\tau_2^c) n^{1/2}(\hat{\sigma}^{(0)} - \sigma) + \mathrm{o}_{\mathsf{P}}(1), \tag{B.7}$$

where the initial estimators $\hat{\beta}^{(0)}, \hat{\sigma}^{(0)}$ have expansions given in Lemma B.6.

The least trimmed squares estimator has been analyzed by [21, 12].

Lemma B.8. *Let $\hat{\beta}^{LTS}$, $\hat{\sigma}^{LTS}$, $\hat{c}$ be the LTS, the 1-step variance and the quantile estimators, respectively. Suppose Assumptions 2.2, 4.2 hold. Then*

$$N^{-1}(\hat{\beta}^{LTS} - \beta) = \{\tau_0^c - 2c\mathsf{f}(c)\}^{-1} (N' \sum_{i=1}^n x_i x_i' N)^{-1} N' \sum_{i=1}^n x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + \mathrm{o}_{\mathsf{P}}(1), \tag{B.8}$$
$$n^{1/2}(\hat{c} - c) = -\{2\mathsf{f}(c)\}^{-1} n^{-1/2} \sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + \mathrm{o}_{\mathsf{P}}(1), \tag{B.9}$$

$$n^{1/2}(\hat{\sigma}^{LTS} - \sigma) = (\sigma/2\tau_2^c) n^{-1/2}$$
$$\times \sum_{i=1}^n [\{(\varepsilon_i/\sigma)^2 1_{(|\varepsilon_i| \leq \sigma c)} - \tau_2^c\} - c^2\{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}] + \mathrm{o}_{\mathsf{P}}(1). \tag{B.10}$$

Proof of Lemma B.8: See [12, Theorems 4, 5] □

B.4 Proof of results for robustified least squares estimators

We represent the truncated moments (4.2) in terms of the two sided processes introduced in §B.1. In this section the superscript ^{RLS} is ignored. Let $\hat\theta_1 = (\tilde a, \tilde b)$ where $\tilde a = n^{1/2}(\tilde\sigma - \sigma)$ and $\tilde b = N^{-1}(\tilde\beta - \beta)$ are the full sample least squares estimation errors. Let also $\hat\theta_m = (\hat a, \hat b)$ where $\hat a = n^{1/2}(\hat\sigma - \sigma)$, $\hat b = N^{-1}(\hat\beta - \beta)$ are the least squares estimation errors for the selected sub-sample. In combination we get $\hat\theta = (\hat\theta_1, \hat\theta_m)$. These errors were analyzed in Lemmas B.6, B.7. Then,

$$n^{1/2}\hat\mu^{RLS}_{k,c} = n^{1/2}\mathsf{G}^k_n(\hat\theta, c)/\mathsf{G}^0_n(\hat\theta, c). \tag{B.11}$$

We will expand the third and fourth moment test statistics in terms of the vectors $z^c_{3,i}$ and $z^c_{4,i}$ given in (4.4), which are asymptotically independent.

Lemma B.9. *Let Assumption 4.1 hold. Recall $\zeta_{3,c}$ and $\zeta_{4,c}$ defined in (4.7). Then, uniformly in $c \geq c_0$ for some $c_0 > 0$, we get the expansions*
(a) $\mathsf{G}^0_n(\hat\theta, c) = \tau^c_0 + \mathrm{o_P}(1)$*;*
(b) $n^{1/2}\mathsf{G}^3_n(\hat\theta, c) = \zeta'_{3,c} n^{-1/2}\sum_{i=1}^n z^c_{3,i} + \mathrm{o_P}(1)$*;*
(c) $n^{1/2}\{\mathsf{G}^4_n(\hat\theta, c) - (\tau^c_4/\tau^c_0)\mathsf{G}^0_n(\hat\theta, c)\} = \zeta'_{4,c} n^{-1/2}\sum_{i=1}^n z^c_{4,i} + \mathrm{o_P}(1)$.

Proof of Lemma B.9. (a) Apply Lemmas A.1, B.4(d) with $d_1 = 0$ noting that $\bar{\mathsf{G}}^0_n(0, c) = \tau^c_0$.
(b) Let $N_{3,c} = \mathsf{G}^3_n(\hat\theta, c) - \bar{\mathsf{G}}^3_n(0, c)$ noting that $\bar{\mathsf{G}}^3_n(0, c) = \mathsf{E}\varepsilon^3 1_{(|\varepsilon_i| \leq \sigma c)} = 0$. Due to Lemmas A.1 and B.4(c) with $d_1 = 0$,

$$n^{1/2}N_{3,c} = n^{1/2}\{\mathsf{G}^3_n(0, c) - \bar{\mathsf{G}}^3_n(0, c)\} + \mathcal{G}^3_n(\hat\theta, c) + \mathrm{o_P}(1). \tag{B.12}$$

Lemma B.3 shows that the bias term is

$$\mathcal{G}^3_n(\hat\theta, c) = 2c^3\mathsf{f}(c)\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_{in}\tilde b - 3\tau^c_2\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_{in}\hat b.$$

Hence, given expansions (B.3), (B.6) for $\tilde b$, $\hat b$, respectively, Lemma B.5 shows

$$\begin{aligned}\textstyle\sum_{i=1}^n x'_{in}\tilde b &= \textstyle\sum_{i=1}^n \varepsilon_i = \sigma(0,0,1)n^{-1/2}\sum_{i=1}^n z^c_{3,i},\\ \textstyle\sum_{i=1}^n x'_{in}\hat b &= \textstyle(1/\tau^c_0)\sum_{i=1}^n \varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)} + \{2c\mathsf{f}(c)/\tau^c_0\}\sum_{i=1}^n \varepsilon_i + \mathrm{o_P}(1)\\ &= \textstyle\sigma\{0, 1/\tau^c_0, 2c\mathsf{f}(c)/\tau^c_0\}n^{-1/2}\sum_{i=1}^n z^c_{3,i},\end{aligned}$$

so that

$$\mathcal{G}^3_n(\hat\theta, c) = [2c^3\mathsf{f}(c)(0,0,1) - 3\tau^c_2\{0, 1/\tau^c_0, 2c\mathsf{f}(c)/\tau^c_0\}]n^{-1/2}\sum_{i=1}^n z^c_{3,i} + \mathrm{o_P}(1).$$

Insert this expression in (B.12) along with $\mathsf{G}^3_n(0, c) = (1,0,0)n^{-1/2}\sum_{i=1}^n z^c_{3,i}$ and $\bar{\mathsf{G}}^3_n(0, c) = 0$ to get $n^{1/2}N_{3,c} = \zeta'_{3,c}n^{-1/2}\sum_{i=1}^n z^c_{3,i} + \mathrm{o_P}(1)$, where $\zeta_{3,c} = \{1, -3\tau^c_2/\tau^c_0, 2(c^2 - 3\tau^c_2/\tau^c_0)c\mathsf{f}(c)\}'$ as required in (4.7).
(c) Let $N_{4,c} = n^{1/2}\{\mathsf{G}^4_n(\hat\theta, c) - (\tau^c_4/\tau^c_0)\mathsf{G}^0_n(\hat\theta, c)\}$. Due to Lemmas A.1, B.4$(c)$ with $d_1 = 0$, uniformly in c we get, for $p = 0, 4$,

$$n^{1/2}\{\mathsf{G}^p_n(\hat\theta, c) - \bar{\mathsf{G}}^p_n(0, c)\} = n^{1/2}\{\mathsf{G}^p_n(0, c) - \bar{\mathsf{G}}^p_n(0, c)\} + \mathcal{G}^p_n(\hat\theta, c) + \mathrm{o_P}(1),$$

with compensators $\bar{\mathsf{G}}^p_n(0, c) = \mathsf{E}\varepsilon^p 1_{(|\varepsilon_i|\leq\sigma c)} = \tau^c_p$. We note the relation

$$\bar{\mathsf{G}}^4_n(0, c) - (\tau^c_4/\tau^c_0)\bar{\mathsf{G}}^0_n(0, c) = \tau^c_4 - \tau^c_0\tau^c_4/\tau^c_0 = 0.$$

Therefore we can write

$$n^{1/2}N_{4,c} = \{\mathbb{G}^4_n(0, c) + \mathcal{G}^4_n(\hat\theta, c)\} - (\tau^c_4/\tau^c_0)\{\mathbb{G}^0_n(0, c) - \mathcal{G}^0_n(\hat\theta, c)\} + \mathrm{o_P}(1).$$

The first components of each term satisfy

$$\mathbb{G}^4_n(0, c) = (1,0,0,0)n^{-1/2}\sum_{i=1}^n z^c_{4,i}, \quad \mathbb{G}^0_n(0, c) = (0,0,1,0)n^{-1/2}\sum_{i=1}^n z^c_{4,i}.$$

From Lemma B.3 the bias terms are

$$\mathcal{G}_n^4(\hat\theta,c) = 2c^5\mathsf{f}(c)\sigma^{-1}\tilde a - 4\tau_4^c\sigma^{-1}\hat a, \quad \mathcal{G}_n^0(\hat\theta,c) = 2c\mathsf{f}(c)\sigma^{-1}\tilde a.$$

Since $\tilde a$ and $\hat a$ satisfy (B.4) and (B.7) we get

$$\begin{aligned}\mathcal{G}_n^4(\hat\theta,c) &= 2c^5\mathsf{f}(c)(0,0,0,1/2)n^{-1/2}\textstyle\sum_{i=1}^n z_{4,i}^c \\ &\quad - 2(\tau_4^c/\tau_2^c)\{0,1,-(\tau_2^c/\tau_0^c), c(c^2-\tau_2^c/\tau_0^c)\mathsf{f}(c)\}n^{-1/2}\textstyle\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1), \\ \mathcal{G}_n^0(\hat\theta,c) &= 2c\mathsf{f}(c)(0,0,0,1/2)n^{-1/2}\textstyle\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1).\end{aligned}$$

Add the expansions for $\mathbb{G}_n^4(0,c)$ and $\mathcal{G}_n^4(\hat\theta,c)$ and subtract τ_4^c/τ_0^c times the sum of $\mathbb{G}_n^0(0,c)$ and $\mathcal{G}_n^0(\hat\theta,c)$ to get $N_{4,c} = (\zeta_{4,c})'n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1)$ where $\zeta_{4,c} = [1, -2\tau_4^c/\tau_2^c, \tau_4^c/\tau_0^c, \{c^4 - 2(\tau_4^c/\tau_2^c)c^2 + \tau_4^c/\tau_0^c\}c\mathsf{f}(c)]'$ as in (4.7). □

Proof of Theorem 4.1: 1. *Empirical process representation.* Recall from (B.11) that, for $p = 3,4$, then $\hat\mu_{p,c} = \mathsf{G}_n^k(\hat\theta,c)/\mathsf{G}_n^0(\hat\theta,c)$.
2. *Denominator.* Lemma B.9(a) shows $\sup_{c\ge c_0}\{\mathsf{G}_n^0(\hat\theta,c) - \tau_0^c\} = \mathsf{o_P}(1)$.
3. *Third moment.* Lemma B.9(b) shows $n^{1/2}\mathsf{G}_n^3(\hat\theta,c)$ has uniform expansion $\zeta_{3,c}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathsf{o_P}(1)$ for $c \ge c_0$. Noting that $(\tau_0^c)^2\lambda_{6,c} = \mathsf{Var}\{(\zeta_{3,c})'z_{3,i}^c\}$ we then get $\hat T_{3,c} = n^{1/2}\hat\mu_{3,c}/\lambda_{6,c}^{1/2} = T_{3,c,n} + \mathsf{o_P}(1)$ uniformly in $c \ge c_0$.
4. *Fourth moment.* Write

$$n^{1/2}(\hat\mu_{4,c} - \tau_4^c/\tau_0^c) = n^{1/2}\{\mathsf{G}_n^4(\hat\theta,c) - (\tau_4^c/\tau_0^c)\mathsf{G}_n^0(\hat\theta,c)\}/\mathsf{G}_n^0(\hat\theta,c).$$

Lemma B.9(c) expands the numerator as $\zeta_{4,c}'n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1)$. Proceed as in item 3 to see that $\hat T_{4,c} = T_{4,c,n} + \mathsf{o_P}(1)$ uniformly in $c \ge c_0$.
5. *Distributions.* The Central Limit Theorem shows that the finite dimensional distributions of $T_{3,c,n}$, $T_{4,c,n}$ converge jointly to zero mean normal distributions with unit marginal variances. Comparing the definitions of $z_{p,i}^c$ and $\mathsf{G}_n^{w,m}(\theta,c)$ in (4.4), (B.1), respectively, it is seen that each coordinate of $T_{p,c,n}$ is of the form $\mathbb{G}_n^{1,m}(0,c)$ so tightness follows from Corollary B.2. □

B.5 Proof of results for least trimmed squares estimators

We represent the truncated moments (4.2) in terms of the two sided processes introduced in §B.1. In this section the superscript ^{LTS} is ignored. Let $\hat\theta = (\hat\theta_1, \hat\theta_m)$ with $\hat\theta_1 = (0,\hat b)$ and $\hat\theta_m = (\hat a, \hat b)$ where $\hat b = N^{-1}(\hat\beta_{LTS} - \beta)$ and $\hat a = n^{1/2}(\hat\sigma_{LTS} - \sigma)$. Let also $\hat d = n^{1/2}(\hat\xi_{(h)} - \sigma c)$ where $h/n = \mathsf{P}(\varepsilon_1^2 < \sigma^2c^2)$. These errors were analyzed in Lemmas B.6, B.8. Then,

$$n^{1/2}\hat\mu_{k,h} = n^{1/2}\mathsf{G}_n^k(\hat\theta, c + n^{-1/2}\sigma^{-1}\hat d)/\mathsf{G}_n^0(\hat\theta, c + n^{-1/2}\sigma^{-1}\hat d). \tag{B.13}$$

Lemma B.10. *Suppose Assumptions 4.1, 4.2 hold. Recall $\zeta_{3,c}$, $\zeta_{4,c}$ from (4.11). Then*
(a) $\mathsf{G}_n^0(\hat\theta, c + n^{-1/2}\hat d) = \tau_0^c + \mathsf{o_P}(1)$;
(b) $n^{1/2}\mathsf{G}_n^3(\hat\theta, c + n^{-1/2}\sigma^{-1}\hat d) = \zeta_{3,c}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathsf{o_P}(1)$;
(c) $n^{1/2}\{\mathsf{G}_n^4(\hat\theta, c + n^{-1/2}\hat d) - \tau_4^c/\tau_0^c\mathsf{G}_n^0(\hat\theta, c + n^{-1/2}\hat d)\}$

$$= \zeta_{4,c}'n^{-1/2}\textstyle\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1).$$

Proof of Lemma B.10. (a) Apply Lemmas A.1, B.4(d) and $\bar{\mathsf{G}}_n^0(0,c) = \tau_0^c$.
(b) Let $N_{3,\hat c} = \mathsf{G}_n^3(\hat\theta, c + n^{-1/2}\hat d) - \bar{\mathsf{G}}_n^3(0,c)$ noting that $\bar{\mathsf{G}}_n^3(0,c) = 0$. Let $\hat\theta_1^d = (n^{-1/2}\sigma^{-1}\hat d, 0)$ and $\hat\theta^d = (\hat\theta_1 + \hat\theta_1^d, \hat\theta_m) = (n^{-1/2}\sigma^{-1}\hat d, \hat b, \hat a, \hat b)$. Then by Lemmas A.1, B.4(c), $n^{1/2}N_{3,\hat c}^{LTS} = \mathbb{G}_n^3(0,c) + \mathcal{G}_n^3(\hat\theta^d, c) + \mathsf{o_P}(1)$. Lemma B.3 shows that $\mathcal{G}_n^3(\hat\theta^d,c) = \{2c^3\mathsf{f}(c) - 3\tau_2^c\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n x_{in}'\hat b$. Given that $\hat b = N^{-1}(\hat\beta - \beta)$ has expansion (B.8), then Lemma B.5 shows $\sum_{i=1}^n x_{in}'\hat b = (\tau_2^c)^{-1}\sum_{i=1}^n \varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + \mathsf{o_P}(1)$, noting that $\tau_2^c = \tau_0^c - 2c\mathsf{f}(c)$ by (3.6). Thus,

$$\mathcal{G}_n^3(\hat\theta^d,c) = (\tau_2^c)^{-1}\{2c^3\mathsf{f}(c) - 3\tau_2^c\}(0,1,0)n^{-1/2}\textstyle\sum_{i=1}^n z_{3,i}^c + \mathsf{o_P}(1).$$

Add $\mathsf{G}_n^3(0,c) = (1,0,0)n^{-1/2}\sum_{i=1}^n z_{3,i}^c$ to get $n^{1/2}N_{3,c} = \zeta_{3,c}' n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathsf{o_P}(1)$ recalling that $\zeta_{3,c} = [1, \{2c^3\mathsf{f}(c) - 3\tau_2^c\}/\tau_2^c, 0]'$ in (4.11).

(*c*) Let $N_{4,\hat{c}} = \{\mathsf{G}_n^4(\hat{\theta}, c+n^{-1/2}\hat{d}) - \tau_4^c/\tau_0^c \mathsf{G}_n^0(\hat{\theta}, c+n^{-1/2}\hat{d})\}$. Due to Lemmas A.1, B.4(*c*) we get, for $j = 0, 4$,

$$n^{1/2}\{\mathsf{G}_n^j(\hat{\theta}^{LTS}, c + n^{-1/2}\hat{d}^{(0)}) - \bar{\mathsf{G}}_n^j(0,c)\} = \mathbb{G}_n^j(0,c) + \mathcal{G}_n^j(\hat{\theta}^d, c) + \mathsf{o_P}(1).$$

The compensators satisfy the identity (B.4) so that

$$n^{1/2}N_{4,\hat{c}} = \{\mathbb{G}_n^4(0,c) + \mathcal{G}_{1n}^4(\hat{\theta}^d, c)\} - (\tau_4^c/\tau_0^c)\{\mathbb{G}_n^0(0,c) + \mathcal{G}_{1n}^0(\hat{\theta}^d, c)\} + \mathsf{o_P}(1).$$

The first component of each term satisfy

$$\mathbb{G}_n^4(0,c) = (1,0,0,0)n^{-1/2}\sum_{i=1}^n z_{4,i}^c, \quad \mathbb{G}_n^0(0,c) = (0,0,1,0)n^{-1/2}\sum_{i=1}^n z_{4,i}^c.$$

From Lemma B.3 the bias terms are

$$\mathcal{G}_{1n}^4(\hat{\theta}^d, c) = 2c^4\mathsf{f}(c)\sigma^{-1}\hat{d} - 4\tau_4^c\sigma^{-1}\hat{a}, \qquad \mathcal{G}_{1n}^0(\hat{\theta}^d, c) = 2\mathsf{f}(c)\sigma^{-1}\hat{d}.$$

Given the expansions for $\hat{a}, \hat{d}$ in (B.9), (B.10) we get

$$\mathcal{G}_{1n}^4(\hat{\theta}^d, c) = \{(0,0,-c^4,0) - 2(\tau_4^c/\tau_2^c)(0,1,-c^2,0)\}n^{-1/2}\sum_{i=1}^n z_{4,i}^c,$$
$$\mathcal{G}_{1n}^0(\hat{\theta}^d, c) = (0,0,-1,0)n^{-1/2}\sum_{i=1}^n z_{4,i}^c.$$

Add the expansions for $\mathbb{G}_n^4(0,c)$ and $\mathcal{G}_n^4(\hat{\theta}^d, c)$ and subtract τ_4^c/τ_0^c times the sum of $\mathbb{G}_n^0(0,c)$ and $\mathcal{G}_n^0(\hat{\theta}^d, c)$ to get $N_{4,c} = \zeta_{4,c}' n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathsf{o_P}(1)$ where $\zeta_{4,c} = \{1, -2\tau_4^c/\tau_2^c, -\tau_4^c/\tau_0^c + c^2 2\tau_4^c/\tau_2^c - c^4, 0\}'$ as in (4.11). □

Proof of Theorem 4.2. As the proof of Theorem 4.1 replacing Lemma B.10 by Lemma B.9. □

References

[1] Barr, D. R., and Sherrill, E. T. Mean and variance of truncated normal distributions. *American Statistician 53* (1999), 357–361.

[2] Bercu, B., Delyon, B., and Rio, E. *Concentration inequalities for sums and martingales.* Springer, Cham, 2015.

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

[4] Croux, C., Dhaene, G., and Hoorelbeke, D. Testing the information matrix equality with robust estimators. *Journal of Statistical Planning and Inference 136* (2006), 3583–3613.

[5] Delyon, B. Exponential inequalities for sums of weakly dependent variables. *Electronic Journal of Probability 14* (2009), 752–779.

[6] Escanciano, J. C. Model checks using residual marked empirical processes. *StatistSinica 17* (2007), 115–138.

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

[8] Jiao, X., and Nielsen, B. Asymptotic analysis of iterated 1-step huber-skip m-estimators with varying cut-offs. In *Analytic Methods in Statistics*, J. Antoch, J. Jurečková, M. Maciak, and M. Pešta, Eds., vol. 193 of *Springer Proceedings in Mathematics & Statistics.* Springer, 2017.

[9] Johansen, S., and Nielsen, B. Saturation by indicators in regression models. In *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*, J. L. Castle and N. Shephard, Eds. Oxford University Press, Oxford, 2009.

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

[11] Johansen, S., and Nielsen, B. Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli 22* (2016), 1131–83.

[12] Johansen, S., and Nielsen, B. Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics 43* (2016), 321–81.

[13] Koul, H. L. *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. Springer, New York, 2002.

[14] Koul, H. L., and Ossiander, M. Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics 22* (1994), 540–562.

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

[16] Lai, T. L., and Wei, C. Z. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics 10* (1982), 154–166.

[17] Rousseeuw, P. J. Least median of squares regressions. 871–880.

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

[19] Serfling, R. J. *Approximation Theorems of Mathematical Statistics.* Wiley, New York, 1980.

[20] Stute, W. Nonparametric model checks for regression. *Annals of Statistics 25* (1997), 613–641.

[21] Víček, J. A. The least trimmed squares; part iii: Asymptotic normality. *Kybernetika 42* (2006), 203–224.

[22] Welsh, A. H., and Ronchetti, E. A journey in single steps: robust one-step m-estimation in linear regression. *Journal of Statistical Planning and Inference 103* (2006), 287–310.