Nuffield College
UNIVERSITY OF OXFORD

ECONOMICS DISCUSSION PAPERS

Paper No: 2023-W02

Normality testing after outlier removal

By Vanessa Berenguer-Rico and Bent Nielsen

June 2023

Normality testing after outlier removal

Vanessa Berenguer-Rico* & Bent Nielsen†

1 June 2023

Abstract

The cumulant based normality test after outlier removal is analyzed. It is shown that the standard least squares normalizations can be misleading in this context. The sample cumulants should be standardized according to the truncation imposed at the removal stage and the estimation method being used. New standardizations that lead to chi-squared inference are derived.

1 Introduction

Misspecification tests are frequently used in empirical studies. When the presence of outliers is suspected, these tests are also performed after the removal of outliers. This is useful as valid inference in regression analysis after outlier removal depends on the distributional assumptions on the good errors. Normality is usually tested. We show that standard cumulant-based normality tests on the clean sub-sample are not valid in i.i.d. settings and develop test statistics that deliver χ^2 inference.

Two procedures for outlier removal are considered. First, we study the robustified least squares (RLS) procedure, where the model is first estimated using ordinary least squares (OLS). Least squares residuals are then used to identify outliers and remove them from the sample. Finally, OLS is applied again on the clean sub-sample. This methodology is commonly used although it is not fully robust. It has been labelled as the 'data analytic strategy' (Welsh and Ronchetti, 2002), 'rejection-plus-least squares' (Hadi and Simonoff, 1993), or 'rejection-estimation procedures' (Hampel, 1985).

Second, we consider the least trimmed squares procedure (LTS), where the model is estimated by the LTS estimator of Rousseeuw (1984). For a given number of good observations, say h, in a sample of size n, the LTS estimator is least squares on the h sub-sample that minimizes the squared residuals, delivering in this way an estimated set of outliers. The robustness properties of the LTS estimator makes this second procedure more appealing when outliers are suspected in the first place.

Asymptotic theory for the RLS estimator with i.i.d. errors has been studied by Johansen and Nielsen (2009). They show that asymptotic inference requires consistency

*Mansfield College & Department of Economics, University of Oxford. Address for correspondence: Mansfield College, Oxford OX1 3TF, UK. E-mail:vanessa.berenguer-rico@economics.ox.ac.uk.

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

and efficiency corrections in order to be valid. These correction factors depend on the underlying distribution of the error term.

Asymptotic theory for the LTS estimator with i.i.d. errors has been studied by Butler (1982), Rousseeuw (1985), Croux and Rousseeuw (1992), Čížek (2005), Víšek (2006), Johansen and Nielsen (2016b). These papers show that appropriate (albeit different from RLS) correction factors are also required in this case. Again, these correction factors depend on the underlying distribution of the error term.

One important aspect of the results derived in this paper is that the standard moment based normality test on the RLS or LTS residuals is not valid in an i.i.d. setting. Specifically, we show that the standardization of the sample moments depend on the truncation imposed at the outlier removal stage and the estimation method being used. The intuition behind this result is easily illustrated when there is no contamination and all errors are normally distributed. In that case, removing outliers from the sample implies that the regression errors are truncated and their underlying distribution is no longer normal but truncated normal. Hence, the standardizations used when assuming (untruncated) normality are not the right ones. We derive the correct standardizations, which bring back χ^2 inference. In this sense, the analyzed statistics can be seen as tests for truncated normality.

The intuition from the non-contamination case actually extends to contaminated samples where the retained observations have truncated normal errors. This means that inferences are valid under a particular type of contamination, which we term ϵ-tail contamination. Specifically, in an i.i.d. setting, the errors have a distribution which is normal in the middle but can have non-normal tails. It is a special case of an ϵ-Lévy neighbourhood (Huber and Ronchetti, 2009, p. 18), but differs from the gross error model or ϵ-contamination (Huber, 1964). The ϵ-tail contamination scheme represents the model behind standard practice when using the LTS estimator. In practical applications of LTS, it is common to implement correction factors for non-contaminated, normal errors. This imposes, de facto, an ϵ-tail contamination structure: normality in the retained central observations with unmodelled tails. We develop normality tests in the LTS context that provide guidance on the validity of this choice.

In practice, one encounters many types of contamination. Bad leverage points are particularly worrysome. The LTS regression estimator is robust with respect to such points, whereas the RLS regression estimator is not. However, bad leverage points cannot be generated through an i.i.d. model. Instead, Berenguer-Rico et al. (2023); Berenguer-Rico and Nielsen (2022) propose and analyze a model termed the LTS model. This model has a proportion of good observations with i.i.d. normal errors, while the remaining errors have support outside the realized range of the good errors. This model permits bad leverage points. The LTS estimator is maximum likelihood in this model and has the same asymptotic distribution as the infeasible least squares estimator on the good observations. These properties are rather different from those discussed above for the i.i.d. model. It is therefore desirable to test which, if any, of the models are relevant for the data at hand. The normality test presented in this paper addresses this empirical need. The analyzed statistics will detect deviations from ϵ-tail contamination (or truncated normality), hence, guiding applied researchers in their data analysis allowing them to conduct valid subsequent inference.

We examine the theoretical results through simulation. We show that the normality test statistics that account for truncation (or outlier removal) have empirical sizes approaching the nominal size in large samples. This confirms the χ^2 asymptotics.

We study, analytically and by simulation, the power of these statistics to detect deviations from (truncated) normality. Specifically, we study the power of the tests to detect Cauchy distributions, the ϵ-contamination scheme and the LTS model. The simulation results show that the tests analyzed in this paper have empirical power approaching to one (as the sample size grows) in the different models considered.

In practice, one would also need misspecification tests for other aspects of the maintained model. In a related analysis, Berenguer-Rico and Wilms (2021) study the effect of outlier removal on heteroscedasticity testing and show that standard inference can be applied if the errors are symmetric.

Outline: §2 describes the model and test statistics. §3 derives the asymptotic properties of the test statistics. The theory is explored through simulations in §4. §5 contains an empirical illustration. Finally, §6 concludes. Proofs are collected in the Appendix.

2 Model and test statistics

We consider the linear model for

$$y_i = \beta' x_i + \varepsilon_i \qquad i = 1, \ldots, n, \tag{2.1}$$

where β and x_i are k vectors. The variables satisfy the following structure.

Assumption 2.1. *Let $\mathcal{F}_{in}$ be an array of filtrations so that $\mathcal{F}_{i-1,n} \subset \mathcal{F}_{in}$ and ε_{i-1}, x_i are $\mathcal{F}_{i-1,n}$-adapted. Let ε_i/σ be independent of $\mathcal{F}_{i-1,n}$ with distribution function F and scale σ.*

Assumption 2.1 jointly with Assumptions 3.1 or 3.2 below allow for a wide variety of regressors that can be both dependent and/or heterogeneously distributed. These include cross-sections, stationary, random walk and fractionally integrated time series. Indicator variables and structural breaks are also allowed. In the time series context, these regressors can be lagged dependent variables, hence, covering autoregressions and error correction models. While Assumption 2.1 allows for a wide range of regressor types, it assumes that the standardized errors are i.i.d. and independent of $\mathcal{F}_{i-1,n}$, hence, avoiding endogeneity and heteroscedasticity.

We define the data analytic strategy for removing outliers. Given initial estimators $\tilde{\beta}, \tilde{\sigma}$, residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ are formed. Observations satisfying $|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c$, for a user chosen cut-off c, are selected and a regression is run on those observations so that

$$\hat{\beta} = \{\textstyle\sum_{i=1}^n x_i x_i' 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c)}\}^{-1} \textstyle\sum_{i=1}^n x_i y_i 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c)} \tag{2.2}$$

This leads to updated residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$ and residual variance estimator,

$$\hat{\sigma}^2 = \varsigma_c^{-2} \{\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c)}\}^{-1} \textstyle\sum_{i=1}^n \hat{\varepsilon}_i^2 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}| \leq c)}, \tag{2.3}$$

where the consistency factor ς_c^2 is defined as follows. Let

$$\tau_p^c = \mathsf{E}(\varepsilon_1/\sigma)^p 1_{(|\varepsilon_1/\sigma| \leq c)} \qquad \text{for } c > 0, \tag{2.4}$$

denote two-sided truncated moments. In particular, $\tau_0^c = \mathsf{P}(|\varepsilon_1/\sigma| \leq c)$. Let $\tau_p = \tau_p^\infty$. The consistency factor ς_c^2 in (2.3) is then defined as $\varsigma_c^2 = \tau_2^c/\tau_0^c$.

We note that when $\mathsf{F} = \Phi$ is the standard normal distribution function then,

$$\tau_{2p+1}^c = 0, \qquad \tau_{2p}^c = \{(2p-1)!!\}\mathsf{P}(\chi_{2p+1}^2 \leq c^2) \qquad \text{for} \quad p \in \mathbb{N}_0, \tag{2.5}$$

where the odd factorial is $(2p-1)!! = \prod_{\ell=1}^{p}(2\ell-1)$ with the convention that $(2p-1)!! = 1$ for $p = 0$. To see this, integrate u^p with respect to Φ, substitute $u^2 = v$ and note $\Gamma\{(p+1)/2\} = \Gamma(1/2)\prod_{\ell=1}^{p/2}\{(2\ell-1)/2\}$ by the gamma functional equation. Barr and Sherrill (1999) has similar formulas for τ_1^c, τ_2^c. Insert $c = \infty$ in (2.5) to get the usual moments: $\tau_0 = \tau_2 = 1$, $\tau_4 = 3$, $\tau_6 = 15$, $\tau_8 = 105$. The normal density satisfies $(\partial/\partial u)\{-u\varphi(u)\} = (u^2-1)\varphi(u)$, so that $\tau_2^c = \int_{-c}^{c} u^2\varphi(u)du = \tau_0^c - 2c\varphi(c)$.

Table 1 gives numerical values for ς_c^2 under the hypothesis of normal errors without outliers. The above estimators are referred to as 1-step Huber-skip estimators and are analyzed in Johansen and Nielsen (2009, 2013, 2016a,b).

In §3.1, we initialize the data analytic strategy with the full sample least squares estimator so that $\tilde{\beta} = \tilde{\beta}_{OLS}$. This gives the robustified least squares estimator and we write $\hat{\beta}_{RLS}$ for $\hat{\beta}$. We note that the least squares estimator arises when $c = \infty$. In §3.3, we initialize with the Least Trimmed Squares estimator. In that case, we choose the indicators in (2.2) differently, which we ignore while establishing notation.

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

$$\hat{\mu}_{p,c}^s = \{\textstyle\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}_s| \leq c)}\}^{-1}\textstyle\sum_{i=1}^n (\hat{\varepsilon}_i/\hat{\sigma}_s)^p 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}_s| \leq c)} \qquad \text{for } p \in \mathbb{N}. \tag{2.6}$$

We then study the following truncated normality test statistics

$$\hat{T}_{3,c}^s = n^{1/2}\hat{\mu}_{3,c}^s/(\lambda_{6,c}^s)^{1/2}, \qquad \hat{T}_{4,c}^s = n^{1/2}(\hat{\mu}_{4,c}^s - \lambda_{3,c}^s)/(\lambda_{24,c}^s)^{1/2}, \tag{2.7}$$

where $\lambda_{3,c}^s, \lambda_{6,c}^s, \lambda_{24,c}^s$ (to be formally defined in the next section) are normalizing constants that depend on the selection method.

We note that when $c = \infty$ there is no selection and the statistics reduce to the standard cumulant based normality test statistics based on least squares with $\lambda_{3,\infty}^{OLS} = 3$, $\lambda_{6,\infty}^{OLS} = 6$ and $\lambda_{24,\infty}^{OLS} = 24$. The resulting test has a long history going back to Thiele, Pearson and Fisher. In econometrics it is often called the Jarque-Bera test.

3 Asymptotic properties

In this section, we study the effect of removing outliers from the sample on the cumulant based normality test described in §2. In practice, it is unknown whether the data are uncontaminated or not. Therefore, we first study the uncontaminated case in §3.1. In this context, we analyze the properties of the test when the procedure is initialized by the OLS estimator, what we call robustified least squares (RLS). In §3.2, we introduce a new contamination scheme, which we term ϵ-tail contamination. This has non-normal tails while the central part of the distribution is normal. Then, in §3.3, we study the properties of the test in this contaminated setting, when the procedure is initialized robustly using the LTS estimator. Power of both tests is discussed analytically in §3.4.

3.1 Robustified least squares

We consider the normality test based on the truncated empirical moments in (2.6) where $\tilde{\beta}, \tilde{\sigma}$ are full sample least squares estimators and $\hat{\beta}, \hat{\sigma}$ are the 1-step Huber skip estimators with residuals $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ and $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$.

In the context of i.i.d. normal errors, Johansen and Nielsen (2009, 2016b) study the asymptotic properties of the RLS estimator, $\hat{\beta}$, and show that $n^{1/2}(\hat{\beta}-\beta)$ is asymptotically $\mathsf{N}(0, \eta_\beta \sigma^2 \Sigma^{-1})$ where $(\tau_0^c)^2\eta_\beta = \tau_2^c + \{4c\varphi(c)\tau_2^c\} + \{2c\varphi(c)\}^2$ depends on the cut-off value c. This dependence is carried into the test statistics $\hat{T}_{3,c}^{RLS}$, $\hat{T}_{4,c}^{RLS}$.

The normalizing constants $\lambda_{3,c}^{RLS}, \lambda_{6,c}^{RLS}, \lambda_{24,c}^{RLS}$, for the test statistics $\hat{T}_{3,c}^{RLS}$, $\hat{T}_{4,c}^{RLS}$ in (2.7) are computed as follows. Define the vectors

$$z_{3,i}^c = \left\{ \begin{array}{c} (\varepsilon_i/\sigma)^3 1_{(|\varepsilon_i/\sigma| \le c)} \\ (\varepsilon_i/\sigma) 1_{(|\varepsilon_i/\sigma| \le c)} \\ (\varepsilon_i/\sigma) \end{array} \right\}, \quad z_{4,i}^c = \left\{ \begin{array}{c} (\varepsilon_i/\sigma)^4 1_{(|\varepsilon_i/\sigma| \le c)} - \tau_4^c \\ (\varepsilon_i/\sigma)^2 1_{(|\varepsilon_i/\sigma| \le c)} - \tau_2^c \\ 1_{(|\varepsilon_i/\sigma| \le c)} - \tau_0^c \\ (\varepsilon_i/\sigma)^2 - 1 \end{array} \right\}. \tag{3.1}$$

For standard normal ε_i, these vectors are uncorrelated. The Central Limit Theorem shows that $n^{-1/2}\sum_{i=1}^n z_{3,i}^c$ and $n^{-1/2}\sum_{i=1}^n z_{4,i}^c$ are asymptotically normal and independent with variances Ω_3^c, Ω_4^c given by

$$\left(\begin{array}{ccc} \tau_6^c & \tau_4^c & \tau_4^c \\ \tau_4^c & \tau_2^c & \tau_2^c \\ \tau_4^c & \tau_2^c & 1 \end{array} \right), \quad \left(\begin{array}{cccc} \tau_8^c - \tau_4^c\tau_4^c & \tau_6^c - \tau_2^c\tau_4^c & \tau_4^c(1-\tau_0^c) & \tau_6^c - \tau_4^c \\ \tau_6^c - \tau_2^c\tau_4^c & \tau_4^c - \tau_2^c\tau_2^c & \tau_2^c(1-\tau_0^c) & \tau_4^c - \tau_2^c \\ \tau_4^c(1-\tau_0^c) & \tau_2^c(1-\tau_0^c) & \tau_0^c(1-\tau_0^c) & \tau_2^c - \tau_0^c \\ \tau_6^c - \tau_4^c & \tau_4^c - \tau_2^c & \tau_2^c - \tau_0^c & 2 \end{array} \right).$$

We compute the vectors

$$\zeta_{3,c}^{RLS} = \{1, -3\tau_2^c/\tau_0^c, 2(c^2 - 3\tau_2^c/\tau_0^c)c\varphi(c)\}', \tag{3.2}$$
$$\zeta_{4,c}^{RLS} = \{1, -2\tau_4^c/\tau_2^c, \tau_4^c/\tau_0^c, (c^4 - c^2 2\tau_4^c/\tau_2^c + \tau_4^c/\tau_0^c)c\varphi(c)\}', \tag{3.3}$$

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

$$\lambda_{3,c}^s = \tau_4^c/\tau_0^c, \quad \lambda_{6,c}^s = \zeta_{3,c}^{s\prime}\Omega_3^c\zeta_{3,c}^s/(\tau_0^c)^2, \quad \lambda_{24,c}^s = \zeta_{4,c}^{s\prime}\Omega_4^c\zeta_{4,c}^s/(\tau_0^c)^2. \tag{3.4}$$

Numerical values are given in Table 1. We note that these normalizations depend substantially on the choice of c.

We introduce a deterministic normalization matrix N and define $x_{in} = N'x_i$. The normalization N is chosen so that $\sum_{i=1}^n x_{in}x_{in}'$ has a positive definite limit. Examples include $N = n^{-1/2}I_k$ for stationary regressors, $N = n^{-1}I_k$ for random walk regressors, while $N = \mathrm{diag}(n^{-1/2}, n^{-3/2})$ if $x_i = (1, i)'$.

Assumption 3.1. *Suppose (i) ε_i/σ are i.i.d. $\mathsf{N}(0,1)$;*
(ii) $\max_{i\le n} \mathsf{E}|n^{1/2}x_{in}|^{2+\kappa} = \mathrm{O}(1)$ for some $\kappa > 0$;
(iii) $(\sum_{i=1}^n x_{in}x_{in}', \sum_{i=1}^n x_{in}\varepsilon_i) \overset{\mathsf{D}}{\to} (\Sigma, U)$, where $\Sigma \overset{a.s.}{>} 0$ may be random.

Theorem 3.1. *Let Assumptions 2.1, 3.1 hold. Let $c > 0$ be fixed. Then, for $p = 3, 4$,*

$$\hat{T}_{p,c}^{RLS} = \{(\zeta_{p,c}^{RLS})'\Omega_p^c(\zeta_{p,c}^{RLS})\}^{-1/2}(\zeta_{p,c}^{RLS})'n^{-1/2}\sum_{i=1}^n z_{p,i}^c + \mathrm{o}_{\mathsf{P}}(1)$$

are asymptotically independent standard normal and $\sum_{j=3}^4(\hat{T}_{j,c}^{RLS})^2$ is asymptotically χ_2^2.

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

$\tau_0^c = \mathsf{P}(\lvert\varepsilon_1/\sigma\rvert < c)$	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
$\lambda_{24,c}^{LTS}$	0.0013	1.6066	6.9538	16.5596	21.8304	23.5115	24

Remark 3.1. *The normalizations $\lambda_{3,c}^{RLS}$, $\lambda_{6,c}^{RLS}$, $\lambda_{24,c}^{RLS}$ in (3.4) differ from the traditional values 3, 6, 24. Those values are commonly applied in practice. This leads to severe size distortions, when there are no outliers, as we compare $\hat\mu_{3,c}^{RLS}$ 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 $\mathsf{N}(\lambda_{3,c}^{RLS}, \lambda_{24,c}^{RLS}/n)$. The 3rd moment test is under-sized while the 4th moment test has asymptotic size of unity. Indeed, suppose we set $c = 2.58$ corresponding to a 1% trimming and let $n = 100$. Incorrect normalizations give 95% sampling regions of $[-0.48, 0.48]$ and $[2.04, 3.96]$ instead of the correct $[-0.30, 0.30]$ and $[1.86, 2.69]$, leading to sizes of 0.24% and 13.5%, respectively. For $n = 200, 400$ the 4th moment test has size 62.0% and 98.9%, respectively.*

3.2 ϵ-tail contamination

We introduce a new contamination scheme. We term this ϵ-tail contamination since the tails of the distribution are left unspecified while the central part is assumed normal.

Definition 1. *Let $0 \le \epsilon < 1$ and let $c_\epsilon = \Phi^{-1}(1 - \epsilon/2)$ be the standard normal $1 - \epsilon/2$ quantile. A distribution function that is differentiable on an open interval containing $[-c_\epsilon, c_\epsilon]$ with standard normal density on that interval is an ϵ-tail contaminated normal distribution function.*

The ϵ-tail contaminated normal distribution allows for outliers, while preserving truncated normality. This provides an appropriate theoretical framework to test for normality after the removal of outliers, in which both contaminated or uncontaminated settings are allowed. It is worth noting that the definition can be extended to other reference distributions. For instance, one could be interested in ϵ-tail contaminated t-distributions. We also note that an ϵ-tail contaminated normal distribution need neither be continuous nor symmetric, while the support can be bounded. It is a special case of an ϵ-Lévy neighbourhood (Huber and Ronchetti, 2009, p. 18), but differs from the gross error model, also called ϵ-contamination, which has support on $\mathbb{R}$ (Huber, 1964).

3.3 Least trimmed squares

Next, we initialize the data analytic strategy robustly with the LTS estimator. The LTS estimator is defined as follows (Rousseeuw, 1984). The user chooses a $h \le n$. For a given β compute the absolute residuals $\xi_i(\beta) = \lvert y_i - x_i'\beta\rvert$ with increasing order statistics $\xi_{(i)}(\beta)$. The LTS estimator is then the minimizer $\tilde\beta_{LTS} = \arg\min_\beta \sum_{i=1}^h \xi_{(i)}^2(\beta)$.

If we let $\tilde{\xi}_i = \xi_i(\tilde{\beta}_{LTS})$ with order statistics $\tilde{\xi}_{(i)}$ we can write the LTS estimator as

$$\tilde{\beta}_{LTS} - \beta = \Big[\sum_{i=1}^n x_i x_i' 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}\Big]^{-1} \sum_{i=1}^n x_i \varepsilon_i 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}. \tag{3.5}$$

The corresponding scale estimator includes the consistency factor $\varsigma_c^2 = \tau_2^c/\tau_0^c$:

$$\tilde{\sigma}^2_{LTS} = (\tau_0^c/\tau_2^c)\Big[\sum_{i=1}^n 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}\Big]^{-1} \sum_{i=1}^n (y_i - x_i'\tilde{\beta}_{LTS})^2 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}. \tag{3.6}$$

We now consider the data analytic strategy initialized by the LTS estimator. Replace $\tilde{\beta}$ and $\tilde{\sigma} c$ with $\tilde{\beta}_{LTS}$ and $\tilde{\xi}_{(h)} = \xi_{(h)}(\tilde{\beta}_{LTS})$ in (2.2) and (2.3), respectively. This selects the same observations as before, so that $\hat{\beta}_{LTS} = \tilde{\beta}_{LTS}$ and $\hat{\sigma}_{LTS} = \tilde{\sigma}_{LTS}$.

Robust estimators are often scaled to be consistent in normal samples. The validity of this scaling depends on the assumed model for the regression errors. As LTS trims the tails, the scaling is valid when the central part of the error distribution is *truncated* normal. This is the case of the ϵ-tail contamination where the full set of errors are i.i.d. as described in §3.2. In contrast, if the retained observations are i.i.d. *untruncated* normal as in the LTS model of Berenguer-Rico et al. (2023), scaling should not be used. Here, we focus on the ϵ-tail contaminated case, but return to the LTS model in the power simulations in §4.2.3.

The available theory for the LTS estimator in an i.i.d. setting shows that under certain regularity conditions, $n^{1/2}(\tilde{\beta}_{LTS} - \beta)$ is asymptotically normal with a variance depending on the error distribution. In particular, for the case of normal errors the limiting distribution is $\mathsf{N}(0, \Sigma^{-1}\sigma^2/\tau_2^c)$. This is proved by Butler (1982) for the location-scale case where the errors have a smooth distribution function. The case with regressors is analyzed by Čížek (2005), Víšek (2006), requiring that the errors are symmetric with smooth distribution function and fourth moments, fixed regressors and boundedness of the estimator. All these papers have i.i.d. errors and allow ϵ-tail contamination. Recently, Berenguer-Rico and Nielsen (2022) have given general conditions for boundedness. In line with these results, we assume the following high level asymptotic expansion.

Assumption 3.2. *Let h be the largest integer not exceeding $n\{\Phi(c) - \Phi(-c)\}$. Suppose*
(i) ε_i/σ are ϵ-tail contaminated normal and $0 < c < c_\epsilon$;
(ii) $\max_{i\le n} \mathsf{E}|n^{1/2}x_{in}|^{2+\kappa} = \mathrm{O}(1)$ for some $\kappa > 0$;
(iii) $(\Sigma_n, U_n) = \{\sum_{i=1}^n x_{in}x_{in}', \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i/\sigma| \le c)}\} \overset{\mathsf{D}}{\to} (\Sigma, U)$ and $\Sigma \overset{a.s.}{>} 0$;
(iv) The LTS estimator has expansion $N^{-1}(\tilde{\beta}_{LTS} - \beta) = (\tau_2^c \Sigma_n)^{-1} U_n + \mathrm{o}_{\mathsf{P}}(1)$.

In the LTS case, we define $\hat{T}^{LTS}_{3,c}$, $\hat{T}^{LTS}_{4,c}$ in (2.7) using the cut-off $\tilde{\xi}_{(h)}$ instead of $\tilde{\sigma} c$,

$$\begin{aligned} \zeta^{LTS}_{3,c} &= \{1, 2c^3\varphi(c)/\tau_2^c - 3, 0\}', & (3.7)\\ \zeta^{LTS}_{4,c} &= \{1, -2\tau_4^c/\tau_2^c, 2c^2\tau_4^c/\tau_2^c - c^4, 0\}', & (3.8)\end{aligned}$$

and normalizations λ^{LTS} as in (3.4) that are tabulated in Table 1. For larger values of c, the values for RLS and LTS are not that different.

Theorem 3.2. *Suppose Assumption 2.1, 3.2. Let c be fixed. Then, for $p = 3, 4$,*

$$\hat{T}^{LTS}_{p,c} = \{(\zeta^{LTS}_{p,c})'\Omega_p^c(\zeta^{LTS}_{p,c})\}^{-1/2}(\zeta^{LTS}_{p,c})' n^{-1/2} \textstyle\sum_{i=1}^n z^c_{p,i} + \mathrm{o}_{\mathsf{P}}(1)$$

are asymptotically independent standard normal and $\sum_{j=3}^4 (\hat{T}^{LTS}_{j,c})^2$ is asymptotically χ^2_2.

3.4 Power

We consider the power of the kurtosis test based on $\hat{T}^s_{4,c} = n^{1/2}(\hat{\mu}^s_{4,c} - \lambda^s_{3,c})/(\lambda^s_{24,c})^{1/2}$. Suppose the alternative hypothesis of interest is a distribution F. Write $\lambda^s_{3c\Phi}$ and $\lambda^s_{24c\Phi}$ for $\lambda^s_{3,c}$ and $\lambda^s_{24,c}$, respectively, and let $\lambda^s_{3c\mathsf{F}}$ be the corresponding limiting term under F. Then, rewrite the kurtosis statistic as

$$\hat{T}^s_{4c} = \frac{n^{1/2}(\hat{\mu}^s_{4c} - \lambda^s_{3c\mathsf{F}})}{(\lambda^s_{24c\Phi})^{1/2}} + \frac{n^{1/2}(\lambda^s_{3c\mathsf{F}} - \lambda^s_{3c\Phi})}{(\lambda^s_{24c\Phi})^{1/2}}. \tag{3.9}$$

The first term in (3.9) is properly demeaned under the alternative distribution F. Hence, this term converges. The second term in (3.9) is a non-centrality term. The test is consistent when $\lambda^s_{3c\mathsf{F}} \neq \lambda^s_{3c\Phi}$.

Next, we study local power by analyzing the non-centrality term for the two procedures considered above, RLS and LTS. Suppose that the alternative of interest is that the innovations ε_i/σ are i.i.d. with a symmetric, continuous, ϵ-contaminated distribution function $\mathsf{F} = (1-\epsilon)\Phi + \epsilon\mathsf{G}$ with four moments and satisfying a local Lipschitz condition in neighbourhoods of the cut-off c, see Remark A.1. Such distributions are covered by the LTS theory by Čížek (2005) and Víšek (2006) and the present appendix.

Let $\tau^c_{p\Phi}$ and $\tau^c_{p\mathsf{G}}$ denote the truncated moments under normality and under G. For the RLS procedure, the numerator of the non-centrality term satisfies

$$\begin{aligned} n^{1/2}(\lambda^{RLS}_{3c\mathsf{F}} - \lambda^{RLS}_{3c\Phi}) = n^{1/2}\epsilon\frac{\tau^c_{4\Phi}}{\tau^c_{0\Phi}}\Big[\Big(\frac{\tau^c_{4\mathsf{G}}}{\tau^c_{4\Phi}} - 2\frac{\tau^c_{2\mathsf{G}}}{\tau^c_{2\Phi}} + \frac{\tau^c_{0\mathsf{G}}}{\tau^c_{0\Phi}}\Big) \\ + c\mathsf{f}(c)\big\{(\tau^\infty_{2\mathsf{G}})^{1/2} - 1\big\}\Big(\frac{c^4}{\tau^c_{4\Phi}} - 2\frac{c^2}{\tau^c_{2\Phi}} + \frac{1}{\tau^c_{0\Phi}}\Big)\Big] + \mathrm{o}(n^{1/2}\epsilon), \end{aligned} \tag{3.10}$$

see the derivation in (D) in Appendix D. Similarly, for the LTS procedure the numerator of the non-centrality terms satisfies, see (D.3) in Appendix D,

$$n^{1/2}(\lambda^{LTS}_{3c\mathsf{F}} - \lambda^{LTS}_{3c\Phi}) = n^{1/2}\epsilon\frac{\tau^c_{4\Phi}}{\tau^c_{0\Phi}}\Big(\frac{\tau^c_{4\mathsf{G}}}{\tau^c_{4\Phi}} - 2\frac{\tau^c_{2\mathsf{G}}}{\tau^c_{2\Phi}} + \frac{\tau^c_{0\mathsf{G}}}{\tau^c_{0\Phi}}\Big) + \mathrm{o}(n^{1/2}\epsilon). \tag{3.11}$$

The results show that, for both procedures RLS and LTS, the relevant local power rate for ϵ is $n^{-1/2}$ as in Heretier and Ronchetti (1994). It is interesting to note that the RLS expression has an additional term relative to the LTS expression. This has consequences for power properties. We illustrate them with a few examples.

First, suppose that $\mathsf{G} = \Phi$ so that $\mathsf{F} = \Phi$ and, hence, there is no contamination. In this case $\tau^c_{p\mathsf{G}} = \tau^c_{p\Phi}$ and $\tau^\infty_{2\mathsf{G}} = 1$, so that both non-centrality terms are zero. This matches the results in Theorem 3.1 and Theorem 3.2 with $\epsilon = 0$.

Second, suppose G is ϵ-tail contaminated normal with $\tau^\infty_{2\mathsf{G}} \neq 1$. Thus, G has a normal density on the interval $[-c, c]$. In this case $\tau^c_{p\mathsf{G}} = \tau^c_{p\Phi}$ and the non-centrality parameter for the LTS procedure is zero. This matches the result in Theorem 3.2. Note however that since $\mathsf{f}(c) = \varphi(c) \neq 0$ and $\tau^\infty_{2\mathsf{G}} \neq 1$ the non-centrality parameter for the RLS statistic is non-zero, so that it declares G as contamination.

Third, suppose G only has probability mass in the tails with zero probability for the interval $[-c, c]$ and $c > 1$. In this case $\tau^c_{p\mathsf{G}} = 0$ while $\mathsf{f}(c) = (1-\epsilon)\varphi(c)$ and $\tau^\infty_{2\mathsf{G}} > c^2 > 1$.

Thus, the non-centrality term is zero for the LTS statistic but non-zero for the RLS statistic. The conclusions are the same as in case two.

Finally, suppose G has a general form so that $\tau^c_{p\mathsf{G}}$ is neither zero nor $\tau^c_{p\Phi}$. In this case both tests will have power.

4 Simulations

For $i = 1, ..., n$ let $y_i = 1 + x_i + \varepsilon_i$ where x_i is scalar, $i.i.d.$ $\mathsf{N}(0, 1)$ and independent of ε_i. To illustrate the above results, we consider different models for ε_i. Throughout, we use a significance level of 5%. The number of replications is 10^6 when using the OLS procedure and 10^4 when using the computationally intensive LTS procedure. All simulations are run in Matlab. LTS is implemented using mlts.m (Argullo et al., 2008).

4.1 Size

We start by considering Theorem 3.1, where there is no contamination and the robustified least squares (RLS) procedure is used. Hence, in the first data generating process (DGP 1) ε_i is $i.i.d.\mathsf{N}(0, 1)$. The empirical size of the normality test is reported in the upper panel of Table 2. We consider sample sizes $n = \{50, 100, 200, 400, 800, 1600\}$ and cut-off values $c = \{0.67, 1.03, 1.96, 2.58, 3.29, 3.89, 4.42\}$. In small samples, the empirical size varies with the cut-off values, c, but approches the nominal value of 5% in larger samples. Overall, these results indicate that the test, when properly normalized using the standardizations derived in Theorem 3.1, has the expected size properties.

Next, we consider Theorem 3.2, where the errors are ϵ-tail contaminated normal and the LTS procedure is used. We start by choosing $\epsilon = 0$, so that there is actually no contamination and the errors are standard normal as in DGP 1. We use the LTS procedure with a trimming proportion, $\gamma = \{0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01\}$. This corresponds to retaining $h = n(1 - \gamma)$ observations in the LTS estimation. To avoid rounding in the implementation of the LTS procedure, we only report results when $h = n(1 - \gamma)$ is an integer and omit the combinations (0.05,50) and (0.01,50). The empirical size of the normality test is reported in the lower panel of Table 2. A pattern similar to the RLS procedure is observed, indicating that the size of the test is controlled, for large samples, when using the normalizing constants derived in Theorem 3.2.

With DGPs 2-8 and Table 3, we study the performance of the LTS procedure under ϵ-tail contamination, as analyzed in Theorem 3.2. We consider seven contamination proportions $\epsilon = \{0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01\}$. Let v_i be standard normal and η_i standard exponential. Let $\lambda_\epsilon = \{1 - \Phi(c_\epsilon)\}/\phi(c_\epsilon)$ with Φ and ϕ denoting the standard normal CDF and PDF, respectively. Then, the errors $\varepsilon_i = v_i 1_{(|v_i|<c_\epsilon)} + (c_\epsilon + \lambda_\epsilon \eta_i) sgn(v_i) 1_{(|v_i|>c_\epsilon)}$ have distribution function satisfying $\mathsf{P}(\varepsilon_i \leq v) = \Phi(v)$ for $|v| \leq c_\epsilon$ and $\mathsf{P}(\varepsilon_i \leq v) = 1 - \{1 - \Phi(v_\epsilon)\} \exp\{-(v - c_\epsilon)/\lambda\}$ for $v > c_\epsilon$. The density is normal in the centre and thus ϵ-tail contaminated so that Assumption 3.2(i) holds.

Empirical sizes of the LTS normality test are reported in Table 3. As expected, the empirical size approaches the nominal size as the sample size grows, supporting that the normalizing constants derived in Theorem 3.2 deliver χ^2 inference, also in the ϵ-tail

Table 2: Size of RLS & LTS procedures. DGP 1.

		n=50	100	200	400	800	1600
RLS(c)	0.67	0.142	0.109	0.084	0.068	0.059	0.055
	1.03	0.104	0.081	0.067	0.059	0.055	0.052
	1.96	0.057	0.054	0.053	0.051	0.050	0.050
	2.58	0.043	0.046	0.048	0.049	0.049	0.049
	3.29	0.044	0.043	0.045	0.047	0.048	0.048
	3.89	0.043	0.046	0.046	0.046	0.048	0.048
	4.42	0.038	0.044	0.046	0.047	0.048	0.048
LTS(γ)	0.5	0.085	0.069	0.058	0.045	0.046	0.049
	0.4	0.071	0.059	0.053	0.048	0.053	0.054
	0.3	0.070	0.055	0.056	0.052	0.047	0.050
	0.2	0.062	0.048	0.052	0.049	0.049	0.053
	0.1	0.052	0.050	0.049	0.053	0.050	0.048
	0.05		0.049	0.052	0.051	0.052	0.050
	0.01		0.051	0.051	0.051	0.052	0.050

ε_i is $i.i.d.\mathsf{N}(0,1)$. c is the cut-off value for the RLS procedure. γ is the trimming proportion for the LTS procedure.

contaminated case. Values of γ that trim the sample more than ϵ will deliver more favourable sizes. Here, we focus on the most stringent case where $\gamma = \epsilon$.

Table 3: Size of the LTS(γ) procedure. DGPs 2-8.

DGPs	γ	$n = 50$	100	200	400	800	1600
2	0.5	0.088	0.069	0.057	0.054	0.055	0.054
3	0.4	0.079	0.060	0.060	0.051	0.051	0.050
4	0.3	0.071	0.060	0.053	0.053	0.050	0.048
5	0.2	0.063	0.058	0.056	0.054	0.049	0.050
6	0.1	0.053	0.051	0.051	0.048	0.053	0.053
7	0.05		0.051	0.051	0.050	0.047	0.048
8	0.01		0.053	0.054	0.054	0.051	0.051

ε_i is i.i.d. ϵ-tail contaminated normal with $\epsilon = \gamma$.

Table 4: RLS(c) procedure. DGP 2.

DGPs	c	$n = 50$	100	200	400	800	1600	5000
2	0.67	0.139	0.110	0.089	0.080	0.082	0.098	0.193

ε_i is i.i.d. ϵ-tail contaminated normal with $\epsilon = 2\{1 - \Phi(c)\}$.

The power analysis in Section 3.4 revealed that the RLS procedure declares the ϵ-tail model as contamination. The simulations of DGP 2 reported in Table 4 confirm this.

4.2 Power

Previous studies have simulated the power of the standard cumulant normality test for full sample OLS residuals. For instance, Jarque and Bera (1987) considered Beta, Student's t, Gamma and Log-normal distributions while Thadewald and Büning (2007) considered the ϵ-contamination model of Huber (1964). Thus inspired, we consider the power of the RLS and LTS procedures for the following error distributions: Cauchy; the ϵ-contamination model; and the LTS model of Berenguer-Rico et al. (2023). We use the asymptotic critical values from Theorems 3.1 and 3.2.

4.2.1 Cauchy distribution

In DGP 9, ε_i is Cauchy distributed. Table 5 reports power results. The RLS procedure is very powerful even in small samples. In contrast, the empirical power of the LTS procedure depends highly on the trimming parameter γ. For $\gamma = 0.5$ or $\gamma = 0.4$, the procedure requires larger samples to achieve adequate levels of rejection frequencies. When $\gamma \leq 0.1$ more observations of the fat tails of the distribution are retained and the power is high in small samples.

Table 5: Power of RLS & LTS procedures. DGP 9.

		$n = 50$	100	200	400	800	1600
RLS(c)	0.67	0.914	0.994	1.000	1.000	1.000	1.000
	1.03	0.971	0.999	1.000	1.000	1.000	1.000
	1.96	0.989	1.000	1.000	1.000	1.000	1.000
	2.58	0.983	0.999	1.000	1.000	1.000	1.000
	3.29	0.973	0.999	1.000	1.000	1.000	1.000
	3.89	0.972	0.999	1.000	1.000	1.000	1.000
	4.42	0.975	0.999	1.000	1.000	1.000	1.000
LTS(γ)	0.5	0.106	0.110	0.156	0.275	0.519	0.845
	0.4	0.144	0.215	0.395	0.684	0.951	0.999
	0.3	0.284	0.497	0.808	0.981	1.000	1.000
	0.2	0.591	0.867	0.989	1.000	1.000	1.000
	0.1	0.904	0.993	1.000	1.000	1.000	1.000
	0.05		0.999	1.000	1.000	1.000	1.000
	0.01		1.000	1.000	1.000	1.000	1.000

ε_i is $i.i.d.\mathsf{t}_1$.

4.2.2 ϵ-contamination

In DGP 10, $\varepsilon_i \sim (1-\epsilon)\mathsf{N}(0,1) + \epsilon\mathsf{N}(2,9)$ with $\epsilon = 0.2$, so the errors are ϵ-contaminated in the sense of Huber (1964). Table 6 reports power results. First, as expected, the empirical power of the RLS procedure increases with sample size and cut-off, c. When $n = \{50, 100\}$ the procedure has low power for small values of c. In larger samples, say $n \geq 800$, the procedure attains empirical power of (nearly) one for all c. Second, the empirical power of the LTS procedure depends highly on the trimming parameter γ. For $\gamma = 0.5$, the empirical power remains low even when $n = 1600$. When $\gamma = 0.5$, the

Table 6: Power RLS & LTS procedures. DGP 10.

		$n = 50$	100	200	400	800	1600
RLS(c)	0.67	0.248	0.330	0.493	0.749	0.956	0.999
	1.03	0.389	0.583	0.825	0.978	0.999	1.000
	1.96	0.650	0.877	0.986	0.999	1.000	1.000
	2.58	0.708	0.922	0.994	1.000	1.000	1.000
	3.29	0.757	0.953	0.998	1.000	1.000	1.000
	3.89	0.804	0.971	0.999	1.000	1.000	1.000
	4.42	0.837	0.980	0.999	1.000	1.000	1.000
LTS(γ)	0.5	0.085	0.071	0.061	0.059	0.065	0.064
	0.4	0.080	0.069	0.066	0.067	0.085	0.114
	0.3	0.077	0.077	0.083	0.111	0.171	0.290
	0.2	0.113	0.131	0.190	0.308	0.535	0.820
	0.1	0.359	0.517	0.725	0.921	0.995	1.000
	0.05		0.867	0.979	0.999	1.000	1.000
	0.01		0.985	0.999	1.000	1.000	1.000

$\varepsilon_i \sim (1-\epsilon)\mathsf{N}(0,1) + \epsilon\mathsf{N}(2,9)$ with $\epsilon = 0.2$.

LTS procedure trims 50% of the sample although there is only $\epsilon = 20\%$ contamination. Hence, the low power. For $\gamma = 0.3$, the empirical power grows only slowly with the sample size. For smaller trimming proportions, $\gamma \leq 0.2$, the LTS procedure performs much better with empirical power close to one for $n = 100$ or larger.

4.2.3 LTS Model

In DGP 11, the error term follows the LTS model of Berenguer-Rico et al. (2023). This is a model where LTS is maximum likelihood. Errors satisfy the following structure. Let ζ be a set with $h = n\epsilon$ elements from $1, ..., n$ with $\epsilon = 0.8$. For $i \in \zeta$, let ε_i be *i.i.d.* $\mathsf{N}(0,1)$. For $j \notin \zeta$, let ξ_j be *i.i.d.* with distribution function $\mathsf{G_j}(x)$ for $x \in \mathbb{R}$ where G_j is continuous at 0. The outlier errors are

$$\varepsilon_j = (\max_{i \in \zeta} \varepsilon_i + \xi_j) 1_{(\xi_j > 0)} + (\min_{i \in \zeta} \varepsilon_i + \xi_j) 1_{(\xi_j < 0)}. \tag{4.1}$$

The LTS model differs from ϵ-tail contamination as introduced in §3.2. The ϵ-tail contamination model has i.i.d. errors and the uncontaminated part is truncated normal. The LTS models does not have i.i.d. errors due to the construction (4.1), but the uncontaminated part is (untruncated) normal.

To study the power, we set $\xi_j - \nu^+ 1_{(\xi_j > 0)} + \nu^- 1_{(\xi_j < 0)}$ to be i.i.d. $\mathsf{N}(0,1)$. We consider two cases. First, we let $\nu^+ = \nu^- = 0$ so that there is no separation between good and outlier observations. Second, we let $\nu^+ = 3$ and $\nu^- = -1$ to allow for separation.

Table 7 reports power results for DGP 11, when $\nu^+ = \nu^- = 0$. The RLS procedure is not very powerful in small samples but power tends to one as the sample size increases for all cut-off values c. The empirical power is low for the LTS procedure when $\gamma = 0.5$, even when $n = 1600$. For smaller values of γ, the power approaches one in the larger sample sizes considered.

Table 7: Power of RLS & LTS procedures. DGP 11.

		$n = 50$	100	200	400	800	1600
RLS(c)	0.67	0.186	0.204	0.278	0.479	0.812	0.990
	1.03	0.287	0.456	0.748	0.973	0.999	1.000
	1.96	0.384	0.807	0.993	1.000	1.000	1.000
	2.58	0.187	0.512	0.922	0.999	1.000	1.000
	3.29	0.100	0.234	0.558	0.943	0.999	1.000
	3.89	0.066	0.153	0.357	0.780	0.995	1.000
	4.42	0.057	0.133	0.308	0.705	0.989	1.000
LTS(γ)	0.5	0.084	0.078	0.069	0.072	0.092	0.138
	0.4	0.085	0.079	0.099	0.127	0.224	0.421
	0.3	0.105	0.143	0.239	0.435	0.732	0.960
	0.2	0.258	0.610	0.949	0.999	1.000	1.000
	0.1	0.307	0.787	0.993	1.000	1.000	1.000
	0.05		0.668	0.983	1.000	1.000	1.000
	0.01		0.320	0.778	0.996	1.000	1.000

$\varepsilon_i \sim$ LTS model with $h = n\epsilon$, $\epsilon = 0.8$. No separation: $\nu^+ = \nu^- = 0$

Table 8: Power of RLS & LTS procedures. DGP 11.

		$n = 50$	100	200	400	800	1600
RLS(c)	0.67	0.362	0.554	0.835	0.990	1.000	1.000
	1.03	0.620	0.880	0.993	1.000	1.000	1.000
	1.96	0.987	1.000	1.000	1.000	1.000	1.000
	2.58	0.973	0.999	1.000	1.000	1.000	1.000
	3.29	0.865	0.998	1.000	1.000	1.000	1.000
	3.89	0.742	0.985	1.000	1.000	1.000	1.000
	4.42	0.698	0.973	1.000	1.000	1.000	1.000
LTS(γ)	0.5	0.084	0.078	0.069	0.073	0.092	0.138
	0.4	0.088	0.079	0.100	0.127	0.224	0.421
	0.3	0.112	0.144	0.239	0.435	0.732	0.960
	0.2	0.505	0.823	0.986	0.999	1.000	1.000
	0.1	0.986	1.000	1.000	1.000	1.000	1.000
	0.05		1.000	1.000	1.000	1.000	1.000
	0.01		0.999	1.000	1.000	1.000	1.000

$\varepsilon_i \sim$ LTS model with $h = n\epsilon$, $\epsilon = 0.8$. Separation: $\nu^+ = 3$, $\nu^- = -1$.

Table 8 reports power results for DGP 11, when $\nu^+ = 3$ and $\nu^- = -1$. RLS is markedly more powerful with separation than without for all values of c, n. The LTS procedure is also notably more powerful for $\gamma \leq 0.2$. For most values of c, in the RLS case, or γ, in the LTS case, the power approaches one even in moderate sample sizes.

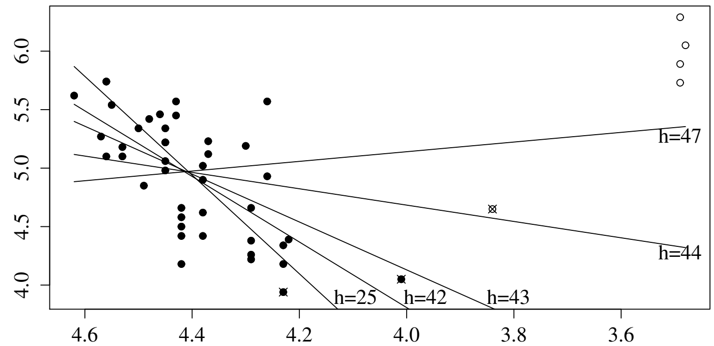

Figure 1: Star data and fit by LTS for different h. Log light intensity against log temperature. Bullets are estimated good observations for $h = 42$. Circle with cross is the F-star. Two bullets with crosses are declared outliers by RLS procedure.

5 Empirical illustration

We illustrate the test for truncated normality using the stars data of Rousseeuw and Leroy (1987, Table 2.3). For further discussion, see also Berenguer-Rico et al. (2023), BR23 henceforth. Figure 1 shows observations on log light intensity and log temperature for the Hertzsprung-Russell diagram of the star cluster CYG OB1 containing $n = 47$ stars. From the right, the first four stars are giant of M-type, the fifth star is of F-type, the next 31 stars (1 doublet) are of B-type, and the last 11 stars (1 doublet) are of O-type. We apply the suggested tests for truncated normality noting that the power will be low in a sample as small as this. Hence, detecting departures from the null requires strong evidence against truncated normal errors.

We start with the robustified least squares procedure, RLS. The initial least squares estimators are

$$\underset{\substack{(\mathsf{se}_{OLS}) \\ [t\text{−}stat_{OLS}]}}{log.light} = \underset{\substack{(1.21) \\ [5.61]}}{6.79} - \underset{\substack{(0.28) \\ [-1.48]}}{0.41}\ log.Te. \tag{5.1}$$

The full sample OLS estimation is influenced by the M-stars. Proceeding with a cut-off of $c = 1.96$, which is the normal 97.5% quantile, RLS declares that observations 14 and 17, marked with circles and crosses in Figure 1, are outliers. The RLS estimates are

$$\underset{\substack{(\mathsf{se}_{RLS}) \\ [t\text{−}stat_{RLS}]}}{log.light} = \underset{\substack{(1.44) \\ [5.08]}}{7.34} - \underset{\substack{(0.33) \\ [-1.58]}}{0.53}\ log.Te, \qquad T_{RLS}^{norm} = 4.83. \tag{5.2}$$

Thus, the RLS estimation remains influenced by the M-stars in line with the analysis of Welsh and Ronchetti (2002). The test statistic for truncated normality is asymptotically χ_2^2 with 90% quantile of 4.60 and 95% quantile of 5.99. Hence, the test $T_{RLS}^{norm} = 4.83$ rejects at the 10% significance level, showing some evidence against the null.

We now turn to the LTS procedure. Figure 1 shows lines fitted by LTS for different values of h. There is not much difference between the fits for $h = 25$ and $h = 42$. The slope starts turning from $h = 42$ onwards. The four M-stars are arguebly bad leverage points. The F-star may also be an outlier, but can have a masking effect (BR23).

For inference, we will refer to two models, both depending on the choice of h. The *truncated normal model* is an i.i.d. ϵ-tail contaminated model where the central h observations are truncated normal. In the *LTS model*, the central h observations are untruncated i.i.d. normal (BR23). These models require different scale estimators. We let $\hat{\sigma}_{trunc}$ be the scale estimator in the ϵ-tail contaminated model and $\hat{\sigma}_{LTS}$ the scale estimator in the *LTS model*. More specifically, $\hat{\sigma}_{trunc}$ is the standard LTS estimates with a consistency correction as in (2.3), while $\hat{\sigma}_{LTS}$ has no consistency correction.

				Trunc. model		LTS model	
h	ψ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{trunc}$	T^{norm}_{trunc}	$\hat{\sigma}_{LTS}$	T^{norm}_{LTS}
25	0.50000	-13.62	4.22	0.48	2.45	0.18	1.72
36	0.75000	-11.49	3.71	0.46	1.97	0.27	1.98
37	0.77274	-9.00	3.16	0.46	3.60	0.28	2.49
40	0.84092	-8.58	3.07	0.45	2.44	0.31	2.13
41	0.83365	-8.50	3.05	0.46	2.35	0.33	1.26
42	0.88638	-7.40	2.80	0.49	5.82	0.37	0.39
43	0.90910	-4.06	2.05	0.51	0.52	0.40	0.69
44	0.93183	1.89	0.70	0.59	2.78	0.49	0.49
45	0.95456	7.34	-0.53	0.60	5.54	0.51	2.94
46	0.97728	6.92	-0.44	0.59	4.99	0.53	2.74
47	1.00000	6.79	-0.41	0.56	3.40	0.55	2.75

Table 9: Estimates by LTS for the full sample.

				Trunc. model		LTS model	
h	ψ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{trunc}$	T^{norm}_{trunc}	$\hat{\sigma}_{LTS}$	T^{norm}_{LTS}
25	0.52274	-13.62	4.22	0.48	3.11	0.18	1.72
36	0.77274	-11.49	3.71	0.45	2.12	0.27	1.98
37	0.79546	-9.00	3.16	0.44	3.78	0.28	2.49
40	0.86365	-8.58	3.07	0.43	2.67	0.31	2.13
41	0.88637	-8.50	3.05	0.44	1.04	0.33	1.26
42	0.90910	-7.40	2.80	0.47	4.57	0.37	0.39
43	0.93183	7.88	-0.65	0.60	4.79	0.49	2.57
44	0.95456	7.74	-0.62	0.59	5.20	0.51	2.76
45	0.97728	7.58	-0.59	0.59	4.99	0.53	2.73
46	1.00000	7.12	-0.49	0.56	3.45	0.55	2.83

Table 10: Estimates by LTS for the sample excluding the F-star.

Table 9 shows the estimated coefficients when fitting LTS for different h values. Two test statistics are reported, T^{norm}_{trunc} and T^{norm}_{LTS}. Both combine third and fourth residual cumulants. T^{norm}_{trunc} uses the new normalizations for the truncated normal model. T^{norm}_{LTS} has the standard normalizations and tests for untruncated normality of the good errors in the LTS model. Both test statistics are asymptotically χ^2_2 with 90% quantile of 4.60 and 95% quantile of 5.99. It should be noted that T^{norm}_{LTS} has not been analyzed formally under the LTS model. The test statistics should be interpreted in a pointwise fashion.

Table 10 is applied to the sample where the F-star is removed. Otherwise, it has the same structure as Table 9. This is to disentangle the masking effect of the F star already pointed out in BR23.

We need to assume a model to conduct inference. Given the doublets, the data are not consistent with the assumption of a continuous distribution. Most likely, the doublets arise from rounding, so we disregard this point. Given the graphical evidence in Figure 1 of having four potential outliers in this data set, the M-stars, we start by considering a truncated normal model with $h = 43$. The LTS estimator declares, precisely, the four M-stars as outliers in this case. The test for truncated normality in Table 9 with $h = 43$ is $T_{trunc}^{norm} = 0.52$, therefore, it does not reject. Removing the F-star from the sample alters this conclusion. The test based on Table 10 with $h = 42$ gives $T_{trunc}^{norm} = 4.57$, which rejects the null hypothesis at the 10% significance level, giving some evidence against the null. This suggests a masking effect of the F-star. Including the possibly outlying F-star as good introduces noise and may explain these differences.

Given these results, next we consider a truncated normal model with $h = 42$ leaving the four M-stars and the F-star as outliers. Table 9 has $T_{trunc}^{norm} = 5.82$. Again, it rejects the null hypothesis of truncated normality at the 10% significance level. The test statistic is actually very close to the critical value at 5% significance level, showing stronger evidence against the null.

Finally, we consider an (untruncated) LTS normal model with $h = 42$. Tables 9, 10 both have $T_{LTS}^{norm} = 0.39$, so that normality cannot be rejected. Moreover, BR23 suggest that for LTS location-scale models, h can be estimated consistently by minimizing T_{LTS}^{norm} over h. Both tables have $h = 42$ as minimizer. This conclusion is clearest in Table 10 and somewhat fragile in Table 9, possibly due to a masking effect of the F-star.

Overall, there is some evidence against the two truncated normal models, whereas the LTS model cannot be rejected. With $h = 42$, the estimated truncated normal model and LTS model along with both sets of standard errors and t-statistics are

$$\underset{\substack{(\mathsf{se}_{LTS})/[t-stat_{LTS}] \\ (\mathsf{se}_{trunc})/[t-stat_{trunc}]}}{log.light} = - \underset{\substack{(2.09)/[-3.54] \\ (3.43)/[-2.16]}}{7.40} + \underset{\substack{(0.48)/[5.09] \\ (0.78)/[3.59]}}{2.80} \; log.Te. \tag{5.3}$$

Going along with the suggestion that the tests for normality and truncated normality give more confidence in the LTS model than the ϵ-tail contamination model, we should favour the smaller standard errors and larger t-statistic marked LTS, which gives more confidendence that the slope is significant than those marked trunc.

6 Discussion

Conducting inference on the unknown parameters of regression models when accounting for the presence of outliers requires knowledge of the distributional properties of the data at hand. Normality of the good errors is often considered in practice. Yet, the good errors could be truncated normal, as implicitly assumed by standard practice when using the LTS estimator of Rousseeuw (1984), or untruncated normal as in the LTS model of Berenguer-Rico et al. (2023); Berenguer-Rico and Nielsen (2022). Test statistics that deliver valid inference differ in each model. Hence, assessing which model better describes a given dataset is key in applied work. We have derived a test for truncated normality of the good errors that delivers standard χ^2 inference. We have applied the test statistic to the stars data of Rousseeuw and Leroy (1987, Table 2.3) and found some evidence against truncated normality and in favour of untruncated normal good errors.

A Empirical processes

A.1 The main empirical process results

We are interested in the weighted and marked empirical distribution functions, for $c > 0$,

$$\hat{\mathsf{G}}_n^{q,p}(c) = n^{-1}\textstyle\sum_{i=1}^n (n^{1/2}x_{in})^{\otimes q}(\hat{\varepsilon}_i/\hat{\sigma})^p 1_{(|\tilde{\varepsilon}_i/\tilde{\sigma}|\le c)}, \tag{A.1}$$

where $v^{\otimes 0} = 1$, $v^{\otimes 1} = v$, $v^{\otimes 2} = vv'$ for the vector $v = n^{1/2}x_{in}$. We refer to $w_{in} = (n^{1/2}x_{in})^{\otimes q}$ as the *weight* and to $(\hat{\varepsilon}_i/\hat{\sigma})^p$ as the *mark*. We will use the (q,p)-combinations

$$\mathcal{Q} = \{(0,0),(0,2),(0,3),(0,4),(1,1)\} \quad \text{as well as} \quad (2,0). \tag{A.2}$$

Define the normalized estimation errors $\tilde{a} = n^{1/2}(\tilde{\sigma}-\sigma)/\sigma$ and $\tilde{b} = N^{-1}(\tilde{\beta}-\beta)/\sigma$, so that $x_i'(\tilde{\beta}-\beta) = x_{in}'\tilde{b}\sigma$. Similarly, define $\hat{a} = n^{1/2}(\hat{\sigma}-\sigma)/\sigma$ and $\hat{b} = N^{-1}(\hat{\beta}-\beta)/\sigma$. The standardized residuals satisfy

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

Let $\tilde{\theta} = (\tilde{a},\tilde{b},\hat{a},\hat{b})$ and $\mathsf{G}_n^{q,p}(\tilde{\theta},c) = \hat{\mathsf{G}}_n^{q,p}(c)$. When analyzing $\mathsf{G}_n^{q,p}(\tilde{\theta},c)$, we can replace $\tilde{\theta}$ with deterministic values $\theta = (a_1,b_1,a_p,b_p)$ varying in some set due to the next result. Subscripts indicate association with indicator or mark.

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

Proof. Intersect the set $\{|\mathsf{G}_n(\tilde{\theta},c)| > \epsilon\}$ with the set $(\tilde{\theta}\in\Theta)$ and its complement. $\square$

The processes of interest are therefore, with $w_{in} = (n^{1/2}x_{in})^{\otimes q}$ and $\varepsilon_i^\sigma = \varepsilon_i/\sigma$,

$$\mathsf{G}_n^{q,p}(\theta,c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\left(\frac{\varepsilon_i^\sigma - x_{in}'b_p}{1+n^{-1/2}a_p}\right)^p 1_{(|\varepsilon_i^\sigma - x_{in}'b_1|\le c+n^{-1/2}a_1c)}, \tag{A.3}$$

$$\overline{\mathsf{G}}_n^{q,p}(\theta,c) = n^{-1}\textstyle\sum_{i=1}^n w_{in}\mathsf{E}_{i-1}\left(\frac{\varepsilon_i^\sigma - x_{in}'b_p}{1+n^{-1/2}a_p}\right)^p 1_{(|\varepsilon_i^\sigma - x_{in}'b_1|\le c+n^{-1/2}a_1c)}, \tag{A.4}$$

where E_{i-1} is the $\mathcal{F}_{i-1,n}$ conditional expectation. The weights w_{in} are $\mathcal{F}_{i-1,n}$ adapted. In particular, using (2.4), we have

$$\overline{\mathsf{G}}_n^{0,p}(0,c) = \mathsf{E}(\varepsilon_i^\sigma)^p 1_{(|\varepsilon_i^\sigma|\le c)} = \tau_p^c. \tag{A.5}$$

Next, define the empirical process

$$\mathbb{G}_n^{q,p}(\theta,c) = n^{1/2}\{\mathsf{G}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(\theta,c)\}, \tag{A.6}$$

which is a martingale. Define also the bias terms

$$\mathcal{G}_{1n}^{q,p}(\theta,c) = 2c^p\varphi(c)n^{-1}\textstyle\sum_{i=1}^n(n^{1/2}x_{in})^{\otimes q}\{1_{(p\text{ even})}ca_1 + 1_{(p\text{ odd})}n^{1/2}x_{in}'b_1\}, \tag{A.7}$$
$$\mathcal{G}_{mn}^{q,p}(\theta,c) = pn^{-1}\textstyle\sum_{i=1}^n(n^{1/2}x_{in})^{\otimes q}\{1_{(p\text{ even})}\tau_p^c a_p + 1_{(p\text{ odd})}\tau_{p-1}^c n^{1/2}x_{in}'b_p\}. \tag{A.8}$$

The asymptotic analysis requires the next assumption. Remark A.1 below outlines how part (i) can be relaxed.

Assumption A.1. *Suppose Assumption 2.1 and*
(i) $\varepsilon_i^\sigma = \varepsilon_i/\sigma$ *are* ϵ*-tail contaminated normal and* $0 < c < c_\epsilon$*,*
(ii) $\max_{1\le i\le n}\mathsf{E}|n^{1/2}x_{in}|^{2+\kappa} = \mathrm{O}(1)$ *for some* $\kappa > 0$*.*

We will need the following asymptotic results.

Theorem A.2. *Suppose Assumption A.1(ii). Let* $q = 0,1,2$ *and* $p = 0,1,\dots,4$*. Then*
(a) $\mathbb{G}_n^{q,p}(0,c) = \mathrm{o}_\mathsf{P}(n^{1/2})$*;*
(b) $\overline{\mathsf{G}}_n^{q,p}(0,c) = n^{-1}\sum_{i=1}^n w_{in}\tau_p^c = \mathrm{O}_\mathsf{P}(1)$*.*

Theorem A.3. *Suppose Assumption A.1, then for all* $B > 0$*,*
(a) $\sup_{|\theta_1|<B}|\mathsf{G}_n^{2,0}(\theta_1,c) - \mathsf{G}_n^{2,0}(0,c)| = \mathrm{o}_\mathsf{P}(1)$*;*
(b) $\sup_{|\theta_1|<B}|\overline{\mathsf{G}}_n^{2,0}(\theta_1,c) - \overline{\mathsf{G}}_n^{2,0}(0,c)| = \mathrm{o}_\mathsf{P}(1)$*.*

Theorem A.4. *Suppose Assumption A.1. Let* $(q,p)\in\mathcal{Q}$*. Then,* $\forall B > 0$*,*
(a) $\sup_{|\theta|\le B}|\mathbb{G}_n^{q,p}(\theta,c) - \mathbb{G}_n^{q,p}(0,c)| = \mathrm{o}_\mathsf{P}(1)$*,*
(b) $\sup_{|\theta_1|\le B}|n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} - \{\mathcal{G}_{1n}^{q,p}(\theta,c) - \mathcal{G}_{mn}^{q,p}(\theta,c)\}| = \mathrm{o}_\mathsf{P}(1)$*.*

To see the usefulness of Theorem A.4 decompose

$$\begin{aligned} n^{1/2}\{\mathsf{G}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} &= n^{1/2}\{\mathsf{G}_n^{q,p}(0,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} \\ &\quad + \mathbb{G}_n^{q,p}(\theta,c) - \mathbb{G}_n^{q,p}(0,c) + n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\}. \end{aligned} \tag{A.9}$$

Combining the two statements of Theorem A.4, we get

$$n^{1/2}\{\mathsf{G}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} = n^{1/2}\{\mathsf{G}_n^{q,p}(0,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} + \{\mathcal{G}_{1n}^{q,p}(\theta,c) - \mathcal{G}_{mn}^{q,p}(\theta,c)\} + \mathrm{o}_\mathsf{P}(1). \tag{A.10}$$

In turn, normalizing and applying Theorem A.2, we get

$$\mathsf{G}_n^{q,p}(\theta,c) = \overline{\mathsf{G}}_n^{q,p}(0,c) + \mathrm{o}_\mathsf{P}(1). \tag{A.11}$$

In the case of LTS estimation, the cut-off is the order statistics $\tilde{\xi}_{(h)}$. We will show that $\tilde{\xi}_{(h)}$ is consistent for σc for h being the largest integer not exceeding $n\{\Phi(c) - \Phi(-c)\}$. We can always write $\tilde{\xi}_{(h)}/\sigma = c + n^{-1/2}\tilde{d}$ where $\tilde{d} = n^{1/2}\{\tilde{\xi}_{(h)}/\sigma - c\}$. In our analysis, the cut-off c is fixed. It is therefore equivalent to think of the estimation uncertainty in the order statistic as a scale estimation error since $\tilde{\xi}_{(h)}/\sigma = c(1 + n^{-1/2}\tilde{d}/c)$. Thus, introduce the notation $c_d = c + n^{-1/2}d$ and $\theta_d = (d/c, 0, 0, 0)$ to get $\overline{\mathsf{G}}_n^{q,p}(0,c_d) = \overline{\mathsf{G}}_n^{q,p}(\theta_d,c)$ and $\mathsf{G}_n^{q,p}(\theta,c_d) = \mathsf{G}_n^{q,p}(\theta+\theta_d,c)$. The uncertainty d will show up in the bias term $\mathcal{G}_{1n}^{q,p}$, but not in $\mathcal{G}^{q,p}_{mn}$. This results in the following expansions, uniformly in $|\theta|, |d| \leq B$. First, Theorem A.4(a) gives

$$n^{1/2}\{\overline{\mathsf{G}}^{q,p}_n(0, c_d) - \overline{\mathsf{G}}^{q,p}_n(0, c)\} = \mathcal{G}^{q,p}_{1n}(\theta_d; c) + \mathrm{o}_{\mathsf{P}}(1); \tag{A.12}$$

Next, the expansions (A.10), (A.11) imply

$$\begin{aligned} n^{1/2}\{\mathsf{G}^{q,p}_n(\theta, c_d) - \overline{\mathsf{G}}^{q,p}_n(0, c)\} &= n^{1/2}\{\mathsf{G}^{q,p}_n(0, c) - \overline{\mathsf{G}}^{q,p}_n(0, c)\} \\ &\quad + \{\mathcal{G}^{q,p}_{1n}(\theta + \theta_d, c) - \mathcal{G}^{q,p}_{mn}(\theta, c)\} + \mathrm{o}_{\mathsf{P}}(1); \qquad \text{(A.13)} \\ \mathsf{G}^{q,p}_n(\theta, c_d) &= \overline{\mathsf{G}}^{q,p}_n(0, c) + \mathrm{o}_{\mathsf{P}}(1). \qquad \text{(A.14)} \end{aligned}$$

A.2 Preliminary Lemmas

The following lemmas are useful in proving the main empirical processes results.

Lemma A.5. *Suppose* $\max_{1\leq i\leq n} \mathsf{E}|n^{1/2}x_{in}|^{2+\kappa} = \mathrm{O}(1)$ *for some* $\kappa > 0$. *Define the sets* $\mathcal{D}_i = (|n^{1/2}x_{in}| \leq n^{\lambda})$ *where* $1/(2+\kappa) < \lambda < 1/2$. *Let* $v_{in}(\theta_1)$ *be random variables. Then, for all* $\epsilon > 0$ *and large* n,

$$\mathsf{P}\Big\{\sup_{|\theta_1|\leq B}\Big|\sum_{i=1}^n v_{in}(\theta_1)\Big| > \epsilon\Big\} \leq \mathsf{P}\Big\{\sup_{|\theta_1|\leq B}\Big|\sum_{i=1}^n v_{in}(\theta_1)1_{\mathcal{D}_i}\Big| > \epsilon\Big\} + \epsilon.$$

Proof. Let $\mathcal{A} = \{\sup_{|\theta_1|\leq B}|\sum_{i=1}^n v_{in}(\theta_1)| > \epsilon\}$ and define $\mathcal{D} = \cap_{i=1}^n \mathcal{D}_i$, so that

$$\mathsf{P}(\mathcal{A}) = \mathsf{P}(\mathcal{A}\cap\mathcal{D}) + \mathsf{P}(\mathcal{A}\cap\mathcal{D}^c) \leq \mathsf{P}(\mathcal{A}\cap\mathcal{D}) + \mathsf{P}(\mathcal{D}^c). \tag{A.15}$$

We find $\mathsf{P}(\mathcal{D}^c)$. Note that $\mathcal{D}^c = \cup_{i=1}^n \mathcal{D}_i^c$. By Boole's and Markov's inequalities

$$\mathsf{P}(\mathcal{D}^c) = \mathsf{P}\{\bigcup_{i=1}^n (|n^{1/2}x_{in}| > n^{\lambda})\} \leq \sum_{i=1}^n \mathsf{P}(|n^{1/2}x_{in}| > n^{\lambda}) \leq n^{-\lambda(2+\kappa)}\sum_{i=1}^n \mathsf{E}|n^{1/2}x_{in}|^{2+\kappa}.$$

Taking maximum over the summands gives $\mathsf{P}(\mathcal{D}^c) \leq n^{1-\lambda(2+\kappa)}\max_{1\leq i\leq n}\mathsf{E}|n^{1/2}x_{in}|^{2+\kappa}$. Since the maximum of expectations is assumed bounded while $\lambda > 1/(2+\kappa)$, we get $\mathsf{P}(\mathcal{D}^c) \to 0$. Thus, $\mathsf{P}(\mathcal{D}^c) \leq \epsilon$ for large n. Insert this in (A.15). Rewrite $(\mathcal{A}\cap\mathcal{D}) = \{\sup_{|\theta_1|\leq B}|\sum_{i=1}^n v_{in}(\theta_1)|1_{\mathcal{D}} > \epsilon\}$. As $\mathcal{D} = \cap_{i=1}^n \mathcal{D}_i$ then $\mathcal{D} \subset \mathcal{D}_i$ for all i. Thus, $(\mathcal{A}\cap\mathcal{D}) \subset \{\sup_{|\theta_1|\leq B}|\sum_{i=1}^n v_{in}(\theta_1)1_{\mathcal{D}_i}| > \epsilon\}$. Insert in (A.15). □

Lemma A.6. *Let* $I_i(\theta_1) = 1_{(|\varepsilon_i^{\sigma} - x_{in}'b_1| \leq c + cn^{-1/2}a_1)} - 1_{(|\varepsilon_i^{\sigma}|\leq c)}$ *so that* $I_i(0) = 0$. *Let*

$$\begin{aligned} J_i(\theta_0, \theta_1) = {} & 1_{\{c - n^{-1/2}a_0c - s_{in}(\theta_0,\theta_1) \leq \varepsilon_i^{\sigma} - x_{in}'b_0 \leq c + n^{-1/2}a_0c + s_{in}(\theta_0,\theta_1)\}} \\ & + 1_{\{-c - n^{-1/2}a_0c - s_{in}(\theta_0,\theta_1) \leq \varepsilon_i^{\sigma} - x_{in}'b_0 \leq -c + n^{-1/2}a_0c + s_{in}(\theta_0,\theta_1)\}}, \end{aligned}$$

where $s_{in}(\theta_0, \theta_1) = cn^{-1/2}|a_1 - a_0| + |x_{in}||b_1 - b_0|$. *Then*

$$|I_i(\theta_1) - I_i(\theta_0)| \leq J_i(\theta_0, \theta_1), \tag{A.16}$$

Proof. The second indicator in the definition of I_i cancels when taking difference. Thus,

$$I_i(\theta_1) - I_i(\theta_0) = \{1_{(\varepsilon_i^\sigma \le c + n^{-1/2}a_1 c + x_{in}' b_1)} - 1_{(\varepsilon_i^\sigma \le c + n^{-1/2}a_0 c + x_{in}' b_0)}\} \\ - \{1_{(\varepsilon_i^\sigma \le -c - n^{-1/2}a_1 c + x_{in}' b_1)} - 1_{(\varepsilon_i^\sigma \le -c - n^{-1/2}a_0 c + x_{in}' b_0)}\}.$$

For the first term, we note that $c + n^{-1/2}a_1 c + x_{in}' b_1$ is located in the interval with midpoint $c + n^{-1/2}a_0 c + x_{in}' b_0$ and radius $s_{in}(\theta_0, \theta_1)$. Thus, the first term is bounded in absolute value by the indicator on that interval. This is the first term of $J_i(\theta_0, \theta_1)$. The second term is bounded in a similar fashion. □

Lemma A.7. *Suppose* F *is* ϵ*-tail contaminated normal with* $0 < c < c_\epsilon$*. Let* $p = 0, 1, 2, \ldots, 8$*. Let* $s = c$ *or* $s = -c$*. Then,* $\forall 0 < \delta < c_\epsilon - c$*,* $\exists C > 0$*,* $\forall s - \delta \le s_1 \le s_2 \le s + \delta$*, we have*
(a) $\mathsf{E}_{i-1}|\varepsilon_i^\sigma|^p 1_{(s_1 \le \varepsilon_i^\sigma \le s_2)} \le (s_2 - s_1)C$*;*
(b) $|\mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p\{1_{(\varepsilon_i^\sigma \le s_1)} - 1_{(\varepsilon_i^\sigma \le s)}\} - (s_1 - s)s^p\varphi(s)| \le (s_1 - s)^2 C$*.*

Remark A.1. *Lemma A.7 has the only argument requiring Assumption A.1(i) that* F *is* ϵ*-tail contaminated normal and* $0 < c < c_\epsilon$*. Inspection of the proof shows that it suffices that* $(1 + |s|^8)\mathsf{f}(s)$ *is locally bounded and Lipschitz in neighbourhoods of* c *and* $-c$*, see also Berenguer-Rico et al. (2019, Lemma A.6). Under those conditions the result holds with the normal density* φ *replaced by* f *in part (ii).*

Proof. Part (a). For a function H with derivative h, the mean value theorem gives $\mathsf{H}(s_2) = \mathsf{H}(s_1) + (s_2 - s_1)\mathsf{h}(s^*)$ for an intermediate point $s_1 \le s_* \le s_2$. Thus

$$\mathsf{E}_{i-1}|\varepsilon_i^\sigma|^p 1_{(s_1 \le \varepsilon_i^\sigma \le s_2)} = \int_{s_1}^{s_2} |v|^p \mathsf{f}(v)dv = (s_2 - s_1)|s_*|^p \mathsf{f}(s_*).$$

Since $s - \delta \le s_* \le s + \delta$ we can take f as the normal density φ. We bound $|s_*|^p \mathsf{f}(s_*) \le \sup_{s-\delta \le s_* \le s+\delta} |s_*|^p \varphi(s_*) < \infty$.

Part (b). Follow the same steps and apply a second order mean value theorem, so that $\mathsf{H}(s_1) = \mathsf{H}(s) + (s_1 - s)\mathsf{h}(s) + (1/2)(s_1 - s)^2\dot{\mathsf{h}}(s_*)$. Use that for the normal density, $|s_*|^p \mathsf{f}(s_*)$ has a bounded derivative locally around s. □

Lemma A.8. *Let* $\max_{1 \le i \le n} \mathsf{E}|n^{1/2}x_{in}|^2 = \mathrm{O}(1)$*. Let* $\mathcal{D}_i = (|n^{1/2}x_{in}| \le n^\lambda)$ *for* $\lambda < 1/2$*. Recall* $w_{in} = (n^{1/2}x_{in})^{\otimes q}$*. Then*
(a) $\mathsf{E}n^{-1}\sum_{i=1}^n (1 + |n^{1/2}x_{in}|^2) = \mathrm{O}(1)$*;*
(b) $\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}| = \mathrm{O}(1)$ *for* $q = 0, 1, 2$*;*
(c) $\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|^2 = \mathrm{O}(1)$ *for* $q = 0, 1$*;*
(d) $\mathsf{E}n^{-2}\sum_{i=1}^n |w_{in}|^2 1_{\mathcal{D}_i} = \mathrm{o}(1)$ *for* $q = 0, 1, 2$*.*

Proof. Part (a). Swap expectation and summation and take maximum over expectations to bound $\mathsf{E}n^{-1}\sum_{i=1}^n (1 + |n^{1/2}x_{in}|^2) \le 1 + \max_{1 \le i \le n} \mathsf{E}|n^{1/2}x_{in}|^2$, which is bounded.

Part (b). For $q = 0, 1, 2$, we get that $|w_{in}| \le 2(1 + |n^{1/2}x_{in}|^2)$. Apply part (a).

Part (c). For $q = 0, 1$, we get that $|w_{in}|^2 \le 2(1 + |n^{1/2}x_{in}|^2)$. Apply part (a).

Part (d). For $q = 0, 1, 2$, we get $|w_{in}|1_{\mathcal{D}_i} \le 2(1 + |n^{1/2}x_{in}|^2)1_{\mathcal{D}_i} \le Cn^{2\lambda}$. Thus, we can bound $\mathsf{E}n^{-2}\sum_{i=1}^n |w_{in}|^2 1_{\mathcal{D}_i} \le Cn^{2\lambda - 1}\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|$. This vanishes as $2\lambda < 1$ and the expecation is bounded by part (b). □

A.3 Proofs of empirical process results

Proof of Theorem A.2. Part (a). Let $n^{-1/2}\mathbb{G}_n^{q,p}(0,c) = n^{-1}\sum_{i=1}^n w_{in}v_i$ with summands given by $v_i = (\varepsilon_i^\sigma)^p 1_{(|\varepsilon_i^\sigma|\le c)} - \tau_p^c$. This is a martingale

Lemma A.5 using Assumption A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$ shows that it suffices that the martingale $n^{-1}\sum_{i=1}^n w_{in}v_i 1_{\mathcal{D}_i}$ vanishes.

The Chebyshev inequality shows that it suffices that $\mathcal{E} = \mathsf{E}|n^{-1}\sum_{i=1}^n w_{in}v_i 1_{\mathcal{D}_i}|^2$ vanishes. By the martingale property, $\mathcal{E} = n^{-2}\sum_{i=1}^n \mathsf{E}|w_{in}|^2 v_i^2 1_{\mathcal{D}_i}$. Apply the law of iterated expectations and note $\mathsf{E}_{i-1}v_i^2$ is constant and finite by the truncation. Thus, $\mathcal{E} \le C\mathsf{E}n^{-2}\sum_{i=1}^n |w_{in}|^2 1_{\mathcal{D}_i}$, which vanishes by Lemma A.8(d) using Assumption A.1(ii).

Part (b). The identity $\overline{\mathsf{G}}_n^{q,p}(0,c) = n^{-1}\sum_{i=1}^n w_{in}\tau_p^c$ follows from (A.4), (A.5).

Lemma A.8(b) using Assumption A.1(ii), shows that $\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|$ is bounded. □

Proof of Theorem A.3. Let $V_n(\theta_1) = \mathsf{G}_n^{2,0}(\theta_1,c) - \mathsf{G}_n^{2,0}(0,c) = \sum_{i=1}^n v_{in}(\theta_1)$ and $\overline{V}_n(\theta_1) = \overline{\mathsf{G}}_n^{2,0}(\theta_1,c) - \overline{\mathsf{G}}_n^{2,0}(0,c) = \sum_{i=1}^n \mathsf{E}_{i-1}v_{in}(\theta_1)$ with $v_{in}(\theta_1) = n^{-1}nx_{in}x_{in}'I_i(\theta_1)$ and $I_i(\theta_1) = 1_{(|\varepsilon_i/\sigma - x_{in}'b_1|\le c+cn^{-1/2}a_1)} - 1_{(|\varepsilon_i/\sigma|\le c)}$. We need to show that $V_n(\theta_1)$ and $\overline{V}_n(\theta_1)$ vanish uniformly in $|\theta_1| \le B$. Throughout, $C > 0$ denotes a generic constant.

Apply Lemma A.5 using Assumption A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$. It suffices to show $V_n^{\mathcal{D}}(\theta_1) = \sum_{i=1}^n |v_{in}(\theta_1)|1_{\mathcal{D}_i}$ and $\overline{V}_n^{\mathcal{D}}(\theta_1) = \sum_{i=1}^n \mathsf{E}_{i-1}|v_{in}(\theta_1)|1_{\mathcal{D}_i}$ vanish uniformly. We will find a bound $|v_{in}(\theta_1)|1_{\mathcal{D}_i} \le v_{in}$ uniformly in θ_1. Thus, $\mathsf{E}\sup_{\theta_1} V_n^{\mathcal{D}}(\theta_1)$ and $\mathsf{E}\sup_{\theta_1} \overline{V}_n^{\mathcal{D}}(\theta_1)$ are both bounded by $\mathsf{E}\sum_{i=1}^n v_{in} = \mathsf{E}\sum_{i=1}^n \mathsf{E}_{i-1}v_{in}$, which we will show to be vanishing.

By Lemma A.6 with $\theta_0 = 0$ and defining $s_{in}(\theta_1) = n^{-1/2}|a_1|c + |x_{in}||b_1|$, we have

$$|I_i(\theta_1)| \le J_i(\theta_1) = 1_{\{c-s_{in}(\theta_1)\le\varepsilon_i^\sigma\le c+s_{in}(\theta_1)\}} + 1_{\{-c-s_{in}(\theta_1)\le\varepsilon_i^\sigma\le -c+s_{in}(\theta_1)\}}.$$

On $\mathcal{D}_i$ we have that $|x_{in}| < n^{\lambda-1/2}$ with $\lambda < 1/2$. Since $|\theta_1| \le B$, c is fixed, we get $s_{in}(\theta_1) \le Cn^{\lambda-1/2} = s_n$. Having exploited $\mathcal{D}_i$, we then bound $1_{\mathcal{D}_i} \le 1$ to get

$$|I_i(\theta_1)|1_{\mathcal{D}_i} \le J_i = 1_{(c-s_n\le\varepsilon_i^\sigma\le c+s_n)} + 1_{(-c-s_n\le\varepsilon_i^\sigma\le -c+s_n)},$$

uniformly in θ_1. Thus, $|v_{in}(\theta_1)|1_{\mathcal{D}_i} \le n^{-1}|n^{1/2}x_{in}|^2 J_i = v_{in}$, uniformly in θ_1. Now, apply Lemma A.7 using Assumption A.1(i) to get $\mathsf{E}_{i-1}J_i \le Cs_n = Cn^{\lambda-1/2}$. In turn, we find that $\mathsf{E}\sum_{i=1}^n \mathsf{E}_{i-1}v_{in} \le Cn^{\lambda-1/2}\mathsf{E}n^{-1}\sum_{i=1}^n |n^{-1/2}x_{in}|^2$ vanishes since the expectation is bounded by Lemma A.8(a) using Assumption A.1(ii) while $\lambda < 1/2$. □

Theorem A.4 compares the empirical process and the compensator at θ and 0. We introduce an intermediate point $\theta_1 = (a_1, b_1, 0, 0)$ representing the situation with estimation error in the indicator but not in the mark and $\theta_p = (0, 0, a_p, b_p)$ representing the situation with estimation error in the mark but not in the indicator. We decompose

$$\mathbb{G}_n^{q,p}(\theta,c) - \mathbb{G}_n^{q,p}(0,c) = \{\mathbb{G}_n^{q,p}(\theta,c) - \mathbb{G}_n^{q,p}(\theta_1,c)\} + \{\mathbb{G}_n^{q,p}(\theta_1,c) - \mathbb{G}_n^{q,p}(0,c)\}. \quad (A.17)$$

We analyze the two terms in (A.17) separately. For the compensator term in Theorem A.4, we decompose

$$\begin{aligned} &n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} - \{\mathcal{G}_{1n}^{q,p}(\theta,c) - \mathcal{G}_{mn}^{q,p}(\theta,c)\} \\ &= n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta,c) - \overline{\mathsf{G}}_n^{q,p}(\theta_1,c)\} + \mathcal{G}_{mn}^{q,p}(\theta_p,c) + n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta_1,c) - \overline{\mathsf{G}}_n^{q,p}(0,c)\} - \mathcal{G}_{1n}^{q,p}(\theta_1,c). \end{aligned} \quad (A.18)$$

As with (A.17), we analyze the compensator comparing θ to θ_1 and the one comparing θ_1 to 0 in (A.18) separately.

Lemma A.9. *Suppose Assumption A.1. Let $q = 0, 1$ and $p = 0, 1, \ldots, 4$. Then, $\forall B > 0$ and for $\theta_1 = (a_1, b_1, 0, 0)$,*
(a) $\sup_{|\theta_1|\leq B} |n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta_1, c) - \overline{\mathsf{G}}_n^{q,p}(0, c)\} - \mathcal{G}_{1n}^{q,p}(\theta_1, c)| = \mathrm{o}_{\mathsf{P}}(1)$;
(b) $\sup_{|\theta_1|\leq B} |\mathbb{G}_n^{q,p}(\theta_1, c) - \mathbb{G}_n^{q,p}(0, c)| = \mathrm{o}_{\mathsf{P}}(1)$.

The proof adapts that of Theorem 1.17 of Johansen and Nielsen (2009). More general results that are also uniform in the cut-off c are given by Johansen and Nielsen (2016a), Jiao and Nielsen (2017), Berenguer-Rico et al. (2019).

Proof. Let $I_i(\theta_1) = 1_{(|\varepsilon_i^\sigma - x_{in}'b_1|\leq c+n^{-1/2}a_1c)} - 1_{(|\varepsilon_i^\sigma|\leq c)}$, while $C > 0$ is a generic constant.

Part (*a*). We show that $\overline{V}_n(\theta_1) = n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta_1, c) - \overline{\mathsf{G}}_n^{q,p}(0, c)\} - \mathcal{G}_{1n}^{q,p}(\theta_1, c)$ vanishes uniformly in $|\theta_1| \leq B$. Write $\overline{V}_n(\theta_1) = \sum_{i=1}^n \overline{v}_{in}(\theta_1)$ with summands given by $\overline{v}_{in}(\theta_1) = n^{-1/2}w_{in}[\mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p I_i(\theta_1) - 2c^p\varphi(c)\{n^{-1/2}a_1c1_{(p \text{ even})} + x_{in}'b_1 1_{(p \text{ odd})}\}]$.

Apply Lemma A.5 using Assumption A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$. We show $\overline{V}_n^{\mathcal{D}}(\theta_1) = \sum_{i=1}^n \overline{v}_{in}(\theta_1)1_{\mathcal{D}_i}$ vanishes uniformly in $|\theta_1| \leq B$.

Consider $\mathcal{E}_i = \mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p I_i(\theta_1)1_{\mathcal{D}_i}$. Write $I_i(\theta_1) = J_{i1}(\theta_1) - J_{i2}(\theta_1)$ where $J_{i1}(\theta_1) = 1_{(\varepsilon_i^\sigma\leq c+n^{-1/2}a_1c+x_{in}'b_1)} - 1_{(\varepsilon_i^\sigma\leq c)}$ and $J_{i2}(\theta_1) = 1_{(\varepsilon_i^\sigma\leq -c-n^{-1/2}a_1c+x_{in}'b_1)} - 1_{(\varepsilon_i^\sigma\leq -c)}$. Since $|\theta_1| \leq B$, c is fixed, and on $\mathcal{D}_i$ we have that $|x_{in}| < n^{\lambda-1/2}$, then $n^{-1/2}|a_1|c+|x_{in}||b_1| \leq Cn^{\lambda-1/2}$ for $\lambda < 1/2$. Lemma A.7 using Assumption A.1(i) then gives

$$\mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p J_{i1}(\theta_1)1_{\mathcal{D}_i} = (n^{-1/2}a_1c + x_{in}'b_1)c^p\varphi(c)1_{\mathcal{D}_i} + R_{i1}(\theta_1)1_{\mathcal{D}_i} \tag{A.19}$$
$$\mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p J_{i2}(\theta_1)1_{\mathcal{D}_i} = (-n^{-1/2}a_1c + x_{in}'b_1)(-c)^p\varphi(-c)1_{\mathcal{D}_i} + R_{i2}(\theta_1)1_{\mathcal{D}_i} \tag{A.20}$$

where $R_{ij}(\theta_1)1_{\mathcal{D}_i} \leq C(n^{-1/2}|a_1|c + |x_{in}||b_1|)^2$. We now collect the first order terms on the right hand side of (A.19), (A.20). We note that the normal density is symmetric so that $\varphi(c) = \varphi(-c)$ and write $(-c)^p = c^p\{1_{(p \text{ even})} - 1_{(p \text{ odd})}\}$. This gives

$$(n^{-1/2}a_1c + x_{in}'b_1)c^p\varphi(c) - (-n^{-1/2}a_1c + x_{in}'b_1)(-c)^p\varphi(-c)$$
$$= 2c^p\varphi(c)\{n^{-1/2}a_1c1_{(p \text{ even})} + x_{in}'b_1 1_{(p \text{ odd})}\},$$

which matches the bias term in $\overline{v}_{in}(\theta_1)$. Thus, we can bound

$$|\overline{V}_n^{\mathcal{D}}(\theta_1)| = |\sum_{i=1}^n \overline{v}_{in}(\theta_1)1_{\mathcal{D}_i}| \leq n^{-1/2}\sum_{i=1}^n |w_{in}|(|R_{i1}(\theta_1)| + |R_{i2}(\theta_1)|)1_{\mathcal{D}_i}.$$

We bound the sum of remainder terms. For $q = 0, 1$, then $|w_{in}| \leq (1 + |n^{1/2}x_{in}|)$, so that $|w_{in}| \leq Cn^\lambda$ on $\mathcal{D}_i$. By the Jensen inequality and the construction $|a|, |b| \leq B$, then $|R_{ij}(\theta_1)| \leq Cn^{-1}(1+|n^{1/2}x_{in}|^2)$. Thus, $|\overline{V}_n^{\mathcal{D}}(\theta_1)| \leq Cn^{\lambda-1/2}n^{-1}\sum_{i=1}^n(1+|n^{1/2}x_{in}|^2)$. This vanishes since the average is bounded in expectation by Lemma A.8(a) using Assumption A.1(ii) while $\lambda < 1/2$.

Part (*b*). Consider $\tilde{V}_n(\theta_1) = \mathbb{G}_n^{q,p}(\theta_1, c) - \mathbb{G}_n^{q,p}(0, c) = \sum_{i=1}^n \tilde{v}_{in}(\theta_1)$ with summands $\tilde{v}_{in}(\theta_1) = n^{-1/2}w_{in}\{(\varepsilon_i^\sigma)^p I_i(\theta_1) - \mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p I_i(\theta_1)\}$. Apply Lemma A.5 using Assumption

A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$. We show $\tilde{V}_n^{\mathcal{D}}(\theta_1) = \sum_{i=1}^n \tilde{v}_{in}(\theta_1)1_{\mathcal{D}_i}$ vanishes uniformly in $|\theta_1| \leq B$.

To tackle the uniformity in θ_1, we use the following chaining argument and inequality. Given a small $\epsilon > 0$, we can choose a (small) radius of size M according to (A.26) below and cover the set $|\theta_1| \leq B$ with a finite number, K, of balls with centres θ_{1k} for $k = 1, ..., K$. The balls are given by

$$B_k = (\theta_1 : |\theta_1 - \theta_{1k}| \leq M, |\theta_1| \leq B).$$

The chaining inequality uses that any θ_1 belongs to some ball with index k. Thus,

$$\begin{aligned} |\tilde{V}_n^{\mathcal{D}}(\theta_1)| &\leq |\tilde{V}_n^{\mathcal{D}}(\theta_{1k})| + |\tilde{V}_n^{\mathcal{D}}(\theta_1) - \tilde{V}_n^{\mathcal{D}}(\theta_{1k})| \\ &\leq \max_k |\tilde{V}_n^{\mathcal{D}}(\theta_{1k})| + \max_k \sup_{\theta_1 \in B_k} |\tilde{V}_n^{\mathcal{D}}(\theta_1) - \tilde{V}_n^{\mathcal{D}}(\theta_{1k})|. \quad \text{(A.21)} \end{aligned}$$

The term $\max_k |\tilde{V}_n^{\mathcal{D}}(\theta_{1k})|$ in (A.21). We show $\mathcal{P}_n = \mathsf{P}\{\max_k |\tilde{V}_n^{\mathcal{D}}(\theta_{1k})| \geq \epsilon\} \to 0$, for any $\epsilon > 0$. Here max is a union of events. The Boole and Chebyshev inequalities give

$$\mathcal{P}_n = \mathsf{P}\bigcup_{k=1}^K \{|\tilde{V}_n^{\mathcal{D}}(\theta_{1k})| > \epsilon\} \leq \sum_{k=1}^K \mathsf{P}\{|\tilde{V}_n^{\mathcal{D}}(\theta_{1k})| > \epsilon\} \leq \frac{1}{\epsilon^2}\sum_{k=1}^K \mathsf{E}|\tilde{V}_n^{\mathcal{D}}(\theta_{1k})|^2. \quad \text{(A.22)}$$

Here, $\tilde{V}_n^{\mathcal{D}}(\theta_{1k})$ is a scalar for $q = 0$ and a vector for $q = 1$. Moreover, it is a sum of martingale differences $\tilde{v}_{in}^{\mathcal{D}}(\theta_{1k}) = \tilde{v}_{in}(\theta_{1k})1_{\mathcal{D}_i}$ and thus a sum of mean zero and uncorrelated terms. Therefore, by iterated expectations,

$$\mathsf{E}|\tilde{V}_n^{\mathcal{D}}(\theta_{1k})|^2 = \sum_{i=1}^n \mathsf{E}|\tilde{v}_{in}^{\mathcal{D}}(\theta_{1k})|^2 = \sum_{i=1}^n \mathsf{E}\mathsf{E}_{i-1}|\tilde{v}_{in}^{\mathcal{D}}(\theta_{1k})|^2. \quad \text{(A.23)}$$

Using the definition of $\tilde{v}_{in}^{\mathcal{D}}(\theta_{1k})$, we find

$$\begin{aligned} \mathsf{E}_{i-1}|\tilde{v}_{in}^{\mathcal{D}}(\theta_{1k})|^2 &= n^{-1}|w_{in}|^2 1_{\mathcal{D}_i}\mathsf{E}_{i-1}\{(\varepsilon_i^\sigma)^p I_i(a_{1k}, b_{1k}) - \mathsf{E}_{i-1}(\varepsilon_i^\sigma)^p I_i(a_{1k}, b_{1k})\}^2 \\ &\leq n^{-1}|w_{in}|^2 1_{\mathcal{D}_i}\mathsf{E}_{i-1}[(\varepsilon_i^\sigma)^{2p}\{I_i(a_{1k}, b_{1k})\}^2]. \quad \text{(A.24)} \end{aligned}$$

Lemma A.6 with $\theta_0 = 0$ shows that $|I_i(a_{1k}, b_{1k})| \leq J_i(\theta_{1k}) = 1_{\{|\varepsilon_i^\sigma - c| \leq s_{in}(\theta_{1k})\}} + 1_{\{|\varepsilon_i^\sigma + c| \leq s_{in}(\theta_{1k})\}}$ with $s_{in}(\theta_{1k}) = cn^{-1/2}|a_{1k}| + |x_{in}||b_{1k}|$. Since $|\theta_{1k}| \leq B$, c fixed and on $\mathcal{D}_i$, we have that $s_{in}(\theta_{1k}) \leq Cn^{\lambda-1/2}$, uniformly in θ_{1k}. The Jensen inequality shows $\{I_i(a_{1k}, b_{1k})\}^2 \leq 2J_i(\theta_{1k})$. Lemma A.7$(a)$ using Assumption A.1(i) then shows

$$\mathsf{E}_{i-1}[(\varepsilon_i^\sigma)^{2p}\{I_i(a_{1k}, b_{1k})\}^2] \leq Cn^{\lambda-1/2} \quad \text{(A.25)}$$

Insert (A.25) in (A.24), (A.23), (A.22) to get $\mathcal{P}_n \leq \epsilon^{-2}KCn^{\lambda-1/2}\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|^2 1_{\mathcal{D}_i}$, which vanishes since ϵ and K are fixed, $\lambda < 1/2$ and the expectation is bounded by Lemma A.8(c) using Assumption A.1(ii).

The term $\max_k \sup_{\theta_1 \in B_k} |Z_n(\theta_{1k}, \theta_1)|$ in (A.21) where $Z_n(\theta_{1k}, \theta_1) = \tilde{V}_n^{\mathcal{D}}(\theta_1) - \tilde{V}_n^{\mathcal{D}}(\theta_{1k})$. and write $Z_n(\theta_{1k}, \theta_1) = \sum_{i=1}^n \{z_{in}(\theta_{1k}, \theta_1) - \mathsf{E}_{i-1}z_{in}(\theta_{1k}, \theta_1)\}$ with summands

$$z_{in}(\theta_{1k}, \theta_1) = n^{-1/2}w_{in}(\varepsilon_i^\sigma)^p\{I_i(a_1, b_1) - I_i(a_{1k}, b_{1k})\}.$$

Apply Lemma A.5 using Assumption A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$. We show $Z_n^{\mathcal{D}}(\theta_{1k},\theta_1) = \sum_{i=1}^n \{z_{in}^{\mathcal{D}}(\theta_{1k},\theta_1) - \mathsf{E}_{i-1}z_{in}^{\mathcal{D}}(\theta_{1k},\theta_1)\}$ vanishes uniformly in θ_{1k},θ_1, where $z_{in}^{\mathcal{D}}(\theta_{1k},\theta_1) = z_{in}(\theta_{1k},\theta_1)1_{\mathcal{D}_i}$. By Lemma A.6, then $|z_{in}^{\mathcal{D}}(\theta_{1k},\theta_1)| \leq n^{-1/2}|w_{in}||\varepsilon_i/\sigma|^p J_i(\theta_{ik},\theta_1)1_{\mathcal{D}_i}$, where

$$J_i(\theta_{ik},\theta_1) = 1_{\{c-n^{-1/2}a_{1k}c-s_{in}(\theta_{1k},\theta_1)\leq\varepsilon_i^\sigma - x_{in}'b_{1k}\leq c+n^{-1/2}a_{1k}c+s_{in}(\theta_{1k},\theta_1)\}} + 1_{\{-c-n^{-1/2}a_{1k}c-s_{in}(\theta_{1k},\theta_1)\leq\varepsilon_i^\sigma - x_{in}'b_{1k}\leq -c+n^{-1/2}a_{1k}c+s_{in}(\theta_{1k},\theta_1)\}}$$

with $s_{in}(\theta_{1k},\theta_1) \leq n^{-1/2}|a_{1k}-a_1|c+|x_{in}||b_{1k}-b_1|$. Since $|\theta_{1k}-\theta_1| \leq M$, c fixed and on $\mathcal{D}_i$, we have that $s_{in}(\theta_{1k},\theta) \leq s_{in}$ uniformly in θ_{1k},θ_1, where $s_{in} = Cn^{-1/2}M(1+|n^{1/2}x_{in}|)$, Thus, $J_i(\theta_{ik},\theta_1) \leq J_{ik}$, where

$$J_{ik} = 1_{(c-n^{-1/2}a_{1k}c-s_{in}\leq\varepsilon_i^\sigma - x_{in}'b_{1k}\leq c+n^{-1/2}a_{1k}c+s_{in})} + 1_{(-c-n^{-1/2}a_{1k}c-s_{in}\leq\varepsilon_i^\sigma - x_{in}'b_{1k}\leq -c+n^{-1/2}a_{1k}c+s_{in})},$$

uniformly in $\theta_1 \in \mathcal{B}_k$. We then get $|z_{in}^{\mathcal{D}}(\theta_{1k},\theta_1)| \leq z_{ik}^J = n^{-1/2}|w_{in}||\varepsilon_i^\sigma|^p J_{ik}1_{\mathcal{D}_i}$. By the triangle inequality

$$Z_n^{\mathcal{D}}(\theta_{1k},\theta_1) \leq \sum_{i=1}^n (z_{ik}^J + \mathsf{E}_{i-1}z_{ik}^J) = \sum_{i=1}^n (z_{ik}^J - \mathsf{E}_{i-1}z_{ik}^J) + \sum_{i=1}^n \mathsf{E}_{i-1}z_{ik}^J = \tilde{Z}_{nk}^J + \overline{Z}_{nk}^J,$$

say. It suffices to show that each of $\tilde{Z}_{nk}^J$ and $\overline{Z}_{nk}^J$ vanishes uniformly in k.

The term $\overline{Z}_{nk}^J$. On $\mathcal{D}_i$, then $s_{in} \leq Cn^{1/2-\lambda}$, which vanishes uniformly in k. Thus, Lemma A.7(a) using Assumption A.1(i) shows that $\mathsf{E}_{i-1}z_{ik}^J \leq Cn^{-1/2}|w_{in}|1_{\mathcal{D}_i}s_{in}$. The weight w_{in} is 1 or $n^{1/2}x_{in}$ so that $|w_{in}| \leq 1+|n^{1/2}x_{in}|$. Then the Jensen inequality shows $|w_{in}|s_{in}1_{\mathcal{D}_i} \leq CMn^{-1/2}(1+|n^{1/2}x_{in}|^2)$ and we get $\mathsf{E}_{i-1}z_{ik}^J \leq CMn^{-1}(1+|n^{1/2}x_{in}|^2)$. Thus, $\overline{Z}_{nk}^J \leq CMn^{-1}\sum_{i=1}^n(1+|n^{1/2}x_{in}|^2)$ uniformly in k. The Markov inequality shows that

$$\mathsf{P}(\max_k \overline{Z}_{nk}^J > \epsilon) \leq \frac{1}{\epsilon}\mathsf{E}\max_k \overline{Z}_{nk}^J \leq \frac{CM}{\epsilon}\mathsf{E}n^{-1}\sum_{i=1}^n(1+|n^{1/2}x_{in}|^2) < \epsilon, \tag{A.26}$$

since the expectation is bounded by Lemma A.8(a) using Assumption A.1(ii) and since, for given $\epsilon > 0$, we can choose M freely.

The term $\tilde{Z}_{nk}^J$. We show $\mathcal{P}_Z = \mathsf{P}\{\max_k |\tilde{Z}_{nk}^J| \geq \epsilon\} \to 0$ for an $\epsilon > 0$. As in (A.22), write $\max_k$ as a union then use Boole's and Chebyshev's inequalities to get

$$\mathcal{P}_Z = \mathsf{P}\bigcup_{k=1}^K(|\tilde{Z}_{nk}^J| \geq \epsilon) \leq \sum_{k=1}^K \mathsf{P}(|\tilde{Z}_{nk}^J| \geq \epsilon) \leq \frac{1}{\epsilon^2}\sum_{k=1}^K \mathsf{E}(\tilde{Z}_{nk}^J)^2. \tag{A.27}$$

We note that $\tilde{Z}_{nk}^J = \sum_{i=1}^n(z_{ik}^J - \mathsf{E}_{i-1}z_{ik}^J)$ is a martingale with $z_{ik}^J = n^{-1/2}|w_{in}||\varepsilon_i^\sigma|^p J_{ik}1_{\mathcal{D}_i}$. Thus it has uncorrelated summands, which shows

$$\mathsf{E}(\tilde{Z}_{nk}^J)^2 = \sum_{i=1}^n \mathsf{E}(z_{ik}^J)^2 = n^{-1}\sum_{i=1}^n \mathsf{E}n\mathsf{E}_{i-1}(z_{ik}^J)^2. \tag{A.28}$$

We proceed as for $\overline{Z}^J_{nk}$. Note that $J^2_{ik} \leq 2J_{ik}$ by the Jensen inequality. Thus, Lemma A.7(a). using Assumption A.1(i) shows that $n\mathsf{E}_{i-1}(z^J_{ik})^2 \leq C|w_{in}|^2 1_{\mathcal{D}_i} s_{in}$. As before, $s_{in}1_{\mathcal{D}_i} \leq Cn^{\lambda-1/2}$. Thus, $n\mathsf{E}_{i-1}|z^J_{nk}|^2 \leq Cn^{\lambda-1/2}|w_{in}|^2$. Insert in (A.28), (A.27), to get $\mathcal{P}_Z\epsilon^{-2}KCn^{\lambda-1/2}\mathsf{E}n^{-1}\sum_{i=1}^n |w_{in}|^2$. This vanishes for ϵ, K, since $\lambda < 1/2$ and the expectation is bounded by Lemma A.8(c) using Assumption A.1(ii). $\square$

For the first term in (A.17) with mark estimation error, we need a further result.

Lemma A.10. *Suppose Assumption A.1. Let* $(q,p) \in \mathcal{Q}$. *Then,* $\forall B > 0$ *and for* $\theta = (a_1, b_1, a_m, b_m)$ *and* $\theta_1 = (a_1, b_1, 0, 0)$,
(a) $\sup_{|\theta|\leq B} |\mathbb{G}^{q,p}_n(\theta, c) - \mathbb{G}^{q,p}_n(\theta_1, c)| = \mathsf{o}_\mathsf{P}(1)$;
(b) $\sup_{|\theta|\leq B} |n^{1/2}\{\overline{\mathsf{G}}^{q,p}_n(\theta, c) - \overline{\mathsf{G}}^{q,p}_n(\theta_1, c)\} + \mathcal{G}^{q,p}_{mn}(\theta, c)| = \mathsf{o}_\mathsf{P}(1)$.

Proof. Notation. Let $v_i(\theta) = \{(\varepsilon^{ab\sigma}_i)^p - (\varepsilon^\sigma_i)^p\}1_{(|\varepsilon^\sigma_i - x'_{in}b_1| \leq c + n^{-1/2}a_1 c)}$ where $\varepsilon^\sigma_i = \varepsilon_i/\sigma$ and $\varepsilon^{ab\sigma}_i = (\varepsilon_i/\sigma - x'_{in}b_m)/(1 + n^{-1/2}a_m)$. Define a^*_m and b^*_m through

$$\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i = \frac{\varepsilon^\sigma_i - x'_{in}b_m}{1 + n^{-1/2}a_m} - \varepsilon^\sigma_i = \frac{-n^{-1/2}a_m}{1 + n^{-1/2}a_m}\varepsilon^\sigma_i - \frac{x'_{in}b_m}{1 + n^{-1/2}a_m} = n^{-1/2}a^*_m\varepsilon^\sigma_i + x'_{in}b^*_m. \tag{A.29}$$

Note that given a $B > 0$ a $B^* > 0$ exists so that $|a^*_m|, |b^*_m| \leq B^*$ for $|a_m|, |b_m| \leq B$.

Finally, the mean value theorem with $|\varepsilon^{\sigma*}_i - \varepsilon^\sigma_i| \leq |\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i|$ shows that

$$(\varepsilon^{ab\sigma}_i)^p - (\varepsilon^\sigma_i)^p = 1_{(p\geq 1)}p(\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i)(\varepsilon^\sigma_i)^{p-1} + 1_{(p\geq 2)}\tfrac{1}{2}p(p-1)(\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i)^2(\varepsilon^{\sigma*}_i)^{p-2}. \tag{A.30}$$

Part (a). We must show that $\mathcal{V}_n(\theta) = n^{1/2}\{\mathbb{G}^{q,p}_n(\theta, c) - \mathbb{G}^{q,p}_n(\theta_1, c)\}$ vanishes uniformly in θ. We have $\mathcal{V}_n(\theta) = n^{-1/2}\sum_{i=1}^n w_{in}\{v_i(\theta) - \mathsf{E}_{i-1}v_i(\theta)\}$.

Decomposition. Using the above expansions write $v_i(\theta) = \sum_{s=1}^2 v_{si}(\theta)$ where

$$v_{1i}(\theta) = 1_{(p\geq 1)}p(\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i)(\varepsilon^\sigma_i)^{p-1}1_{(|\varepsilon^\sigma_i - x'_{in}b_1| \leq c + n^{-1/2}a_1 c)},$$
$$v_{2i}(\theta) = 1_{(p\geq 2)}\tfrac{1}{2}p(p-1)(\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i)^2(\varepsilon^{\sigma*}_i)^{p-2}1_{(|\varepsilon^\sigma_i - x'_{in}b_1| \leq c + n^{-1/2}a_1 c)}.$$

Let $\mathcal{V}_n(\theta) = \sum_{s=1}^2 \mathcal{V}_{sn}(\theta)$ with $\mathcal{V}_{sn}(\theta) = n^{-1/2}\sum_{i=1}^n w_{in}\{v_{si}(\theta) - \mathsf{E}_{i-1}v_{si}(\theta)\}$. By the triangle inequality, it suffices to show that each $\mathcal{V}_{sn}$ is $\mathsf{o}_\mathsf{P}(1)$ uniformly in θ.

The term $\mathcal{V}_{1n}(\theta)$. Since $\varepsilon^{ab\sigma}_i - \varepsilon^\sigma_i = n^{-1/2}a^*_m\varepsilon^\sigma_i + x'_{in}b^*_m$ by (A.29), we can write

$$\mathcal{V}_{1n}(\theta) = n^{-1/2}p\{\mathbb{G}^{q,p}_n(\theta_1, c)a^*_m + \mathbb{G}^{q+1,p-1}_n(\theta_1, c)b^*_m\}$$

for $p \geq 1$. We argue that $\mathbb{G}_n(\theta_1, c) = \mathbb{G}_n(0, c) + \mathsf{o}_\mathsf{P}(n^{1/2})$ uniformly in θ_1. Apply Theorem A.3 for $(q+1, p-1) = (2, 0)$ and Lemma A.9(b) for all other cases. Both use Assumption A.1. Theorem A.2 using Assumption A.1(ii) shows $\mathbb{G}^{q,\ell}_n(0, c) = \mathsf{o}_\mathsf{P}(n^{1/2})$. Thus, $\mathcal{V}_{1n}(\theta)$. vanishes due to the factor $n^{-1/2}$ and since $|a^*_m|, |b^*_m| \leq B^*$.

The term $\mathcal{V}_{2n}(\theta)$. Since $p \geq 2$ then $q = 0$ for all $(q,p) \in \mathcal{Q}$, see (A.2), thus we can set $w_{in} = 1$ and $\mathcal{V}_{2n}(\theta) = n^{-1/2}\sum_{i=1}^n\{v_{2i}(\theta) - \mathsf{E}_{i-1}v_{3i}(\theta)\}$. Apply Lemma A.5 using Assumption A.1(ii) and where $\mathcal{D}_i = (|n^{1/2}x_{in}| < n^\lambda)$ with $1/(2+\kappa) < \lambda < 1/2$. We show $\mathcal{V}^{\mathcal{D}}_{2n}(\theta) = n^{-1/2}\sum_{i=1}^n\{v_{2i}(\theta) - \mathsf{E}_{i-1}v_{2i}(\theta)\}1_{\mathcal{D}_i}$ vanishes.

By the triangle inquality, $|\mathcal{V}^{\mathcal{D}}_{2n}(\theta)| \leq n^{-1/2}\sum_{i=1}^n\{|v_{2i}(\theta)| + \mathsf{E}_{i-1}|v_{2i}(\theta)|\}1_{\mathcal{D}_i}$.

We bound $v_{2i}(\theta)$. By Jensen's inequality $|\varepsilon_i^{\sigma *}|^{p-2} \leq C(|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}|^{p-2} + |\varepsilon_i^{\sigma}|^{p-2})$. Thus,

$$|v_{2i}(\theta)| \leq C|\varepsilon_i^{ab\sigma} - \varepsilon_i^{\sigma}|^2(|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}|^{p-2} + |\varepsilon_i^{\sigma}|^{p-2})1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)}.$$

By (A.29) then $|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}| \leq |\varepsilon_i^{ab\sigma} - \varepsilon_i^{\sigma}| \leq n^{-1/2}|a_m^*||\varepsilon_i^{\sigma}| + |x_{in}||b_m^*|$. Here, $|a_m^*|, |b_m^*| < B^*$, so that $|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}| \leq Cn^{-1/2}(1+|n^{1/2}x_{in}|)(1+|\varepsilon_i^{\sigma}|)$. We need two further bounds. First, by the Jensen inequality, $|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}|^2 \leq Cn^{-1}(1+|n^{1/2}x_{in}|^2)(1+|\varepsilon_i^{\sigma}|^2)$. Second, $|x_{in}| \leq n^{\lambda - 1/2}$ on $\mathcal{D}_i$ where $\lambda < 1/2$, so that $|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}| \leq n^{\lambda - 1/2}C \leq 1$ for large n. In combination

$$|v_{2i}(\theta)|1_{\mathcal{D}_i} \leq Cn^{-1}(1 + |n^{1/2}x_{in}|^2)(1 + |\varepsilon_i^{\sigma}|^2)(1 + |\varepsilon_i^{\sigma}|^{p-2})1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)}1_{\mathcal{D}_i}$$

By Jensen's inequality, $(1 + |\varepsilon_i^{\sigma}|^2)(1 + |\varepsilon_i^{\sigma}|^{p-2}) \leq C(1 + |\varepsilon_i^{\sigma}|^p)$. Further, on $\mathcal{D}_i$ we have $n^{-1/2}|a_1^*|c + |x_{in}||b_1^*| \leq Cn^{\lambda - 1/2}$, so that $1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)} \leq 1_{(|\varepsilon_i^{\sigma}| \leq c + Cn^{\lambda - 1/2})}$. Finally, we bound $1_{\mathcal{D}_i} \leq 1$. Thus

$$\sup_{|\theta| \leq B^*} |v_{2i}(\theta)|1_{\mathcal{D}_i} \leq Cn^{-1}(1 + |n^{1/2}x_{in}|^2)(1 + |\varepsilon_i^{\sigma}|^p)1_{(|\varepsilon_i^{\sigma}| \leq c + Cn^{\lambda - 1/2})}.$$

Take conditional expectation, apply Lemma A.7 using Assumption A.1(i). We get

$$\mathsf{E}_{i-1} \sup_{|\theta| \leq B^*} |v_{2i}(\theta)|1_{\mathcal{D}_i} \leq Cn^{-1}(1 + |n^{1/2}x_{in}|^2).$$

Return to the sum and bound

$$\mathsf{E} \sup_{|\theta| \leq B^*} |\mathcal{V}_{2i}(\theta)| \leq \mathsf{E}n^{-1/2} \sum_{i=1}^{n} \mathsf{E}_{i-1} \sup_{|\theta| \leq B^*} |v_{2i}(\theta)|1_{\mathcal{D}_i} \leq Cn^{-1/2}\mathsf{E}n^{-1} \sum_{i=1}^{n}(1 + |n^{1/2}x_{in}|^2),$$

which vanishes as the expecation is bounded by Lemma A.8(a) with Assumption A.1(ii).

Part (b). *Decomposition*. We show $\overline{\mathcal{V}}_n(\theta) = n^{1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta, c) - \overline{\mathsf{G}}_n^{q,p}(\theta_1, c)\} + \mathcal{G}_{mn}^{q,p}(\theta, c)$ vanishes uniformly in θ. Use (A.5), (A.8) and note $\tau_p^c = 0$ when p is odd and $\tau_{p-1}^c = 0$ when p is even and write $\overline{\mathcal{V}}_n(\theta) = n^{-1/2} \sum_{i=1}^n w_{in}\overline{v}_i(\theta)$ where

$$\overline{v}_i(\theta) = \mathsf{E}_{i-1}\big[\{(\varepsilon_i^{ab\sigma})^p - (\varepsilon_i^{\sigma})^p\}1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)} + s_i p(\varepsilon_i^{\sigma})^{p-1}1_{(|\varepsilon_i^{\sigma}| \leq c)}\big],$$

with $s_i = n^{-1/2}a_m\varepsilon_i^{\sigma} + x_{in}' b_m$. Apply the expansion (A.30) to $v_i(\theta)$ and add and subtract $s_i p(\varepsilon_i^{\sigma})^{p-1}1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)}$ to get $\overline{v}_i(\theta) = \sum_{s=1}^3 \overline{v}_{si}(\theta)$ where

$$\begin{aligned}
\overline{v}_{1i}(\theta) &= -\mathsf{E}_{i-1}\big[s_i p(\varepsilon_i^{\sigma})^{p-1}\big\{1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)} - 1_{(|\varepsilon_i^{\sigma}| \leq c)}\big\}\big], \\
\overline{v}_{2i}(\theta) &= 1_{(p \geq 1)}\mathsf{E}_{i-1}\big\{\big(\varepsilon_i^{ab\sigma} - \varepsilon_i^{\sigma} + s_i\big)p(\varepsilon_i^{\sigma})^{p-1}1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)}\big\}, \\
\overline{v}_{3i}(\theta) &= 1_{(p \geq 2)}\tfrac{1}{2}p(p-1)\mathsf{E}_{i-1}(\varepsilon_i^{ab\sigma} - \varepsilon_i^{\sigma})^2(\varepsilon_i^{\sigma *})^{p-2}1_{(|\varepsilon_i^{\sigma} - x_{in}' b_1| \leq c + n^{-1/2} a_1 c)},
\end{aligned}$$

for $|\varepsilon_i^{\sigma *} - \varepsilon_i^{\sigma}| \leq |\varepsilon_i^{ab\sigma} - \varepsilon_i^{\sigma}|$. We analyze the terms $\overline{\mathcal{V}}_{sn}(\theta) = n^{-1/2} \sum_{i=1}^n w_{in}\overline{v}_{si}(\theta)$ in turn.

The term $\overline{\mathcal{V}}_{1n}(\theta)$. We note $p \geq 1$ so that $q \leq 1$. Thus, by the definition of s_i, we have $\overline{\mathcal{V}}_{1n}(\theta) = \{\overline{\mathsf{G}}_n^{q,p}(\theta_1, c) - \overline{\mathsf{G}}_n^{q,p}(0, c)\}a_m + \{\overline{\mathsf{G}}_n^{q+1,p-1}(\theta_1, c) - \overline{\mathsf{G}}_n^{q+1,p-1}(0, c)\}b_m$. We find that $\overline{\mathsf{G}}_n(0, c) = \mathsf{o_P}(1)$ uniformly in θ_1 by applying Theorem A.3(b) for $(q+1, p-1) = (2, 0)$ and Lemma A.9(b) for all other cases. Both use Assumption A.1. Since $|\theta| \leq B$, we find that $\overline{\mathcal{V}}_{1n}(\theta)$ vanishes.

The term $\overline{\mathcal{V}}_{2n}(\theta)$. Expand using (A.29) to get

$$\varepsilon_i^{ab\sigma} - \varepsilon_i^\sigma + s_i = n^{-1/2}\{\varepsilon_i^\sigma(a_m^* + a_m) + n^{1/2}x'_{in}(b_m^* + b_m)\}.$$

From (A.29), we get $a_m^* = -a_m/(1+n^{-1/2}a_m)$ so that $a_m^*+a_m = -n^{-1/2}a_m^* a_m$. Similarly $b_m^* + b_m = -n^{-1/2}a_m^* b_m$. Thus, $\varepsilon_i^{ab\sigma} - \varepsilon_i^\sigma + s_i = -n^{-1}\{\varepsilon_i^\sigma a_m + n^{1/2}x'_{in}b_m\}a_m^*$.

The sum of interest is $\overline{\mathcal{V}}_{2n}(\theta) = -n^{-1/2}\{\overline{\mathsf{G}}_n^{q,p}(\theta_1, c)a_m + \overline{\mathsf{G}}_n^{q+1,p-1}(\theta_1, c)b_m\}a_m^*$. We find $\overline{\mathcal{V}}_{2n}(\theta) = -n^{-1/2}\{\overline{\mathcal{V}}_{1n}(\theta) + \overline{\mathsf{G}}_n^{q,p}(0, c)a_m + \overline{\mathsf{G}}_n^{q+1,p-1}(0, c)b_m\}a_m^*$ by adding and subtracting $\overline{\mathsf{G}}_n^{q,p}(0, c)$. Here, $\overline{\mathcal{V}}_{1n}(\theta)$ was found to vanish above; Theorem A.2(b) using Assumption A.1(ii) shows that $\overline{\mathsf{G}}_n^{\ell,p}(0, c)$ is bounded; and $|a_m|, |a_m^*|, |b_m|$ are bounded. Thus, $\overline{\mathcal{V}}_{2n}(\theta)$ vanishes due to the $n^{-1/2}$ factor.

The term $\overline{\mathcal{V}}_{3n}(\theta)$. Since $p \geq 2$ then $w_{in} = 1$. Note that $\overline{v}_{3i}(\theta) = \mathsf{E}_{i-1}v_{2i}(\theta)$. We get that $\mathsf{E}\sup_\theta|\overline{\mathcal{V}}_{3n}(\theta)| = \mathsf{E}\sup_\theta|\sum_{i=1}^n \mathsf{E}_{i-1}v_{2i}(\theta)| \leq \mathsf{E}\sum_{i=1}^n \mathsf{E}_{i-1}\sup_\theta|v_{2i}(\theta)|$, which was found to vanish for the term $\mathcal{V}_{2n}(\theta)$ above. □

Proof of Theorem A.4. Use the decompositions (A.17), (A.18) for $\mathbb{G}_n^{q,p}$ and $\overline{\mathsf{G}}_n^{q,p}$ along with Lemmas A.9, A.10. □

B Normality testing initialized by OLS

B.1 Preliminary Results on Estimators

Lemma B.1. *Let $x_i = (1, z'_i)'$ while m_n, v_n are random sequences and*

$$N^{-1}(\hat{\beta} - \beta) = (\textstyle\sum_{i=1}^n x_{in}x'_{in})^{-1}\sum_{i=1}^n x_{in}m_i + v_n. \tag{B.1}$$

Then, $\sum_{i=1}^n x'_i(\hat{\beta} - \beta) = \sum_{i=1}^n m_i + v_n\sum_{i=1}^n x'_{in}$.

Proof. Use that $x_{in} = N'x_i$ and $x_i = (1, z'_i)'$. We get $\sum_{i=1}^n x'_i = (1, 0)\sum_{i=1}^n x_i x'_i$ so that $\sum_{i=1}^n x'_i(\hat{\beta} - \beta) = (1, 0)\sum_{i=1}^n x_i x'_i N N^{-1}(\hat{\beta} - \beta)$. Insert expansion (B.1) for $\hat{\beta}$. Cancel normalizations and sums of squares of x_i. Use that $(1, 0)x_i = 1$. □

Lemma B.2. *Let $\tilde{\beta}$, $\tilde{\sigma}$ be full sample least squares estimators of β, σ. Suppose Assumption 3.1(iii). Then*

$$N^{-1}(\tilde{\beta} - \beta)/\sigma = (\textstyle\sum_{i=1}^n x_{in}x'_{in})^{-1}\sum_{i=1}^n x_{in}\varepsilon_i^\sigma = \mathrm{O}_{\mathsf{P}}(1), \tag{B.2}$$

$$n^{1/2}(\tilde{\sigma} - \sigma) = (\sigma/2)n^{-1/2}\textstyle\sum_{i=1}^n\{(\varepsilon_i^\sigma)^2 - 1\} + \mathrm{o}_{\mathsf{P}}(1) = \mathrm{O}_{\mathsf{P}}(1). \tag{B.3}$$

Proof. (B.2) follows from Assumption 3.1(iii). For (B.3) note that $n^{1/2}(\tilde{\sigma}^2 - \sigma^2) = n^{-1/2}\sum_{i=1}^n(\varepsilon_i^2 - \sigma^2) - n^{-1/2}Q_n$ where $Q_n = \sum_{i=1}^n \varepsilon_i x'_i(\sum_{i=1}^n x_i x'_i)^{-1}\sum_{i=1}^n x_i\varepsilon_i$. The first term is asymptotically normal since ε_i are independent normal by Assumption 3.1(i). The term $n^{-1/2}Q_n$ is $\mathrm{o}_{\mathsf{P}}(1)$ by (B.2) and Assumption 3.1(iii). A Taylor expansion of $(1 + x)^{1/2}$ with $x = \hat{\sigma}^2/\sigma^2 - 1$ shows that $\hat{\sigma} - \sigma = (\sigma/2)(\hat{\sigma}^2/\sigma^2 - 1) + \mathrm{o}_{\mathsf{P}}(\hat{\sigma}^2 - \sigma^2)$. The main term is asymptotically normal. □

The estimators satisfy an improved version of Jiao and Nielsen (2017, Theorem 1).

Lemma B.3. *Let $c > 0$. Suppose Assumptions 2.1, 3.1 hold. Then*

$$
\begin{aligned}
N^{-1}(\hat{\beta}^{RLS} - \beta)/\sigma &= \{2c\varphi(c)/\tau_0^c\}N^{-1}(\tilde{\beta} - \beta)/\sigma \\
&\quad + (\tau_0^c \textstyle\sum_{i=1}^n x_{in}x_{in}')^{-1}\sum_{i=1}^n x_{in}\varepsilon_i^\sigma 1_{(|\varepsilon_i^\sigma| \le c)} + \mathrm{o}_{\mathsf{P}}(1), \qquad (B.4)\\
n^{1/2}(\hat{\sigma}^{RLS} - \sigma) &= \{c(c^2 - \tau_2^c/\tau_0^c)\varphi(c)/\tau_2^c\}n^{1/2}(\tilde{\sigma} - \sigma) \\
&\quad + \{\sigma/(2\tau_2^c)\}n^{-1/2}\textstyle\sum_{i=1}^n \{(\varepsilon_i^\sigma)^2 - \tau_2^c/\tau_0^c\}1_{(|\varepsilon_i^\sigma| \le c)} + \mathrm{o}_{\mathsf{P}}(1), \qquad (B.5)
\end{aligned}
$$

where the initial estimators $\tilde{\beta}, \tilde{\sigma}$ have expansions given in Lemma B.2.

Proof. We apply Theorems A.2, A.3, A.4 using Assumptions 2.1, 3.1(i, ii)

Expression (B.4). Write $N^{-1}(\hat{\beta}^{RLS} - \beta)/\sigma = \{\hat{\mathsf{G}}_n^{2,0}(c)\}^{-1}\{n^{1/2}\hat{\mathsf{G}}_n^{1,1}(c)\}$. By Assumption 3.1($iii$), the initial estimator converges in probability. Thus, by Lemma A.1 it suffices to analyze $\{\mathsf{G}_n^{2,0}(\theta_1, c)\}^{-1}\{n^{1/2}\mathsf{G}_n^{1,1}(\theta_1, c)\}$ uniformly in $|\theta_1| < B$, where $\theta_1 = (a_1, b_1, 0, 0)$ with $a_1 = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$ and $b_1 = N^{-1}(\tilde{\beta} - \beta)/\sigma$.

The denominator. By Theorem A.3, $\mathsf{G}_n^{2,0}(\theta_1, c) = \mathsf{G}_n^{2,0}(0, c) + \mathrm{o}_{\mathsf{P}}(1)$. By Theorem A.2, $\mathsf{G}_n^{2,0}(0, c) = \overline{\mathsf{G}}_n^{2,0}(0, c) + \mathrm{o}_{\mathsf{P}}(1)$, where $\overline{\mathsf{G}}_n^{2,0}(0, c) = \tau_0^c \sum_{i=1}^n x_{in}x_{in}'$.

The numerator. By Theorem A.2 and since $\tau_1^c = 0$ then $\overline{\mathsf{G}}_n^{1,1}(0, c) = 0$. By Theorem A.4, see also (A.9), (A.10), $n^{1/2}\mathsf{G}_n^{1,1}(0, c) = \mathbb{G}_n^{1,1}(0, c) + \mathcal{G}_{1n}^{1,1}(\theta_1, c) - \mathcal{G}_{mn}^{1,1}(\theta_1, c)$. Here, $\mathbb{G}_n^{1,1}(0, c) = \sum_{i=1}^n x_{in}\varepsilon_i^\sigma 1_{(|\varepsilon_i^\sigma| \le c)}$ while $\mathcal{G}_{1n}^{1,1}(\theta_1, c) = 2c\varphi(c)\sum_{i=1}^n x_{in}x_{in}'b_1$ by (A.7), and $\mathcal{G}_{mn}^{1,1}(\theta_1, c) = 0$ by (A.8). Combine these elements and scale by σ to get (B.4).

Expression (B.5). Proceed along the same lines. See also the proof of Jiao and Nielsen (2017, Theorem 1). ∎

B.2 Proof of results for the RLS procedure

Consider the truncated moments (2.6). Here, the superscript ^{RLS} is ignored. Let $\tilde{\theta}_1 = (\tilde{a}, \tilde{b})$ where $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$, $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)/\sigma$ are full sample least squares estimation errors. Let also $\tilde{\theta}_p = (\hat{a}, \hat{b})$ where $\hat{a} = n^{1/2}(\hat{\sigma} - \sigma)/\sigma$, $\hat{b} = N^{-1}(\hat{\beta} - \beta)/\sigma$ are the least squares estimation errors for the selected sub-sample. In combination, $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_p)$, which was analyzed in Lemmas B.2, B.3. We expand $\mathsf{G}_n^{0,p}(\tilde{\theta}, c)$ for $p = 3, 4$ in terms of the vectors $z_{p,i}^c$ given in (3.1).

Lemma B.4. *Let Assumptions 2.1, 3.1, hold. Recall $\zeta_{3,c}^{RLS}$, $\zeta_{4,c}^{RLS}$ defined in (3.2), (3.3). Then, uniformly in $c \ge c_0$ for some $c_0 > 0$, we get*
(a) $\mathsf{G}_n^{0,0}(\tilde{\theta}, c) = \tau_0^c + \mathrm{o}_{\mathsf{P}}(1)$;
(b) $n^{1/2}\mathsf{G}_n^{0,3}(\tilde{\theta}, c) = (\zeta_{3,c}^{RLS})' n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathrm{o}_{\mathsf{P}}(1)$;
(c) $n^{1/2}\{\mathsf{G}_n^{0,4}(\tilde{\theta}, c) - (\tau_4^c/\tau_0^c)\mathsf{G}_n^{0,0}(\tilde{\theta}, c)\} = (\zeta_{4,c}^{RLS})' n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathrm{o}_{\mathsf{P}}(1)$.

Proof. Throughout, we use Theorem A.4 using Assumptions 2.1, 3.1(i, ii) and Lemma A.1 noting that $\hat{a}, \hat{b}$ are bounded by Assumption 3.1(iii), while $\tilde{a}, \tilde{b}$ are bounded by Lemma B.3 using Assumptions 2.1, 3.1.
(a) Apply Lemma A.1 and Theorem A.4 with (A.11) as well as $\overline{\mathsf{G}}_n^{0,0}(0, c) = \tau_0^c$.
(b) Let $N_{3,c} = \mathsf{G}_n^{0,3}(\tilde{\theta}, c) - \overline{\mathsf{G}}_n^{0,3}(0, c)$ where $\overline{\mathsf{G}}_n^{0,3}(0, c) = \mathsf{E}(\varepsilon_i^\sigma)^3 1_{(|\varepsilon_i^\sigma| \le c)} = 0$. By Lemma A.1, Theorem A.4 and (A.10),

$$n^{1/2}N_{3,c} = n^{1/2}\{\mathsf{G}_n^{0,3}(0, c) - \overline{\mathsf{G}}_n^{0,3}(0, c)\} + \mathcal{G}_n^{0,3}(\tilde{\theta}, c) + \mathrm{o}_{\mathsf{P}}(1). \qquad (B.6)$$

Theorem A.4 and (A.7), (A.8) show that the bias term is

$$\mathcal{G}_n^{0,3}(\tilde{\theta},c) = 2c^3\varphi(c)n^{-1/2}\textstyle\sum_{i=1}^n x_{in}'\tilde{b} - 3\tau_2^c n^{-1/2}\sum_{i=1}^n x_{in}'\hat{b}.$$

Let $v_{\mathsf{G}} = (1,0,0)'$ $\tilde{v}_{\mathcal{G}} = (0,0,1)'$ and $\hat{v}_{\mathcal{G}} = \{0, 1/\tau_0^c, 2c\varphi(c)/\tau_0^c\}'$, so that $\zeta_{3,c}^{RLS} = v_{\mathsf{G}} + 2c^3\varphi(c)\tilde{v}_{\mathcal{G}} - 3\tau_2^c\hat{v}_{\mathcal{G}} = \{1, -3\tau_2^c/\tau_0^c, 2(c^2 - 3\tau_2^c/\tau_0^c)c\varphi(c)\}'$ as in (3.2). We show that $n^{1/2}N_{3,c} = (\zeta_{3,c}^{RLS})'n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathrm{o}_{\mathsf{P}}(1)$.

We have that $\mathsf{G}_n^{0,3}(0,c) = v_{\mathsf{G}}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c$ and $\overline{\mathsf{G}}_n^{0,3}(0,c) = 0$. For the bias terms, given expansions for $\tilde{b}$, $\hat{b}$ in (B.2), (B.4), Lemma B.1 implies $\sum_{i=1}^n x_{in}'\tilde{b} = \sum_{i=1}^n \varepsilon_i^\sigma = \tilde{v}_{\mathcal{G}}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c$ and $\sum_{i=1}^n x_{in}'\hat{b} = (1/\tau_0^c)\sum_{i=1}^n \varepsilon_i^\sigma 1_{(|\varepsilon_i^\sigma|\le c)} + \{2c\varphi(c)/\tau_0^c\}\sum_{i=1}^n\varepsilon_i^\sigma + \mathrm{o}_{\mathsf{P}}(1)$ so that $\sum_{i=1}^n x_{in}'\hat{b} = \hat{v}_{\mathcal{G}}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c$. Insert these expressions in (B.6).

(*c*) Let $N_{4,c} = \mathsf{G}_n^{0,4}(\tilde{\theta},c) - (\tau_4^c/\tau_0^c)\mathsf{G}_n^{0,0}(\tilde{\theta},c)$. Due to Lemma A.1 and Theorem A.4 with (A.10), we get, for $p = 0, 4$,

$$n^{1/2}\{\mathsf{G}_n^{0,p}(\tilde{\theta},c) - \overline{\mathsf{G}}_n^{0,p}(0,c)\} = n^{1/2}\{\mathsf{G}_n^{0,p}(0,c) - \overline{\mathsf{G}}_n^{0,p}(0,c)\} + \mathcal{G}_n^{0,p}(\tilde{\theta},c) + \mathrm{o}_{\mathsf{P}}(1),$$

with compensators $\overline{\mathsf{G}}_n^{0,p}(0,c) = \mathsf{E}\varepsilon_i^\sigma 1_{(|\varepsilon_i^\sigma|\le c)} = \tau_p^c$. We note the equation $\overline{\mathsf{G}}_n^{0,4}(0,c) - (\tau_4^c/\tau_0^c)\overline{\mathsf{G}}_n^{0,0}(0,c) = \tau_4^c - \tau_0^c\tau_4^c/\tau_0^c = 0$. Therefore we can write

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

Proceed as in (*b*). Let $v_{\mathbb{G},4} = (1,0,0,0)'$ and $v_{\mathbb{G},0} = (0,0,1,0)'$, so that $\mathbb{G}_n^j(0,c) = v_{\mathbb{G},j}'n^{-1/2}\sum_{i=1}^n z_{4,i}^c$ for $j = 0, 4$. Let $v_{\mathcal{G},0} = 2c\varphi(c)(0,0,0,1/2)'$. From Theorem A.4 we have $\mathcal{G}_n^{0,0}(\tilde{\theta},c) = 2c\varphi(c)\tilde{a}$. As $\tilde{a}$ satisfies (B.3) we get $\mathcal{G}_n^{0,0}(\tilde{\theta},c) = v_{\mathcal{G},0}'n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathrm{o}_{\mathsf{P}}(1)$. Let $v_{\mathcal{G},4} = 2c^5\varphi(c)(0,0,0,1/2)' - 2(\tau_4^c/\tau_2^c)\{0, 1, -(\tau_2^c/\tau_0^c), c(c^2 - \tau_2^c/\tau_0^c)\varphi(c)\}$. Theorem A.4 shows $\mathcal{G}_n^{0,4}(\tilde{\theta},c) = 2c^5\varphi(c)\tilde{a} - 4\tau_4^c\hat{a}$. Since $\tilde{a}$ and $\hat{a}$ satisfy (B.3) and (B.5) we get $\mathcal{G}_n^{0,4}(\tilde{\theta},c) = v_{\mathcal{G},4}'n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathrm{o}_{\mathsf{P}}(1)$. Insert these expressions in the expansion of $n^{1/2}N_{4,c}$ noting that $\zeta_{4,c}^{RLS} = v_{\mathbb{G},4} + v_{\mathcal{G},4} - (\tau_4^c/\tau_0^c)(v_{\mathbb{G},0} + v_{\mathcal{G},0})$ to get (3.3). □

Proof of Theorem 3.1. Throughout, we use Theorem A.4 using Assumptions 2.1, 3.1.
1. Write $\hat{\mu}_{p,c} = \mathsf{G}_n^{0,p}(\tilde{\theta},c)/\mathsf{G}_n^{0,0}(\tilde{\theta},c)$ for $p = 3, 4$. Let

$$T_{p,c,n}^{RLS} = \{(\zeta_{p,c}^{RLS})'\Omega_p^c(\zeta_{p,c}^{RLS})\}^{-1/2}(\zeta_{p,c}^{RLS})'n^{-1/2}\textstyle\sum_{i=1}^n z_{p,i}^c.$$

2. *Denominator.* Lemma B.4(*a*) shows $\mathsf{G}_n^{0,0}(\tilde{\theta},c) - \tau_0^c = \mathrm{o}_{\mathsf{P}}(1)$.
3. *Third moment.* Lemma B.4(*b*) shows $n^{1/2}\mathsf{G}_n^{0,3}(\tilde{\theta},c) = \zeta_{3,c}'n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathrm{o}_{\mathsf{P}}(1)$. Note that $(\tau_0^c)^2\lambda_{6,c} = \mathsf{Var}\{(\zeta_{3,c})'z_{3,i}^c\}$ to get $\hat{T}_{3,c} = n^{1/2}\hat{\mu}_{3,c}/\lambda_{6,c}^{1/2} = T_{3,c,n} + \mathrm{o}_{\mathsf{P}}(1)$.
4. *Fourth moment.* Expand the demeaned moment $n^{1/2}(\hat{\mu}_{4,c} - \tau_4^c/\tau_0^c)$ as $n^{1/2}\{\mathsf{G}_n^{0,4}(\tilde{\theta},c) - (\tau_4^c/\tau_0^c)\mathsf{G}_n^{0,0}(\tilde{\theta},c)\}/\mathsf{G}_n^{0,0}(\tilde{\theta},c)$. Expand the numerator as $\zeta_{4,c}'n^{-1/2}\sum_{i=1}^n z_{4,i}^c + \mathrm{o}_{\mathsf{P}}(1)$ using Lemma B.4(*c*). Proceed as in item 3 to see that $\hat{T}_{4,c} = T_{4,c,n} + \mathrm{o}_{\mathsf{P}}(1)$.
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. □

C Normality testing initialized by LTS

C.1 Preliminary Results on Estimators

We analyze the order statistics of the LTS residuals and the LTS variance estimator. We follow the analysis in §D.4 of Johansen and Nielsen (2016a), henceforth JN16. Let $\tilde{c}_{LTS} = \tilde{\xi}_{(h)}/\sigma$ be the hth smallest order statistic of $\tilde{\xi}_i = |y_i - x_{in}'\hat{\beta}_{LTS}|$, where $\tilde{b}_{LTS} = N^{-1}(\hat{\beta}_{LTS} - \beta)/\sigma$. Let $\tilde{\theta}_{LTS} = (0, \tilde{b}_{LTS}, 0, \tilde{b}_{LTS})$. Then,

$$\tilde{c}_{LTS} = \inf\left\{c : \frac{1}{n}\sum_{i=1}^{n} 1_{(|\varepsilon_i^\sigma - x_{in}'\tilde{b}_{LTS}| \le c)} \ge \frac{h}{n}\right\} = \inf\left\{c : \mathsf{G}_n^{0,0}(\tilde{\theta}_{LTS}, c) \ge \frac{h}{n}\right\}$$

and $\mathsf{G}_n^{0,0}(\tilde{\theta}_{LTS}, \tilde{c}_{LTS}) = h/n$. Similarly, if $\hat{c}_0$ is the hth order statistic of $|\varepsilon_i^\sigma|$ then $\mathsf{G}_n^{0,0}(0, \hat{c}_0) = h/n$. Finally, let G be the distribution function of $|\varepsilon_i^\sigma|$ and let $\tilde{\theta}_d = \{n^{1/2}(\tilde{c}_{LTS}/c - 1), 0, 0, 0)\}$.

Lemma C.1. *Let $c \in (0, c_\epsilon)$ and $h = \lfloor n\mathsf{G}(c)\rfloor$. Suppose Assumptions 2.1, 3.2. Then*

$$n^{1/2}(\tilde{c}_{LTS} - c) = -\{2\mathsf{f}(c)\}^{-1}\mathbb{G}_n^{0,0}(0, c) + \mathrm{o}_{\mathsf{P}}(1), \tag{C.1}$$

$$n^{1/2}(\tilde{\sigma}_{LTS} - \sigma) = (\sigma/2\tau_2^c)n^{-1/2}\{\mathbb{G}_n^{0,2}(0, c) - c^2\mathbb{G}_n^{0,0}(0, c)\} + \mathrm{o}_{\mathsf{P}}(1). \tag{C.2}$$

Proof. Suppress the index ^{LTS}. Lemma A.1 and Theorem A.4 are used repeatedly. This requires Assumption 3.2(i, ii).

Quantiles of ε_i/σ. From Bahadur (1966), we have $2\mathsf{f}(c)n^{1/2}(\hat{c}_0 - c) = -\mathbb{G}_n^{0,0}(0, c) + \mathrm{o}_{\mathsf{P}}(1)$, which is then $\mathrm{O}_{\mathsf{P}}(1)$ by the Central Limit Theorem.

Initial assessment of $\tilde{c}$. We argue that $n^{1/2}(\tilde{c} - c) = \mathrm{O}_{\mathsf{P}}(1)$. Lemma D.6 of JN16 shows that $n^{1/2}|\tilde{c} - c| \le 2|\tilde{b}|\max_{1\le i\le n}|x_{in}|$. Assumption 3.2$(ii, iii, iv)$ gives the bounds $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)/\sigma = \mathrm{O}_{\mathsf{P}}(1)$ and $\max_{1\le i\le n}|x_{in}| = \mathrm{o}_{\mathsf{P}}(1)$, see also Lemma A.5.

Result (C.1). Follow the proof of Theorem D.7 in JN16 for fixed c. By construction

$$h/n = \mathsf{G}_n^{0,0}(\tilde{\theta}, \tilde{c}) = \mathsf{G}_n^{0,0}(\tilde{\theta} + \tilde{\theta}_d, c), \qquad \text{and} \qquad h/n = \mathsf{G}_n^{0,0}(0, \hat{c}_0) = \mathsf{G}_n^{0,0}(\tilde{\theta}_d, c).$$

Equating the two expressions we have

$$0 = n^{1/2}\{\mathsf{G}_n^{0,0}(\tilde{\theta}, \tilde{c}) - \mathsf{G}_n^{0,0}(\tilde{\theta}_d, c)\}.$$

Here, $\tilde{b} = \mathrm{O}_{\mathsf{P}}(1)$ by assumption while $n^{1/2}(\tilde{c} - c), n^{1/2}(\hat{c}_0 - c) = \mathrm{O}_{\mathsf{P}}(1)$ for fixed c as argued above. Thus, using Lemma A.1, we can replace these estimation errors with deterministic terms and apply Theorem A.4 with the expansion (A.13) to each of the $\mathsf{G}_n^{0,0}$ functions. Deleting common terms in the two expansions then shows that

$$0 = \mathcal{G}_{1n}^{0,0}(\tilde{\theta} + \tilde{\theta}_d, c) + \mathcal{G}_{1n}^{0,0}(\tilde{\theta}_d, c) + \mathrm{o}_{\mathsf{P}}(1).$$

Thus, by the expression for $\mathcal{G}_{1n}^{0,0}$ in (A.7) we get $\mathrm{o}_{\mathsf{P}}(1) = 2c\mathsf{f}(c)\{(\tilde{c}/c - 1) - (\hat{c}_0/c - 1)\}$ so that $2c\mathsf{f}(c)(\tilde{c} - c) = 2c\mathsf{f}(c)(\hat{c}_0 - c) + \mathrm{o}_{\mathsf{P}}(1)$. Last, insert the expansion for $\hat{c}_0$.

The result (C.2). Recall that $\xi_i = |\varepsilon_i^\sigma - x_{in}'(\hat{\beta} - \beta)|$. We have that

$$\tilde{\sigma}^2 = \left(\frac{\tau_0^c}{\tau_2^c}\right)\frac{n^{-1}\sum_{i=1}^n \tilde{\varepsilon}_i^2 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}}{n^{-1}\sum_{i=1}^n 1_{\{\tilde{\xi}_i \le \tilde{\xi}_{(h)}\}}} = \sigma^2\left(\frac{\tau_0^c}{\tau_2^c}\right)\frac{\mathsf{G}_n^{0,2}(\tilde{\theta}, \tilde{c})}{\mathsf{G}_n^{0,0}(\tilde{\theta}, \tilde{c})},$$

when using the empirical process notation with $\tilde{b}=n^{1/2}(\hat{\beta}-\beta)/\sigma$ as well as $\tilde{\theta}=(0,\tilde{b},0,\tilde{b})$ and $\tilde{\theta}_d=\{n^{1/2}(\tilde{c}-c),\tilde{b},0,0\}$. Normalize to get

$$n^{1/2}(\tilde{\sigma}^2-\sigma^2)=\sigma^2\left(\frac{\tau_0^c}{\tau_2^c}\right)\frac{n^{1/2}\{\mathsf{G}_n^{0,2}(\tilde{\theta},\tilde{c})-(\tau_2^c/\tau_0^c)\mathsf{G}_n^{0,0}(\tilde{\theta},\tilde{c})\}}{\mathsf{G}_n^{0,0}(\tilde{\theta},\tilde{c})}.$$

By assumption $\tilde{b}=\mathrm{O}_{\mathsf{P}}(1)$. Applying Lemma A.1, Theorem A.4 with (A.14) to the denominator shows $\mathsf{G}_n^{0,0}(\tilde{\theta},\tilde{c})=\overline{\mathsf{G}}_n^{0,0}(0,c)+\mathrm{o}_{\mathsf{P}}(1)$. By (A.5) we have that $\overline{\mathsf{G}}_n^{0,p}(0,c)=\tau_p^c$ so that $\mathsf{G}_n^{0,0}(\tilde{\theta},\tilde{c})=\tau_0^c+\mathrm{o}_{\mathsf{P}}(1)$ as well as $\overline{\mathsf{G}}^{0,2}(0,c)-(\tau_2^c/\tau_0^c)\overline{\mathsf{G}}^{0,0}(0,c)=0$. In combination,

$$n^{1/2}(\tilde{\sigma}^2-\sigma^2)=\sigma^2\left(\frac{\tau_0^c}{\tau_2^c}\right)\frac{n^{1/2}\{\mathsf{G}_n^{0,2}(\tilde{\theta},\tilde{c})-\overline{\mathsf{G}}^{0,2}(0,c)\}-(\tau_2^c/\tau_0^c)n^{1/2}\{\mathsf{G}_n^{0,0}(\tilde{\theta},\tilde{c})-\overline{\mathsf{G}}^{0,0}(0,c)\}}{\tau_0^c+\mathrm{o}_{\mathsf{P}}(1)}.$$

Now, the expansion in Lemma A.1, Theorem A.4 with (A.13) shows that

$$n^{1/2}\{\mathsf{G}_n^{0,p}(\tilde{\theta},\tilde{c})-\overline{\mathsf{G}}^{0,p}(0,c)\}=\mathbb{G}_n^{0,p}(0,c)+\mathcal{G}_{1n}^{0,p}(\tilde{\theta}+\tilde{\theta}_d,c)-\mathcal{G}_{mn}^{0,p}(\tilde{\theta}+\tilde{\theta}_d,c).$$

For $p=0,2$ we get from (A.7), (A.8) that $\mathcal{G}_{1n}^{0,p}(\tilde{\theta}+\tilde{\theta}_d,c)=2c^{p+1}\varphi(c)n^{1/2}(\tilde{c}/c-1)$ and $\mathcal{G}_{mn}^{0,p}(\tilde{\theta}+\tilde{\theta}_d,c)=0$. The expansion in (C.1) shows that $2\varphi(c)n^{1/2}(\tilde{c}-c)=-\mathbb{G}_n^{0,0}(0,c)+\mathrm{o}_{\mathsf{P}}(1)$. Insert all this above to get

$$n^{1/2}(\tilde{\sigma}^2-\sigma^2)=\left(\frac{\sigma^2}{\tau_2^c}\right)\Big[\mathbb{G}_n^{0,2}(0,c)-\left(\frac{\tau_2^c}{\tau_0^c}\right)\mathbb{G}_n^{0,0}(0,c)-\left(c^2-\frac{\tau_2^c}{\tau_0^c}\right)\mathbb{G}_n^{0,0}(0,c)\Big]+\mathrm{o}_{\mathsf{P}}(1).$$

Cancel the (τ_2^c/τ_0^c) terms and use that $n^{1/2}(\tilde{\sigma}-\sigma)=n^{1/2}(\tilde{\sigma}^2-\sigma^2)/(2\sigma)+\mathrm{o}_{\mathsf{P}}(1)$ by the δ-method. □

C.2 Proof of results for the LTS procedure

Consider the truncated moments (2.6). Define estimation errors $\tilde{a}_{LTS}=n^{1/2}(\tilde{\sigma}_{LTS}-\sigma)/\sigma$ and $\tilde{b}_{LTS}=N^{-1}(\tilde{\beta}_{LTS}-\beta)/\sigma$. Let $\tilde{c}_{LTS}$ be the h quantile of $|y_i-x_{in}'\tilde{\beta}_{LTS}|$.

Lemma C.2. *Suppose Assumptions 2.1, 3.2 hold. Recall $\zeta_{3,c}^{LTS}$, $\zeta_{4,c}^{LTS}$ from (3.7) and (3.8) and $z_{3,i}^c$, $z_{4,i}^c$ from (3.1). Let $\tilde{\vartheta}_{LTS}=(0,\tilde{b}_{LTS},\tilde{a}_{LTS},\tilde{b}_{LTS})$. Then*
(a) $\mathsf{G}_n^{0,0}(\tilde{\vartheta}_{LTS},\tilde{c}_{LTS})=\tau_0^c+\mathrm{o}_{\mathsf{P}}(1)$;
(b) $n^{1/2}\mathsf{G}_n^{0,3}(\tilde{\vartheta}_{LTS},\tilde{c}_{LTS})=(\zeta_{3,c}^{LTS})'n^{-1/2}\sum_{i=1}^n z_{3,i}^c+\mathrm{o}_{\mathsf{P}}(1)$;
(c) $n^{1/2}\{\mathsf{G}_n^{0,4}(\tilde{\vartheta}_{LTS},\tilde{c}_{LTS})-\tau_4^c/\tau_0^c\mathsf{G}_n^{0,0}(\tilde{\vartheta}_{LTS},\tilde{c}_{LTS})\}=(\zeta_{4,c}^{LTS})'n^{-1/2}\sum_{i=1}^n z_{4,i}^c+\mathrm{o}_{\mathsf{P}}(1)$.

Proof. Let $\tilde{\theta}_d=\{n^{1/2}(\tilde{c}_{LTS}/c-1),0,0,0)\}$. Note $\mathsf{G}_n^{0,p}(\tilde{\theta}_{LTS},\tilde{c}_{LTS})=\mathsf{G}_n^{0,p}(\tilde{\theta}_{LTS}+\tilde{\theta}_d,c)$. Note that $\tilde{b}_{LTS}$ is $\mathrm{O}_{\mathsf{P}}(1)$ due to Assumptions 3.2(iii,iv), while $\tilde{\theta}_d$, $\tilde{a}_{LTS}$ are $\mathrm{O}_{\mathsf{P}}(1)$ due to Lemma C.1 using Assumptions 2.1, 3.2. Lemma A.1 shows that we can replace random estimation errors with determiniatic quanties in a compact set. We then apply Theorem A.4 using Assumptions 3.2(i,ii). Suppress the sub-index ^{LTS} throughout.

(a) Apply Theorem A.4 with (A.14) and $\overline{\mathsf{G}}_n^{0,0}(0,c)=\tau_0^c$.

(b) Let $N_{3,\hat{c}}=\mathsf{G}_n^{0,3}(\tilde{\vartheta},\tilde{c})-\overline{\mathsf{G}}_n^{0,3}(0,c)$ with $\overline{\mathsf{G}}_n^{0,3}(0,c)=0$. Lemma A.1 and Theorem A.4 with (A.13) show $n^{1/2}N_{3,\hat{c}}=\mathbb{G}_n^{0,3}(0,c)+\mathcal{G}_n^{0,3}(\tilde{\vartheta}+\tilde{\theta}_d,c)+\mathrm{o}_{\mathsf{P}}(1)$.

Define $v_{\mathsf{G}} = (1,0,0)'$ and $v_{\mathcal{G}} = (\tau_2^c)^{-1}\{2c^3\varphi(c) - 3\tau_2^c\}(0,1,0)'$ so that $\zeta_{3,c} = v_{\mathsf{G}} + v_{\mathcal{G}} = [1, \{2c^3\varphi(c) - 3\tau_2^c\}/\tau_2^c, 0]'$ as in (3.7). We show that $n^{1/2}N_{3,c} = \zeta_{3,c}' n^{-1/2}\sum_{i=1}^n z_{3,i}^c + \mathrm{o}_{\mathsf{P}}(1)$.

First, $\mathsf{G}_n^{0,3}(0,c) = v_{\mathsf{G}}' n^{-1/2}\sum_{i=1}^n z_{3,i}^c$. Second, Theorem A.4 with (A.10), (A.11) shows $\mathcal{G}_n^{0,3}(\tilde{\vartheta} + \tilde{\theta}_d, c) = \{2c^3\varphi(c) - 3\tau_2^c\} n^{-1/2}\sum_{i=1}^n x_{in}'\tilde{b}$. The estimation error $\tilde{b} = N^{-1}(\tilde{\beta} - \beta)$ has an expansion given in Assumption 3.2(iv) and is of the form considered in Lemma B.1. Therefore, $\sum_{i=1}^n x_{in}'\tilde{b} = (\tau_2^c)^{-1}\sum_{i=1}^n (\varepsilon_i^\sigma) 1_{(|\varepsilon_i^\sigma| \le c)} + \mathrm{o}_{\mathsf{P}}(1)$. In turn, $\mathcal{G}_n^{0,3}(\tilde{\vartheta} + \tilde{\theta}_d, c) = v_{\mathcal{G}}'\sum_{i=1}^n z_{3,i}^c + \mathrm{o}_{\mathsf{P}}(1)$.

(c) Let $N_{4,\hat{c}} = \{\mathsf{G}_n^{0,4}(\tilde{\theta}, \tilde{c}) - \tau_4^c/\tau_0^c \mathsf{G}_n^{0,0}(\tilde{\theta}, \tilde{c})\}$. Lemma A.1 and Theorem A.4 with (A.13) give, for $j = 0, 4$, that $n^{1/2}\{\mathsf{G}_n^j(\tilde{\theta}, \tilde{c}) - \overline{\mathsf{G}}_n^j(0,c)\} = \mathbb{G}_n^j(0,c) + \mathcal{G}_n^j(\tilde{\vartheta} + \tilde{\theta}_d, c) + \mathrm{o}_{\mathsf{P}}(1)$. Due to the identity $\overline{\mathsf{G}}_n^{0,4}(0,c) - (\tau_4^c/\tau_0^c)\overline{\mathsf{G}}_n^{0,0}(0,c) = \tau_4^c - \tau_0^c\tau_4^c/\tau_0^c = 0$, we write

$$n^{1/2}N_{4,\hat{c}} = \big\{\mathbb{G}_n^{0,4}(0,c) + \mathcal{G}_{1n}^{0,4}(\tilde{\vartheta} + \tilde{\theta}_d, c)\big\} - (\tau_4^c/\tau_0^c)\big\{\mathbb{G}_n^{0,0}(0,c) + \mathcal{G}_{1n}^{0,0}(\tilde{\vartheta} + \tilde{\theta}_d, c)\big\} + \mathrm{o}_{\mathsf{P}}(1).$$

Let $v_{\mathbb{G},4} = (1,0,0,0)'$ and $v_{\mathbb{G},0} = (0,0,1,0)'$ for $p = 0,4$, so that $\mathbb{G}_n^{0,p}(0,c) = v_{\mathbb{G},p}' n^{-1/2}\sum_{i=1}^n z_{4,i}^c$. From Theorem A.4 with (A.10), (A.11) we get biases $\mathcal{G}_{1n}^{0,0}(\tilde{\vartheta} + \tilde{\theta}_d, c) = 2\varphi(c) n^{1/2}(\tilde{c} - c)$ and $\mathcal{G}_{1n}^{0,4}(\tilde{\vartheta} + \tilde{\theta}_d, c) = 2c^4\varphi(c) n^{1/2}(\tilde{c} - c) - 4\tau_4^c\sigma^{-1}\tilde{a}$.

Let $v_{\mathcal{G},0} = (0,0,-1,0)'$ and $v_{\mathcal{G},4} = \{(0,0,-c^4,0) - 2(\tau_4^c/\tau_2^c)(0,1,-c^2,0)\}'$. Then, the expansions for $\tilde{a}, n^{1/2}(\tilde{c} - c)$ in (C.1), (C.2) give $\mathcal{G}_{1n}^{0,p}(\tilde{\vartheta} + \tilde{\theta}_d, c) = v_{\mathcal{G},p}'\sum_{i=1}^n z_{4,i}^c$.

Insert these expressions in the above expansion of $n^{1/2}N_{4,c}$ noting that $\zeta_{4,c} = v_{\mathbb{G},4} + v_{\mathcal{G},4} - (\tau_4^c/\tau_0^c)(v_{\mathbb{G},0} + v_{\mathcal{G},0})$ giving the expression in (3.8). □

Proof of Theorem 3.2. As the proof of Theorem 3.1 replacing Lemma B.4 by Lemma C.2. □

D Power expansions for the kurtosis statistics

The kurtosis statistic $\hat{T}_{4c}^s$ was expanded in (3.9). Here, we consider the numerator of the non-centrality term, that is $\lambda_{3c\Phi}^s - \lambda_{3c\mathsf{F}}^s$, where $\lambda_{3c\mathsf{F}}^s$ is the limiting value of $\hat{\mu}_{4c}^s$ as defined in (2.6). We let $\tau_{p\Phi}^c$ and $\tau_{p\mathsf{F}}^c$ denote the truncated moments under normality and F.

Least trimmed squares limits. The initial and updated LTS scale estimators defined in (3.6) satisfy

$$\tilde{\sigma}_{LTS}^2 = \hat{\sigma}_{LTS}^2 \overset{\mathsf{P}}{\to} \sigma^2/\varpi_c^2 \qquad \text{where} \qquad \varpi_c^2 = (\tau_{0\mathsf{F}}^c/\tau_{2\mathsf{F}}^c)(\tau_{2\Phi}^c/\tau_{0\Phi}^c),$$

see Lemma C.1. The fourth moment estimator defined in (2.6) then satisfies

$$\hat{\mu}_{4c}^{LTS} = \varpi_c^4 \frac{\sum_{i=1}^n \{\hat{\varepsilon}_i/(\hat{\sigma}_{LTS}\varpi_c)\}^4 1_{(|\tilde{\varepsilon}_i| \le \tilde{\xi}_h)}}{\sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i| \le \tilde{\xi}_h)}} \overset{\mathsf{P}}{\to} \lambda_{3c\mathsf{F}}^{LTS} = \varpi_c^4 \frac{\tau_{4\mathsf{F}}^c}{\tau_{0\mathsf{F}}^c},$$

by Lemma C.2. We expand the numerator of the non-centrality in (3.9) as

$$\lambda_{3c\mathsf{F}}^{LTS} - \lambda_{3c\Phi}^{LTS} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big\{\varpi_c^4\Big(\frac{\tau_{4\mathsf{F}}^c}{\tau_{4\Phi}^c}\Big)\Big(\frac{\tau_{0\Phi}^c}{\tau_{0\mathsf{F}}^c}\Big) - 1\Big\} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big\{\Big(\frac{\tau_{0\mathsf{F}}^c}{\tau_{0\Phi}^c}\Big)\Big(\frac{\tau_{4\mathsf{F}}^c}{\tau_{4\Phi}^c}\Big)\Big(\frac{\tau_{2\Phi}^c}{\tau_{2\mathsf{F}}^c}\Big)^2 - 1\Big\} \tag{D.1}$$

Since $\mathsf{F} = (1-\epsilon)\Phi + \epsilon\mathsf{G}$, we get that $\tau_{p\mathsf{F}}^c = (1-\epsilon)\tau_{p\Phi}^c + \epsilon\tau_{p\mathsf{G}}^c$. Rearrange to get

$$\frac{\tau_{p\mathsf{F}}^c}{\tau_{p\Phi}^c} = 1 + \epsilon\Big(\frac{\tau_{p\mathsf{G}}^c}{\tau_{p\Phi}^c} - 1\Big). \tag{D.2}$$

Insert the expression (D.2) in (D.1) to get

$$\lambda_{3c\mathsf{F}}^{LTS} - \lambda_{3c\Phi}^{LTS} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big[\frac{\{1+\epsilon(\tau_{0\mathsf{G}}^c/\tau_{0\Phi}^c-1)\}\{1+\epsilon(\tau_{4\mathsf{G}}^c/\tau_{4\Phi}^c-1)\}}{1+2\epsilon(\tau_{2\mathsf{G}}^c/\tau_{2\Phi}^c-1)} - 1\Big]$$
$$= \epsilon\frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big(\frac{\tau_{4\mathsf{G}}^c}{\tau_{4\Phi}^c} - 2\frac{\tau_{2\mathsf{G}}^c}{\tau_{2\Phi}^c} + \frac{\tau_{0\mathsf{G}}^c}{\tau_{0\Phi}^c}\Big) + \mathrm{o}(\epsilon). \qquad \text{(D.3)}$$

Robustified least squares limits. The initial least squares estimator satisfies

$$\tilde{\sigma}_{OLS}^2 \overset{\mathsf{P}}{\to} \sigma^2/\varpi^2 \qquad \text{where} \qquad \varpi^2 = \varpi_\infty^2 = 1/\tau_{2\mathsf{F}}^\infty.$$

The updated least squares scale estimator satisfies

$$\hat{\sigma}_{RLS}^2 = \frac{\tau_{0\Phi}^c \sum_{i=1}^n \hat{\varepsilon}_i^2 1_{\{|\tilde{\varepsilon}_i|<(\tilde{\sigma}_{OLS}\varpi)(c/\varpi)\}}}{\tau_{2\Phi}^c \sum_{i=1}^n 1_{\{|\tilde{\varepsilon}_i|<(\tilde{\sigma}_{OLS}\varpi)(c/\varpi)\}}} \overset{\mathsf{P}}{\to} \frac{\sigma^2}{\tilde{\varpi}_c^2} \quad \text{where} \quad \tilde{\varpi}_c^2 = \Big(\frac{\tau_{0\mathsf{F}}^{c/\varpi}}{\tau_{2\mathsf{F}}^{c/\varpi}}\Big)\Big(\frac{\tau_{2\Phi}^c}{\tau_{0\Phi}^c}\Big).$$

The fourth moment estimator defined in (2.6) then satisfies

$$\hat{\mu}_{4c}^{RLS} = \tilde{\varpi}_c^4 \frac{\sum_{i=1}^n \{\hat{\varepsilon}_i/(\hat{\sigma}_{RLS}\tilde{\varpi}_c)\}^4 1_{\{|\tilde{\varepsilon}_i|\le(\tilde{\sigma}_{OLS}\varpi)(c/\varpi)\}}}{\sum_{i=1}^n 1_{\{|\tilde{\varepsilon}_i|\le(\tilde{\sigma}_{OLS}\varpi)(c/\varpi)\}}} \overset{\mathsf{P}}{\to} \lambda_{3c\mathsf{F}}^{RLS} = \tilde{\varpi}_c^4 \frac{\tau_{4\mathsf{F}}^{c/\varpi}}{\tau_{0\mathsf{F}}^{c/\varpi}}, \qquad \text{(D.4)}$$

by Lemma B.4. We expand the numerator of the non-centrality term in (3.9) as

$$\lambda_{3c\mathsf{F}}^{RLS} - \lambda_{3c\Phi}^{RLS} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big\{\tilde{\varpi}_c^4\Big(\frac{\tau_{4\mathsf{F}}^{c/\varpi}}{\tau_{4\Phi}^c}\Big)\Big(\frac{\tau_{0\Phi}^c}{\tau_{0\mathsf{F}}^{c/\varpi}}\Big) - 1\Big\} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big\{\Big(\frac{\tau_{0\mathsf{F}}^{c/\varpi}}{\tau_{0\Phi}^c}\Big)\Big(\frac{\tau_{4\mathsf{F}}^{c/\varpi}}{\tau_{4\Phi}^c}\Big)\Big(\frac{\tau_{2\Phi}^c}{\tau_{2\mathsf{F}}^{c/\varpi}}\Big)^2 - 1\Big\} \quad \text{(D.5)}$$

We expand the truncated moments evaluated in a distorted cut-off. First, note that p is even and, then, apply the mean value theorem to get

$$\tau_{p\mathsf{F}}^{c/\varpi} = \int_{-c/\varpi}^{c/\varpi} u^p d\mathsf{F}(u) = \tau_{p\mathsf{F}}^c + 2\int_c^{c/\varpi} u^p d\mathsf{F}(u) = 1 + \frac{2}{\tau_{p\mathsf{F}}^c} c^{p+1}\mathsf{f}(c^*)\Big(\frac{1}{\varpi} - 1\Big) + \mathrm{o}(\epsilon).$$

It is convenient to let $y = (1/\varpi - 1) = (\tau_{2\mathsf{F}}^\infty)^{1/2} - 1$. Combine with (D.2) to get

$$\frac{\tau_{p\mathsf{F}}^{c/\varpi_d}}{\tau_{p\Phi}^c} = \Big(\frac{\tau_{p\mathsf{F}}^{c/\varpi_d}}{\tau_{p\mathsf{F}}^c}\Big)\Big(\frac{\tau_{p\mathsf{F}}^c}{\tau_{p\Phi}^c}\Big) = \big\{1 + (\epsilon/\tau_{p\mathsf{F}}^c)c^{p+1}\mathsf{f}(c)y + \mathrm{o}(\epsilon)\big\}\big[1 + \epsilon\{(\tau_{p\mathsf{G}}^c/\tau_{p\Phi}^c) - 1\}\big]$$
$$= 1 + \epsilon\Big\{\frac{\tau_{p\mathsf{G}}^c}{\tau_{p\Phi}^c} - 1 + \frac{1}{\tau_{p\Phi}^c}c^{p+1}\mathsf{f}(c)y\Big\} + \mathrm{o}(\epsilon). \quad \text{(D.6)}$$

Insert the expression (D.6) in (D.5) to get

$$\lambda_{3c\mathsf{F}}^{RLS} - \lambda_{3c\Phi}^{RLS} = \frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big([1 + \epsilon\{(\tau_{0\mathsf{G}}^c/\tau_{0\Phi}^c) - 1 + c\mathsf{f}(c)y/\tau_{0\Phi}^c\}]$$
$$\times \frac{1 + \epsilon\{(\tau_{4\mathsf{G}}^c/\tau_{4\Phi}^c) - 1 + c^5\mathsf{f}(c)y/\tau_{4\Phi}^c\}}{[1 + \epsilon\{(\tau_{2\mathsf{G}}^c/\tau_{2\Phi}^c) - 1 + c^3\mathsf{f}(c)y/\tau_{2\Phi}^c\}]^2} - 1\Big) + \mathrm{o}(\epsilon).$$

Expand for small ϵ and use that $y = (\tau_{2\mathsf{F}}^\infty)^{1/2} - 1$ to get the final expression

$$\lambda_{3c\mathsf{F}}^{RLS} - \lambda_{3c\Phi}^{RLS} = \epsilon\frac{\tau_{4\Phi}^c}{\tau_{0\Phi}^c}\Big\{\Big(\frac{\tau_{4\mathsf{G}}^c}{\tau_{4\Phi}^c} - 2\frac{\tau_{2\mathsf{G}}^c}{\tau_{2\Phi}^c} + \frac{\tau_{0\mathsf{G}}^c}{\tau_{0\Phi}^c}\Big) + c\mathsf{f}(c)\big\{(\tau_{2\mathsf{G}}^\infty)^{1/2} - 1\big\}\Big(\frac{c^4}{\tau_{4\Phi}^c} - 2\frac{c^2}{\tau_{2\Phi}^c} + \frac{1}{\tau_{0\Phi}^c}\Big)\Big\} + \mathrm{o}(\epsilon).$$

References

Argullo, J., Croux, C., and Van Aelst, S. (2008). The multivariate least trimmed squares estimator. *J. Multivariate Anal.*, 99:311–318.

Bahadur, R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.*, 37:577–580.

Barr, D. R. and Sherrill, E. T. (1999). Mean and variance of truncated normal distributions. *Amer. Statist.*, 53:357–361.

Berenguer-Rico, V., Johansen, S., and Nielsen, B. (2019). The analysis of marked and weighted empirical processes of estimated residuals. Discussion Paper 19-05, Dept. Econ., Copenhagen.

Berenguer-Rico, V., Johansen, S., and Nielsen, B. (2023). A model where the least trimmed squares estimator is maximum likelihood. *J. Roy. Statist. Soc. Ser. B.* To appear.

Berenguer-Rico, V. and Nielsen, B. (2022). Least trimmed squares asymptotics: Regression with leverage. mimeo, Oxford.

Berenguer-Rico, V. and Wilms, I. (2021). Heteroscedasticity testing after outlier removal. *Econometric Reviews*, 40:51–85.

Butler, R. (1982). Nonparametric interval and point prediction using data trimmed by a Grubbs-type outlier rule. *Ann. Statist.*, 10:197–204.

Čížek, P. (2005). Least trimmed squares in nonlinear regression under dependence. *J. Statist. Plann. Inference*, 136:3967–3988.

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

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

Hampel, F. R. (1985). The breakdown points of the mean combined with some rejection rules. *Technometrics*, 27:98–107.

Heretier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.*, 89:897–904.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, 35:73–101.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics.* John Wiley & Sons, Hoboken, NJ, 2nd edition.

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55:163–172.

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

Johansen, S. and Nielsen, B. (2009). Saturation by indicators in regression models. In *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*, pages 1–36. Oxford University Press, Oxford.

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

Johansen, S. and Nielsen, B. (2016a). Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli*, 22:1131–1183. Corrigendum (2019) 25, 3201.

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

Rousseeuw, P. (1985). Least median of squares regressions. In Grossmann, W., Pflug, G., Vincze, I., and Wertz, W., editors, *Mathematical Statistics and Applications*, pages 283–297. Reidel, Dordrecht.

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

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

Thadewald, T. and Büning, H. (2007). Jarque-Bera test and its competitors for testing normality - a power comparison. *Journal of Applied Statistics*, 34:87–105.

Víšek, J. Á. (2006). The least trimmed squares; part III: Asymptotic normality. *Kybernetika*, 42:203–224.

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