



Department of Economics Discussion Paper Series

White heteroscedasticity testing after outlier removal

Vanessa Berenguer-Rico and Ines Wilms

Number 853
June, 2018

White heteroscedasticity testing after outlier removal

Vanessa Berenguer-Rico* and Ines Wilms†

18 June 2018

Abstract

Given the effect that outliers can have on regression and specification testing, a vastly used robustification strategy by practitioners consists in: (i) starting the empirical analysis with an outlier detection procedure to deselect atypical data values; then (ii) continuing the analysis with the selected non-outlying observations. The repercussions of such robustifying procedure on the asymptotic properties of subsequent specification tests are, however, underexplored. We study the effects of such a strategy on the White test for heteroscedasticity. Using weighted and marked empirical processes of residuals theory, we show that the White test implemented after the outlier detection and removal is asymptotically chi-square if the underlying errors are symmetric. Under asymmetric errors, the standard chi-square distribution will not always be asymptotically valid. In a simulation study, we show that - depending on the type of data contamination - the standard White test can be either severely undersized or oversized, as well as have trivial power. The statistic applied after deselecting outliers has good finite sample properties under symmetry but can suffer from size distortions under asymmetric errors.

Keywords: Asymptotic theory; Empirical processes; Heteroscedasticity; Marked and Weighted Empirical processes; Outlier detection; Robust Statistics; White test.

JEL classification: C01; C10.

1 Introduction

Regression analysis and specification testing are, in general, not robust to the presence of outlying or extreme observations. Because of this, it is a widely and commonly applied strategy to start empirical analyses by detecting atypical values in the data, deselecting or dummifying them out, then reestimating the model with the selected observations and finally conducting inferences on the model with the selected observations. Welsh and Ronchetti (2006) refer to this applied methodology as the “*data-analytic strategy*” –*DAS* henceforth. The asymptotic properties of inferential procedures implemented after such a strategy are, in general, underexplored, although recent contributions have been made on this regard. Chen and Bien (2017) show that conducting standard inferences on the parameters of the model after detecting and removing outliers can be misleading. In the case of testing for normality of residuals after implementing the *DAS*, Berenguer-Rico and Nielsen (2018) have recently shown that the outlier detection and deselection stage affects the moment-based normality test. More precisely, the usual normalizing constants of the test statistic change and depend on the truncation imposed at the deselection stage and the estimation method being used. The consequences of the *DAS* on other testing procedures are, however, still unknown. In this paper we focus on the effect of the *DAS* on the White test for heteroscedasticity (White, 1980) as this statistic remains the cornerstone of much empirical work in economics (Kim, Morse and Zingales, 2006). It is worth emphasizing that rather than proposing a procedure to test for heteroscedasticity, our goal is to theoretically, as well as, computationally analyze the properties of

*Department of Economics, University of Oxford, Mansfield College, and Programme for Economic Modelling. Email address: vanessa.berenguer-rico@economics.ox.ac.uk

†Faculty of Economics and Business, KU Leuven, Belgium. Email address: ines.wilms@kuleuven.be

the White heteroscedasticity test after having implemented the *DAS* –an extensively used procedure by practitioners.

Studying the consequences that the presence of outliers and non-normal errors may have when testing for heteroscedasticity has a long tradition in econometrics and statistics. Bickel (1978) introduces a heteroscedasticity test in the spirit of Ascombe’s (1961) test that is robust to the presence of non-normal errors. Carroll and Ruppert (1981) generalize Bickel’s test in such a way that new estimation procedures can be considered. Koenker (1981) shows that the Breusch and Pagan (1979) heteroscedasticity test has correct asymptotic size only under normality. Koenker and Bassett (1982) develop a heteroscedasticity test based on quantile regression which is robust to departures from normality. Through simulations, Evans (1992) studies the effect of non-normal errors on a battery of tests for heteroscedasticity and shows that they tend to be very sensitive to the violation of the kurtosis assumption. Godfrey (1996) shows that the Glesjer (1969) heteroscedasticity test is not asymptotically valid when the distribution of the error term is asymmetric whereas the Koenker (1981) test for heteroscedasticity is susceptible to skewness in finite samples. Lyon and Tsai (1996) study the effect of outliers in a set of eight different tests for heteroscedasticity showing that the White test seems to be highly affected. Van Dijk, Franses and Lucas (1999) show that the asymptotic as well as the finite sample size and power of the Lagrange Multiplier test for autoregressive conditional heteroscedasticity (ARCH) is severely affected by the presence of additive outliers. Im (2000) and Machado and Santos Silva (2000) propose alternative robustifications of the Glesjer tests to non-normal errors. Lebreton and Peguin-Feissolle (2007) propose two tests for heteroscedasticity –one based on artificial neural networks, the other on a Taylor expansion– which after a correction are robust to non-normality. Kalina (2011) analyzes the properties of the Breusch and Pagan (1979) and Goldfeld and Quandt (1965) tests for heteroscedasticity in robust regressions such as quantile or least trimmed squares regressions. Alih and Ong (2015) suggest an outlier-resistant version of the Goldfeld and Quandt (1965) test for heteroscedasticity which is in accordance with the *DAS* described above, but no theoretical results or finite sample size analysis are presented.

In this paper, we contribute to this literature by being, to the best of our knowledge, the first paper to theoretically and computationally investigate the effects that the *DAS* has on the White test for heteroscedasticity. Specifically, by using marked and weighted empirical processes of residuals techniques, we show that the *DAS* version of the White test statistic is asymptotically χ^2 under symmetric errors. Under an asymmetric distribution of the error term at least two effects are noted. First, the estimators implemented after deselecting outliers may be inconsistent. Second, the detection and deselection of outliers introduce biases in the asymptotic expansions of the White test statistic. These two issues can change the asymptotic distribution of the test statistic under asymmetric errors and the usual critical values become invalid.

In a number of Monte Carlo experiments we show that (a) the standard White test statistic can be either severely oversized or extremely undersized, as well as have trivial power, depending on the configuration of contamination in the data; (b) the *DAS* version of the White test enjoys good finite sample properties in terms of both size and power under normality but (c) may suffer from size distortions under asymmetric errors.

The paper is organized as follows. In Section 2 the model and assumptions are put forward. Section 3 analyzes the asymptotic properties of the *DAS* version of the White heteroscedasticity test and shows that it is asymptotically chi-square under symmetric errors. A discussion of the effect of asymmetric errors follows. Section 4 contains the results from the Monte Carlo experiments conducted to study the size and power properties of both the standard and the *DAS* version of the White test under various types of data contamination, symmetric and asymmetric error distributions. Section 5 concludes. All the proofs are collected in the Appendix.

2 Model and Assumptions

We consider the following linear model

$$y_i = \mu + z_i' \alpha + \varepsilon_i = \beta' x_i + \varepsilon_i, \quad (2.1)$$

where $\beta = (\mu, \alpha)'$ and $x_i = (1, z_i)'$ are $k \times 1$. The error term is assumed to be independent and identically distributed with unknown distribution function $F(c) = \mathbb{P}(\varepsilon_i \leq \sigma c)$ and unknown scale $\sigma > 0$. The *DAS* can be formalized as follows. Let $\tilde{\beta}, \tilde{\sigma}$ be some initial estimators of the unknown β and σ . From these estimates compute the first step residuals $\tilde{\varepsilon}_i = y_i - x_i' \tilde{\beta}$ and select observations satisfying $|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c$ for a given cut-off value, c , set by the investigator. Once the outlying observations have been identified, a new regression is conducted, which only includes the retained observations. Denote the estimator in the second stage $\hat{\beta}$. For instance, one could run a least squares regression on the truncated sample, in which case,

$$\hat{\beta} = \left\{ \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma} c)} \right\}^{-1} \sum_{i=1}^n x_i y_i \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma} c)}.$$

This gives updated residuals $\hat{\varepsilon}_{i,c} = (y_i - x_i' \hat{\beta}) \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma} c)} = \hat{\varepsilon}_i \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma} c)}$ for a second stage estimator $\hat{\beta}$.

We consider a standard cross-sectional regression model where the vector (ε_i, x_i) is independent and identically distributed and the innovations ε_i are independent of x_i . To proof some of our main results, we use the marked and weighted empirical processes of residuals theory developed in Berenguer-Rico and Nielsen (2018) –BN18 henceforth. In order to apply this machinery, we describe our cross-sectional setup using the following assumptions.

Assumption 2.1 *Let \mathcal{F}_i be an increasing sequence of σ -fields so that ε_{i-1}, x_i are \mathcal{F}_{i-1} -measurable and ε_i/σ is independent of \mathcal{F}_{i-1} with a positive density f on \mathbb{R} with derivative \dot{f} . Suppose:*

- (a) *moments: $\mathbb{E}(\varepsilon_1^8) < \infty$;*
- (b) *boundedness: $\sup_{u \in \mathbb{R}} \{(1 + |u|^9) f(u) + (1 + u^8) |\dot{f}(u)|\} < \infty$;*
- (c) *tail monotonicity of $u^8 f(u)$.*

Assumption 2.2 *Let z_i be independent and identically distributed. Let $Z_i = \{z_i', \text{vech}(z_i z_i')'\}'$ where the *vech* operator vectorizes the upper triangle of its symmetric argument so that $\mathbb{E}|Z_i Z_i'|^{2+\omega} < \infty$ for some $\omega > 0$.*

Assumption 2.3 *The first and second stage estimation errors, properly normalized, are bounded in probability, that is: $n^{1/2}(\tilde{\beta} - \beta) = \text{O}_{\mathbb{P}}(1)$, $n^{1/2}(\tilde{\sigma} - \sigma) = \text{O}_{\mathbb{P}}(1)$, $n^{1/2}(\hat{\beta} - \beta) = \text{O}_{\mathbb{P}}(1)$.*

The requirement on the existence of moments on the error term in Assumption 2.1(a) is due to the use in our proofs of the marked and weighted empirical processes of residuals theory developed in BN18. In particular, the authors derive an iterated exponential martingale inequality –based on Bercu and Touati (2008) and Johansen and Nielsen (2016) inequalities– which requires, in our setting, that the fourth moment of ε_i^2 exists. The boundedness and monotonicity conditions in Assumption 2.1(b, c) are used in BN18 –albeit in a more general setting– to control the marked and weighted empirical processes via a chaining argument. Overall, Assumption 2.1 on the distribution of the error term is satisfied by, for instance, the normal and Student t distributions.

Assumption 2.2 considers independent and identically distributed regressors. The required moment assumption is due to the fact that Z_i appears in the auxiliary regression of the White test. Both Z_i and $Z_i Z_i'$ are then part of the weights in the marked and weighted empirical processes theory used in the Appendix to derive the asymptotic distribution of the *DAS* version of the White test statistic. This requires that the second moments of the weights exist.

The boundedness of the rescaled estimators in Assumption 2.3 applies to a variety of procedures. In our simulations section, special attention will be devoted to the robustified least squares procedure, that is, the case in which the least squares estimator is used in the first stage of the *DAS*, that is $\tilde{\beta}, \tilde{\sigma}$ are least squares estimates of β and σ . Other robust estimation procedures, such as the least trimmed squares estimator (Rousseeuw, 1984) or the MM-estimator (Yohai, 1987), will also be considered.

3 White Heteroscedasticity Test after Outlier Removal

The *DAS* version of the White heteroscedasticity test consists in applying the standard White's statistic to the updated “cleaned” residuals

$$\hat{\varepsilon}_{i,c} = (y_i - x_i' \hat{\beta}) 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} = \hat{\varepsilon}_i 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)}.$$

In this section, the asymptotic properties of the *DAS* White heteroscedasticity test are analyzed. This can be formalized as follows. The test is based on the multiple correlation coefficient R^2 from the auxiliary regression

$$\hat{\varepsilon}_{i,c}^2 = \gamma 0 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} + \gamma' \tilde{Z}_{i,c} + u_{i,c}, \quad (3.1)$$

with $\tilde{Z}_{i,c} = Z_i 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)}$ and $Z_i = \{z_i', \text{vech}(z_i z_i')'\}'$ where the *vech* operator vectorizes the upper triangle of its symmetric argument. The auxiliary hypothesis is $\gamma = 0$ which amounts to $q = k - 1 + k(k - 1)/2$ restrictions.

Let $\tilde{Z}_c = n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c}$, $\hat{T}_k^c = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^k 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)}$ and

$$\begin{aligned} \hat{N}_{n,c} &= \sum_{i=1}^n \{\hat{\varepsilon}_{i,c}^2 - 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} (\hat{T}_2^c / \hat{T}_0^c)\} \{\tilde{Z}_{i,c} - 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\}, \\ \hat{M}_{n,c} &= \sum_{i=1}^n \{\tilde{Z}_{i,c} - 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\} \{\tilde{Z}_{i,c} - 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\}', \\ \hat{D}_{n,c} &= \sum_{i=1}^n \{\hat{\varepsilon}_{i,c}^2 - 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma}_c)} (\hat{T}_2^c / \hat{T}_0^c)\}^2. \end{aligned}$$

The White test applied to the robustified residuals satisfying $|\hat{\varepsilon}_i / \bar{\sigma}| \leq c$ is based on the statistic

$$\hat{T}_0^c n \hat{R}_{n,c}^2 = \hat{T}_0^c n \frac{\hat{N}_{n,c}' \hat{M}_{n,c}^{-1} \hat{N}_{n,c}}{\hat{D}_{n,c}}, \quad (3.2)$$

which we refer to as the *DASWhite* test statistic.

Let $\tau_k^c = \mathbb{E}(\varepsilon_i / \sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)}$. Let also $Z_{i,c} = Z_i 1_{(|\varepsilon_i| \leq \sigma c)}$ and $\bar{Z}_c = n^{-1} \sum_{i=1}^n Z_{i,c}$. For later reference let $T_k^c = n^{-1} \sum_{i=1}^n \varepsilon_i^k 1_{(|\varepsilon_i| \leq \sigma c)}$ and $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$. Define

$$\begin{aligned} \mathcal{N}_{n,c} &= \sum_{i=1}^n \{(\varepsilon_{i,c} / \sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_2^c / \tau_0^c)\} \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_0^c)^{-1} \bar{Z}_c\}, \\ \mathcal{M}_{n,c} &= \sum_{i=1}^n \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_0^c)^{-1} \bar{Z}_c\} \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_0^c)^{-1} \bar{Z}_c\}', \\ \mathcal{D}_{n,c} &= \sum_{i=1}^n \{(\varepsilon_{i,c} / \sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_2^c / \tau_0^c)\}^2. \end{aligned}$$

Our first result is about the difference between $\hat{T}_0^c n \hat{R}_{n,c}^2$ and

$$\tau_0^c n R_{n,c}^2 = \tau_0^c n \frac{\mathcal{N}_{n,c}' \mathcal{M}_{n,c}^{-1} \mathcal{N}_{n,c}}{\mathcal{D}_{n,c}}. \quad (3.3)$$

Theorem 3.1 *Consider model (2.1) and suppose that Assumptions 2.1 with a symmetric density f , 2.2 and 2.3 are satisfied. Then, uniformly in c ,*

$$\hat{T}_0^c n \hat{R}_{n,c}^2 = \tau_0^c n R_{n,c}^2 + o_{\mathbb{P}}(1).$$

Theorem 3.1 shows that the White test applied to the truncated squared residuals, that is the *DASWhite* test statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$, is asymptotically equivalent to the test applied to the true truncated squared errors, $\tau_0^c n R_{n,c}^2$. Therefore, to derive the asymptotic distribution of $\hat{T}_0^c n \hat{R}_{n,c}^2$ it suffices to analyze $\tau_0^c n R_{n,c}^2$. Using this fact, the next result then shows that $\hat{T}_0^c n \hat{R}_{n,c}^2$ is asymptotically χ_q^2 under a symmetric error distribution.

Theorem 3.2 *Consider model (2.1) and suppose that Assumptions 2.1 with a symmetric density f , 2.2 and 2.3 are satisfied. Then, for each $c > 0$, $\hat{T}_0^c n \hat{R}_{n,c}^2 \sim \chi_q^2$.*

The symmetry assumption on the density f of the standardized error term is fundamental in the derivation of the previous two theorems, specially in Theorem 3.1 when showing that $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\mathcal{N}_{n,c} + o_{\mathbf{P}}(1)$. In particular, under symmetry, the biases in the asymptotic expansion of $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c}$ coming from the estimation errors in the first and second stages of the *DAS* will cancel out so that the asymptotic equivalence $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\mathcal{N}_{n,c} + o_{\mathbf{P}}(1)$ holds.

Under an asymmetric distribution two issues arise. First, the second stage estimator $\hat{\beta}$ could be inconsistent for the intercept as the example in the following result shows.

Theorem 3.3 *Consider model (2.1) and suppose Assumptions 2.1 and 2.2 are satisfied. Let also $n^{1/2}(\tilde{\sigma} - \sigma) = O_{\mathbf{P}}(1)$, $n^{1/2}(\hat{\beta} - \beta) = O_{\mathbf{P}}(1)$,*

$$\hat{\beta} = \{n^{-1}\sum_{i=1}^n x_i x_i' 1_{(|\varepsilon_i| < \tilde{\sigma}c)}\}^{-1} \sum_{i=1}^n x_i y_i 1_{(|\varepsilon_i| < \tilde{\sigma}c)},$$

and $\Sigma_{n,c}^{xx} = n^{-1}\sum_{i=1}^n x_i x_i' 1_{(|\varepsilon_i| < \sigma c)}$. Then,

$$\hat{\beta} - \beta = (\Sigma_{n,c}^{xx})^{-1} n^{-1} \sum_{i=1}^n x_i \varepsilon_i 1_{(|\varepsilon_i| < \sigma c)} + O_{\mathbf{P}}(n^{-1/2}). \quad (3.4)$$

From expansion (3.4) in Theorem 3.3 we see that the term $n^{-1}\sum_{i=1}^n x_i \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)}$ may not converge to zero since the first element is $n^{-1}\sum_{i=1}^n \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} = \tau_0^c + o_{\mathbf{P}}(1)$. Hence, the intercept μ in model (2.1) is not going to be consistently estimated, unless $\tau_0^c = 0$. This inconsistency could make Assumption 2.3 invalid since $n^{1/2}(\hat{\beta} - \beta)$ may not be bounded in probability. Even if Assumption 2.3 is satisfied so that the second stage estimator satisfies $n^{1/2}(\hat{\beta} - \beta) = O_{\mathbf{P}}(1)$, a second issue arises under asymmetric errors. In particular, the result $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\mathcal{N}_{n,c} + o_{\mathbf{P}}(1)$ would not hold since the estimation errors coming from the first and second stages of the *DAS* will induce biases in the asymptotic expansion of $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c}$. More precisely, under an asymmetric f , we have the following result.

Theorem 3.4 *Consider model (2.1) and suppose Assumptions 2.1, 2.2, and 2.3 are satisfied. Then,*

$$n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\mathcal{N}_{n,c} + \mathcal{G}_{n,c} + o_{\mathbf{P}}(1),$$

where $\mathcal{G}_{n,c} = \mathcal{G}_{1n}^c - \mathcal{G}_{mn}^c$ and

$$\begin{aligned} \mathcal{G}_{1n}^c &= c^2 \{f(c) - f(-c)\} \sigma^{-1} n^{-1/2} \sum_{i=1}^n (Z_i - \bar{Z}) x_i' (\hat{\beta} - \beta), \\ \mathcal{G}_{mn}^c &= 2\tau_1^c \sigma^{-1} n^{-1/2} \sum_{i=1}^n (Z_i - \bar{Z}) x_i' (\hat{\beta} - \beta). \end{aligned}$$

Theorem 3.4 derives an asymptotic expansion for $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c}$ under asymmetric errors. \mathcal{G}_{1n}^c and \mathcal{G}_{mn}^c are the biases coming from the first and second stages of the *DAS*, respectively. Under symmetry $\mathcal{G}_{1n}^c = \mathcal{G}_{mn}^c = 0$, since $f(c) = f(-c)$ and $\tau_1^c = 0$, and hence the biases cancel out. Under asymmetric errors, however, these biases may not cancel and $\mathcal{G}_{n,c}$ will affect the asymptotic distribution. Notice that in that case, the *DAS* White test is not self-normalized. This suggests that one could perhaps correct the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ to make the *DAS* White test robust to asymmetries, though this is not a trivial task. Also, the potential inconsistency of the second stage estimator, $\hat{\beta}$, should be addressed in the asymmetric case. Hence, deriving a *DAS* White test that is robust to asymmetries is beyond the scope of this paper and left for future research.

It is also worth mentioning that in the asymmetric case, if the distribution of the error term were known, then it would be possible to choose asymmetric cut-off values to cancel out the biases. To be more precise, instead of truncating using $1_{(|\varepsilon_i| \leq \sigma c)}$, one could use $1_{(\sigma c_1 \leq \varepsilon_i \leq \sigma c_2)}$. Then, by choosing the cut-off values c_1, c_2 , so that $f(c_1) = f(c_2)$ and $\tau_1^{c_1, c_2} = \mathbf{E}(\varepsilon_i / \sigma) 1_{(\sigma c_1 \leq \varepsilon_i \leq \sigma c_2)} = 0$, the biases would also disappear. However, knowing the error distribution will be quite unlikely in practice.

4 Finite Sample Performance

We consider the multiple regression model

$$y_i = \mu_0 + \mu_1' z_i + \varepsilon_i,$$

for $i = 1, \dots, n$ with $\mu_0 = 0$, and μ_1 a p -dimensional vector of zeros. The p predictor variables are drawn from a standard multivariate normal distribution and independently of $\varepsilon_i \sim i.i.d.N(0, \sigma_i^2)$. The sample sizes are $n \in \{50, 100, 500, 1000\}$ and the number of predictors $p \in \{1, 2, 4\}$.

We consider the following four test statistics to test the null hypothesis of homoscedasticity: (i) *White*: the standard White test, (ii) *DASWhite LS*: the *DAS* White test with the least squares as initial estimator, (iii) *DASWhite LTS*: the *DAS* White test with the least trimmed squares as initial estimator, (iv) *DASWhite MM*: the *DAS* White test with the MM estimator as initial estimator. The later three follow the data-analytic strategy to obtain the “*robustified*” White test of equation (3.2). Observations in the initial regression are flagged as outlying if their absolute standardized residual exceeds $c = 1.96$ (the 97.5th quantile of the standard normal). Then 5% of the observations are expected to be flagged as outliers in the normal model. Other cut-off values, such as $c = 2.24$ and $c = 1.64$, are also considered.

The least trimmed squares estimator is computed with the “*ltsReg*” function in *R* using standard settings and 25% breakdown point. The MM-estimator is computed with the “*lmrob*” function in *R* using standard settings of 50% breakdown point and an efficiency of 95% at the normal model, using the biweight loss function. For more details on these robust estimators, see e.g. Maronna, Martin and Yohai (2006).

4.1 Size. No contamination

To study the size of the test statistics, we take a data generating process under the null hypothesis and set $\sigma_i^2 = 1$ for $1 \leq i \leq n$. We simulate $M = 10000$ of these data generating processes and compute the empirical size, i.e. the percentage of simulation runs where the null is rejected. Table 1 gives the empirical sizes of the four test statistics for a nominal size of $\alpha = 0.05$ and cut-off values $c = 1.96$, $c = 2.24$ and $c = 1.64$. Results for nominal sizes $\alpha = 0.01$ and $\alpha = 0.1$ are similar and available from the authors upon request.

Table 1: Empirical sizes of the four tests in uncontaminated settings, nominal size 5%.

Constant	Test Statistic	$n = 50$			$n = 100$			$n = 500$			$n = 1000$		
		$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$
	<i>White</i>	4.58	5.11	4.81	4.57	5.37	5.96	5.04	5.44	5.87	4.98	5.11	5.24
$c = 1.96$	<i>DASWhite LS</i>	4.20	4.15	3.13	4.52	4.39	4.05	4.89	4.45	4.95	4.85	4.62	4.83
	<i>DASWhite LTS</i>	2.96	2.48	1.62	3.75	3.12	2.92	4.92	4.41	4.54	4.94	4.50	4.68
	<i>DASWhite MM</i>	3.77	3.83	2.68	4.46	4.33	3.89	4.83	4.43	4.72	4.89	4.75	4.59
$c = 2.24$	<i>DASWhite LS</i>	4.53	4.84	3.83	4.79	4.89	4.70	4.96	4.94	5.24	4.73	4.98	5.05
	<i>DASWhite LTS</i>	3.25	3.01	2.19	4.04	3.72	2.97	4.65	4.78	4.69	4.72	4.92	4.88
	<i>DASWhite MM</i>	4.20	4.35	3.28	4.44	4.54	4.31	4.87	4.96	4.94	4.68	5.04	5.10
$c = 1.64$	<i>DASWhite LS</i>	3.97	3.99	2.78	4.69	4.61	4.07	4.88	4.86	4.74	4.87	4.87	4.85
	<i>DASWhite LTS</i>	3.10	2.59	1.38	4.37	3.66	2.99	4.83	4.66	4.14	4.85	5.02	4.77
	<i>DASWhite MM</i>	3.85	3.72	2.59	4.59	4.40	3.85	4.72	4.74	4.68	5.02	4.98	4.70

As expected, the standard *White* test has the overall best performance in these settings without outliers. Next, consider the results for cut-off $c = 1.96$. The *DAS* White test statistics are, overall, competitive to the standard *White*. For a small sample size $n = 50$, size distortions occur in *DASWhite LTS* and *DASWhite MM* due to the loss in efficiency. These distortions are most

severe for *DASWhite LTS*, which is known to suffer from a statistical efficiency problem. As the sample size increases, these size distortions disappear. All test procedures result in empirical sizes close to the nominal size for $n = 1000$.

The same Monte Carlo experiments were conducted for different values of the cut-off c to investigate the sensitivity of the results to this choice. These results, collected in Table 1, are summarized as follows. For larger values of c (i.e. $c = 2.24$, the 98.75th quantile of the standard normal), less observations are expected to be flagged as outlying and the results indicate that (i) the performance of *DASWhite LS* gets closer to that of the standard *White*, (ii) size distortions for *DASWhite LTS* and *DASWhite MM* are still present for $n = 50$, though less severe compared to the $c = 1.96$ case.

For smaller values of c (i.e. $c = 1.64$, the 95th quantile of the standard normal), more observations are expected to be flagged as outlying and we find that *DASWhite LS*, *DASWhite LTS* and *DASWhite MM* require a larger sample size (compared to the $c = 1.96$ case) for the size distortions to disappear. For $n = 1000$, all empirical sizes are close to the nominal size.

4.2 Power. No contamination

To evaluate the power of the test statistics, we take a data generating process under the alternative hypothesis and set

$$\sigma_i^2 = 1 + \delta \sum_{j=1}^p x_{ij}^2,$$

where the parameter $\delta \neq 0$ measures the deviation from the null. We take a logarithmic spaced grid of ten δ -values between 0.01 and 5. We simulate $M = 10000$ data generating processes under the alternative and compute the empirical power, i.e. the percentage of simulation runs where the null hypothesis is rejected, for the different δ -values. We discuss the results for significance level of $\alpha = 0.05$, cut-off $c = 1.96$, sample sizes $n \in \{100, 500, 1000\}$ and number of predictors $p \in \{1, 2, 4\}$. Figure 1 gives the power curves of the four test statistics for the different combinations of sample sizes n and number of predictors p .

The results for $n = 100$ indicate that the power of all test statistics increases as the deviation from the null increases, as expected. The standard *White* is considerably more powerful than the other test procedures even for large deviations from the null. The *DAS White* test statistics result in an efficiency loss, and as a consequence a power loss compared to the standard *White*, due to their outlier flagging in this uncontaminated setting.

As the sample size increases to $n = 500$ and $n = 1000$, the standard *White* remains more powerful than the other test statistics for small deviations from the null. However, this power loss disappears for larger deviations from the null. Specifically, for larger sample sizes, all test statistics detect the deviations from the null in each simulation run, resulting in a perfect power.

The results of the *DAS White* test statistics are based on the cut-off value $c = 1.96$. We also conducted the Monte Carlo experiment for $c = 2.24$ and $c = 1.64$. Results are available from the authors upon request. Overall, the main findings remain unchanged: the standard *White* is the best performing, and the robustified test procedures are equally powerful for large deviations from the null as long as the sample size is large. However, if more observations are expected to be flagged as outlying (i.e. smaller values of c), the *DAS* procedures lose in power compared to the results from Figure 1. The robustified procedures now require larger deviations from the null (i.e. $\delta > 2$) to obtain perfect power. The reverse occurs for larger values of the cut-off c .

4.3 Size. Contamination

Although our theoretical results do not cover the contaminated settings considered here, we also compare the empirical sizes (and power in the next section) of the four test statistics in the presence

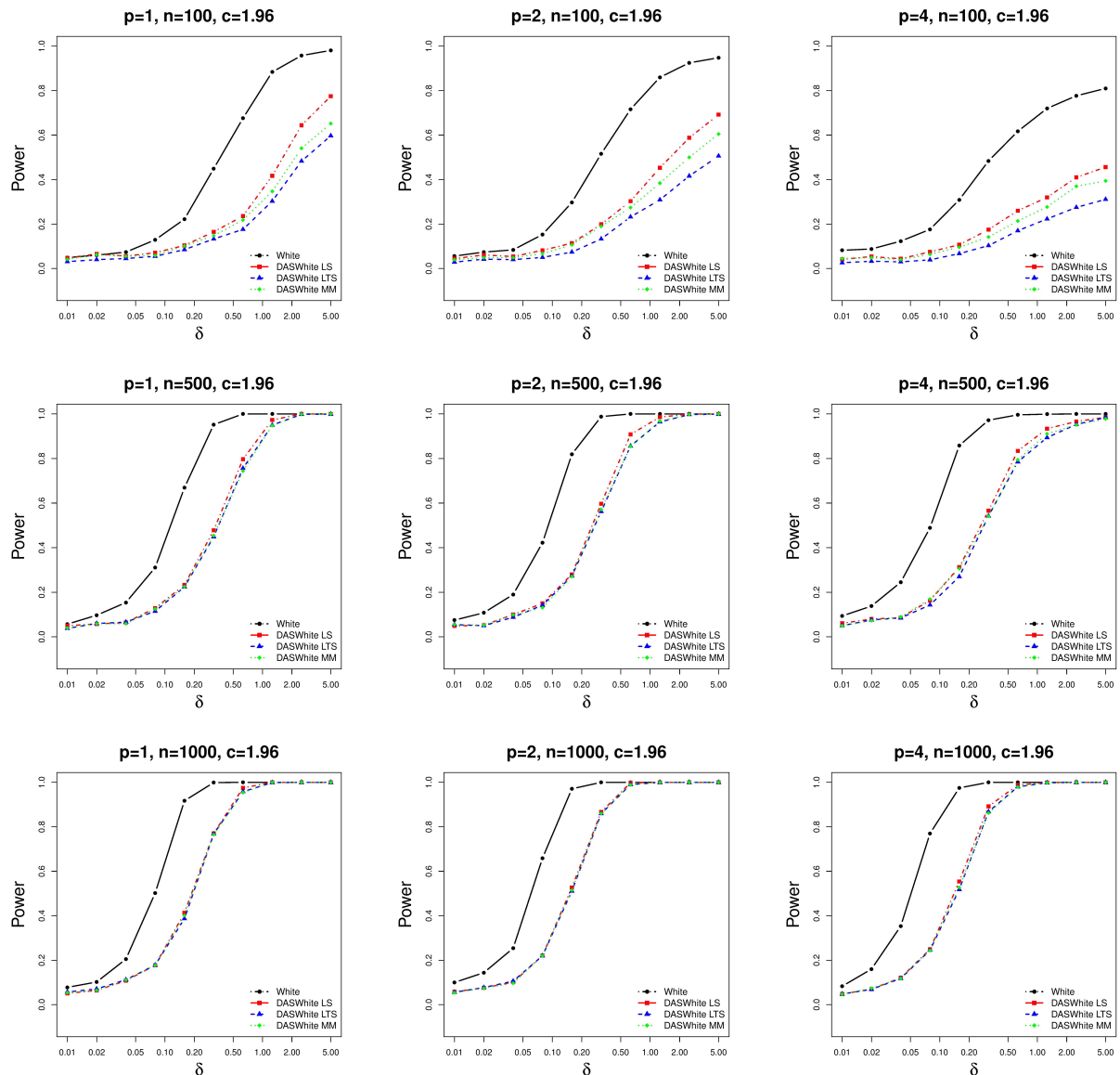


Figure 1: Power of the four tests in uncontaminated settings, significance level 5%.

of outliers. To our knowledge, no Monte Carlo results on the *DAS White* test statistics under contaminated settings have been discussed in the literature and hence, it is worth analyzing their relative performance in depth. To this end, we consider three contamination schemes where we add contamination to the clean simulation set-up. In the first two schemes, we only contaminate the first two observations: the first observation is located at $(x, y) = (x, 10)$, the second at $(x, y) = (x, -10)$. In the “bad leverage points” scheme, we take $x = 5$, giving observations outlying in both the x - and y -direction, known as bad leverage points in the robustness literature. In the second “vertical outliers” scheme, we take $x = 0$, giving observations outlying in the y -direction only, known as vertical outliers. In the third “masked outliers” contamination scheme, we replace the first 5% of the observations by a point mass contamination located at $(x, y) = (10, 20)$. This last scheme will be used to compare the outlier detection performance among the different *DAS White* test statistics. Table 2 gives the empirical sizes of the four test statistics for nominal size $\alpha = 0.05$, $p = 1$, and $c = 1.96$. To focus the discussion on the type of outlier contamination we only report the results for the case $p = 1$ and $c = 1.96$. Results for other values of p and c are all similar, hence omitted, but available upon request.

First, consider the “bad leverage points” contamination scheme. The standard *White* test is heavily affected by the introduction of the outliers: it rejects the null of homoscedasticity in each

simulation run, resulting in empirical sizes of 100%. The chosen outlier configuration artificially inflates the variance for large predictor values. The standard *White* test has power to detect this behavior, but confuses the outliers with heteroscedasticity in doing so. We did consider other values for the y -coordinate (other than $|y| = 10$) of the outliers. As soon as $|y| > 3$, the standard *White* test becomes oversized.

Table 2: Empirical sizes of the four tests in contaminated settings, nominal size 5%.

Outlier configuration	Test Statistic	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
Bad leverage points	<i>White</i>	100.00	100.00	100.00	100.00
	<i>DASWhite LS</i>	4.47	5.00	5.04	5.03
	<i>DASWhite LTS</i>	3.34	3.95	5.08	4.91
	<i>DASWhite MM</i>	4.10	4.39	5.25	4.79
Vertical outliers	<i>White</i>	0.00	0.00	0.00	0.00
	<i>DASWhite LS</i>	4.13	4.60	4.66	5.00
	<i>DASWhite LTS</i>	2.74	4.25	4.84	5.19
	<i>DASWhite MM</i>	3.72	4.62	4.86	5.27
Masked outliers	<i>White</i>	87.86	99.53	100.00	100.00
	<i>DASWhite LS</i>	40.35	63.99	100.00	100.00
	<i>DASWhite LTS</i>	3.13	4.34	5.04	5.05
	<i>DASWhite MM</i>	3.96	4.43	5.12	4.83

The *DAS White* test statistics, in contrast, maintain their size. The initial estimators correctly flag the first two observations as outliers, and hence these observations do not influence the test statistic of equation (3.2). As a result, they remain accurately sized even in the presence of the considered outliers. Similar to the results without contamination, there are only minor size distortions for small sample sizes.

Next, consider the “vertical outliers” contamination scheme. The standard *White* test is again heavily affected by the presence of the outliers. However, under this configuration, it behaves very differently compared to the first outlier configuration: the outliers remain unnoticed and the standard *White* test is unable to reject the null of homoscedasticity, resulting in zero empirical sizes. The robustified test statistics, in contrast, correctly detect the outlying observations as atypical, and remove them from the data set. As the sample size increases, their empirical sizes get closer to the nominal size of 5%.

Finally, consider the “masked outliers” contamination scheme. Similar to the “bad leverage points” contamination, the standard *White* test is severely oversized. However, in contrast to the previous contamination schemes, *DASWhite LS* is also severely oversized. This occurs since the outliers interact in such a way that they are not detected by the initial least squares estimation. This effect is known as the masking effect (Maronna, Martin and Yohai; 2006). As outlined in Section 2, the outlier detection method compares standardized residual values with a user-specified cut-off value to separate good observations (with small standardized residuals) from outlying ones (with large standardized residuals). However, in this contamination scheme, the least squares regression line is drawn towards the outliers. This results in large standardized residuals for clean observations and small ones for the outliers that thus remain unnoticed. As a result of this masking effect, *DASWhite LS* is severely oversized. The masking effect can be avoided by using an initial robust estimator - such as the *LTS* or *MM* - to separate good observations from outlying observations. Indeed, both *DASWhite LTS* and *DASWhite MM* remain accurately sized under this contamination scheme.

4.4 Power. Contamination

The contamination settings outlined when discussing the size of the test statistic were chosen to illustrate the severe size distortions that the standard *White* test statistic can suffer from. We now focus on a contamination setting that severely impacts its power. To this end, we add contamination to the clean simulation set-up by replacing the response value of each of the first 5% observations by a draw from a normal distribution with mean 100 and standard deviation 1.

Figure 2 gives the power curves of the four test statistics under the contaminated simulation settings for cut-off $c = 1.96$. The results for $c = 1.64$ and $c = 2.24$ are similar and available from the authors upon request. In all considered settings, the performance of the standard *White* test is severely influenced by the presence of the outliers: the outliers mask the heteroscedasticity and, as a consequence, the standard *White* test no longer has power to reject the null. All outlying observations have large residual values. The variance of the residuals is thus inflated regardless of their predictor values. As a consequence, the null hypothesis of homoscedasticity is only rejected in a small minority of the simulation runs, and the power does not increase with larger deviations from the null. Even when increasing the sample size from $n = 100$ to $n = 1000$, the standard *White* test continues to have no power.

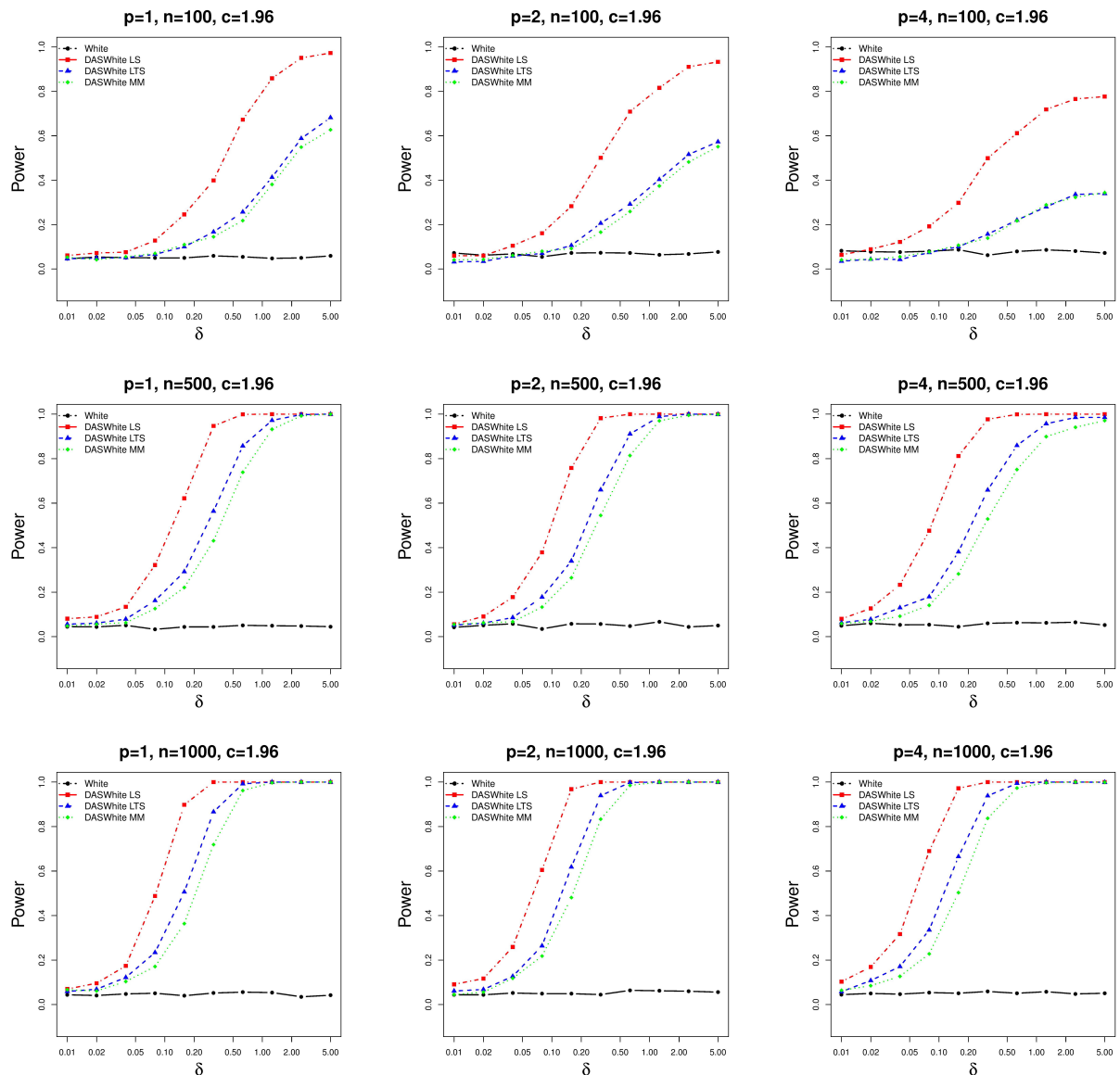


Figure 2: Power of the four tests in contaminated settings, significance level 5%.

The *DAS* White procedures, on the other hand, maintain their power. Their power increases as the sample size increases and/or the deviation from the null is larger. *DASWhite LS* is more powerful than the other two procedures, and this regardless of the deviation from the null.

4.5 Asymmetric Error Distribution

In this section, we discuss the finite sample performance of the test statistic under asymmetric error distributions. As discussed above, two issues arise in this case: (i) the possible inconsistency of the second stage estimate for the intercept, and (ii) the appearance of biases in the asymptotic expansion as a result of the first and second stages of the *DAS* approach. As a consequence, the finite sample performance of the tests can be severely impacted as we illustrate with a simple Monte Carlo simulation experiment.

We consider the data generating process without contamination but with the following seven asymmetric error distributions: (i-ii) chi-square with one and four degrees of freedom, respectively; (iii) a lognormal where $\epsilon_i = u_i + 1.27$ with u_i a lognormal with mean 0 and standard deviation 0.70 –then, similar to Im (2000) and Ali and Giaccotto (1984), the expected value and variance of the errors are approximately zero and one, respectively; (iv) a normal mixture where the errors are $\epsilon_i \sim 0.5N(-2, 1) + 0.5N(2, 4)$, (v) an exponential with parameter one; (vi-vii) beta densities with shape parameters 2 and 20 (right skewed), and 20 and 2 (left skewed).

In Table 3, we give the empirical sizes of the four test statistics (for a nominal size of $\alpha = 0.05$) under the considered asymmetric error distributions. Compared to the symmetric distribution case, we have added results for $n = 10000$ to illustrate that, in most cases, the size distortions do not disappear even for very large finite sample sizes. For simplicity, we use a cut-off $c = 1.96$ for the *DAS* test statistics since in practice we do not know the underlying error distribution.

Table 3: Empirical sizes of the four tests under asymmetric error distributions, nominal size 5%, $c = 1.96$.

Error Distribution	Test Statistic	$n = 50$			$n = 100$			$n = 500$			$n = 1000$			$n = 10000$		
		$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$
$\chi^2(1)$	<i>White</i>	7.10	9.81	12.76	6.22	9.26	12.25	5.75	7.07	10.19	5.30	6.61	9.68	4.82	5.75	6.50
	<i>DASWhite LS</i>	8.32	10.35	11.60	8.02	9.77	12.18	8.00	8.57	9.84	7.29	8.08	9.06	7.46	8.32	8.90
	<i>DASWhite LTS</i>	7.21	6.98	5.45	7.84	7.97	7.70	8.86	10.05	10.90	9.92	10.44	11.66	9.57	10.92	12.07
	<i>DASWhite MM</i>	9.63	10.57	9.66	9.69	10.95	11.64	9.85	11.24	12.18	9.92	10.44	12.40	10.25	10.76	12.26
$\chi^2(4)$	<i>White</i>	6.02	8.00	9.20	5.92	7.59	10.37	5.20	7.05	9.14	5.18	5.76	7.89	5.02	5.49	5.89
	<i>DASWhite LS</i>	6.80	7.62	6.62	7.04	7.78	8.41	7.10	8.26	8.40	7.12	7.88	8.56	7.61	7.99	8.70
	<i>DASWhite LTS</i>	4.55	4.68	3.01	5.74	5.99	5.28	6.74	6.99	7.05	6.99	7.18	7.64	7.27	7.72	8.43
	<i>DASWhite MM</i>	6.78	7.49	5.30	6.92	7.90	7.50	7.55	8.15	8.86	7.65	8.51	9.02	7.94	8.35	9.55
Lognormal	<i>White</i>	6.55	9.25	12.19	6.14	9.03	12.02	5.57	7.83	11.22	5.54	7.28	10.36	5.33	6.47	7.97
	<i>DASWhite LS</i>	7.08	9.38	8.69	7.26	8.02	9.85	6.65	7.71	9.03	7.28	7.81	8.42	7.16	7.89	8.14
	<i>DASWhite LTS</i>	4.92	5.01	2.99	5.69	6.06	5.63	7.14	6.93	8.15	7.42	8.40	8.18	7.49	7.91	8.66
	<i>DASWhite MM</i>	6.96	8.30	6.16	7.77	8.73	8.65	7.93	8.57	9.15	8.36	9.43	9.82	8.21	9.60	9.57
Normal Mixture	<i>White</i>	5.15	5.53	5.17	5.05	5.48	5.58	4.91	5.63	5.50	4.87	5.32	5.46	5.29	5.15	4.84
	<i>DASWhite LS</i>	6.04	5.99	4.06	6.73	6.88	5.95	8.02	7.96	8.39	8.02	7.73	8.33	7.37	7.43	8.70
	<i>DASWhite LTS</i>	4.58	4.22	2.29	5.78	5.81	4.14	7.38	7.60	7.20	7.58	7.60	7.47	7.56	7.38	8.56
	<i>DASWhite MM</i>	5.91	5.65	3.71	6.75	6.85	5.58	7.61	8.27	8.47	8.14	7.76	8.52	7.52	7.83	8.96
Exponential	<i>White</i>	6.01	8.79	10.90	5.58	8.36	11.50	4.73	6.77	9.51	5.36	6.77	8.53	4.99	5.47	6.00
	<i>DASWhite LS</i>	7.74	9.72	8.70	7.22	9.55	10.29	7.91	8.33	9.75	7.90	8.64	9.70	8.00	8.57	9.05
	<i>DASWhite LTS</i>	5.78	5.78	3.88	7.13	7.14	6.08	8.55	8.49	8.93	8.23	9.50	9.50	8.56	9.64	10.23
	<i>DASWhite MM</i>	7.75	9.24	6.99	9.00	9.20	9.50	9.36	9.82	11.44	9.08	9.92	11.16	8.95	10.22	10.68
Beta (left skewed)	<i>White</i>	5.57	7.87	8.37	5.09	6.51	9.46	5.31	6.38	8.20	5.00	6.15	7.46	4.81	5.24	5.50
	<i>DASWhite LS</i>	6.18	7.52	6.14	7.31	7.47	7.63	6.82	8.08	8.51	7.49	8.03	8.67	7.37	8.11	8.09
	<i>DASWhite LTS</i>	4.57	4.46	2.64	5.96	5.70	4.51	6.70	7.15	6.55	7.17	7.07	8.01	7.49	7.83	8.13
	<i>DASWhite MM</i>	6.33	6.91	5.25	7.06	7.39	7.14	6.93	8.38	8.68	7.74	8.26	9.04	7.94	8.42	8.86
Beta (right skewed)	<i>White</i>	5.40	7.51	8.51	5.87	6.92	9.41	4.88	6.21	7.63	5.13	5.60	6.88	5.03	5.15	5.39
	<i>DASWhite LS</i>	6.16	7.42	5.70	6.83	7.46	7.50	7.12	7.85	8.38	7.87	7.62	8.87	7.94	7.70	8.51
	<i>DASWhite LTS</i>	4.41	4.12	2.46	5.21	5.66	4.79	6.37	7.13	7.40	7.34	7.03	7.90	7.37	7.62	7.96
	<i>DASWhite MM</i>	6.25	6.84	4.57	6.80	7.65	7.18	7.17	8.11	8.80	7.80	8.12	9.30	8.06	8.58	9.05

First, consider the standard *White* test. We find that size distortions can occur. The standard *White* test tends to reject the null too often when the underlying errors are asymmetric, and these distortions are most prevalent for the considered distributions whose skewness is the largest, i.e., the $\chi^2(1)$, lognormal and exponential. Size distortions are most severe for larger number of predictors and/or small sample sizes, and also persist for some error distributions even if the sample size increases. As an example, the empirical size for the standard *White* in the lognormal case is still

around 8% for $p = 4$ even if $n = 10000$. For the normal mixture, on the other hand, the empirical sizes are still close to the nominal size.

Next, we summarize the results for the *DAS* test statistics. We notice that size distortions that are still present for $n = 10000$ are, for most considered error distributions, more severe for the *DAS White* tests than for the standard *White*. While the standard *White* is only affected by the \mathcal{G}_{mn}^c bias in Theorem 3.4 (i.e., the bias coming from the estimation error of the residuals in the mark function), the *DAS White* tests are also affected by the \mathcal{G}_{1n}^c bias (i.e., the bias coming from the first outlier flagging stage of the *DAS*). The interplay between both biases can potentially result in more severe size distortions that persist even for large sample sizes.

In sum, this simple Monte Carlo experiment reveals that both the standard *White* and the *DAS White* tests may suffer from size distortions under asymmetric errors, even for large sample sizes.

5 Concluding Remarks

We have derived the asymptotic distribution of the White test for heteroscedasticity implemented after an initial data analytic strategy used to detect and remove atypical data values. Under symmetric errors, the standard chi-square distribution is recovered. Under asymmetric errors, the standard distribution will not always be asymptotically valid as the second stage estimate can be inconsistent for the intercept and/or additional biases appear in the asymptotic expansions.

Our Monte Carlo experiments indicate that the size and power of the standard *White* test can be severely impacted by the presence of atypical data values. In contrast, the *DAS* versions of the White test have, overall, good finite sample properties in terms of size and power in the presence of outliers although, as expected, the *DASWhite LS* may suffer from the masking effect. Finally, all *DAS White* test statistics may suffer from size distortions under asymmetric errors, even for large sample sizes. A more detailed study of the test statistics under asymmetric error distributions is left for future research.

A Empirical Processes Results

The *DASWhite* test statistic in equation (3.2) is constructed by means of two sided marked and weighted empirical distribution functions of the form

$$\hat{G}_n^{Z,k}(c) = \frac{1}{n} \sum_{i=1}^n Z_i \hat{\varepsilon}_{i,c}^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)},$$

where the weights and the marks are Z_i and $\hat{\varepsilon}_{i,c}^k$, respectively. In a more general setting, BN18 have recently developed an asymptotic theory for marked and weighted empirical processes which will be useful in proving the theoretical results in this paper.

A.1 Preliminary Lemmas

The following two results are from BN18. The first one is the well-known result that in order to analyze statistics depending on certain estimation errors, it is enough to study the behavior of these statistics for deterministic estimation errors that vary in a certain set. The second result describes the asymptotic behavior of the marked and weighted empirical processes of residuals, which forms the basis to the proof of the main results on the *DASWhite* test for heteroscedasticity analyzed in the above sections.

Lemma A.1 (BN18 Lemma A.1) *Let $\epsilon > 0$. Suppose a compact set Θ exists so $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}| \in \Theta) > 1 - \epsilon$. Then,*

$$\mathbb{P}\{|M_n(\hat{\theta}, c)| > \epsilon\} \leq \mathbb{P}\{\sup_{\theta \in \Theta} |M_n(\theta, c)| > \epsilon\} + \epsilon.$$

Let $\theta = (a_1, b_1, a_m, b_m)'$, $\theta_1 = (a_1, b_1)'$ and $\theta_m = (a_m, b_m)$. The two sided empirical distribution function of interest is defined as

$$\mathbb{G}_n^{w,k}(\theta, c) = \frac{1}{n} \sum_{i=1}^n w_{in} \left(\frac{\varepsilon_i - x'_{in} b_m}{\sigma + n^{-1/2} a_m} \right)^k 1_{(|\varepsilon_i - x'_{in} b_1| \leq \sigma c + n^{-1/2} a_1 c)},$$

with compensator $\bar{\mathbb{G}}_n^{w,k}(\theta, c) = \mathbb{E}_{i-1} \mathbb{G}_n^{w,k}(\theta, c)$. The corresponding two sided empirical process is

$$\mathbb{G}_n^{w,k}(\theta, c) = n^{1/2} \{ \mathbb{G}_n^{w,k}(\theta, c) - \bar{\mathbb{G}}_n^{w,k}(\theta, c) \}.$$

Assumption A.1 Let \mathcal{F}_i be an increasing sequence of σ -fields so that ε_{i-1}, x_i are \mathcal{F}_{i-1} -measurable and ε_i/σ is independent of \mathcal{F}_{i-1} with a positive density f on \mathbb{R} with derivative \dot{f} . Set $0 \leq \kappa \leq 1/4$ and $k = 0, 1, 2$. Suppose:

(i) density and marks satisfy, with $\tilde{m}(u)$ representing each of $m(u) = u^k$, $w\dot{m}(u) = ku^k$, $\dot{m}(u) = ku^{k-1}$:

(a) moments: $\int_{-\infty}^{\infty} \tilde{m}^4(u) f(u) du < \infty$;

(b1) boundedness: $\sup_{u \in \mathbb{R}} |u| \{1 + m^4(u)\} f(u) < \infty$;

(b2) boundedness: $\sup_{u \in \mathbb{R}} \left| \frac{\partial}{\partial u} [\{1 + \tilde{m}^4(u)\} f(u)] \right| < \infty$;

(b3) boundedness: $\sup_{u \in \mathbb{R}} (1 + u^2) |\dot{m}(u) f(u)| < \infty$;

(c) let $\tilde{h}(u) = \{1 + \tilde{m}^4(u)\} f(u)$ so that

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

(ii) regressors: $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = O_{\mathbb{P}}(1)$ where N is a non-stochastic normalization matrix;

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

In BN18 the weighted and marked empirical distribution function is defined in terms of a general mark function $m(u)$, which in the present case is $m(u) = u^k$. Note that Assumption A.1 considers $k = 0, 1, 2$. BN18 make the following assumption on their general mark function $m(u)$.

Assumption A.2 Suppose that $\exists \omega > 0$ and a function $\hat{m}(u) \geq 0$ so that $\forall u^*, u$ satisfying $|u^* - u| \leq (1 + |u|)\omega$ then $|\hat{m}(u^*) - \hat{m}(u)| \leq |u^* - u| \hat{m}(u)$ where $\int_{-\infty}^{\infty} (1 + |u|^2) \hat{m}(u) f(u) du < \infty$.

Assumption A.2 is trivially satisfied for $m(u) = u^k$ with $k = 0, 1$. For $k = 2$ we have $\dot{m}(u) = 2u$. Take $|u^* - u| \leq (1 + |u|)\omega$ so that $|\dot{m}(u^*) - \dot{m}(u)| = |2u^* - 2u| = 2|u^* - u| \leq 2(1 + |u|)\omega$. Thus we choose $\hat{m}(u) = 2$ and see that $\int_{-\infty}^{\infty} (1 + |u|^2) 2f(u) du < \infty$ by Assumption A.1(a). This shows that the local Lipschitz Assumption A.2 in BN18 is satisfied in the present setting. Hence, it is not included in Assumption A.1.

Lemma A.2 (BN18 Corollary B.1) Suppose Assumption A.1 holds. Then, for any $B > 0$ and $n \rightarrow \infty$,

$$\sup_{|\theta| \leq B} \sup_{c \in \mathbb{R}} |\mathbb{G}_n^{w,k}(\theta, c) - \mathbb{G}_n^{w,k}(0, c)| = o_{\mathbb{P}}(1),$$

and

$$\sup_{|\theta| \leq B} \sup_{c \in \mathbb{R}} |n^{1/2} \{ \bar{\mathbb{G}}_n^{w,k}(\theta, c) - \bar{\mathbb{G}}_n^{w,k}(0, c) \} - \mathcal{G}_n^{w,k}(\theta, c)| = o_{\mathbb{P}}(1),$$

where

$$\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_{1n}^{w,k}(\theta_1, c) - \mathcal{G}_{mn}^{w,k}(\theta_m, c), \tag{A.1}$$

and, for $j = \{1, m\}$,

$$\mathcal{B}_{jn}^{w,k}(\theta_j, c) = \mathcal{B}_{jn}^{w,k}(\theta_j, c) - \lim_{\tilde{c} \downarrow -c} \mathcal{B}_{jn}^{w,k}(\theta_j, \tilde{c}),$$

with

$$\mathcal{B}_{1n}^{w,k}(\theta_1, c) = \sigma^{-1} c^k f(c) n^{-1/2} \sum_{i=1}^n w_{in} (n^{-1/2} a_1 c + x'_{in} b_1), \tag{A.2}$$

$$\mathcal{B}_{mn}^{w,k}(\theta_m, c) = \sigma^{-1} n^{-1/2} \sum_{i=1}^n w_{in} \left\{ (n^{-1/2} a_m k \mathbb{E} \left(\frac{\varepsilon_i}{\sigma} \right)^k 1_{(\varepsilon_i \leq \sigma c)} + x'_{in} b_m k \mathbb{E} \left(\frac{\varepsilon_i}{\sigma} \right)^{k-1} 1_{(\varepsilon_i \leq \sigma c)}) \right\}.$$

The last lemma in this section is due to Berenguer-Rico, Johansen and Nielsen (2018) and derives a Law of Large Numbers result for the marked and weighted empirical distribution function of residuals –as such, the conditions are much weaker in this case. The result is used when analyzing the denominator of the *DASWhite* test.

Assumption A.3 Let \mathcal{F}_i be an increasing sequence of σ fields so that ε_{i-1}, x_i are \mathcal{F}_{i-1} -measurable and ε_i is independent of \mathcal{F}_{i-1} with continuous density f . Suppose

- (i) boundedness: (1) $\sup_{c \in \mathbb{R}} f(c) < \infty$ and (2) $\sup_{c \in \mathbb{R}} |c|f(c) < \infty$;
(ii) regressors x_i satisfy, for some non-stochastic normalization matrix N ,

$$n^{-1} \mathbf{E} \sum_{i=1}^n |n^{1/2} N' x_i| = O(1).$$

Lemma A.3 (Berenguer-Rico, Johansen and Nielsen, 2018) Let $\mathbf{E} |\varepsilon_i|^{k(1+\omega)} < \infty$ and

$$\mathbf{E} n^{-1} \sum_{i=1}^n |w_{in}|^{1+\omega} = O(1),$$

for some $\omega > 0$. Suppose Assumption A.3 holds. Then, for any $B > 0$ and $n \rightarrow \infty$,

$$\sup_{c \in \mathbb{R}} \sup_{|a|, |b| \leq B} |n^{-1} \sum_{i=1}^n w_{in} \varepsilon_i^k \{1_{(\varepsilon_i \leq \sigma c + n^{-1/2} a c + x'_{in} b)} - 1_{(\varepsilon_i \leq \sigma c)}\}| = o_{\mathbf{P}}(1).$$

A.2 Main Theorems on Empirical Processes and their Corollaries

For the remainder of Appendix A, let (unless otherwise stated) $x_{in} = N x_i$ with $N = n^{-1/2}$ and either $w_{in} = 1$, $w_{in} = Z_i$ or $w_{in} = Z_i Z'_i$.

Theorem A.4 Suppose Assumptions 2.1 and 2.2 hold. Then, for $k = 0, 1, 2$, all $B > 0$ and $n \rightarrow \infty$,

$$\sup_{|\theta| \leq B} \sup_{c \in \mathbb{R}} |\mathbb{G}_n^{w,k}(\theta, c) - \mathbb{G}_n^{w,k}(0, c)| = o_{\mathbf{P}}(1),$$

and

$$\sup_{|\theta| \leq B} \sup_{c \in \mathbb{R}} |n^{1/2} \{\bar{\mathbb{G}}_n^{w,k}(\theta, c) - \bar{\mathbb{G}}_n^{w,k}(0, c)\} - \mathcal{G}_n^{w,k}(\theta, c)| = o_{\mathbf{P}}(1).$$

Proof of Theorem A.4: We just need to show that our Assumptions 2.1 and 2.2 imply Assumption A.1 of Lemma A.2.

1. Conditions on the error term and mark functions:

(i) The marks $\tilde{m}(u)$ in Assumption A.1 of Lemma A.2 are, for $k = 0, 1, 2$, $m(u) = u^k$, $um(u) = ku^k$, $\dot{m}(u) = ku^{k-1}$. It suffices to check the case $k = 2$. Hence:

(a) By Assumption 2.1(a) : $\int_{-\infty}^{\infty} u^8 f(u) du < \infty$ hence $\int_{-\infty}^{\infty} \tilde{m}^4(u) f(u) du < \infty$.

(b1) By Assumption 2.1(b) : $\sup_{u \in \mathbb{R}} |u| \{1 + u^8\} f(u) < \infty$, hence, $\sup_{u \in \mathbb{R}} |u| \{1 + m^4(u)\} f(u) < \infty$.

(b2) By Assumption 2.1(b) : $\sup_{u \in \mathbb{R}} \{|u^7 f(u) + (1 + u^8) f(u)|\} < \infty$, so, $\sup_{u \in \mathbb{R}} |\frac{\partial}{\partial u} \{1 + \tilde{m}^4(u)\} f(u)| < \infty$.

(b3) By Assumption 2.1(b) : $\sup_{u \in \mathbb{R}} (1 + u^2) |u f(u)| < \infty$, hence $\sup_{u \in \mathbb{R}} (1 + u^2) |\dot{m}(u) f(u)| < \infty$.

(c) The tail monotonicity condition in Assumption 2.1(c) imply that for $\tilde{m}(u)$ being equal to $m(u) = u^2$, $um(u) = 2u^2$, $\dot{m}(u) = 2u$, the smoothness assumption (ic) in Assumption A.1 holds.

2. Conditions on the regressors and weights:

(ii) regressors: choose $N = n^{-1/2}$ and $\kappa = 1/5$ so that $|n^{1/2-\kappa} N' x_i| = |n^{-1/5} x_i|$. Let $M > 0$. By Boole and Markov inequalities

$$\mathbf{P}(\max_{1 \leq i \leq n} |n^{-1/5} x_i| > M) = \mathbf{P}(\cup_{i=1}^n (|n^{-1/5} x_i| > M)) \leq \sum_{i=1}^n \mathbf{P}(|x_i| > n^{1/5} M) \leq \frac{\sum_{i=1}^n \mathbf{E} |x_i|^6}{n^{6/5} M^6}.$$

By Assumption 2.2, this vanishes so that $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = \max_{1 \leq i \leq n} |n^{-\kappa} x_i| = o_{\mathbf{P}}(1)$.

(iii) weights: set $w_{in} = 1$, $w_{in} = X_i$ or $w_{in} = X_i X'_i$. Then, by Assumption 2.2 we have that $\mathbf{E} \sum_{i=1}^n |w_i|^{2+\omega} = O(n)$ as required by Assumption A.1. \square

Theorem A.5 *Suppose Assumptions 2.1 and 2.2 hold. Then, for any $B > 0$ and $n \rightarrow \infty$,*

$$\sup_{c \in \mathbb{R}} \sup_{|a|, |b| \leq B} |n^{-1} \sum_{i=1}^n \varepsilon_i^4 \{1_{(\varepsilon_i \leq \sigma c + n^{-1/2} a c + x'_{in} b)} - 1_{(\varepsilon_i \leq \sigma c)}\}| = o_{\mathbb{P}}(1).$$

Proof of Theorem A.5: We just need to show that our Assumptions 2.1 and 2.2 imply all the assumptions of Lemma A.3 with $w_{in} = 1$ and $k = 4$.

1. *Conditions on the error term:*

(a) By Assumption 2.1(a) : $\mathbb{E}(\varepsilon_i^8) < \infty$ hence $\mathbb{E}\{|\varepsilon_i|^{4(1+\omega)}\} < \infty$ for some $\omega > 0$ as required in Lemma A.3 when $k = 4$.

(b) By Assumption 2.1(b) : $\sup_{c \in \mathbb{R}} (1 + |c|^9) f(c) < \infty$, hence, $\sup_{c \in \mathbb{R}} f(c) < \infty$ and $\sup_{c \in \mathbb{R}} |c| f(c) < \infty$ as required by Assumption A.3.

2. *Conditions on the regressors and weights:*

(c) *regressors:* Recall, $N = n^{-1/2}$, hence, by Assumption 2.2 $n^{-1} \mathbb{E} \sum_{i=1}^n |x_i| = O(1)$ as required by Assumption A.3.

(d) *weights:* in our case $w_{in} = 1$. Then, trivially $\mathbb{E} n^{-1} \sum_{i=1}^n |w_{in}|^{1+\omega} = O(1)$ as required in Lemma A.3. \square

A.3 Corollaries

Under a general density f , by definition of the biases in (A.1) and (A.2) we have

$$\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_{1n}^{w,k}(\theta_1, c) - \mathcal{G}_{mn}^{w,k}(\theta_m, c), \quad (\text{A.3})$$

with

$$\begin{aligned} \mathcal{G}_{1n}^{w,k}(\theta_1, c) &= \{c^{k+1} f(c) - (-c)^{k+1} f(-c)\} \sigma^{-1} a_1 n^{-1} \sum_{i=1}^n w_{in} \\ &\quad + \{c^k f(c) - (-c)^k f(-c)\} \sigma^{-1} n^{-1/2} \sum_{i=1}^n w_{in} x'_{in} b_1, \end{aligned} \quad (\text{A.4})$$

and

$$\mathcal{G}_{mn}^{w,k}(\theta_m, c) = k \tau_k^c \sigma^{-1} a_m n^{-1} \sum_{i=1}^n w_{in} + k \tau_{k-1}^c \sigma^{-1} n^{-1/2} \sum_{i=1}^n w_{in} x'_{in} b_m, \quad (\text{A.5})$$

where recall $\tau_k^c = \mathbb{E}(\varepsilon_i / \sigma)^k \{1_{(\varepsilon_i \leq \sigma c)} - 1_{(\varepsilon_i \leq -\sigma c)}\}$.

Corollary A.6 *Let Assumptions 2.1 and 2.2 hold. Then, for $k = 0, 1, 2$, uniformly in c and $|\theta| \leq B$:*

(a) $\mathcal{G}_{1n}^{w,k}(\theta, c) = O_{\mathbb{P}}(1)$;

(b) $n^{1/2} \{\mathcal{G}_n^{w,k}(\theta, c) - \bar{\mathcal{G}}_n^{w,k}(0, c)\} = n^{1/2} \{\mathcal{G}_n^{w,k}(0, c) - \bar{\mathcal{G}}_n^{w,k}(0, c)\} + \mathcal{G}_n^{w,k}(\theta, c) + o_{\mathbb{P}}(1)$;

(c) $\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_n^{w,k}(0, c) + O_{\mathbb{P}}(n^{-1/2})$.

Proof of Corollary A.6:

(a) Recall the expressions for $\mathcal{G}_{1n}^{w,k}(\theta_1, c)$ and $\mathcal{G}_{mn}^{w,k}(\theta_m, c)$ in (A.4) and (A.5) and apply the triangle and norm inequalities to get

$$\begin{aligned} |\mathcal{G}_{1n}^{w,k}(\theta_1, c)| &\leq \{|c^{k+1} f(c)| + |(-c)^{k+1} f(-c)|\} \sigma^{-1} |a_1| n^{-1} \sum_{i=1}^n |w_{in}| \\ &\quad + \{|c^k f(c)| + |(-c)^k f(-c)|\} \sigma^{-1} n^{-1/2} \sum_{i=1}^n |w_{in}| |x_{in}| |b_1|, \\ |\mathcal{G}_{mn}^k(\theta_m, c)| &\leq k |\tau_k^c| \sigma^{-1} |a_m| n^{-1} \sum_{i=1}^n |w_{in}| + k |\tau_{k-1}^c| \sigma^{-1} n^{-1/2} \sum_{i=1}^n |w_{in}| |x_{in}| |b_m|. \end{aligned}$$

Note that $|\tau_k^c|$, $|\tau_{k-1}^c|$, $|c^{k+1} f(c)|$, $|c^k f(c)|$, $|(-c)^{k+1} f(-c)|$, $|(-c)^k f(-c)|$ are bounded by Assumption 2.1(a, b); $\sum_{i=1}^n |w_{in}| = O_{\mathbb{P}}(n)$ and $\sum_{i=1}^n |w_{in}| |x_{in}| = O_{\mathbb{P}}(n^{1/2})$ for $w_{in} = 1$, $w_{in} = Z_i$ or $w_{in} = Z_i Z'_i$, $x_{in} = n^{-1/2} x_i$ by Assumption 2.2 and the Law of Large Numbers, while the estimation errors vary in $|\theta| \leq B$ so that $\theta = O(1)$. Hence, the result follows.

(b) Expand

$$\begin{aligned} n^{1/2}\{\mathbb{G}_n^{w,k}(\theta, c) - \bar{\mathbb{G}}_n^{w,k}(0, c)\} &= [n^{1/2}\{\mathbb{G}_n^{w,k}(0, c) - \bar{\mathbb{G}}_n^{w,k}(0, c)\} + \mathcal{G}_n^{w,k}(\theta, c)] \\ &\quad + [n^{1/2}\{\mathbb{G}_n^{w,k}(\theta, c) - \mathbb{G}_n^{w,k}(0, c)\}] \\ &\quad + [n^{1/2}\{\bar{\mathbb{G}}_n^{w,k}(\theta, c) - \bar{\mathbb{G}}_n^{w,k}(0, c)\} - \mathcal{G}_n^{w,k}(\theta, c)]. \end{aligned}$$

By Theorem A.4, the second and third terms vanish and the desired result follows.

(c) By part (b) we have

$$\mathbb{G}_n^{w,k}(\theta, c) = \mathbb{G}_n^{w,k}(0, c) + n^{-1/2}\mathcal{G}_n^{w,k}(\tilde{\theta}, c) + n^{-1/2}\text{O}_{\mathbb{P}}(1).$$

By part (a), $n^{-1/2}\mathcal{G}_n^{w,k}(\tilde{\theta}, c)$ is $\text{O}_{\mathbb{P}}(n^{-1/2})$. Hence, we get $\mathbb{G}_n^{w,k}(\theta, c) = \mathbb{G}_n^{w,k}(0, c) + \text{O}_{\mathbb{P}}(n^{-1/2})$ as desired. \square

The next result derives the bias term $\mathcal{G}_n^{w,k}(\theta, c)$ under a symmetric density.

Corollary A.7 *If the density f is symmetric, then the biases in (A.1) are*

$$\mathcal{G}_{1n}^{w,k}(\theta_1, c) = 1_{(k \text{ even})}2c^{k+1}f(c)\sigma^{-1}a_1n^{-1}\sum_{i=1}^nw_{in} + 1_{(k \text{ odd})}2c^kf(c)\sigma^{-1}n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_1, \quad (\text{A.6})$$

$$\mathcal{G}_{mn}^{w,k}(\theta_m, c) = 1_{(k \text{ even})}k\tau_k^c\sigma^{-1}a_mn^{-1}\sum_{i=1}^nw_{in} + 1_{(k \text{ odd})}k\tau_{k-1}^c\sigma^{-1}n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_m. \quad (\text{A.7})$$

Proof of Corollary A.7. Notice that $\tau_k^c = \mathbb{E}(\varepsilon_i/\sigma)^k\{1_{(\varepsilon_i \leq \sigma c)} - 1_{(\varepsilon_i \leq -\sigma c)}\}$. Also, under a symmetric density $f(c) = f(-c)$ and $\tau_k^c = 0$ for k odd. Applying these results to (A.4) and (A.5) we get the desired expressions. \square

Corollary A.8 *Let Assumptions 2.1, 2.2 and 2.3 hold. Then, for $k = 0, 1, 2$, uniformly in c ,*

- (a) $n^{-1}\sum_{i=1}^nw_i(\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)} = n^{-1}\sum_{i=1}^nw_i(\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + \text{O}_{\mathbb{P}}(n^{-1/2});$
- (b) $n^{-1/2}\sum_{i=1}^nw_i(\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)} = n^{-1/2}\sum_{i=1}^nw_i(\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{w,k}(\hat{\theta}, c) + \text{o}_{\mathbb{P}}(1);$
- (c) $\bar{Z}_c = n^{-1}\sum_{i=1}^nZ_i 1_{(|\varepsilon_i| \leq \sigma c)} = \hat{T}_0^c n^{-1}\sum_{i=1}^nZ_i + \text{o}_{\mathbb{P}}(1) = \tau_0^c n^{-1}\sum_{i=1}^nZ_i + \text{o}_{\mathbb{P}}(1).$

Proof of Corollary A.8: Let $\hat{\theta} = (\tilde{a}, \tilde{b}, 0, \hat{b})'$ and $\tilde{\theta} = (\tilde{a}, \tilde{b})$ where $\tilde{b} = n^{1/2}(\hat{\beta} - \beta)$, $\hat{b} = n^{1/2}(\hat{\beta} - \beta)$ and $\tilde{a} = n^{1/2}(\hat{\sigma} - \sigma)$. By Assumption 2.3 we have that $\tilde{b}, \hat{b}, \tilde{a}$ are bounded in probability. Let $\mathbb{G}_n^{w,k}(\hat{\theta}, c) = n^{-1}\sum_{i=1}^nw_i(\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)}$.

(a) By Lemma A.1 it suffices to analyze $\mathbb{G}_n^{w,k}(\theta, c)$ uniformly over both θ and c . Then, by Corollary A.6(c), uniformly in θ and c ,

$$\mathbb{G}_n^{w,k}(\theta, c) = \mathbb{G}_n^{w,k}(0, c) + \text{O}_{\mathbb{P}}(n^{-1/2}) = n^{-1}\sum_{i=1}^nw_i(\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + \text{O}_{\mathbb{P}}(n^{-1/2}).$$

(b) By Lemma A.1 it suffices to analyze $n^{1/2}\mathbb{G}_n^{w,k}(\theta, c)$ uniformly over both θ and c . Then, by Corollary A.6(b), uniformly in θ and c ,

$$n^{1/2}\mathbb{G}_n^{w,k}(\theta, c) = n^{1/2}\mathbb{G}_n^{w,k}(0, c) + \mathcal{G}_n^{w,k}(\theta, c) + \text{o}_{\mathbb{P}}(1),$$

and the result follows.

(c) Add and subtract \hat{T}_0^c to $1_{(|\varepsilon_i| \leq \sigma c)}$ so that

$$\bar{Z}_c = n^{-1}\sum_{i=1}^nZ_i 1_{(|\varepsilon_i| \leq \sigma c)} = \hat{T}_0^c n^{-1}\sum_{i=1}^nZ_i - n^{-1}\sum_{i=1}^nZ_i\{1_{(|\varepsilon_i| \leq \sigma c)} - \hat{T}_0^c\}.$$

Add and subtract τ_0^c to \hat{T}_0^c in $\{1_{(|\varepsilon_i| \leq \sigma c)} - \hat{T}_0^c\}$ so that

$$\bar{Z}_c = \hat{T}_0^c n^{-1}\sum_{i=1}^nZ_i - n^{-1}\sum_{i=1}^nZ_i\{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} - (\hat{T}_0^c - \tau_0^c)n^{-1}\sum_{i=1}^nZ_i.$$

Then, (i) applying the Law of Large Numbers to the second term on the right hand side; (ii) noting that by item (a) with $w_{in} = 1$ and $k = 0$ $\hat{T}_0^c = T_0^c + o_{\mathbf{P}}(1)$ and by the Law of Large Numbers $T_0^c = \tau_0^c + o_{\mathbf{P}}(1)$ and (iii) since $\bar{Z} = O_{\mathbf{P}}(1)$ by Assumption 2.2 and the Law of Large Numbers, we get

$$\bar{Z}_c = n^{-1} \sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)} = \hat{T}_0^c n^{-1} \sum_{i=1}^n Z_i + o_{\mathbf{P}}(1).$$

Again, by item (a) we have $\hat{T}_0^c = T_0^c + o_{\mathbf{P}}(1)$, by the Law of Large Numbers $T_0^c = \tau_0^c + o_{\mathbf{P}}(1)$ and by Assumption 2.2 and the Law of Large Numbers $n^{-1} \sum_{i=1}^n Z_i = O_{\mathbf{P}}(1)$, therefore,

$$\bar{Z}_c = \{\tau_0^c + o_{\mathbf{P}}(1)\} n^{-1} \sum_{i=1}^n Z_i + o_{\mathbf{P}}(1) = \tau_0^c n^{-1} \sum_{i=1}^n Z_i + o_{\mathbf{P}}(1),$$

as desired. \square

Corollary A.8, items (a) and (b), give sufficient conditions under which \hat{T}_k^c and $n^{1/2}(\hat{T}_k^c - \sigma^k \tau_k^c)$ are equivalent to T_k^c and $n^{1/2}(T_k^c - \sigma^k \tau_k^c)$, respectively, for $k = 0, 1, 2$. Notice, however, that the denominator of the *DAS White* test involves \hat{T}_k^c for $k = 4$. To derive an asymptotic expansion for $n^{1/2}(\hat{T}_4^c - \sigma^4 \tau_4^c)$ using the approach of Corollary A.8 would imply an increase in the amount of moments required for ε_i . Since the *DAS White* test involves \hat{T}_4^c but not $n^{1/2}(\hat{T}_4^c - \sigma^4 \tau_4^c)$, in the next Corollary, we derive the asymptotic equivalence between \hat{T}_4^c and T_4^c using Theorem A.5. In this way, the result can be derived without further assumptions.

Corollary A.9 : *Under Assumptions 2.1, 2.2 and 2.3,*

$$n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_{i,c}/\sigma)^4 = n^{-1} \sum_{i=1}^n (\varepsilon_{i,c}/\sigma)^4 + o_{\mathbf{P}}(1).$$

Proof of Corollary A.9: Let

$$\mathcal{M}_{n,c} = |n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_{i,c}^4 - \varepsilon_{i,c}^4)| = |n^{-1} \sum_{i=1}^n \{\hat{\varepsilon}_i^4 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)} - \varepsilon_i^4 1_{(|\varepsilon_i| \leq \sigma c)}\}|.$$

We show that $\mathcal{M}_{n,c} = o_{\mathbf{P}}(1)$ uniformly in c . Add and subtract $\varepsilon_i^4 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)}$ so that

$$\mathcal{M}_{n,c} = |n^{-1} \sum_{i=1}^n [(\hat{\varepsilon}_i^4 - \varepsilon_i^4) 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)} + \varepsilon_i^4 \{1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)} - 1_{(|\varepsilon_i| \leq \sigma c)}\}]|.$$

By the triangle inequality then $\mathcal{M}_{n,c} \leq \mathcal{M}_{1,n,c} + \mathcal{M}_{2,n,c}$ where

$$\mathcal{M}_{1,n,c} = |n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_i^4 - \varepsilon_i^4) 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)}|, \quad \mathcal{M}_{2,n,c} = |n^{-1} \sum_{i=1}^n \varepsilon_i^4 \{1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)} - 1_{(|\varepsilon_i| \leq \sigma c)}\}|.$$

It suffices to show that $\mathcal{M}_{1,n,c}, \mathcal{M}_{2,n,c}$ are $o_{\mathbf{P}}(1)$ uniformly in c .

1.1. The term $\mathcal{M}_{1,n,c}$: Applying the triangle inequality and bounding $1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)}$ by 1 we get

$$\mathcal{M}_{1,n,c} \leq n^{-1} \sum_{i=1}^n |\hat{\varepsilon}_i^4 - \varepsilon_i^4| = n^{-1} \sum_{i=1}^n |\{\varepsilon_i - (\hat{\beta} - \beta)' x_t\}^4 - \varepsilon_i^4|.$$

Using the binomial expansion $(a - b)^4 - a^4 = b^4 - 4b^3a + 6b^2a^2 - 4ba^3$ we can write

$$\mathcal{M}_{1,n,c} \leq n^{-1} \sum_{i=1}^n |\{(\hat{\beta} - \beta)' x_t\}^4 - 4\{(\hat{\beta} - \beta)' x_t\}^3 \varepsilon_i + 6\{(\hat{\beta} - \beta)' x_t\}^2 \varepsilon_i^2 - 4\{(\hat{\beta} - \beta)' x_t\} \varepsilon_i^3|.$$

By the triangle and norm inequalities

$$\begin{aligned} \mathcal{M}_{1,n,c} &\leq |\hat{\beta} - \beta|^4 n^{-1} \sum_{i=1}^n |x_t|^4 + 4|\hat{\beta} - \beta|^3 n^{-1} \sum_{i=1}^n |x_t|^3 \varepsilon_i \\ &\quad + 6|\hat{\beta} - \beta|^2 n^{-1} \sum_{i=1}^n x_t^2 \varepsilon_i^2 + 4|\hat{\beta} - \beta| n^{-1} \sum_{i=1}^n |x_t| |\varepsilon_i|^3. \end{aligned}$$

By Assumption $\hat{\beta} - \beta = o_{\mathbf{P}}(1)$. Therefore, by Assumptions 2.1, 2.2, 2.3 and the Law of Large Numbers, we get that $\mathcal{M}_{1,n,c} = o_{\mathbf{P}}(1)$.

1.2. The term $\mathcal{M}_{2,n,c}$: Recall

$$\mathcal{M}_{2,n,c} = |n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(|\hat{\varepsilon}_i| \leq \bar{\sigma} c)} - n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(|\varepsilon_i| \leq \sigma c)}|.$$

Note that

$$n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)} = n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)} - \lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{\{\tilde{\varepsilon}_i \leq \tilde{\sigma}(-c-h)\}}.$$

Hence,

$$\begin{aligned} \mathcal{M}_{2,n,c} &= |n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)} - \lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{\{\tilde{\varepsilon}_i \leq \tilde{\sigma}(-c-h)\}} \\ &\quad - [n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(\varepsilon_i \leq \sigma c)} - \lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{\{\varepsilon_i \leq \sigma(-c-h)\}}]|. \end{aligned}$$

Rearranging and using the triangle inequality we get

$$\begin{aligned} \mathcal{M}_{2,n,c} &= |n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)} - n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{(\varepsilon_i \leq \sigma c)}| \\ &\quad + |\lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{\{\tilde{\varepsilon}_i \leq \tilde{\sigma}(-c-h)\}} - \lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 1_{\{\varepsilon_i \leq \sigma(-c-h)\}}|. \end{aligned}$$

By Lemma A.1 and Theorem A.5 we get, given Assumptions 2.1, 2.2, 2.3, that

$$\sup_c |n^{-1} \sum_{i=1}^n \varepsilon_i^4 \{1_{(\tilde{\varepsilon}_i \leq \tilde{\sigma}c)} - 1_{(\varepsilon_i \leq \sigma c)}\}| = o_{\mathbf{P}}(1),$$

and

$$\sup_c |\lim_{h \downarrow 0} n^{-1} \sum_{i=1}^n \varepsilon_i^4 [1_{\{\tilde{\varepsilon}_i \leq \tilde{\sigma}(-c-h)\}} - 1_{\{\varepsilon_i \leq \sigma(-c-h)\}}]| = o_{\mathbf{P}}(1).$$

Hence, $\mathcal{M}_{2,n,c} = o_{\mathbf{P}}(1)$ uniformly in c .

1.3. *Combine items 1.1 and 1.2:* to get the desired result. \square

B Appendix B: Results on Test

Proof of Theorem 3.1: Write the test statistic in (3.2) as $\hat{T}_0^c n \hat{R}_{n,c}^2 = \widehat{Num} / \widehat{Den}$ where

$$\begin{aligned} \widehat{Num} &= \sigma^{-2} n^{-1/2} \hat{\mathcal{N}}'_{n,c} (n^{-1} \hat{\mathcal{M}}_{n,c})^{-1} \sigma^{-2} n^{-1/2} \hat{\mathcal{N}}_{n,c}, \\ \widehat{Den} &= \sigma^{-4} (\hat{T}_0^c n)^{-1} \sum_{i=1}^n \{\hat{\varepsilon}_{i,c}^2 - 1_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}c)} (\hat{T}_2^c / \hat{T}_0^c)\}^2. \end{aligned}$$

1. *The term \widehat{Den} :* Expanding the curly bracket in \widehat{Den} we get

$$\widehat{Den} = (\hat{T}_0^c)^{-1} n^{-1} \sum_{i=1}^n \left(\frac{\hat{\varepsilon}_{i,c}}{\sigma}\right)^4 - \{(T_2^c / \sigma^2) / \hat{T}_0^c\}^2.$$

By Corollary A.8(a) with $w_{in} = 1$ we have $\hat{T}_k^c / \sigma^k = n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_{i,c} / \sigma)^k = n^{-1} \sum_{i=1}^n (\varepsilon_{i,c} / \sigma)^k + o_{\mathbf{P}}(n^{-1/2}) = T_k^c / \sigma^k + o_{\mathbf{P}}(n^{-1/2})$ for $k = 0, 2$. By Assumption 2.1 and the Law of Large Numbers $T_k^c = O_{\mathbf{P}}(1)$. Hence,

$$\widehat{Den} = (\hat{T}_0^c)^{-1} n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_{i,c} / \sigma)^4 - \{(T_2^c / \sigma^2) / T_0^c\}^2 + o_{\mathbf{P}}(n^{-1/2}).$$

By Corollary A.9, we have $n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_{i,c} / \sigma)^4 = n^{-1} \sum_{i=1}^n (\varepsilon_{i,c} / \sigma)^4 + o_{\mathbf{P}}(1)$. Hence, since

$$n^{-1} \sum_{i=1}^n (\varepsilon_{i,c} / \sigma)^4 = O_{\mathbf{P}}(1),$$

by Assumption 2.1 and the Law of Large Numbers and given that $\hat{T}_0^c = T_0^c + o_{\mathbf{P}}(n^{-1/2})$ as already shown, we get

$$\widehat{Den} = (T_0^c)^{-1} n^{-1} \sum_{i=1}^n (\varepsilon_{i,c} / \sigma)^4 - \{(T_2^c / \sigma^2) / T_0^c\}^2 + o_{\mathbf{P}}(1).$$

By the Law of Large Numbers, $T_k^c = \sigma^k \tau_k^c + o_{\mathbf{P}}(1)$ where $\tau_k^c > 0$, hence,

$$\widehat{Den} = (T_0^c)^{-1} n^{-1} \sum_{i=1}^n (\varepsilon_{i,c} / \sigma)^4 - (\tau_2^c / \tau_0^c)^2 + o_{\mathbf{P}}(1).$$

Using again the fact that $T_k^c/\sigma^k = \tau_k^c + \text{op}(1)$ we can write

$$\widehat{Den} = (\tau_0^c)^{-1} n^{-1} \sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_2^c/\tau_0^c)\}^2 + \text{op}(1).$$

2. The term \widehat{Num} :

2.1. The term $n^{-1}\widehat{\mathcal{M}}_{n,c}$. By definition of $\widehat{\mathcal{M}}_{n,c}$

$$n^{-1}\widehat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n \{\tilde{Z}_{i,c} - 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\} \{\tilde{Z}_{i,c} - 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\}'.$$

Expanding $n^{-1}\widehat{\mathcal{M}}_{n,c}$ we have

$$n^{-1}\widehat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c} \tilde{Z}_{i,c}' - (\hat{T}_0^c)^{-1} \tilde{Z}_c \tilde{Z}_c'.$$

By Corollary A.8(a) with $w_{in} = Z_i Z_i'$ and $k = 0$,

$$n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c} \tilde{Z}_{i,c}' = n^{-1} \sum_{i=1}^n Z_i Z_i' 1_{(|\varepsilon_i| < \sigma c)} + \text{Op}(n^{-1/2}) = n^{-1} \sum_{i=1}^n Z_i Z_i' + \text{Op}(n^{-1/2}).$$

By Corollary A.8(a) with $w_{in} = Z_i$ and $k = 0$,

$$\tilde{Z}_c = n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c} = n^{-1} \sum_{i=1}^n Z_i 1_{(|\varepsilon_i| < \sigma c)} + \text{Op}(n^{-1/2}) = \bar{Z}_c + \text{Op}(n^{-1/2}).$$

By Corollary A.8(a) with $w_{in} = 1$ and $k = 0$ and the Law of Large Numbers $\hat{T}_0^c = T_0^c + \text{Op}(n^{-1/2}) = \tau_0^c + \text{op}(1)$. Hence, since $\bar{Z}_c = \text{Op}(1)$ by Assumptions 2.1 and 2.2 and the Law of Large Numbers we can write

$$n^{-1}\widehat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n Z_i Z_i' - \{\tau_0^c + \text{op}(1)\}^{-1} \bar{Z}_c \bar{Z}_c' + \text{Op}(n^{-1/2}).$$

Since $\bar{Z}_c \bar{Z}_c' = \text{Op}(1)$ by Assumptions 2.1 and 2.2 and the Law of Large Numbers

$$n^{-1}\widehat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n Z_i Z_i' - (\tau_0^c)^{-1} \bar{Z}_c \bar{Z}_c' + \text{op}(1).$$

Given that $T_0^c = n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| < \sigma c)} = \tau_0^c + \text{op}(1)$ and $\bar{Z}_c = \text{Op}(1)$ by Assumptions 2.1 and 2.2 and the Law of Large Numbers, we can write

$$n^{-1}\widehat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_0^c)^{-1} \bar{Z}_c\} \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)} (\tau_0^c)^{-1} \bar{Z}_c\}' + \text{op}(1).$$

2.2. The term $\sigma^{-2} n^{-1/2} \hat{\mathcal{N}}_{n,c}$. Recall

$$\hat{\mathcal{N}}_{n,c} = \sum_{i=1}^n \{\hat{\varepsilon}_{i,c}^2 - 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)} (\hat{T}_2^c/\hat{T}_0^c)\} \{\tilde{Z}_{i,c} - 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)} (\hat{T}_0^c)^{-1} \tilde{Z}_c\},$$

which is equivalent to

$$\hat{\mathcal{N}}_{n,c} = \sum_{i=1}^n \{\hat{\varepsilon}_{i,c}^2 - 1_{(|\tilde{\varepsilon}_i| < \tilde{\sigma} c)} (\hat{T}_2^c/\hat{T}_0^c)\} \tilde{Z}_{i,c}.$$

Expanding $\sigma^{-2} n^{-1/2} \hat{\mathcal{N}}_{n,c}$ we can write $\sigma^{-2} n^{-1/2} \hat{\mathcal{N}}_{n,c} = n^{-1/2} (\hat{\mathcal{N}}_{1,n,c} - \hat{\mathcal{N}}_{2,n,c})$ where

$$\hat{\mathcal{N}}_{1,n,c} = \sum_{i=1}^n (\hat{\varepsilon}_{i,c}/\sigma)^2 \tilde{Z}_{i,c}, \quad \hat{\mathcal{N}}_{2,n,c} = \sigma^{-2} (\hat{T}_2^c/\hat{T}_0^c) (\sum_{i=1}^n \tilde{Z}_{i,c}).$$

2.2.1. The term $n^{-1/2} \hat{\mathcal{N}}_{1,n,c}$. Note that $n^{-1/2} \hat{\mathcal{N}}_{1,n,c} = n^{-1/2} \sum_{i=1}^n (\hat{\varepsilon}_{i,c}/\sigma)^2 Z_i$, hence by Corollary A.8(b) with $w_{in} = Z_i$ and $k = 2$ we have

$$n^{-1/2} \hat{\mathcal{N}}_{1,n,c} = n^{-1/2} \sum_{i=1}^n (\varepsilon_{i,c}/\sigma)^2 Z_i + \mathcal{G}_n^{Z,2}(\hat{\theta}, c) + \text{op}(1),$$

where

$$\mathcal{G}_n^{Z,2}(\hat{\theta}, c) = 2c^3 \mathbf{f}(c) \sigma^{-1} \tilde{a} n^{-1} \sum_{i=1}^n Z_i, \quad (\text{B.1})$$

by recalling the bias in (A.3) and then applying Corollary A.7, with $k = 2$, $a_1 = \tilde{a}$, $a_m = 0$, $w_{in} = Z_i$. Add and subtract $\tau_2^c \sqrt{n} \bar{Z}$ to $n^{-1/2} \hat{\mathcal{N}}_{1,n,c}$ so that

$$n^{-1/2} \hat{\mathcal{N}}_{1,n,c} = n^{-1/2} \sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\} Z_i + \tau_2^c \sqrt{n} \bar{Z} + \mathcal{G}_n^{Z,2}(\hat{\theta}, c) + \text{op}(1).$$

2.2.2. *The term $n^{-1/2}\hat{\mathcal{N}}_{2,n,c}$.* Multiply and divide by \sqrt{n} so that $n^{-1/2}\hat{\mathcal{N}}_{2,n,c} = \sqrt{n}\sigma^{-2}(\hat{T}_2^c/\hat{T}_0^c)\tilde{Z}_c$. We expand each of these terms using Corollary A.8(b) as follows: (i) for \hat{T}_0^c set $w_{in} = 1$, $k = 0$ and $a_m = 0$; (ii) for $\sqrt{n}\hat{T}_2^c/\sigma^2$ set $w_{in} = 1$, $k = 2$ and $a_m = 0$; and (iii) for \tilde{Z}_c set $w_{in} = Z_i$, $k = 0$ and $a_m = 0$. We get

$$\begin{aligned}\hat{T}_0^c &= n^{-1/2}\{n^{-1/2}\sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(1)\}, \\ \sqrt{n}\hat{T}_2^c/\sigma^2 &= n^{-1/2}\sum_{i=1}^n (\varepsilon_{i,c}/\sigma)^2 + \mathcal{G}_n^{1,2}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(1), \\ \tilde{Z}_c &= n^{-1/2}\{n^{-1/2}\sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{Z,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(1)\},\end{aligned}$$

where by Corollary A.7,

$$\mathcal{G}_n^{1,0}(\hat{\theta}, c) = 2\text{cf}(c)\sigma^{-1}\tilde{a}, \quad \mathcal{G}_n^{1,2}(\hat{\theta}, c) = 2c^3\text{f}(c)\sigma^{-1}\tilde{a}, \quad \mathcal{G}_n^{Z,0}(\hat{\theta}, c) = 2\text{cf}(c)\sigma^{-1}\tilde{a}\bar{Z}. \quad (\text{B.2})$$

Add and subtract (i) τ_0^c to \hat{T}_0^c ; (ii) $\sqrt{n}\tau_2^c$ to $\sqrt{n}\hat{T}_2^c/\sigma^2$; and (iii) $\tau_0^c\bar{Z}$ to \tilde{Z}_c so that

$$\begin{aligned}\hat{T}_0^c &= n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + \tau_0^c + n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2}), \\ \sqrt{n}\hat{T}_2^c/\sigma^2 &= n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\} + \sqrt{n}\tau_2^c + \mathcal{G}_n^{1,2}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(1), \\ \tilde{Z}_c &= n^{-1}\sum_{i=1}^n Z_i \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + \tau_0^c\bar{Z} + n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2}).\end{aligned}$$

Take common factor τ_0^c in the expansions for \hat{T}_0^c and \tilde{Z}_c so that they cancel out in $n^{-1/2}\hat{\mathcal{N}}_{2,n,c} = \sqrt{n}\sigma^{-2}(\hat{T}_2^c/\hat{T}_0^c)\tilde{Z}_c$ and we get

$$\begin{aligned}n^{-1/2}\hat{\mathcal{N}}_{2,n,c} &= \frac{n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\} + \sqrt{n}\tau_2^c + \mathcal{G}_n^{1,2}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(1)}{\{1 + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2})\}} \\ &\quad \times \{\bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n Z_i \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2})\}.\end{aligned}$$

The expansion for $(\hat{T}_0^c)^{-1}$, which is,

$$(\hat{T}_0^c)^{-1} = \frac{1}{1 + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2})},$$

is of the form $1/(1 + x_n)$ where $x_n = \text{O}_{\mathbf{P}}(n^{-1/2})$ by the Central Limit Theorem and A.6, hence, expanding we can write

$$\frac{1}{1 + x_n} = 1 - x_n + \text{O}(x_n^2) = 1 - x_n + \text{O}_{\mathbf{P}}(n^{-1}).$$

This then gives that

$$(\hat{T}_0^c)^{-1} = 1 - (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} - (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2}) + \text{O}_{\mathbf{P}}(n^{-1}).$$

Combine the expansions in the product $(\hat{T}_0^c)^{-1}\tilde{Z}_c$ so that

$$\begin{aligned}(\hat{T}_0^c)^{-1}\tilde{Z}_c &= \{1 - (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} - (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2})\} \\ &\quad \times \{\bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n Z_i \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2})\}.\end{aligned} \quad (\text{B.3})$$

Expanding this product we have

$$\begin{aligned}(\hat{T}_0^c)^{-1}\tilde{Z}_c &= \bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n Z_i \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) + \text{o}_{\mathbf{P}}(n^{-1/2}) \\ &\quad - (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}\bar{Z} + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}) \\ &\quad - (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c)\bar{Z} + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}) \\ &\quad + \text{O}_{\mathbf{P}}(n^{-1/2}) + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}) + \text{O}_{\mathbf{P}}(n^{-1}).\end{aligned}$$

Noting that the term $\mathcal{G}_n^{Z,0}(\hat{\theta}, c) - \mathcal{G}_n^{1,0}(\hat{\theta}, c)\bar{Z} = 0$, see (B.2), we get

$$(\hat{T}_0^c)^{-1}\tilde{Z}_c = \bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n Z_i\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} - (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\}\bar{Z} + o_{\mathbf{P}}(n^{-1/2}),$$

which simplifies to

$$(\hat{T}_0^c)^{-1}\tilde{Z}_c = \bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} + o_{\mathbf{P}}(n^{-1/2}).$$

Combining the expansions for $(\hat{T}_0^c)^{-1}\tilde{Z}_c$ and $\sqrt{n}\hat{T}_2^c/\sigma^2$ in $\hat{\mathcal{N}}_{2,n,c} = \sqrt{n}\sigma^{-2}(\hat{T}_2^c/\hat{T}_0^c)\tilde{Z}_c$, we get

$$\begin{aligned} n^{-1/2}\hat{\mathcal{N}}_{2,n,c} &= [n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\} + \sqrt{n}\tau_2^c + \mathcal{G}_n^{1,2}(\hat{\theta}, c) + o_{\mathbf{P}}(1)] \\ &\quad \times \{\bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} + o_{\mathbf{P}}(n^{-1/2})\}. \end{aligned}$$

Expanding the product we get

$$\begin{aligned} n^{-1/2}\hat{\mathcal{N}}_{2,n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}\bar{Z} + o_{\mathbf{P}}(n^{-1/2}) + o_{\mathbf{P}}(n^{-1/2}) \\ &\quad + \sqrt{n}\tau_2^c\bar{Z} + \sqrt{n}\tau_2^c(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} + o_{\mathbf{P}}(1) \\ &\quad + \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} + o_{\mathbf{P}}(n^{-1/2}) + o_{\mathbf{P}}(n^{-1/2}) \\ &\quad + o_{\mathbf{P}}(1) + o_{\mathbf{P}}(n^{-1/2}) + o_{\mathbf{P}}(n^{-1/2}), \end{aligned}$$

which simplifies to

$$\begin{aligned} n^{-1/2}\hat{\mathcal{N}}_{2,n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}\bar{Z} + \sqrt{n}\tau_2^c(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} \\ &\quad + \sqrt{n}\tau_2^c\bar{Z} + \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} + o_{\mathbf{P}}(n^{-1/2}). \end{aligned}$$

2.2.3. *Combine items 2.2.1-2.2.2.* Recall $\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}(\hat{\mathcal{N}}_{1,n,c} - \hat{\mathcal{N}}_{2,n,c})$. Inserting the expansions for $n^{-1/2}\hat{\mathcal{N}}_{1,n,c}$ and $n^{-1/2}\hat{\mathcal{N}}_{2,n,c}$ in items 2.2.1 and 2.2.2, we then get

$$\begin{aligned} \sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}Z_i + \tau_2^c\sqrt{n}\bar{Z} + \mathcal{G}_n^{Z,2}(\hat{\theta}, c) + o_{\mathbf{P}}(1) \\ &\quad - [n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}\bar{Z} + \sqrt{n}\tau_2^c(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} \\ &\quad + \sqrt{n}\tau_2^c\bar{Z} + \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} + o_{\mathbf{P}}(n^{-1/2})]. \end{aligned}$$

The terms $\tau_2^c\sqrt{n}\bar{Z}$ cancel while $\mathcal{G}_n^{Z,2}(\hat{\theta}, c) = \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z}$, see (B.1) and (B.2), hence,

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}(Z_i - \bar{Z}) - \tau_2^c(\tau_0^c)^{-1}n^{-1/2}\sum_{i=1}^n(Z_i - \bar{Z})\{1_{(|\varepsilon_i|\leq\sigma c)} - \tau_0^c\} + o_{\mathbf{P}}(1).$$

Take $\tau_2^c(\tau_0^c)^{-1}$ in the second term inside the curly bracket so that

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}(Z_i - \bar{Z}) - n^{-1/2}\sum_{i=1}^n(Z_i - \bar{Z})\{\tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)} - \tau_2^c\} + o_{\mathbf{P}}(1),$$

which can be simplified as

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)}\}(Z_i - \bar{Z}) + o_{\mathbf{P}}(1).$$

Take common factor $1_{(|\varepsilon_i|\leq\sigma c)}$ so that

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)}\}\{1_{(|\varepsilon_i|\leq\sigma c)}Z_i - 1_{(|\varepsilon_i|\leq\sigma c)}\bar{Z}\} + o_{\mathbf{P}}(1),$$

which we can write as

$$\begin{aligned} \sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)}\}1_{(|\varepsilon_i|\leq\sigma c)}Z_i \\ &\quad - \bar{Z}n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)}\} \\ &\quad + o_{\mathbf{P}}(n^{-1/2}), \end{aligned}$$

By Corollary A.8(c), $\bar{Z} = (\tau_0^c)^{-1}\bar{Z}_c + o_{\mathbf{P}}(1)$. Hence,

$$\begin{aligned} \sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}Z_i \\ &\quad - \{(\tau_0^c)^{-1}\bar{Z}_c + o_{\mathbf{P}}(1)\}n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\} \\ &\quad + o_{\mathbf{P}}(1). \end{aligned}$$

Notice that the term $n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\}$ is $O_{\mathbf{P}}(1)$ by the Central Limit Theorem, hence,

$$\begin{aligned} \sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}Z_i \\ &\quad - (\tau_0^c)^{-1}\bar{Z}_cn^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\} \\ &\quad + o_{\mathbf{P}}(1). \end{aligned}$$

This can be rewritten as

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\}\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\} + o_{\mathbf{P}}(1).$$

3. Combine items 1 and 2: By item 1,

$$\widehat{Den} = (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_2^c/\tau_0^c)\}^2 + o_{\mathbf{P}}(1).$$

By item 2.1,

$$n^{-1}\hat{\mathcal{M}}_{n,c} = n^{-1}\sum_{i=1}^n\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}' + o_{\mathbf{P}}(1).$$

By item 2.2,

$$\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_2^c/\tau_0^c)\}\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\} + o_{\mathbf{P}}(1).$$

Note that the term $\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c}$ is $O_{\mathbf{P}}(1)$ by the Central Limit Theorem. Combining we get $\hat{T}_0^c n \hat{R}_{n,c}^2 = \tau_0^c n R_{n,c}^2 + o_{\mathbf{P}}(1)$ as stated. \square

Proof of Theorem 3.2: By Theorem 3.1, $\hat{T}_0^c n \hat{R}_{n,c}^2 = \tau_0^c n R_{n,c}^2 + o_{\mathbf{P}}(1)$. Hence, in what follows we analyze the asymptotic properties of $\tau_0^c n R_{n,c}^2$. We write $\tau_0^c n R_{n,c}^2 = Num/Den$ where

$$\begin{aligned} Num &= n^{-1/2}\mathcal{N}'_{n,c}(n^{-1}\mathcal{M}_{n,c})^{-1}n^{-1/2}\mathcal{N}_{n,c}, \\ Den &= (\tau_0^c n)^{-1}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_2^c/\tau_0^c)\}^2. \end{aligned}$$

1. The term Den: Expanding the curly bracket

$$Den = (\tau_0^c n)^{-1}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^4 - 2(\tau_2^c/\tau_0^c)(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^2 + (\tau_2^c/\tau_0^c)^2(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}\}.$$

By Assumption 2.1 and the Law of Large Numbers $n^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^2 = \tau_2^c + o_{\mathbf{P}}(1)$ and $n^{-1}\sum_{i=1}^n\mathbf{1}_{(|\varepsilon_i|\leq\sigma c)} = \tau_0^c + o_{\mathbf{P}}(1)$, where $\tau_0^k > 0$. Therefore,

$$Den = (\tau_0^c n)^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^4 - (\tau_2^c/\tau_0^c)^2 + o_{\mathbf{P}}(1).$$

By the Law of Large Numbers, $n^{-1}\sum_{i=1}^n\varepsilon_{i,c}^4 = \tau_4^c + o_{\mathbf{P}}(1)$, hence, by the Slutsky theorem, $Den = \{\tau_4^c/\tau_0^c - (\tau_2^c/\tau_0^c)^2\} + o_{\mathbf{P}}(1)$.

2. The term Num:

2.1. The term $n^{-1}\mathcal{M}_{n,c}$. Recall

$$n^{-1}\mathcal{M}_{n,c} = n^{-1}\sum_{i=1}^n\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i|\leq\sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}'.$$

Expand so that

$$n^{-1}\mathcal{M}_{n,c} = n^{-1}\sum_{i=1}^n Z_{i,c}Z'_{i,c} - (\tau_0^c)^{-1}\bar{Z}_c\bar{Z}'_c.$$

By definition of $Z_{i,c}$ and \bar{Z}_c we can write

$$n^{-1}\mathcal{M}_n = n^{-1}\sum_{i=1}^n Z_i Z'_i 1_{(|\varepsilon_i| \leq \sigma c)} - (\tau_0^c)^{-1} \{n^{-1}\sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)}\} \{n^{-1}\sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)}\}'.$$

By the Law of Large Numbers and independence of Z_i and ε_i ,

$$n^{-1}\mathcal{M}_n \rightarrow_{\mathbb{P}} \tau_0^c \mathbf{E}Z_1 Z'_1 - (\tau_0^c)^{-1} (\tau_0^c)^2 \mathbf{E}Z_1 \mathbf{E}Z'_1 + o_{\mathbb{P}}(1) = \tau_0^c \mathbf{E}(Z_1 - \mathbf{E}Z_1)(Z_1 - \mathbf{E}Z_1)' + o_{\mathbb{P}}(1).$$

Let $\Sigma_Z = \mathbf{E}(Z_1 - \mathbf{E}Z_1)(Z_1 - \mathbf{E}Z_1)'$ so that

$$n^{-1}\mathcal{M}_n \rightarrow_{\mathbb{P}} \tau_0^c \Sigma_Z.$$

2.2. The term $n^{-1/2}\mathcal{N}_{n,c}$. Recall

$$n^{-1/2}\mathcal{N}_{n,c} = n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\} \{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}.$$

Notice that taking common factor $1_{(|\varepsilon_i| \leq \sigma c)}$ in $\{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\}$ and multiplying it to $\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\}$ we can write

$$n^{-1/2}\mathcal{N}_{n,c} = n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\} \{Z_i - (\tau_0^c)^{-1}\bar{Z}_c\}.$$

By Corollary A.8(c), $\bar{Z}_c = \tau_0^c \bar{Z} + o_{\mathbb{P}}(1)$. Hence, $(\tau_0^c)^{-1}\bar{Z}_c = \bar{Z} + o_{\mathbb{P}}(1)$, so that

$$n^{-1/2}\mathcal{N}_{n,c} = n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\} \{Z_i - \bar{Z} + o_{\mathbb{P}}(1)\}.$$

Since $n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\}$ is $O_{\mathbb{P}}(1)$ by the Central Limit Theorem we get

$$n^{-1/2}\mathcal{N}_{n,c} = n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)\} (Z_i - \bar{Z}) + o_{\mathbb{P}}(1).$$

By independence between ε_i and Z_i we have

$$\mathbf{E}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\} (Z_i - \bar{Z}) = \mathbf{E}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\} \mathbf{E}(Z_i - \bar{Z}) = 0.$$

Hence,

$$\mathbf{V}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\} (Z_i - \bar{Z}) = \mathbf{E}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\}^2 \mathbf{E}(Z_i - \bar{Z})(Z_i - \bar{Z})'.$$

Notice that

$$\mathbf{E}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\}^2 = \mathbf{E}\{(\varepsilon_{i,c}/\sigma)^4 - 2(\varepsilon_{i,c}/\sigma)^2\tau_2^c/\tau_0^c + 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_2^c/\tau_0^c)^2\}.$$

In particular, let $\varsigma_4^c = \tau_4^c - (\tau_2^c)^2/\tau_0^c$, so that,

$$\mathbf{E}\{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\}^2 = \tau_4^c - 2(\tau_2^c)^2/\tau_0^c + \tau_0^c(\tau_2^c/\tau_0^c)^2 = \varsigma_4^c.$$

Hence, by the Central Limit Theorem,

$$n^{-1/2}\mathcal{N}_{n,c} = n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - 1_{(|\varepsilon_i| \leq \sigma c)}\tau_2^c/\tau_0^c\} (Z_i - \bar{Z}) + o_{\mathbb{P}}(1) \rightarrow_{\mathbb{D}} \mathbf{N}(0, \varsigma_4^c \Sigma_Z).$$

3. Combine items 1 and 2: From item 1,

$$\text{Den} = \tau_4^c/\tau_0^c - (\tau_2^c/\tau_0^c)^2 + o_{\mathbb{P}}(1) = \varsigma_4^c/\tau_0^c + o_{\mathbb{P}}(1).$$

From item 2.1,

$$n^{-1}\mathcal{M}_n \rightarrow \tau_0^c \Sigma_Z.$$

From item 2.2,

$$n^{-1/2}\mathcal{N}_{n,c} \rightarrow_{\mathbb{D}} \mathbf{N}(0, \varsigma_4^c \Sigma_Z).$$

Hence, combining via the Slutsky theorem,

$$\tau_0^c n R_{n,c}^2 \rightarrow_{\mathbb{D}} \frac{\mathbf{N}(0, \varsigma_4^c \Sigma_Z)(\tau_0^c \Sigma_Z)^{-1} \mathbf{N}(0, \varsigma_4^c \Sigma_Z)'}{\varsigma_4^c / \tau_0^c} \sim \chi_1^2,$$

as stated. \square

Proof of Theorem 3.3: 1. *Expanding $\hat{\beta}$:* The robustified least squares estimator is

$$\hat{\beta} = \left\{ \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)} \right\}^{-1} \sum_{i=1}^n x_i y_i \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)}.$$

Inserting the model $y_i = x_i' \beta + \varepsilon_i$, we can write

$$\hat{\beta} = \beta + \left\{ n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)} \right\}^{-1} n^{-1} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)}.$$

By Corollary A.6(c) with $w_{in} = x_i x_i'$ and $k = 0$,

$$n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)} = n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\varepsilon_i| < \sigma_c)} + \mathbf{O}_{\mathbb{P}}(n^{-1/2}),$$

and by Corollary A.6(c) with $w_{in} = x_i$ and $k = 1$,

$$n^{-1} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)} = n^{-1} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}_{(|\varepsilon_i| < \sigma_c)} + \mathbf{O}_{\mathbb{P}}(n^{-1/2}).$$

Since $n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\varepsilon_i| < \sigma_c)}$ converges in probability by Assumptions 2.1 and 2.2 and the Law of Large Numbers, then

$$n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\tilde{\varepsilon}_i| < \tilde{\sigma}_c)} = n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\varepsilon_i| < \sigma_c)} \{1 + \mathbf{O}_{\mathbb{P}}(n^{-1/2})\}.$$

Hence, since $n^{-1} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}_{(|\varepsilon_i| < \sigma_c)}$ also converges in probability in a similar fashion, we get

$$\hat{\beta} = \beta + \left\{ n^{-1} \sum_{i=1}^n x_i x_i' \mathbf{1}_{(|\varepsilon_i| < \sigma_c)} \right\}^{-1} n^{-1} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}_{(|\varepsilon_i| < \sigma_c)} + \mathbf{O}_{\mathbb{P}}(n^{-1/2}),$$

as desired. \square

Proof of Theorem 3.4: The proof follows the same steps as those in item 2.2 in the proof of Theorem 3.1. The main difference with this proof resides in the bias terms, which are different from those in Theorem 3.1 where symmetry is assumed.

We start by noting that as in item 2.2 in the proof of Theorem 3.1 $n^{-1/2} \sigma^{-2} \hat{\mathcal{N}}_{n,c} = n^{-1/2} (\hat{\mathcal{N}}_{1,n,c} - \hat{\mathcal{N}}_{2,n,c})$ where $\hat{\mathcal{N}}_{1,n,c} = \sum_{i=1}^n (\hat{\varepsilon}_{i,c}/\sigma)^2 \tilde{Z}_{i,c}$ and $\hat{\mathcal{N}}_{2,n,c} = \sigma^{-2} (\hat{T}_2^c / \hat{T}_0^c) (\sum_{i=1}^n \tilde{Z}_{i,c})$.

1. The term $n^{-1/2} \hat{\mathcal{N}}_{1,n,c}$. As in item 2.2.1 of Theorem 3.1,

$$n^{-1/2} \hat{\mathcal{N}}_{1,n,c} = n^{-1/2} \sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\} Z_i + \tau_2^c \sqrt{n} \bar{Z} + \mathcal{G}_n^{Z,2}(\hat{\theta}, c) + \mathbf{o}_{\mathbb{P}}(1),$$

where in this case

$$\begin{aligned} \mathcal{G}_n^{Z,2}(\hat{\theta}, c) &= \{c^3 \mathbf{f}(c) - (-c)^3 \mathbf{f}(-c)\} \sigma^{-1} \tilde{a} n^{-1} \sum_{i=1}^n Z_i \\ &\quad + \{c^2 \mathbf{f}(c) - (-c)^2 \mathbf{f}(-c)\} \sigma^{-1} n^{-1/2} \sum_{i=1}^n Z_i x_i' (\tilde{\beta} - \beta) \\ &\quad - 2\tau_1^c \sigma^{-1} n^{-1/2} \sum_{i=1}^n Z_i x_i' (\hat{\beta} - \beta), \end{aligned}$$

by definition of the bias in (A.3) and from (A.4) and (A.5) with $k = 2$, $\hat{a}_1 = \tilde{a}$, $\hat{a}_m = 0$, $w_{in} = Z_i$, $x_{in} = n^{-1/2} x_i$, $\hat{b}_1 = n^{1/2} (\tilde{\beta} - \beta)$, $\hat{b}_m = n^{1/2} (\hat{\beta} - \beta)$.

2. The term $n^{-1/2}\hat{\mathcal{N}}_{2,n,c}$. As in item 2.2.2 of Theorem 3.1,

$$\begin{aligned}\hat{T}_0^c &= n^{-1/2}\{n^{-1/2}\sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{1,0}(\hat{\theta}, c) + o_{\mathbf{P}}(1)\}, \\ \sqrt{n}\hat{T}_2^c/\sigma^2 &= n^{-1/2}\sum_{i=1}^n (\varepsilon_{i,c}/\sigma)^2 + \mathcal{G}_n^{1,2}(\hat{\theta}, c) + o_{\mathbf{P}}(1), \\ \tilde{Z}_c &= n^{-1/2}\{n^{-1/2}\sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{Z,0}(\hat{\theta}, c) + o_{\mathbf{P}}(1)\},\end{aligned}$$

where by definition of the bias in (A.3) and from (A.4) and (A.5) with $k = 0$ or $k = 2$, $w_{in} = 1$ or $w_{in} = Z_i$, $\hat{a}_1 = \tilde{a}$, $a_m = 0$, $\hat{b}_1 = n^{1/2}(\hat{\beta} - \beta)$, $\hat{b}_m = n^{1/2}(\hat{\beta} - \beta)$, we get

$$\begin{aligned}\mathcal{G}_n^{1,0}(\hat{\theta}, c) &= \{cf(c) - (-c)f(-c)\}\sigma^{-1}\tilde{a} + \{f(c) - f(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_i(\tilde{\beta} - \beta), \\ \mathcal{G}_n^{1,2}(\hat{\theta}, c) &= \{c^3f(c) - (-c)^3f(-c)\}\sigma^{-1}\tilde{a} \\ &\quad + \{c^2f(c) - (-c)^2f(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_i(\tilde{\beta} - \beta) \\ &\quad - 2\tau_1^c\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_i(\tilde{\beta} - \beta), \\ \mathcal{G}_n^{Z,0}(\hat{\theta}, c) &= \{cf(c) - (-c)f(-c)\}\sigma^{-1}\tilde{a}\bar{Z} + \{f(c) - f(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n Z_i x'_i(\tilde{\beta} - \beta).\end{aligned}$$

Again, following the same steps as in item 2.2.2 of Theorem 3.1, and expanding B.3, we have

$$\begin{aligned}(\hat{T}_0^c)^{-1}\tilde{Z}_c &= \bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n Z_i \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) \\ &\quad - (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}\bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) \\ &\quad - (\tau_0^c)^{-1}n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c)\bar{Z} \\ &\quad + o_{\mathbf{P}}(n^{-1/2}),\end{aligned}$$

where now

$$n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) - n^{-1/2}\mathcal{G}_n^{1,0}(\hat{\theta}, c)\bar{Z} = n^{-1/2}\{f(c) - f(-c)\}\sigma^{-1}n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})x'_i n^{1/2}(\tilde{\beta} - \beta) = o_{\mathbf{P}}(n^{-1/2}).$$

Note also that

$$\begin{aligned}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) &= n^{-1/2}\{cf(c) - (-c)f(-c)\}\sigma^{-1}\tilde{a}\bar{Z} \\ &\quad + n^{-1/2}\{f(c) - f(-c)\}\sigma^{-1}n^{-1}\sum_{i=1}^n Z_i x'_i n^{1/2}(\tilde{\beta} - \beta) \\ &= O_{\mathbf{P}}(n^{-1/2}),\end{aligned}$$

hence,

$$(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n \{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\}n^{-1/2}\mathcal{G}_n^{Z,0}(\hat{\theta}, c) = O_{\mathbf{P}}(n^{-1/2})O_{\mathbf{P}}(n^{-1/2}) = O_{\mathbf{P}}(n^{-1}).$$

Inserting this results in the expression for $(\hat{T}_0^c)^{-1}\tilde{Z}_c$ and simplifying we get

$$(\hat{T}_0^c)^{-1}\tilde{Z}_c = \bar{Z} + (\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})\{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} + o_{\mathbf{P}}(n^{-1/2}),$$

which is the same expression for $(\hat{T}_0^c)^{-1}\tilde{Z}_c$ in item 2.2.2 of Theorem 3.1. Therefore, as in item 2.2.2 of Theorem 3.1

$$\begin{aligned}n^{-1/2}\hat{\mathcal{N}}_{2,n,c} &= n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c\}\bar{Z} + \sqrt{n}\tau_2^c(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})\{1_{(|\varepsilon_i| \leq \sigma c)} - \tau_0^c\} \\ &\quad + \sqrt{n}\tau_2^c\bar{Z} + \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} + O_{\mathbf{P}}(n^{-1/2}).\end{aligned}$$

3. Combine items 1 and 2. All the steps in item 2.2.3 of Theorem 3.1 apply here. The only thing to recall is that the biases $\mathcal{G}_n^{X,2}(\hat{\theta}, c)$ and $\mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z}$ are now different. In particular, we get

$$\begin{aligned}n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i| \leq \sigma c)}\}\{Z_{i,c} - 1_{(|\varepsilon_i| \leq \sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\} \\ &\quad + \mathcal{G}_n^{Z,2}(\hat{\theta}, c) - \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} + o_{\mathbf{P}}(1),\end{aligned}$$

where

$$\begin{aligned} \mathcal{G}_n^{Z,2}(\hat{\theta}, c) - \mathcal{G}_n^{1,2}(\hat{\theta}, c)\bar{Z} &= \{c^2\mathbf{f}(c) - (-c)^2\mathbf{f}(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n Z_i x'_i(\tilde{\beta} - \beta) \\ &\quad - \{c^2\mathbf{f}(c) - (-c)^2\mathbf{f}(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_i(\tilde{\beta} - \beta)\bar{Z} \\ &\quad - 2\tau_1^c \sigma^{-1}n^{-1/2}\sum_{i=1}^n Z_i x'_i(\hat{\beta} - \beta) \\ &\quad + 2\tau_1^c \sigma^{-1}n^{-1/2}\sum_{i=1}^n x'_i(\hat{\beta} - \beta)\bar{Z} \\ &\quad + \mathbf{o}_P(1). \end{aligned}$$

Inserting this expression in the expansion for $n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c}$ we get

$$\begin{aligned} n^{-1/2}\sigma^{-2}\hat{\mathcal{N}}_{n,c} &= n^{-1/2}\sum_{i=1}^n \{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}\mathbf{1}_{(|\varepsilon_i| \leq \sigma c)}\}\{Z_{i,c} - \mathbf{1}_{(|\varepsilon_i| \leq \sigma c)}(\tau_0^c)^{-1}\bar{Z}_c\} \\ &\quad + c^2\{\mathbf{f}(c) - \mathbf{f}(-c)\}\sigma^{-1}n^{-1/2}\sum_{i=1}^n (Z_i - \bar{Z})x'_i(\tilde{\beta} - \beta) \\ &\quad - 2\tau_1^c \sigma^{-1}n^{-1/2}\sum_{i=1}^n (Z_i - \bar{Z})x'_i(\hat{\beta} - \beta) \\ &\quad + \mathbf{o}_P(1). \end{aligned}$$

as stated. □

References

- Alih, E. and C.O. Ong, 2015, An outlier-resistant test for heteroscedasticity in linear models. *Journal of Applied Statistics* 42, 1617-1634.
- Ascombe, F.J., 1961, Examination of residuals. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 1-36.
- Bercu, B. and A.Touati, 2008, Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848–1869.
- Berenguer-Rico, V., S. Johansen, B. and Nielsen, 2018, Some new results on marked and weighted empirical processes of residuals. *Mimeo*
- Berenguer-Rico, V. and B. Nielsen, 2018, Marked and weighted empirical processes of residuals with applications to robust regressions. *Mimeo*.
- Bickel, P.J., 1978, Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Annals of Statistics* 6, 266–291.
- Breusch, T.S. and A.R. Pagan, 1979 A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287-1294.
- Carroll, J.C. and D. Ruppert, 1981, On robust tests for heteroscedasticity. *Annals of Statistics* 9, 206-210.
- Chen, S. and J. Bien, 2017, Valid inference corrected for outlier removal. *Mimeo*.
- Evans, M., 1992, Robustness of size of tests of autocorrelation and heteroscedasticity to nonnormality. *Journal of Econometrics* 51, 7-24.
- Glesjer, H., 1969, A new test for heteroskedasticity. *Journal of the American Statistical Association* 64, 316-323.
- Godfrey L.G., 1996, Some results on the Glejser and Koenker tests for heteroscedasticity. *Journal of Econometrics* 72, 275-299.
- Goldfeld, S.M. and R.E. Quandt, 1965, Some tests for homoscedasticity. *Journal of the American Statistical Association* 61, 539-547.
- Im, K.S., 2000, Robustifying Glejser test of heteroskedasticity. *Journal of Econometrics* 97, 179-188.
- Johansen, S. and B. Nielsen, 2016, Analysis of the Forward Search using some new results for martingales and empirical processes. *Bernoulli* 22, 1131–1183.
- Kalina, J., 2011, Testing heteroscedasticity in robust regression. *Research Journal of Economics Business and ICT* 4, 25-28.
- Kim, E.H., A. Morse and L. Zingales, 2006, What has mattered in economics since 1970. *Journal of economic perspectives* 20, 189-202.
- Koenker, R., 1981, A note on studentizing a test for heteroscedasticity. *Journal of Econometrics* 17, 107-112.

- Koenker, R. and G. Bassett, 1982, Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43-61.
- Koul, H.L. and M. Ossiander, 1994, Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540-562.
- Lebreton, M. and A. Peguin-Feissolle, 2007, Robust tests for heteroscedasticity in a general framework. *Annales d'Economie et de Statistique* 85, 159-187.
- Lyon, J.D. and C. Tsai, 1996, A comparison of tests for heteroscedasticity. *The Statistician* 45, 337-349.
- Machadom, J.A.F. and J.M.C Santos Silva, 2000, Glejser's test revisited. *Journal of Econometrics* 97, 189-202.
- Maronna, R.A., R.D. Martin, and V.J. Yohai, 2006, *Robust Statistics: Theory and Methods*. John Wiley & Sons, Hoboken, New Jersey.
- Rousseeuw, P.J., 1984, Least Median of Squares Regression. *Journal of the American Statistical Association* 79, 871-880.
- Van Dijk, D., P.H. Franses, and A. Lucas, 1999, Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* 14, 539-562.
- Welsh, A.H. and E. Ronchetti, 2006, A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference* 103, 287-310.
- Yohai, V.J., 1987, High breakdown-point and high efficiency estimates for regression. *Annals of Statistics*, 15, 642-656.