

Heteroscedasticity testing after outlier removal

Vanessa Berenguer-Rico* and Ines Wilms†

20 January 2020

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 inferential procedures are, however, underexplored. We study the effects of such a strategy on testing for heteroscedasticity. Specifically, 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. In a simulation study, we show that - depending on the type of outliers - 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. Given these results, we devise an empirical modeling strategy to guide practitioners whose preferred approach is to remove outliers from the sample.

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

JEL classification: C01; C10.

1 Introduction

Regression and specification testing based on classical least squares principles are, in general, not robust to the presence of outlying or extreme observations. To avoid this problem, one could opt for robust regression techniques –e.g., Maronna, Martin and Yohai (2006). Yet, least squares remains, in practice, the predominant approach presumably due to its elegant, simple use and easy accessibility in standard software packages. To address its lack of robustness to the presence of outliers, however, it is a 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. In the case of testing for normality of residuals after implementing the *DAS*, Berenguer-Rico and Nielsen (2018) have 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. Johansen and Nielsen (2009), Chen and Bien (2017) or Berenguer-Rico, Johansen and Nielsen (2019a) show that

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

†Department of Quantitative Economics, Maastricht University, The Netherlands. Email address: i.wilms@maastrichtuniversity.nl

conducting standard inferences on the parameters of the model after detecting and removing outliers can be misleading. These results are reminiscent of the phenomenon of post-selection inference – see, for instance, Pöschner (1991) or Leeb and Pöschner (2005) – which is well-documented in the literature.

However, the consequences of the *DAS* on testing procedures, other than the cases just mentioned, are still underinvestigated. In this paper we focus on the effect of the *DAS* on the White test for heteroscedasticity (White, 1980) although the results that follow could be used to study other heteroscedasticity testing procedures as well. It is worth emphasizing that rather than proposing a procedure to test for heteroscedasticity, our main goal is to theoretically, as well as, computationally analyze the properties of the White heteroscedasticity test after having implemented the *DAS*.

We focus on White’s heteroscedasticity test since it is routinely used and reported in academic work on, for instance, agricultural economics (e.g., Bold et al., 2017), development economics (e.g., Hirvonen et al., 2017), innovation economics (e.g., Rojas et al., 2018; Stanko and Henard, 2017), financial economics (e.g., Bauer and Neuenkirch, 2017), fiscal policy (e.g., Chen et al., 2017) or macro-economics (e.g., Atalla et al. 2016; Montagnoli et al. 2016). Over and above, in some cases, researchers start their empirical analysis by removing outliers and subsequently perform the White heteroscedasticity test –see, for instance, Akhtaruzzaman et al. (2018), Ling and Wahab (2018), Carrera et al. (2017) and Stanko and Henard (2017), amongst others. More generally, practitioners performing standard regression analysis are routinely confronted with the White test via built-in commands in statistical software packages such as Eviews, STATA or OxMetrics. Finally, even though the usage of robust standard errors in ordinary least squares regressions is a widespread practice, Romano and Wolf (2017) show that efficiency gains can be obtained by using an adaptive procedure that ties the choice of estimation method (ordinary versus weighted least squares) to the outcome of a heteroscedasticity test, such as White’s test. All in all, given its frequent occurrence in research and data analysis and its potential in detecting heteroscedasticity, it is important to understand the consequences of the outlier removal stage on the validity of the White test.

Several studies have investigated the consequences of outliers on heteroscedasticity testing procedures. Lyon and Tsai (1996) study the effect of outliers for a set of eight different tests for heteroscedasticity showing that the White test seems to be highly affected. Van Dijk et al. (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. Hendry and Santos (2005) investigate through simulations the effect on the White test of including dummies in the model to account for potential outliers. 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.

We contribute to the literature on heteroscedasticity testing 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. Specifically, by using marked and weighted empirical processes of residuals techniques recently developed in Berenguer-Rico and Nielsen (2018) and Berenguer-Rico, Johansen and Nielsen (2019a,b), we show that the *DAS* version of the White test statistic is asymptotically χ^2 under symmetric errors.

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 outlier configuration 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.

To illustrate how to implement in practice the *DAS* version of the White test, we conduct the following empirical exercise. We use data taken from Wooldridge (2013) and test for het-

eroscedasticity in a simple housing price regression. While the outcome of the standard White test is considerably affected by the presence of outliers, the *DAS* White test statistics consistently point towards the absence of heteroscedasticity.

Several recommendations to practitioners and applied researchers can be made based on our findings. First, implementing the White test for heteroscedasticity after having removed outliers requires a symmetric distribution of the error term. More generally, removing outliers requires certain knowledge of the error distribution since outliers are defined with respect to a reference distribution, usually the normal distribution. Hence, if practitioners pursue the route of removing outliers, we recommend to combine heteroscedasticity testing with testing for normality in the context of the *DAS* –see Berenguer-Rico and Nielsen (2018). Second, a non-rejection of the null hypothesis of homoscedasticity allows practitioners to proceed with least squares, although with a consistency and efficiency correction –see Johansen and Nielsen (2009, 2016b). However, if the null of homoscedasticity is rejected, conducting correct inferences on the parameters of the model requires careful modeling. As a matter of fact, classical heteroscedasticity robust standard errors are not valid in the context of the *DAS*. Since outliers are selected from standardized residuals, an estimate of the residual scale is required. Therefore, under heteroscedasticity a corresponding heteroscedastic function for the standard deviation needs to be modeled and estimated in order to properly identify and remove outliers from the sample.

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. 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. Section 5 contains a discussion on: (i) the effect of asymmetries on both the standard and the *DAS* version of the White test; (ii) how to test for normality in the context of the *DAS*; (iii) the usage of heteroscedasticity robust standard errors in the context of the *DAS*. Section 6 contains an empirical illustration. Section 7 concludes. All the proofs are collected in the Appendix.

2 Model and Assumptions

We consider the following linear model

$$y_i = \mu + z_i' \alpha + \varepsilon_i = \beta' x_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

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 $F(c) = P(\varepsilon_i \leq \sigma c)$ and unknown scale $\sigma > 0$. The *DAS* can be formalized as follows. Let $\hat{\beta}, \tilde{\sigma}$ be some initial estimators of the unknown β and σ . From these estimates, compute the first step residuals $\tilde{\varepsilon}_i = y_i - x_i' \hat{\beta}$ 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,

$$\hat{\beta} = \left\{ \sum_{i=1}^n x_i x_i' 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)} \right\}^{-1} \sum_{i=1}^n x_i y_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma} c)}, \quad (2.2)$$

but other estimators could be used instead. This gives updated residuals

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

for a second stage estimator $\hat{\beta}$. In the context of the *DAS*, these residuals are then used in misspecification testing to analyze the fit of the model after the removal of outliers. Berenguer-Rico and Nielsen (2018) have shown that the standard moment based normality test using $\hat{\varepsilon}_{i,c}$ is invalid since the normalizing constants depend on the cut-off value c and the estimation method

being used. However, the properties of misspecification tests using $\hat{\varepsilon}_{i,c}$ are, in general, unknown. In the next section we study the properties of White's test for heteroscedasticity in this context. We first describe the theoretical framework in the following assumptions.

Assumption 2.1 *Let ε_i/σ be independent and identically distributed, independent of z_i , and with a positive density f on \mathbb{R} . Suppose:*

- (a) *moments: $E(\varepsilon_1^8) < \infty$;*
- (b) *$(1 + |u|^8)f(u)$ is Lipschitz with a weakly unimodal bound.*

Assumption 2.2 *Let z_i be independent and identically distributed so that $E|z_i|^{6+\omega} < \infty$ for some $\omega > 0$.*

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

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 Berenguer-Rico and Nielsen (2018). In particular, the authors derive an iterated exponential martingale inequality –based on Bercu and Touati (2008) and Johansen and Nielsen (2016a) inequalities– which requires, in our setting, that the fourth moment of ε_i^2 exists. The boundedness and continuity condition in Assumption 2.1(b) is used in Berenguer-Rico and Nielsen (2018) –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 squared regressors appear in the auxiliary regression of the White test. These then form 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 theory requires that the second moment of the weights times the regressors squared, that is the sixth moment of the regressors, exist.

The boundedness of the rescaled estimators in Assumption 2.3 applies to a variety of procedures. The \sqrt{n} -consistency property of the initial estimators in an i.i.d. setting holds by, for instance, the least squares estimator, the least trimmed squares estimator of Rousseeuw (1984), or the MM-estimator of Yohai (1987) –see Vřšek (1999) for an asymptotic analysis of the least trimmed squares estimator and Marona, Martin and Yohai (2006) for the MM-estimator. In the second stage, the estimator $\hat{\beta}$ defined in (2.2) will be \sqrt{n} -consistent under symmetry, if the first stage estimators are also \sqrt{n} -consistent (see Johansen and Nielsen, 2016a, 2016b and Berenguer-Rico, Johansen and Nielsen, 2019a,b). In our simulations section, given its widespread use, special attention will be devoted to the robustified least squares procedure, that is, the case in which $\tilde{\beta}, \tilde{\sigma}$ are least squares estimates of β and σ and $\hat{\beta}$ is as defined in (2.2). Other robust estimation procedures, such as the least trimmed squares estimator or the MM-estimator will also be considered.

3 Heteroscedasticity Testing after Outlier Removal

The standard White's heteroscedasticity test is based on the multiple correlation coefficient R^2 from the auxiliary regression

$$\tilde{\varepsilon}_i^2 = \delta_0 + \delta' Z_i + u_i, \quad i = 1, \dots, n,$$

where $Z_i = \{z_i', \text{vech}(z_i z_i')'\}'$ and the *vech* operator vectorizes the upper triangle of its symmetric argument. The auxiliary null hypothesis is $\delta = 0$. White (1980) showed that under standard regularity conditions the statistic nR^2 converges to a chi-square distribution. White's test procedure became very popular and it is nowadays implemented in most of the widely used econometric software.

Given the effect that outliers can have on regression analysis, researchers may start by detecting atypical values in the dataset and removing them from the sample. In order to analyze the effect of outlier removal on heteroscedasticity testing, we model the White test when the *DAS* has been implemented in advance, that is, when outlying observations satisfying $|\tilde{\varepsilon}_i| > \tilde{\sigma}c$ have been previously detected and deleted from the sample. Mathematically, this can be described as follows. Let S be the number of observations for which $|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c$ and denote by $\hat{\varepsilon}_{i,c}^S$ the second stage residuals for which individual i satisfies $|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c$. Let $Z_{i,c}^S$ by the corresponding auxiliary regressors. The *DAS* version of the White heteroscedasticity test is based on the multiple correlation coefficient $\hat{R}_{S,c}^2$ from the auxiliary regression

$$\hat{\varepsilon}_{i,c}^S = \gamma_0 + \gamma' Z_{i,c}^S + u_{i,c}^S, \quad i = 1, \dots, S, \quad (3.1)$$

where $u_{i,c}^S$ is an auxiliary error term. The auxiliary hypothesis is $\gamma = 0$ which amounts to $q = k - 1 + k(k - 1)/2$ restrictions. The statistic is then $S\hat{R}_{S,c}^2$.

When presenting our theoretical results, it is convenient to work with the numerically equivalent statistic

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

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

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

and $\tilde{Z}_c = n^{-1} \sum_{i=1}^n Z_i 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}$.

The statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ is based on two sided marked and weighted empirical distribution functions

$$\hat{\mathcal{G}}_n^{Z,k}(c) = n^{-1} \sum_{i=1}^n Z_i \hat{\varepsilon}_i^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)} \quad \text{or} \quad \hat{\mathcal{G}}_n^{ZZ',k}(c) = n^{-1} \sum_{i=1}^n Z_i Z_i' \hat{\varepsilon}_i^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)},$$

for $k = 0, 1, 2, \dots$. New results by Berenguer-Rico and Nielsen (2018) and Berenguer-Rico, Johansen and Nielsen (2019a,b) –extending previous work by Johansen and Nielsen (2016a) or Koul and Ossiander (1994)– provide uniform asymptotic expansions for $\hat{\mathcal{G}}_n^{Z,k}(c)$ or $\hat{\mathcal{G}}_n^{ZZ',k}(c)$ and their associated empirical processes. These results allow us to express $\hat{\mathcal{G}}_n^{Z,k}(c)$ and $\hat{\mathcal{G}}_n^{ZZ',k}(c)$ in terms of their estimation-error-free version

$$\mathcal{G}_n^{Z,k}(c) = n^{-1} \sum_{i=1}^n Z_i \varepsilon_i^k 1_{(|\varepsilon_i| \leq \sigma c)} \quad \text{and} \quad \mathcal{G}_n^{ZZ',k}(c) = n^{-1} \sum_{i=1}^n Z_i Z_i' \varepsilon_i^k 1_{(|\varepsilon_i| \leq \sigma c)}.$$

In turn, this allows us to show the following result.

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,*

$$\hat{T}_0^c n \hat{R}_{n,c}^2 = \tau_0^c n R_{n,c}^2 + o_P(1), \quad (3.3)$$

where

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

and

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

with $\tau_k^c = E(\varepsilon_i / \sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)}$.

Theorem 3.1 shows that the White test applied to the truncated squared residuals, that is $\hat{T}_0^c n \hat{R}_{n,c}^2$, 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 $\tau_0^c n R_{n,c}^2$. We can then use this fact to proof that $\hat{T}_0^c n \hat{R}_{n,c}^2$ is asymptotically χ_q^2 under a symmetric error distribution, as stated in the following result.

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 $c > 0$, $\hat{T}_0^c n \hat{R}_{n,c}^2 \rightarrow_d \chi_q^2$.*

Theorem 3.2 shows that the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ is asymptotically chi-square under the null hypothesis of homoscedasticity. Next, we investigate the consistency of the test in a heteroscedastic model.

Theorem 3.3 *For $\gamma \in (0, 1)$, let ς_1 be a subset of the set $\{1, \dots, n\}$ with $\nu_1 = \text{int}[\gamma n]$ elements and let ς_2 be the complement of ς_1 with ν_2 elements. Define*

$$\sigma_i^2 = \begin{cases} \sigma_1^2 & \text{for } i \in \varsigma_1 \\ \sigma_2^2 & \text{for } i \in \varsigma_2 \end{cases}.$$

Consider the model

$$y_i = \mu + \alpha 1_{(i \in \varsigma_1)} + \varepsilon_i = \beta' x_i + \varepsilon_i,$$

where ε_i/σ_i are independent and identically distributed with a positive and symmetric density f and $\tau_k^c = E\{(\varepsilon_i/\sigma_i)^k 1_{(|\varepsilon_i| \leq \sigma_i c)}\}$. Let $\lambda_k^c = (\sigma_2^k \tau_k^{c\sigma/\sigma_2})/(\sigma_1^k \tau_k^{c\sigma/\sigma_1})$. Then,

- (i) the least squares estimators $\tilde{\beta}, \tilde{\sigma}^2$ and the second stage estimator $\hat{\beta}$ in (2.2) satisfy Assumption 2.3 with $\beta = (\mu, \alpha)'$ and $\sigma^2 = \gamma \sigma_1^2 + (1 - \gamma) \sigma_2^2$.*
- (ii) for each $c > 0$ and $\lambda_2^c \neq \lambda_0^c$, the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$, computed using the estimators in (i), diverges.*

Theorem 3.3 shows that the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ is consistent in a model with a piecewise pattern of heteroscedasticity. Specifically, different subsets of individuals have both different means and variances. Notice that the statistic has power as long as $\lambda_2^c \neq \lambda_0^c$. In particular, when $c = \infty$, so that there is no removal of outliers and the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ becomes the standard White test, the condition $\lambda_2^c \neq \lambda_0^c$ implies that $\sigma_2^2 \neq \sigma_1^2$.

A key aspect in the derivation of Theorem 3.3 is the piecewise structure of the mean and variance in the model. In particular, this implies that the marked and weighted empirical processes theory in Berenguer-Rico and Nielsen (2018) can be applied to each subsample and these can be combined to show, first, that the second stage estimator $\hat{\beta}$ is \sqrt{n} -consistent for $\beta = (\mu, \alpha)'$, and, second, that the test statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ diverges. However, the existing theory on marked and weighted empirical processes of residuals does not cover general patterns of heteroscedasticity and its generalization is beyond the scope of this paper. Nonetheless, we study the power of the test statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ in more general conditional heteroscedastic models by simulation in the next section.

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 $\varepsilon_i \sim i.i.d. N(0, \sigma_i^2)$. As sample sizes, we take $n \in \{50, 100, 500, 1000\}$. As number of predictors, we take $p \in \{1, 2, 4\}$. These results are a representative selection from an extensive simulation study we have performed.

Additional experiments with other sample sizes and dimensions are available from the authors upon request.

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 (Rousseeuw, 1984) as initial estimator, (iv) *DASWhite MM*: the *DAS* White test with the MM estimator (Yohai, 1987) 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).

We start by analyzing the size and power of the four test statistics when no outliers (“No contamination”) are present. Next, we investigate how their size and power is affected by the presence of outliers (i.e. under “Contamination”).

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 tests in uncontaminated settings, nominal size 5%.

cut-off	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</i>												
	<i>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>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>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>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>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>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>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>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>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, in the sense that its empirical sizes are the closest to the nominal size of 5%. Consider first the results for cut-off $c = 1.96$. The *DAS* White test statistics are, overall, competitive to the standard *White* with empirical sizes close to the nominal size. 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

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

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 $n \in \{50, 100, 500, 1000\}$ 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 $n = 50$ and $n = 100$ 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 framework introduced in Section 2 and the corresponding theoretical results discussed in Section 3 do not cover contaminated settings, we also compare the empirical sizes (and power in the next section) of the four test statistics in the presence of outliers. To our knowledge, no Monte Carlo results on the *DAS White* test statistics under contaminated settings have been discussed 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. The first two schemes are designed to show that the standard White test can be severely influenced in the presence of outliers. Depending on the outlier configurations, its empirical size can explode to 100% (first scheme) or implode to 0% (second scheme). The third scheme is designed to illustrate that also *DASWhite LS* can be severely affected by the presence of outliers since the initial LS estimator is non-robust. Next, we provide specifics on each of the three contamination schemes.

In the first two schemes, we only contaminate the first two observations: the first observation is located at $(z, y) = (z, 10)$, the second at $(z, y) = (z, -10)$. In the “bad leverage points” scheme, we take $z = 5$, giving observations outlying in both the z - and y -direction, known as bad leverage points in the robustness literature. In the second “vertical outliers” scheme, we take $z = 0$, 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 $(z, y) = (10, 20)$. 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 behaviour although it rejects the null of homoscedasticity due to the heteroscedasticity induced by the outliers. 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 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, due to the unbounded influence of the outliers on the least squares estimator, see Huber and Ronchetti (2009). 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, on which outliers have a bounded influence– 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 $n = 50$ to $n = 1000$, the standard *White* test continues to have no power.

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. The reason behind this feature is that under the chosen outlier configuration, the outliers greatly increase the residual variance of the least squares whereas the residual scale of the LTS and MM remains largely unaffected by the presence of the outliers. Specifically, the least squares regression line is drawn towards the outliers which increases the (absolute) residual values of all clean observations and greatly increases the residual variance of the LS. On the contrary, the regression line obtained from using the robust LTS and MM estimators is unaffected by the presence of outliers. This gives additional power to the *DASWhite LS*, relative to the *DASWhite LTS* and *DASWhite MM*, to detect heteroscedasticity.

Even though *DASWhite LS* is more powerful than *DASWhite LTS* and *DASWhite MM*, we recall from Table 2 that the size of the former statistic can be severely affected by the presence of masked outliers while the later statistics do not suffer from the masking effect. We would therefore recommend the use of *DASWhite LTS* and *DASWhite MM* in practice.

5 Discussion

5.1 On asymmetric error distribution

The symmetry assumption on the density f of the standardized error term is fundamental in the derivation of Theorem 3.2. In particular, under symmetry, the biases in the asymptotic expansion of $\hat{T}_0^c n \hat{R}_{n,c}^2$ coming from the estimation errors in the first and second stages of the *DAS* cancel out so that the asymptotic equivalence (3.3) holds. Since the statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$, as well as its asymptotically equivalent version $\tau_0^c n R_{n,c}^2$, are self-normalized, a χ^2 limiting distribution is obtained in the limit.

Under asymmetry, two issues arise in the context of the *DAS*. First, the second stage estimator $\hat{\beta}$ will be inconsistent for the intercept –see Berenguer-Rico, Johansen and Nielsen (2019a). This inconsistency makes Assumption 2.3 invalid since $n^{1/2}(\hat{\beta} - \beta)$ is not bounded in probability. Second, even if Assumption 2.3 were to hold, by bias correcting $\hat{\beta}$ say, a second issue arises under asymmetric errors. In particular, the result $\hat{T}_0^c n \hat{R}_{n,c}^2 = \tau_0^c n R_{n,c}^2 + o_P(1)$ derived in Theorem 3.1 above 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 $\hat{T}_0^c n \hat{R}_{n,c}^2$. These biases make the test statistic $\hat{T}_0^c n \hat{R}_{n,c}^2$ not to be self-normalized. Given the involved nature of the biases, constructing and analyzing a self-normalized statistic in the asymmetric case is beyond the scope of this paper. Instead, in this section, we study the consequences of an asymmetric distribution through simulations. Then, a discussion about how to analyze the underlying density f in the context of the *DAS* follows in the next section.

In the following experiments, we consider the data generating process without contamination studied in Section 4 but with the following seven asymmetric error distributions: (i-ii) χ^2 with one and four degrees of freedom, respectively, and inspired by Im (2000); (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.

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 bias coming from the estimation error of the residuals in the mark function of the numerator of the statistic, the *DAS* *White* tests are also affected by 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. Hence, under asymmetry, asymptotic χ^2 critical values are not valid. Since the underlying density f is unknown in practice, the question then is how to proceed in applications?

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

Error Distribution	Statistic	$n = 50$			$n = 100$			$n = 500$			$n = 1000$			$n = 10000$		
		$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$
$\chi^2(1)$	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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
$\chi^2(4)$	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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	White	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
	LS	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
	LTS	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
	MM	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	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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)	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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)	White	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
	DASWhite															
	LS	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
	LTS	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
	MM	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

5.2 On testing for normal errors in the context of the *DAS*

In standard contexts, where no outliers have been removed from the sample, the White test is usually performed jointly with other specification tests such as normality or functional form tests. In a situation where outliers are taken into consideration, knowledge of the error distribution becomes even more fundamental. Outliers are defined in terms of a reference distribution, often the normal distribution. Hence, if the preferred approach by an applied researcher is to remove outliers from the sample when modeling the data, we suggest to perform the following *DAS* normality test, studied in Berenguer-Rico and Nielsen (2018), simultaneously with the *DAS* White test.

Let

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

where

$$\hat{\mu}_{k,c} = \left\{ \sum_{i=1}^n 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}_c)} \right\}^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}_c)},$$

and the normalizing constants $\lambda_{3,c}, \lambda_{6,c}, \lambda_{24,c}$ depend on the procedure implemented to select outliers. If no outliers are removed so that $c = \infty$ and the least squares estimator is used, then $\lambda_{3,\infty} = 3, \lambda_{6,\infty} = 6, \lambda_{24,\infty} = 24$, which is the usual Jarque-Bera test. However, if the *DAS* is implemented the normalizing constants $\lambda_{3,c}, \lambda_{6,c}, \lambda_{24,c}$ depend on the truncation level c and the estimation method used in the first stage. Table 1 in Berenguer-Rico and Nielsen (2018) provides values for $\lambda_{3,c}, \lambda_{6,c}, \lambda_{24,c}$ for different initial estimators used to remove outliers. Under the assumption of normal errors, the test statistic $(\hat{T}_{3,c})^2 + (\hat{T}_{4,c})^2$ is shown to be asymptotically χ^2_2 —see Theorems 4.1 and 4.2 in Berenguer-Rico and Nielsen (2018).

In practice, applying the *DAS* normality test and the *DAS* White test simultaneously will require an adjustment of the significance level. Following the standard recommendation—see, for instance, chapter 9 in Hendry and Nielsen (2007)—we suggest to use a significance level $\alpha/2$ for each test, where α is the global significance level, usually 5%, and 2 refers to the number of specification tests implemented, two in this case. As in standard econometric modeling practice, if any of the null hypotheses is rejected, a misspecification is flagged and further modeling needs to be undertaken.

5.3 On robust White standard errors

The current common praxis in applied work, when heteroscedasticity is suspected to be an issue, is the calculation of White robust standard errors, which are robust to the presence of heteroscedasticity and also suitable under homoscedasticity. Hence, the appealingness of the White robust standard errors is that they provide valid asymptotic inference irrespective of whether there is homoscedasticity or heteroscedasticity.

However, the detection of outliers in the context of the *DAS* cannot be carried out irrespective of whether there is heteroscedasticity or not. To detect outliers, one needs to use standardized residuals as described in Section 2. To be more precise, the *DAS* approach declares observation i as an outlier if $|\tilde{\varepsilon}_i/\tilde{\sigma}| > c$, for an initial estimation $\tilde{\sigma}$ of the unknown scale σ . Hence, it assumes homoscedasticity in the data. Under heteroscedasticity, one should consider $|\tilde{\varepsilon}_i/\tilde{\sigma}_i| > c$ for a suitably modeled and estimated heteroscedastic function $\tilde{\sigma}_i$; otherwise, the residuals are not properly standardized and the detection of outliers may become misleading. In this sense, the outlier removal stage within the *DAS* approach cannot be carried out irrespective of whether there is heteroscedasticity or not.

If one were to compute classical White standard errors after outlier removal, wrong coverage would be obtained. Indeed, even under homoscedastic errors, wrong coverage occurs due to the omission of (i) a consistency factor needed when calculating the residual variance; and (ii) an efficiency factor needed when calculating the standard errors of the estimators after having removed outliers. To be more precise, Corollary 1 in Johansen and Nielsen (2016b) shows that, under homoscedasticity, when the initial estimators $\tilde{\beta}, \tilde{\sigma}$ are the least squares estimators, then the second stage estimator $\hat{\beta}$ in (2.2) satisfies,

$$n^{1/2}(\hat{\beta} - \beta) \rightarrow_d \mathbf{N}(0, \eta_\beta \sigma^2 \Sigma^{-1}),$$

where $\sigma^2 \Sigma^{-1}$ is the usual limiting variance of the least squares estimator and $\eta_\beta = \tau_2^c/(\tau_0^c)^2 + \{4\mathbf{f}(c)\tau_2^c\}/(\tau_0^c)^2 + \{2\mathbf{f}(c)\}^2/(\tau_0^c)^2$. The efficiency factor η_β depends on the cut-off value c and the underlying distribution of the error term through the density \mathbf{f} and its moments τ_2^c and τ_0^c . Moreover, the estimator of σ^2 after having removed outliers requires a consistency factor. In particular, the estimator

$$\hat{\sigma}^2 = (\tau_0^c/\tau_2^c)n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)},$$

is consistent for σ^2 —see, for instance, Theorem 4.1 in Berenguer-Rico, Johansen and Nielsen (2019). Hence, under homoscedasticity, if an adjusted version of the White standard errors is used, basically by multiplying the classical standard errors by $\sqrt{(\tau_0^c/\tau_2^c)\eta_\beta}$, then correct coverage is obtained. Simulation results confirming this fact are available from the authors upon request.

However, under heteroscedasticity, the adjusted version of the White standard errors is no longer valid. One of the reasons for this failure is that the residuals are not standardized accommodating the presence of heteroscedasticity since a common error scale $\tilde{\sigma}$ for all cross-sectional units is used. Therefore, removing outliers cannot be done irrespective of whether heteroscedasticity is present or not. This feature is inherited by the White standard errors computed after applying the *DAS*. As a consequence, if there is evidence of heteroscedasticity, then removing outliers requires a careful modeling of the heteroscedasticity in the data.

6 Empirical Application

We present an application of the White test after outlier detection. We use the *hprice1* data from Wooldridge (2013), publicly available in the R package `wooldridge` (Shea and Brown, 2018), to test for heteroscedasticity in a housing price equation. The response variable `price` is given by house price in \$1000s and we use the following $p = 3$ predictors: `lotsize`, the size of the lot in square feet; `sqrft`, the size of the house in square feet and `bdrms`, the number of bedrooms. There are $n = 88$ houses included in the analysis.

Following Wooldridge (2013), first, we consider a regression model with the variables in levels and, subsequently, a model with variables in logs. In our empirical analysis, we combine the *DAS*

White test with the *DAS* normality test described in Section 5.2. Following our discussion from that section, to account for the simultaneous implementation of these two specification tests, we use a significance level of $5/2 = 2.5\%$ for each of them.

A question that applied researchers face when implementing the *DAS* is: what cut-off value, c , to choose to detect outliers? Here, we make this choice by using the concept of gauge introduced by Hendry and Santos (2010) which, in turn, is based on an idea by Hoover and Perez (1999) in the context of variable selection –see also Johansen and Nielsen (2016b) and Berenguer-Rico, Johansen and Nielsen (2019a). The gauge is based on the number of falsely detected outliers, when there are none. Notice that deciding which observations, $i = 1, \dots, n$, are outliers is a multiple testing procedure. Choosing the cut-off value c using the gauge permits the investigator to control the overall rate of false rejection when there are no outliers; the same way the size controls the probability of rejection when the null is true in a single hypothesis testing scenario. To be precise, the empirical gauge is defined as

$$\hat{\gamma}_c = n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i/\hat{\sigma}| > c)},$$

which is the empirical frequency of rejected outliers. It has been shown by, for instance, Johansen and Nielsen (2016b) and Berenguer-Rico, Johansen and Nielsen (2019a) that $\hat{\gamma}_c \rightarrow_p \mathbf{P}(|\varepsilon_i/\sigma| > c)$. Choosing $\mathbf{P}(|\varepsilon_i/\sigma| > c)$ gives a cut-off value c that controls the overall rejection rate when there are no outliers. Hence, under normality, if one chooses $\mathbf{P}(|\varepsilon_i/\sigma| > c) = 0.05$, then the corresponding c is 1.96. Following the recommendation by Johansen and Nielsen (2016b) and given the sample size in this empirical exercise, we choose $\mathbf{P}(|\varepsilon_i/\sigma| > c) = 0.01$, which gives a cut-off value of 2.58.

6.1 Housing price regression in levels

We start by using the LS, LTS and MM estimators to detect outliers in the regression

$$\text{price}_i = \mu + \alpha_1 \text{lotsize}_i + \alpha_2 \text{sqrft}_i + \alpha_3 \text{bdrms}_i + \epsilon_i. \quad (6.1)$$

Figure 3 visualizes the absolute standardized residuals from regression 6.1 estimated by the LS, LTS and MM estimators, respectively. Absolute standardized residual values larger than $c = 2.58$ are flagged as outliers and indicated in red. The LS detects less observations as outlying: 3 compared to 5 for the LTS and MM. Note that all outliers detected by the MM are also detected by the LTS estimator.

To test for heteroscedasticity, we use the standard *White*, *DASWhite LS*, *DASWhite LTS* and *DASWhite MM* test statistics. We have simultaneously applied the normality test in the context of the *DAS* described in Section 5.2. The test does not reject the null of normality after outlier removal, hence, we find no evidence against the symmetry assumption. Table 4 (first row) reports the values of the four test statistics for testing the null of homoscedasticity in the model with **price** as dependent variable. The standard White test provides strong evidence against homoscedasticity and rejects the null. By contrast, the *DASWhite* test statistics do not reject the null of homoscedasticity. Similar to the “bad leverage points” simulation design, the White test is heavily influenced by the presence of the outliers and rejects the null due to the heteroscedasticity induced by the outliers.

Table 4: Values of the White and *DASWhite* test statistics for testing the null of homoscedasticity in the model with **price** and **log(price)** as dependent variable. The critical value is the 97.5th percentile of the χ_9^2 distribution.

Model	White	<i>DASWhite</i> LS	<i>DASWhite</i> LTS	<i>DASWhite</i> MM	critical value
price	33.731	11.269	10.433	13.809	19.023
log(price)	9.549	9.635	9.635	9.635	19.023

6.2 Housing price regression in logs

We repeat the same analysis as in Section 6.1 but this time taking `price`, `lotsize` and `sqrft` in logarithmic form in the regression model

$$\log(\text{price})_i = \mu + \alpha_1 \log(\text{lotsize})_i + \alpha_2 \log(\text{sqrft}) + \alpha_3 \text{bdrms} + \epsilon_i, \quad (6.2)$$

such that, in line with Wooldridge (2013), the elasticities of `price` with respect to `logsize` and `sqrft` are constant.

Figure 4 visualizes the absolute standardized residuals from regression (6.2) estimated by the LS, LTS and MM estimators, respectively. After applying the logarithmic transformation, the three considered initial estimators detect the same outliers. In Figure 5, we visualize all 88 observations for `log(price)` (vertical axis) against each of the three independent variables (horizontal axis). The three detected outliers are indicated in red. For the outlying observation numbers 24 and 81 the value of `log(price)` is considerably lower compared to the bulk of clean observations whereas for the outlying observation 76 the value of `log(price)` is higher than for the bulk of clean observations especially when plotted against `log(sqrft)`.

Similar to the analysis in Section 6.1, normality of the residuals after outlier removal can not be rejected using the statistics discussed in Section 5.2. Table 4 (second row) reports the values of the test statistics for testing the null of homoscedasticity in the model with `log(price)` as dependent variable. The outcome of the test changes compared to the analysis in Section 6.1 for the standard White test, but the *DASWhite* test statistics arrive at the same conclusion. While the occurrence of less heteroscedasticity in regression models with a logarithmic transformed response has been noticed in many empirical applications, it is likely that the standard White test mainly detects the increased heteroscedasticity induced by the outliers in Section 6.1.

7 Concluding Remarks

The asymptotic distribution of the White test for heteroscedasticity implemented after an initial data analytic strategy used to detect and remove atypical data values has been derived. Under symmetric errors, the standard chi-square distribution is recovered. 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 message for practitioners arises. Removing outliers from a sample can impact subsequent inferential procedures and the consequences need to be addressed. Hence, if the *DAS* is the preferred route in practice, we make the following recommendations. First, when conducting specification testing in the context of the *DAS*, we advise to combine heteroscedasticity tests with normality tests that properly account for having removed outliers. As it is usual in standard practice, we recommend to adjust the significance level for simultaneous specification testing. Second, if the null hypotheses of normality and homoscedasticity are not rejected subsequent inferences via least squares are advisable after an appropriate consistency and efficiency correction –see Johansen and Nielsen (2009, 2016b). However, if any of the null hypotheses is rejected, then a more comprehensive modeling task is required. Classical heteroscedasticity robust standard errors, widely implemented in practice, are not appropriate in the context of the *DAS*. Detecting outliers requires an estimate of the residual scale. Hence, it cannot be carried out irrespective of whether there is heteroscedasticity or not. In other words, the detection of outliers in the context of the *DAS* requires the explicit modeling of the heteroscedasticity in the data.

8 Acknowledgements

We thank the reviewers for their comments and suggestions which substantially improved the quality of the manuscript. The authors wish to thank Bent Nielsen and Christophe Croux for stimulating discussions and helpful comments.

A The Test Statistic: Notation

The DAS White test statistic described in Section 3

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

can be rewritten using the following equivalent notation

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

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

Accordingly, the statistic

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

can be rewritten using

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

where $\varepsilon_{i,c} = \varepsilon_i 1_{(|\varepsilon_i| \leq \sigma_c)}$, $Z_{i,c} = Z_i 1_{(|\varepsilon_i| \leq \sigma_c)}$ 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 $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$.

B Empirical Processes Results

The *DASWhite* test statistic in equation (A.1) is constructed by means of two sided marked and weighted empirical distribution functions, such as,

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

where the weights and the marks are Z_i and $\hat{\varepsilon}_i^k$, respectively. In a more general setting, Berenguer-Rico and Nielsen (2018) have recently developed an asymptotic theory for marked and weighted empirical processes which will be useful in proving the theoretical results in this paper.

B.1 Preliminary Lemmas

The following two results are from Berenguer-Rico and Nielsen (2018) and are included to make this appendix self-contained. The first lemma 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.

Lemma B.1 (Berenguer-Rico and Nielsen (2018) 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,

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

Given Lemma B.1, we can study objects such as $\hat{\mathbb{G}}_n^{Z,k}(c)$ in (B.1) by means of the following auxiliary empirical distribution function. Let $\theta = (a_1, b_1, a_m, b_m)'$, $\theta_1 = (a_1, b_1)'$ and $\theta_m = (a_m, b_m)$. Let also w_{in} be some arbitrary weights and x_{in} be normalized regressors. The two sided auxiliary empirical distribution function of interest is defined as

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

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

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

When $w_{in} = Z_i$, $x_{in} = n^{-1/2} x_i$ and $\mathbb{G}_n^{w,k}$ is evaluated at $\hat{\theta} = (\tilde{a}, \tilde{b}, 0, \hat{b})$ with $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$, $\tilde{b} = n^{1/2}(\tilde{\beta} - \beta)/\sigma$ and $\hat{b} = n^{1/2}(\hat{\beta} - \beta)/\sigma$, then $\mathbb{G}_n^{w,k}(\hat{\theta}, c) = \hat{\mathbb{G}}_n^{Z,k}(c)$, where $\hat{\mathbb{G}}_n^{Z,k}(c)$ was defined in (B.1). Analyzing the uniformity properties of $\mathbb{G}_n^{w,k}(\theta, c)$ over θ and c allows us, via Lemma B.1, to study the properties of statistics such as $\hat{\mathbb{G}}_n^{Z,k}(c)$ in (B.1) which form the DAS White statistic.

The following lemma derives such uniformity results. It is a Corollary to Theorem 2.1 of Berenguer-Rico and Nielsen (2018). Specifically, it describes the asymptotic behavior of the marked and weighted empirical processes of residuals, $\mathbb{G}_n^{w,k}(\theta, c)$, in a general setting, which the next Assumption B.1 describes. In the next section, we will show that Assumptions 2.1 and 2.2 in our setup are sufficient for Assumption B.1.

Assumption B.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 \mathbf{f} on \mathbb{R} . Set $k = 0, 1, 2$. Suppose:

- (i) density and marks: (a) $u^k \mathbf{f}(u)$, (b) $(1 + u^{4k}) \mathbf{f}(u)$, (c) $(1 + k^4 u^{4(k-1)}) \mathbf{f}(u)$, (d) $(1 + k^4 u^{4k}) \mathbf{f}(u)$ are integrable and Lipschitz, with weakly unimodal bounds, while (e) $u^{k+2} \mathbf{f}(u)$ is Lipschitz and (f) $k|u|^{k+1} \mathbf{f}(u) < \infty$ uniformly in c ;
- (ii) regressors: $\max_{1 \leq i \leq n} |n^{1/2-\kappa} x_{in}| = \mathcal{O}_{\mathbb{P}}(1)$ for some $\kappa \in [0, 1/4)$;
- (iii) weights: $\mathbb{E} \sum_{i=1}^n \{|w_{in}|^{2+\omega} + (1 + |w_{in}|^{2+\omega}) |n^{1/2} x_{in}|^2\} = \mathcal{O}(n)$ for some $\omega > 0$.

In Berenguer-Rico and Nielsen (2018) 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 B.1 considers $k = 0, 1, 2$. Berenguer-Rico and Nielsen (2018) make the following assumption on their general mark function $m(u)$.

Assumption B.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 $\int_{-\infty}^{\infty} (1 + |u|^2) \hat{m}(u) \mathbf{f}(u) du < \infty$.

Assumption B.2 is trivially satisfied for $m(u) = u^k$ with $k = 0, 1$. For $k = 2$ we have $\hat{m}(u) = 2u$. Take $|u^* - u| \leq (1 + |u|)\omega$ so that $|\hat{m}(u^*) - \hat{m}(u)| = |2u^* - 2u| = 2|u^* - u| \leq 2(1 + |u|)\omega$. Thus, we choose $\hat{m}(u) = 2$ and see that $\int_{-\infty}^{\infty} (1 + |u|^2) 2\mathbf{f}(u) du < \infty$ by Assumption B.1(a). This shows that the local Lipschitz Assumption B.2 in Berenguer-Rico and Nielsen (2018) is satisfied in the present setting. Hence, it is not included in Assumption B.1 above.

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

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

and

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

where

$$\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_{1n}^{w,k}(\theta_1, c) - \mathcal{G}_{mn}^{w,k}(\theta_m, c), \quad (\text{B.2})$$

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

$$\mathcal{G}_{jn}^{w,k}(\theta_1, c) = \mathcal{B}_{jn}^{w,k}(\theta_1, c) - \lim_{\tilde{c} \downarrow -c} \mathcal{B}_{jn}^{w,k}(\theta_1, \tilde{c}),$$

with

$$\begin{aligned} \mathcal{B}_{1n}^{w,k}(\theta_1, c) &= c^k f(c) n^{-1/2} \sum_{i=1}^n w_{in} (n^{-1/2} a_1 c + x'_{in} b_1), \\ \mathcal{B}_{mn}^{w,k}(\theta_m, c) &= n^{-1/2} \sum_{i=1}^n w_{in} \{ (n^{-1/2} a_m k E(\frac{\varepsilon_i}{\sigma})^k 1_{(\varepsilon_i \leq \sigma c)} + x'_{in} b_m k E(\frac{\varepsilon_i}{\sigma})^{k-1} 1_{(\varepsilon_i \leq \sigma c)}) \}. \end{aligned} \quad (\text{B.3})$$

The next lemma is due to Berenguer-Rico, Johansen and Nielsen (2019a). It derives a uniform 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 B.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 distribution function F . Suppose

(i) innovations ε_i/σ :

(a) $E|\varepsilon_i|^{k(1+\omega)} < \infty$ for some $\omega > 0$;

(b) F is additively and multiplicatively Lipschitz;

(ii) regressors x_{in} : $E n^{-1} \sum_{i=1}^n n^{1/2} |x_{in}| = O(1)$.

(iii) weights w_{in} : $E n^{-1} \sum_{i=1}^n |w_{in}|^{1+\omega} = O(1)$.

Lemma B.3 (Berenguer-Rico, Johansen and Nielsen, 2019a, Theorem 3.1) Suppose Assumption B.3 holds. Then, for any $B > 0$ and $n \rightarrow \infty$,

$$\sup_{c \in \mathbb{R}} \sup_{|a_1|, |b_1| \leq B} |n^{-1} \sum_{i=1}^n w_{in} \varepsilon_i^k \{1_{(\varepsilon_i/\sigma \leq c + n^{-1/2} a_1 c + x'_{in} b_1)} - 1_{(\varepsilon_i/\sigma \leq c)}\}| = o_P(1).$$

B.2 Main Theorems on Empirical Processes

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

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

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

and

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

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

1. Conditions on the error term and mark functions: First note that by Assumption 2.1, ε_i/σ is independent and identically distributed and it is independent of x_i , hence, this is covered by the main setting in Assumption B.1. Now:

(i) By Assumption 2.1 $(1 + |u|^8)f(u)$ is integrable, Lipschitz with a weakly unimodal bound. Then, by Lemma A.7 in Berenguer-Rico, Johansen and Nielsen (2019b) for $k = 0, 1, 2$ we have that (a) $u^k f(u)$, (b) $(1 + u^{4k})f(u)$, (c) $(1 + k^4 u^{4(k-1)})f(u)$, (d) $(1 + k^4 u^{4k})f(u)$ are integrable and Lipschitz, with weakly unimodal bounds. Moreover, by Lemma A.4, A.5 and A.6 in Berenguer-Rico, Johansen

and Nielsen (2019b) we have that (e) $u^{k+2}f(u)$ is Lipschitz and (f) $k|u|^{k+1}f(u) < \infty$ uniformly in c , as required.

2. *Conditions on the regressors and weights:*

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

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

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

(iii) *weights*: set $w_{in} = 1$, $w_{in} = Z_i$ or $w_{in} = Z_i Z'_i$. Then, by Assumption 2.2 we have that $\mathbb{E} \sum_{i=1}^n \{|w_{in}|^{2+\omega} + (1 + |w_{in}|^{2+\omega})|n^{1/2}x_{in}|^2\} = O(n)$ as required by Assumption B.1. \square

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

$$\sup_{c \in \mathbb{R}} \sup_{|a_1|, |b_1| \leq B} |n^{-1} \sum_{i=1}^n \varepsilon_i^4 \{1_{(\varepsilon_i/\sigma \leq c + n^{-1/2}a_1c + x'_{in}b_1)} - 1_{(\varepsilon_i \leq \sigma c)}\}| = o_{\mathbb{P}}(1).$$

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

1. *Conditions on the error term*: First note that by Assumption 2.1, ε_i/σ is independent and identically distributed and it is independent of x_i , hence, this is covered by the main setting in Lemma B.3. Now:

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

(b) By Assumption 2.1(b), we have that $(1 + |c|^8)f(c)$ is Lipschitz with a weakly unimodal bound. Therefore, Lemma A.4 in Berenguer-Rico, Johansen and Nielsen (2019b) gives $\sup_{c \in \mathbb{R}} (1 + |c|^8)f(c) < \infty$. In turn, this implies that F is additively and multiplicatively Lipschitz –see Berenguer-Rico, Johansen and Nielsen (2019a) remark 3.1.

2. *Conditions on the regressors and weights:*

(c) *regressors*: Recall $x_{in} = n^{-1/2}x_{in}$ so that $n^{-1}\mathbb{E} \sum_{i=1}^n |n^{1/2}x_{in}| = n^{-1}\mathbb{E} \sum_{i=1}^n |x_i|$. By Assumption 2.2 we have $n^{-1}\mathbb{E} \sum_{i=1}^n |x_i| = O(1)$ as required by Assumption B.3.

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

Theorem B.6 *Suppose Assumptions 2.1 and 2.2 hold. Then, for any $B > 0$ and $n \rightarrow \infty$,*

$$\sup_{c \in \mathbb{R}} \sup_{|a_1|, |b_1| \leq B} |n^{-1} \sum_{i=1}^n Z_i Z'_i \{1_{(\varepsilon_i/\sigma \leq c + n^{-1/2}a_1c + x'_{in}b_1)} - 1_{(\varepsilon_i \leq \sigma c)}\}| = o_{\mathbb{P}}(1).$$

Proof of Theorem B.5: We need to show that our Assumptions 2.1 and 2.2 imply all the assumptions of Lemma B.3 with $w_{in} = Z_i Z'_i$ and $k = 0$.

1. *Conditions on the error term*: First note that by Assumption 2.1, ε_i/σ is independent and identically distributed and it is independent of x_i , hence, this is covered by the main setting in Lemma B.3. Now:

(a) Since $k = 0$ in this case, no condition on the moments of ε_i is required.

(b) By Assumption 2.1(b), we have that $(1 + |c|^8)f(c)$ is Lipschitz with a weakly unimodal bound. Therefore, Lemma A.4 in Berenguer-Rico, Johansen and Nielsen (2019b) gives $\sup_{c \in \mathbb{R}} (1 + |c|^8)f(c) < \infty$. In turn, this implies that F is additively and multiplicatively Lipschitz –see Berenguer-Rico, Johansen and Nielsen (2019a) remark 3.1.

2. *Conditions on the regressors and weights:*

(c) *regressors*: Recall $x_{in} = n^{-1/2}x_{in}$ so that $n^{-1}\mathbb{E} \sum_{i=1}^n |n^{1/2}x_{in}| = n^{-1}\mathbb{E} \sum_{i=1}^n |x_i|$. By Assumption 2.2, we have $n^{-1}\mathbb{E} \sum_{i=1}^n |x_i| = O(1)$ as required by Assumption B.3.

(d) *weights*: since $w_{in} = Z_i Z'_i$, then, by Assumption 2.2 $\mathbb{E} n^{-1} \sum_{i=1}^n |w_{in}|^{1+\omega} = O(1)$ as required in Lemma B.3. \square

B.3 Corollaries

Under a general density \mathbf{f} , by definition of the biases in (B.2) and (B.3) we have

$$\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_{1n}^{w,k}(\theta_1, c) - \mathcal{G}_{mn}^{w,k}(\theta_m, c), \quad (\text{B.4})$$

with

$$\begin{aligned} \mathcal{G}_{1n}^{w,k}(\theta_1, c) &= \{c^{k+1}\mathbf{f}(c) - (-c)^{k+1}\mathbf{f}(-c)\}a_1n^{-1}\sum_{i=1}^nw_{in} \\ &\quad + \{c^k\mathbf{f}(c) - (-c)^k\mathbf{f}(-c)\}n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_1, \end{aligned} \quad (\text{B.5})$$

and

$$\mathcal{G}_{mn}^{w,k}(\theta_m, c) = k\tau_k^c a_m n^{-1}\sum_{i=1}^nw_{in} + k\tau_{k-1}^c n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_m, \quad (\text{B.6})$$

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

Corollary B.7 *Let Assumptions 2.1 and 2.2 hold. Then, for $k = 0, 1, 2$, uniformly in c and $|\theta| \leq B$:*

- (a) $\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{O}_P(1)$;
- (b) $n^{1/2}\{\mathcal{G}_n^{w,k}(\theta, c) - \bar{\mathcal{G}}_n^{w,k}(0, c)\} = n^{1/2}\{\mathcal{G}_n^{w,k}(0, c) - \bar{\mathcal{G}}_n^{w,k}(0, c)\} + \mathcal{G}_n^{w,k}(\theta, c) + \mathcal{O}_P(1)$;
- (c) $\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_n^{w,k}(0, c) + \mathcal{O}_P(n^{-1/2})$.

Proof of Corollary B.7:

(a) Recall the expressions for $\mathcal{G}_{1n}^{w,k}(\theta_1, c)$ and $\mathcal{G}_{mn}^{w,k}(\theta_m, c)$ in (B.5) and (B.6) and apply the triangle and norm inequalities to get

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

Note that $|\tau_k^c|$, $|\tau_{k-1}^c|$, $|c^{k+1}\mathbf{f}(c)|$, $|c^k\mathbf{f}(c)|$, $|(-c)^{k+1}\mathbf{f}(-c)|$, $|(-c)^k\mathbf{f}(-c)|$ are bounded by Assumption 2.1(a, b); $\sum_{i=1}^n|w_{in}| = \mathcal{O}_P(n)$ and $\sum_{i=1}^n|w_{in}||x_{in}| = \mathcal{O}_P(n^{1/2})$ 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 $\theta = \mathcal{O}(1)$. Hence, the result follows.

(b) Expand

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

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

(c) By part (b) we have

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

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

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

Corollary B.8 *If the density \mathbf{f} is symmetric, then the biases in (B.2) are*

$$\mathcal{G}_{1n}^{w,k}(\theta_1, c) = 1_{(k \text{ even})}2c^{k+1}\mathbf{f}(c)a_1n^{-1}\sum_{i=1}^nw_{in} + 1_{(k \text{ odd})}2c^k\mathbf{f}(c)n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_1, \quad (\text{B.7})$$

$$\mathcal{G}_{mn}^{w,k}(\theta_m, c) = 1_{(k \text{ even})}k\tau_k^c a_m n^{-1}\sum_{i=1}^nw_{in} + 1_{(k \text{ odd})}k\tau_{k-1}^c n^{-1/2}\sum_{i=1}^nw_{in}x'_{in}b_m. \quad (\text{B.8})$$

Proof of Corollary B.8. Notice that $\tau_k^c = E(\varepsilon_i/\sigma)^k \{1_{(\varepsilon_i \leq \sigma c)} - 1_{(\varepsilon_i \leq -\sigma c)}\}$. Also, under a symmetric density $f(c) = f(-c)$ and $\tau_k^c = 0$ for k odd. Applying these results to (B.5) and (B.6) we get the desired expressions. \square

Corollary B.9 *Let Assumptions 2.1, 2.2 and 2.3 hold. Then, for $k = 0, 1, 2$, uniformly in c ,*

- (a) $n^{-1} \sum_{i=1}^n w_i (\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \tilde{\sigma} c)} = n^{-1} \sum_{i=1}^n w_i (\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + O_P(n^{-1/2});$
- (b) $n^{-1/2} \sum_{i=1}^n w_i (\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \tilde{\sigma} c)} = n^{-1/2} \sum_{i=1}^n w_i (\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{G}_n^{w,k}(\hat{\theta}, c) + o_P(1);$
- (c) $\bar{Z}_c = n^{-1} \sum_{i=1}^n Z_i 1_{(|\varepsilon_i| \leq \sigma c)} = \hat{T}_0^c n^{-1} \sum_{i=1}^n Z_i + o_P(1) = \tau_0^c n^{-1} \sum_{i=1}^n Z_i + o_P(1).$

Proof of Corollary B.9: Recall $\hat{\theta} = (\tilde{a}, \tilde{b}, 0, \hat{b})'$ and $\tilde{\theta} = (\tilde{a}, \tilde{b})$ where $\tilde{b} = n^{1/2}(\tilde{\beta} - \beta)/\sigma$, $\hat{b} = n^{1/2}(\hat{\beta} - \beta)/\sigma$ and $\tilde{a} = n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$. By Assumption 2.3 we have that $\tilde{b}, \tilde{\sigma}, \hat{b}$ are bounded in probability. Let $\mathcal{G}_n^{w,k}(\hat{\theta}, c) = n^{-1} \sum_{i=1}^n w_i (\hat{\varepsilon}_i/\sigma)^k 1_{(|\hat{\varepsilon}_i| \leq \tilde{\sigma} c)}$.

(a) By Lemma B.1 it suffices to analyze $\mathcal{G}_n^{w,k}(\theta, c)$ uniformly over both θ and c . Then, by Corollary B.7(c), uniformly in θ and c ,

$$\mathcal{G}_n^{w,k}(\theta, c) = \mathcal{G}_n^{w,k}(0, c) + O_P(n^{-1/2}) = n^{-1} \sum_{i=1}^n w_i (\varepsilon_i/\sigma)^k 1_{(|\varepsilon_i| \leq \sigma c)} + O_P(n^{-1/2}).$$

(b) By Lemma B.1 it suffices to analyze $n^{1/2} \mathcal{G}_n^{w,k}(\theta, c)$ uniformly over both θ and c . Then, by Corollary B.7(b), uniformly in θ and c ,

$$n^{1/2} \mathcal{G}_n^{w,k}(\theta, c) = n^{1/2} \mathcal{G}_n^{w,k}(0, c) + \mathcal{G}_n^{w,k}(\theta, c) + o_P(1),$$

and the result follows.

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

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

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

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

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 $k = 0$ then $\hat{T}_0^c = T_0^c + o_P(1)$ and that $T_0^c = \tau_0^c + o_P(1)$ by the Law of Large Numbers and (iii) since $\bar{Z} = O_P(1)$ by Assumption 2.2 and the Law of Large Numbers, we get

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

Now, (i) by item (a) we have $\hat{T}_0^c = T_0^c + o_P(1)$, (ii) by the Law of Large Numbers $T_0^c = \tau_0^c + o_P(1)$, and (iii) by Assumption 2.2 and the Law of Large Numbers $n^{-1} \sum_{i=1}^n Z_i = O_P(1)$, therefore,

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

as desired. \square

Corollary B.9, 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 B.9 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 B.5. In this way, the result can be derived without further assumptions.

Corollary B.10 : *Under Assumptions 2.1, 2.2 and 2.3,*

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

Proof of Corollary B.10: Let

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

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

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

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

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

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 $1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}$ 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

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

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

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

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 B.1 and Theorem B.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

C Results on Test

Proof of Theorem 3.1: Write the test statistic in (A.1) 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

$$\widehat{Den} = (\hat{T}_0^c)^{-1} n^{-1} \sum_{i=1}^n \left(\frac{\hat{\varepsilon}_{i,c}}{\sigma} \right)^4 - \{ (\hat{T}_2^c / \sigma^2) / \hat{T}_0^c \}^2.$$

By Corollary B.9(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 B.10, 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 $\hat{T}_0^c = T_0^c + O_{\mathbf{P}}(n^{-1/2})$ by Corollary B.9(a), 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 Assumption 2.1 and the Law of Large Numbers, $T_k^c = \sigma^k \tau_k^c + o_{\mathbf{P}}(1)$ where $\tau_0^c > 0$, 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 $T_k^c / \sigma^k = \tau_k^c + o_{\mathbf{P}}(1)$ we can write

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

2. The term \widehat{Num} :

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

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

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

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

By Theorem B.6 and Lemma B.1,

$$n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c} \tilde{Z}_{i,c}' = n^{-1} \sum_{i=1}^n Z_i Z_i' 1_{(|\varepsilon_i| < \sigma_c)} + o_{\mathbf{P}}(1) = n^{-1} \sum_{i=1}^n Z_{i,c} Z_{i,c}' + o_{\mathbf{P}}(1).$$

By Corollary B.9(a) with $w_{in} = Z_i$ and $k = 0$,

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

By Corollary B.9(a) with $w_{in} = 1$ and $k = 0$ and the Law of Large Numbers $\hat{T}_0^c = T_0^c + O_{\mathbf{P}}(n^{-1/2}) = \tau_0^c + o_{\mathbf{P}}(1)$. Hence, since $\bar{Z}_c = O_{\mathbf{P}}(1)$ by Assumptions 2.1 and 2.2 and the Law of Large Numbers we can write

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

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

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

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

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

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 $\sigma^{-2} n^{-1/2} \hat{\mathcal{N}}_{n,c} = \hat{\mathcal{N}}_{1,n,c} - \hat{\mathcal{N}}_{2,n,c}$ where

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

2.2.1. The term $\hat{\mathcal{N}}_{1,n,c}$. Note that $\hat{\mathcal{N}}_{1,n,c} = n^{-1/2} \sum_{i=1}^n (\hat{\varepsilon}_{i,c} / \sigma)^2 Z_i$, hence by Corollary B.9(b) with $w_{in} = \bar{Z}_i$ and $k = 2$ we have

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

where

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

by recalling (i) the bias in (B.4), (ii) that $\hat{\theta} = (\tilde{a}, \tilde{b}, 0, \hat{b})'$, $\hat{\theta}_1 = (\tilde{a}, \tilde{b})$, $\hat{\theta}_m = (0, \hat{b})$, and then (iii) applying Corollary B.8, 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 $\hat{\mathcal{N}}_{1,n,c}$ so that

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

2.2.2. The term $\hat{\mathcal{N}}_{2,n,c}$. Multiply and divide by \sqrt{n} so that $\hat{\mathcal{N}}_{2,n,c} = \sqrt{n} \sigma^{-2} (\hat{T}_2^c / \hat{T}_0^c) \tilde{Z}_c$. We expand each of these terms using Corollary B.9(b) as follows: (i) for \hat{T}_0^c set $w_{in} = 1$, $k = 0$; (ii) for $\sqrt{n} \hat{T}_2^c / \sigma^2$ set $w_{in} = 1$, $k = 2$; and (iii) for \tilde{Z}_c set $w_{in} = Z_i$, $k = 0$. We then get

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

where by Corollary B.8, recalling that $\hat{\theta} = (\tilde{a}, \tilde{b}, 0, \hat{b})'$, $\hat{\theta}_1 = (\tilde{a}, \tilde{b})$, $\hat{\theta}_m = (0, \hat{b})$ so that $a_m = 0$, we have

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

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

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

Multiply and divide by τ_0^c in $\hat{\mathcal{N}}_{2,n,c}$ so that $\hat{\mathcal{N}}_{2,n,c} = \sqrt{n} \sigma^{-2} \hat{T}_2^c (\hat{T}_0^c / \tau_0^c)^{-1} \tilde{Z}_c (\tau_0^c)^{-1}$ and

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

The expansion for $(\hat{T}_0^c/\tau_0^c)^{-1}$, which is,

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

is of the form $1/(1+x_n)$ where $x_n = O_P(n^{-1/2})$ by the Central Limit Theorem and B.7, hence, expanding we can write

$$\frac{1}{1+x_n} = 1 - x_n + O(x_n^2) = 1 - x_n + O_P(n^{-1}).$$

This then gives that

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

Combine the expansions in the product $(\hat{T}_0^c)^{-1}\tilde{Z}_c = (\hat{T}_0^c/\tau_0^c)^{-1}\tilde{Z}_c(\tau_0^c)^{-1}$ so that

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

Expanding this product we have

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

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

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

which simplifies to

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

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

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

Expanding the product we get, by Assumptions 2.1 and 2.2, the Law of Large Numbers, the Central Limit Theorem and Corollary B.7(a),

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

which simplifies to

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

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

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

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 (C.1) and (C.2), hence,

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

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

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

which can be simplified as

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

Since $1_{(|\varepsilon_i|\leq\sigma c)} = \{1_{(|\varepsilon_i|\leq\sigma c)}\}^2$, we can expand $\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c}$ so that

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

By Corollary B.9(c), $\bar{Z} = (\tau_0^c)^{-1}\bar{Z}_c + o_P(1)$. Hence,

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

Notice that the term $n^{-1/2}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^2 - \tau_2^c(\tau_0^c)^{-1}1_{(|\varepsilon_i|\leq\sigma c)}\}$ is $O_P(1)$ by the Central Limit Theorem given Assumption 2.1, hence,

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

This can be rewritten as

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

3. Combine items 1 and 2: Recall $\hat{T}_0^cn\hat{R}_{n,c}^2 = \widehat{Num}/\widehat{Den}$ where

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

By item 1,

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

so that $\widehat{Den} = (\tau_0^c)^{-1}n^{-1}\mathcal{D}_{n,c} + o_P(1)$. By item 2.1,

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

that is $n^{-1}\hat{\mathcal{M}}_{n,c} = n^{-1}\mathcal{M}_{n,c} + o_P(1)$. By item 2.2,

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

so that $\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c} = n^{-1/2}\mathcal{N}_{n,c} + o_P(1)$. Note that the terms $\sigma^{-2}n^{-1/2}\hat{\mathcal{N}}_{n,c}$ and $n^{-1}\hat{\mathcal{M}}_{n,c}$ are $O_P(1)$ by the Central Limit Theorem and the Law of Large Numbers, respectively. Combining we get $\hat{T}_0^cn\hat{R}_{n,c}^2 = \tau_0^cnR_{n,c}^2 + o_P(1)$ as stated. \square

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

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

1. The term Den: Expanding the curly bracket

$$Den = (\tau_0^cn)^{-1}\sum_{i=1}^n\{(\varepsilon_{i,c}/\sigma)^4 - 2(\tau_2^c/\tau_0^c)(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^2 + (\tau_2^c/\tau_0^c)^2(\tau_0^c)^{-1}n^{-1}\sum_{i=1}^n1_{(|\varepsilon_i|\leq\sigma c)}\}.$$

By Assumption 2.1 and the Law of Large Numbers

$$n^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^2 = \tau_2^c + o_P(1), \quad n^{-1}\sum_{i=1}^n1_{(|\varepsilon_i|\leq\sigma c)} = \tau_0^c + o_P(1),$$

where $\tau_0^k > 0$. Therefore,

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

By the Law of Large Numbers, $n^{-1}\sum_{i=1}^n(\varepsilon_{i,c}/\sigma)^4 = \tau_4^c + o_P(1)$, hence, by the Slutsky theorem,

$$Den = \{\tau_4^c/\tau_0^c - (\tau_2^c/\tau_0^c)^2\} + o_P(1).$$

2. The term Num:

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

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

Expand so that, by the Law of Large Numbers,

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

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

$$n^{-1}\mathcal{M}_n = n^{-1}\sum_{i=1}^nZ_iZ_i'1_{(|\varepsilon_i|\leq\sigma c)} - (\tau_0^c)^{-1}\{n^{-1}\sum_{i=1}^nZ_i1_{(|\varepsilon_i|\leq\sigma c)}\}\{n^{-1}\sum_{i=1}^nZ_i1_{(|\varepsilon_i|\leq\sigma c)}\} + o_P(1).$$

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

$$n^{-1}\mathcal{M}_n \rightarrow_P \tau_0^c\mathbf{E}Z_1Z_1' - (\tau_0^c)^{-1}(\tau_0^c)^2\mathbf{E}Z_1\mathbf{E}Z_1' + o_P(1) = \tau_0^c\mathbf{E}(Z_1 - \mathbf{E}Z_1)(Z_1 - \mathbf{E}Z_1)' + o_P(1).$$

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

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

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 B.9(c), $\bar{Z}_c = \tau_0^c \bar{Z} + o_P(1)$. Hence, $(\tau_0^c)^{-1} \bar{Z}_c = \bar{Z} + o_P(1)$, so that

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

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 $O_P(1)$ by the Central Limit Theorem we get

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

By independence between ε_i and Z_i we have

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

Hence,

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

Notice that

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

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

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

Hence, by the Central Limit Theorem,

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

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

$$Den = \tau_4^c/\tau_0^c - (\tau_2^c/\tau_0^c)^2 + o_P(1) = \varsigma_4^c/\tau_0^c + o_P(1).$$

From item 2.1,

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

From item 2.2,

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

Hence, combining via the Slutsky theorem,

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

as stated. \square

Proof of Theorem 3.3:

(i) The first and second stage estimators.

(a) The first stage estimator $\tilde{\beta}$. We want to show $\tilde{\beta} - \beta = O_P(n^{-1/2})$. It is convenient to reparametrize the model as follows

$$y_i = \theta_1 1_{(i \in \varsigma_1)} + \theta_2 1_{(i \in \varsigma_2)} + \varepsilon_i.$$

It suffices to show that the least squares estimator for $\theta = (\theta_1, \theta_2)'$ is $n^{1/2}$ -consistent. Let $m_i = \{1_{(i \in \varsigma_1)}, 1_{(i \in \varsigma_2)}\}'$. We expand

$$\begin{aligned} n^{-1} \sum_{i=1}^n m_i m_i' &= n^{-1} \begin{pmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{pmatrix} \rightarrow \begin{pmatrix} \gamma & 0 \\ 0 & 1 - \gamma \end{pmatrix}, \\ n^{-1/2} \sum_{i=1}^n m_i \varepsilon_i &= n^{-1/2} \begin{pmatrix} \sum_{i \in \varsigma_1} \varepsilon_i \\ \sum_{i \in \varsigma_2} \varepsilon_i \end{pmatrix} = O_P(1), \end{aligned}$$

where the latter bound follows by noting that each sum is a sum of i.i.d. variables and then applying the Central Limit Theorem. Since $0 < \gamma < 1$, we get that $\tilde{\theta} - \theta = (\sum_{i=1}^n m_i m'_i)^{-1} \sum_{i=1}^n m_i \varepsilon_i = O_P(n^{-1/2})$.

(b) The residual variance $\tilde{\sigma}^2$. We want to show $\tilde{\sigma}^2 - \sigma^2 = O_P(n^{-1/2})$. Expand

$$n^{-1/2} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) = n^{-1/2} \sum_{i \in \mathcal{S}_1} (\varepsilon_i^2 - \sigma_1^2) + n^{-1/2} \sum_{i \in \mathcal{S}_2} (\varepsilon_i^2 - \sigma_2^2) + n^{-1/2} (\nu_1 \sigma_1^2 + \nu_2 \sigma_2^2 - n \sigma^2).$$

Each of the first two terms is $O_P(1)$ by the Central Limit Theorem. For the third term notice that

$$n^{-1/2} (\nu_1 \sigma_1^2 + \nu_2 \sigma_2^2 - n \sigma^2) = n^{1/2} \left(\frac{\text{int}[n\gamma]}{n} - \frac{n\gamma}{n} \right) (\sigma_1^2 - \sigma_2^2) \rightarrow 0.$$

Therefore, $\sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) = O_P(n^{-1/2})$. Using also the above expansions we get that

$$\tilde{\sigma}^2 - \sigma^2 = n^{-1} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) - n^{-1} \sum_{i=1}^n \varepsilon_i m'_i (\sum_{i=1}^n m_i m'_i)^{-1} \sum_{i=1}^n m_i \varepsilon_i = O_P(n^{-1/2}).$$

(c) The second stage estimator $\hat{\beta}$. We want to show $\hat{\beta} - \beta = O_P(n^{-1/2})$. We expand

$$\begin{aligned} n^{-1} \sum_{i=1}^n m_i m'_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} &= n^{-1} \begin{pmatrix} \sum_{i \in \mathcal{S}_1} 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} & 0 \\ 0 & \sum_{i \in \mathcal{S}_2} 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} \end{pmatrix}, \\ n^{-1/2} \sum_{i=1}^n m_i \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} &= n^{-1/2} \begin{pmatrix} \sum_{i \in \mathcal{S}_1} \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} \\ \sum_{i \in \mathcal{S}_2} \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} \end{pmatrix}. \end{aligned}$$

For $j = 1, 2$ and $k = 0, 1$, let

$$\tilde{S}_{k,j}^c = \nu_j^{-1} \sum_{i \in \mathcal{S}_j} \varepsilon_i^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)} = \nu_j^{-1} \sum_{i \in \mathcal{S}_j} \varepsilon_i^k 1_{(|\tilde{\varepsilon}_i| \leq \sigma_j c_j + n^{-1/2} \tilde{a}_j c_j)},$$

where $\tilde{a}_j = \sigma_j n^{1/2} (\tilde{\sigma} - \sigma) / \sigma$ and $c_j = (\sigma / \sigma_j) c$ so that $c_j \in \mathbb{R}$. By Corollary B.9(a), (b) with $w_{in} = 1$,

$$\begin{aligned} \tilde{S}_{0,j}^c &= \nu_j^{-1} \sum_{i \in \mathcal{S}_j} 1_{(|\varepsilon_i| \leq \sigma_j c_j)} + O_P(n^{-1/2}), \\ \nu_j^{1/2} \tilde{S}_{1,j}^c &= \sigma_j \nu_j^{-1/2} \sum_{i \in \mathcal{S}_j} (\varepsilon_i / \sigma_j) 1_{(|\tilde{\varepsilon}_i| \leq \sigma_j c_j + n^{-1/2} \tilde{a}_j c_j)} \\ &= \sigma_j \{ \nu_j^{-1/2} \sum_{i \in \mathcal{S}_j} (\varepsilon_i / \sigma_j) 1_{(|\varepsilon_i| \leq \sigma_j c_j)} + 2c_j f(c_j) \nu_j^{1/2} (\tilde{\theta}_j - \theta_j) + o_P(1) \}. \end{aligned}$$

By the Law of Large Numbers $\tilde{S}_{0,j}^c \rightarrow \mathbf{E} 1_{(|\varepsilon_i| \leq \sigma_j c_j)} = \tau_0^{c_j} = \tau_0^{\sigma c / \sigma_j}$. We argue that $\nu_j^{1/2} \tilde{S}_{1,j}^c = O_P(1)$.

The first term in the expansion is $O_P(1)$ by the Central Limit Theorem, while $\nu_j^{1/2} (\tilde{\theta}_j - \theta_j)$ is $O_P(1)$ by the above results for the full sample least squares estimator.

Inserting the results for $\tilde{S}_{k,j}^c$ above we find that

$$\begin{aligned} n^{-1} \sum_{i=1}^n m_i m'_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} &= \begin{pmatrix} \gamma \tau_0^{c_1} & 0 \\ 0 & (1 - \gamma) \tau_0^{c_2} \end{pmatrix} + o_P(1), \\ n^{-1/2} \sum_{i=1}^n m_i \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} &= O_P(1). \end{aligned}$$

By assumption f is positive so that $\tau_0^c > 0$ for $c > 0$. In addition $0 < \gamma < 1$, so that the denominator matrix is invertible. Thus, we get that $\hat{\theta} - \theta = (\sum_{i=1}^n m_i m'_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})})^{-1} \sum_{i=1}^n m_i \varepsilon_i 1_{(|\tilde{\varepsilon}_i| \leq c\tilde{\sigma})} = O_P(n^{-1/2})$.

(ii) For this part of the proof it is convenient to work with the parametrization

$$y_i = \mu + \alpha 1_{(i \in \mathcal{S}_1)} + \varepsilon_i,$$

since in this case the only regressor entering the auxiliary regression in the White test is $Z_i = 1_{(i \in \mathcal{S}_1)}$.

Write the test statistic in (A.1) as $\hat{T}_0^c n \hat{R}_{n,c}^2 = \widehat{Num} / \widehat{Den}$ where

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

For $j = 1, 2$ and $k = 0, 2, 4$, let

$$\hat{S}_{k,j}^c = \nu_j^{-1} \sum_{i \in \mathcal{S}_j} \hat{\varepsilon}_i^k 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)} = \nu_j^{-1} \sum_{i \in \mathcal{S}_j} \hat{\varepsilon}_i^k 1_{(|\tilde{\varepsilon}_i| \leq \sigma_j c_j + n^{-1/2} \tilde{a}_j c_j)},$$

where $\tilde{a}_j = \sigma_j n^{1/2}(\tilde{\sigma} - \sigma)/\sigma$ and $c_j = (\sigma/\sigma_j)c$ so that $c_j \in \mathbb{R}$. Note that

$$\hat{T}_k^c = \frac{\nu_1}{n} \hat{S}_{k,1}^c + \frac{\nu_2}{n} \hat{S}_{k,2}^c.$$

Let also $S_{k,j}^c = \nu_j^{-1} \sum_{i \in \mathcal{S}_j} \varepsilon_i^k 1_{(|\varepsilon_i| \leq \sigma_j c_j)}$. Since $n^{-1/2} \tilde{a}_j = \sigma_j(\tilde{\sigma} - \sigma)/\sigma = O_P(n^{-1/2})$ by assumption, we can apply the theory as before. Hence, by Corollary B.9(a) with $w_{in} = 1$ and Corollary B.10 applied to each of the subsamples, we get $\hat{T}_k^c = (\nu_1/n) S_{k,1}^c + (\nu_2/n) S_{k,2}^c + O_P(n^{-1/2})$ for $k = 0, 2, 4$. Note that $\sigma_j^k \tau_k^{c_j} = E\{\varepsilon_i^k 1_{(|\varepsilon_i| \leq \sigma_j c_j)}\}$ and let $\omega_c^k = \gamma \sigma_1^k \tau_k^{c_1} + (1 - \gamma) \sigma_2^k \tau_k^{c_2}$. Then, by the Law of Large Numbers $S_{k,j}^c = \sigma_j^k \tau_k^{c_j} + o_P(1)$. In turn,

$$\hat{T}_k^c = \gamma \sigma_1^k \tau_k^{c_1} + (1 - \gamma) \sigma_2^k \tau_k^{c_2} + o_P(1) = \omega_c^k + o_P(1). \quad (\text{C.4})$$

1. The term \widehat{Den} : Following the same steps as in item 1 in the proof of Theorem 3.1 but using (C.4), we get

$$\widehat{Den} = \omega_c^4 / \omega_c^0 - (\omega_c^2 / \omega_c^0)^2 + o_P(1).$$

2. The term \widehat{Num} :

2.1. The term $n^{-1} \hat{\mathcal{M}}_{n,c}$: Since $Z_i = 1_{(i \in \mathcal{S}_1)}$ and $\tilde{Z}_{i,c} = 1_{(i \in \mathcal{S}_1)} 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}$, we get that $n^{-1} \hat{\mathcal{M}}_{n,c} = n^{-1} \sum_{i=1}^n \tilde{Z}_{i,c} \tilde{Z}'_{i,c} - (\hat{T}_0^c)^{-1} \tilde{Z}_c \tilde{Z}'_c$, becomes

$$n^{-1} \hat{\mathcal{M}}_{n,c} = (\nu_1/n) \nu_1^{-1} \sum_{i \in \mathcal{S}_1} 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)} - (\hat{T}_0^c)^{-1} \{(\nu_1/n) \nu_1^{-1} \sum_{i \in \mathcal{S}_1} 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c)}\}^2.$$

We can write it in terms of the $\hat{S}_{k,1}^c$ so that

$$n^{-1} \hat{\mathcal{M}}_{n,c} = (\nu_1/n) \hat{S}_{0,1}^c - (\hat{T}_0^c)^{-1} \{(\nu_1/n) \hat{S}_{0,1}^c\}^2 = \gamma \sigma_1^0 \tau_0^{c_1} + \omega_c^0 (\gamma \sigma_1^0 \tau_0^{c_1})^2 + o_P(1).$$

2.2. The term $n^{-1/2} \hat{\mathcal{N}}_{n,c}$: For $j = 1, 2$ let

$$\hat{\varepsilon}_{i,c_j}^2 = \hat{\varepsilon}_i^2 1_{(|\tilde{\varepsilon}_i| \leq \sigma_j c_j + n^{-1/2} \tilde{a}_j c_j)}, \quad \tilde{Z}_{i,c_j} = Z_i 1_{(|\tilde{\varepsilon}_i| \leq \sigma_j c_j + n^{-1/2} \tilde{a}_j c_j)},$$

where in this case $Z_i = 1_{(i \in \mathcal{S}_1)}$. Hence,

$$n^{-1} \hat{\mathcal{N}}_{n,c} = n^{-1} \sum_{i \in \mathcal{S}_1} \{\hat{\varepsilon}_{i,c_1}^2 - 1_{(|\tilde{\varepsilon}_i| \leq \tilde{\sigma}c_1)} (\hat{T}_2^c / \hat{T}_0^c)\}.$$

This can be rewritten, in terms of $\hat{S}_{k,1}^c$ for $k = 0, 2$, as follows

$$n^{-1} \hat{\mathcal{N}}_{n,c} = (\nu_1/n) \{\hat{S}_{2,1}^c - \hat{S}_{0,1}^c (\hat{T}_2^c / \hat{T}_0^c)\}.$$

As discussed above, for $k = 2, 0$ we have $S_{k,1}^c = \sigma_1^k \tau_k^{c_1} + o_P(1)$ and $\hat{T}_k^c = \gamma \sigma_1^k \tau_k^{c_1} + (1 - \gamma) \sigma_2^k \tau_k^{c_2} + o_P(1)$, hence,

$$n^{-1} \hat{\mathcal{N}}_{n,c} \rightarrow_p \gamma \left(\sigma_1^2 \tau_2^{c_1} - \sigma_1^0 \tau_0^{c_1} \frac{\gamma \sigma_1^2 \tau_2^{c_1} + (1 - \gamma) \sigma_2^2 \tau_2^{c_2}}{\gamma \sigma_1^0 \tau_0^{c_1} + (1 - \gamma) \sigma_2^0 \tau_0^{c_2}} \right).$$

Let $\lambda_k^c = (\sigma_2^k \tau_k^{c_2} / \sigma_1^k \tau_k^{c_1})$ so that we can write

$$n^{-1} \hat{\mathcal{N}}_{n,c} \rightarrow_p \gamma \sigma_1^2 \tau_2^{c_1} \left(1 - \frac{\gamma + (1 - \gamma) \lambda_2}{\gamma + (1 - \gamma) \lambda_0} \right).$$

Hence, the limit is non-zero since $\lambda_2 \neq \lambda_0$. In turn, this implies that $n^{-1/2} \hat{\mathcal{N}}_{n,c}$ diverges.

3. Combine items 1 and 2: By item 1, $\widehat{Den} = \sigma_c^4 / \tau_0^c - (\sigma_c^2 / \tau_0^c)^2 + o_P(1)$. By item 2.1, $n^{-1} \hat{\mathcal{M}}_{n,c} = \gamma \sigma_1^0 \tau_0^{c_1} + \omega_c^0 (\gamma \sigma_1^0 \tau_0^{c_1})^2 + o_P(1)$. By item 2.2 $n^{-1/2} \hat{\mathcal{N}}_{n,c}$ diverges. In combination, $\hat{T}_0^c n \hat{R}_{n,c}^2$ diverges. \square

D Figures

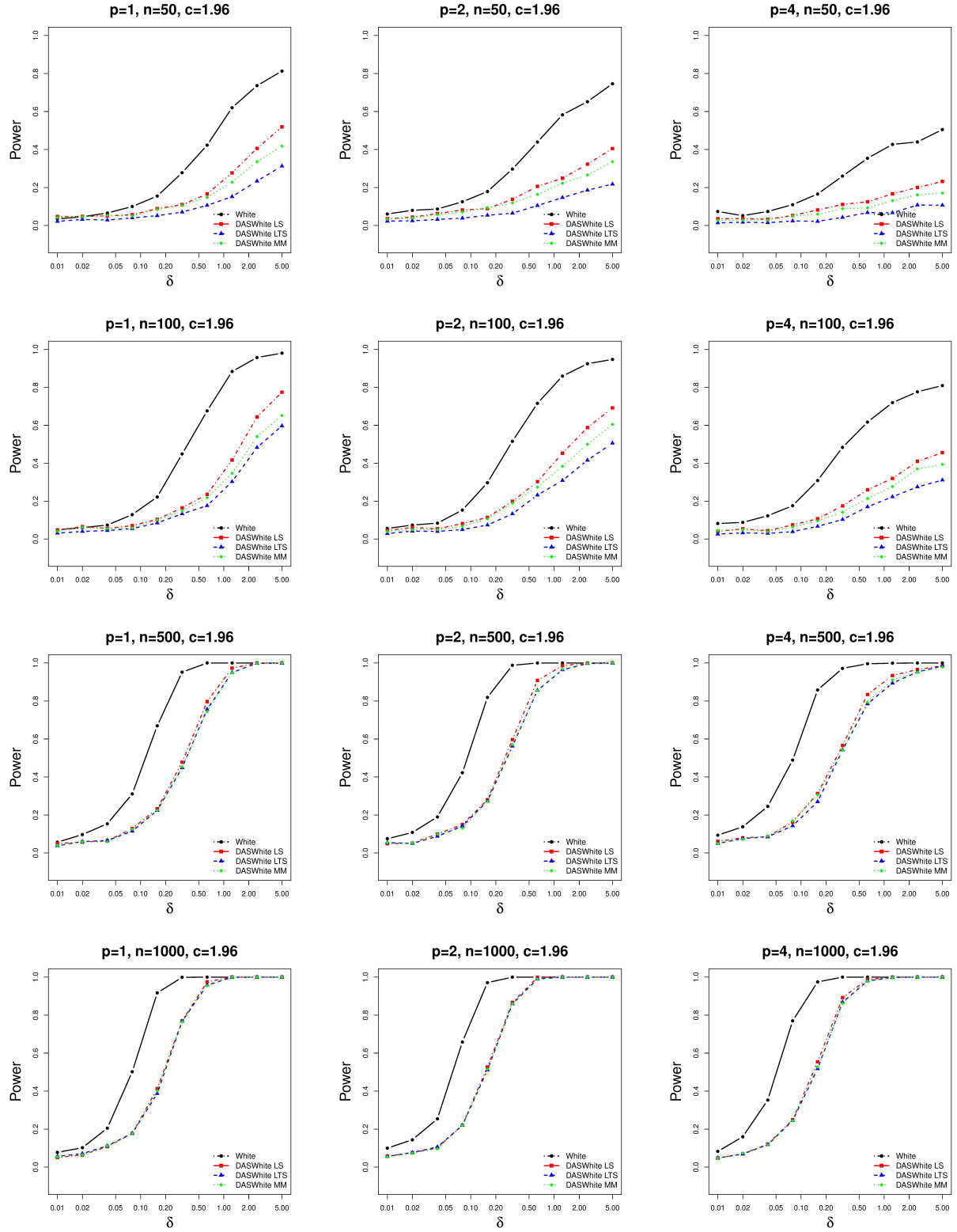


Figure 1: Power of the four tests in uncontaminated settings, significance level 5%.

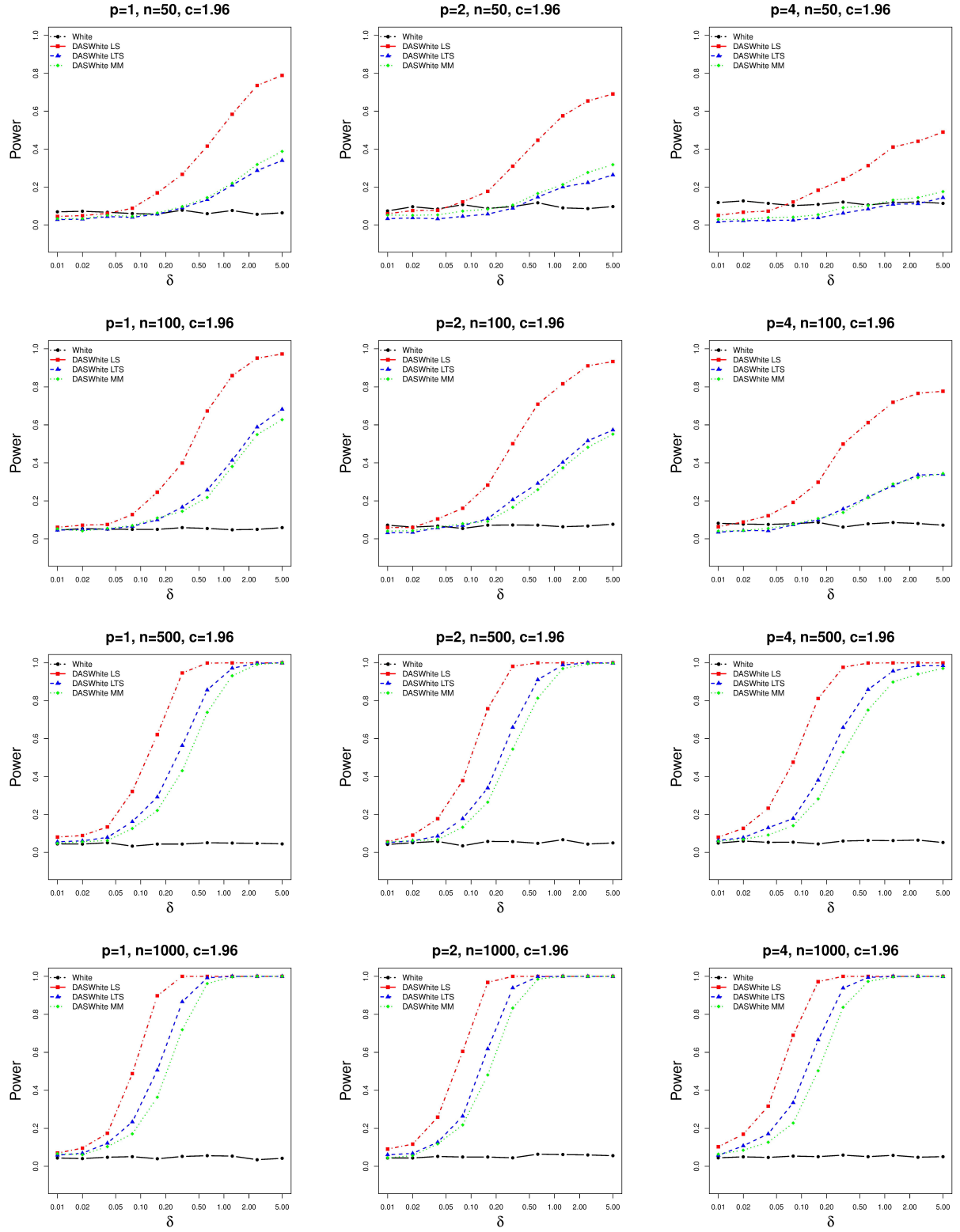


Figure 2: Power of the four tests in contaminated settings, significance level 5%.

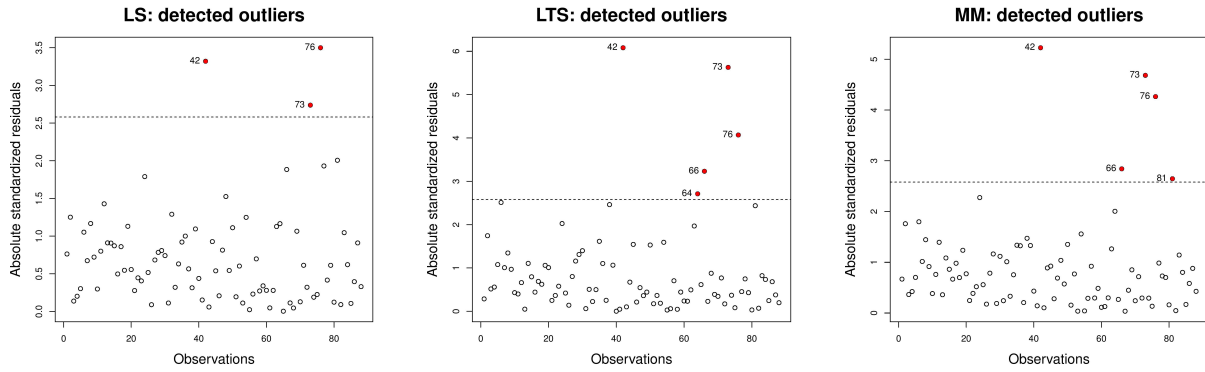


Figure 3: Detected outliers (indicated in red) in regression model for price.

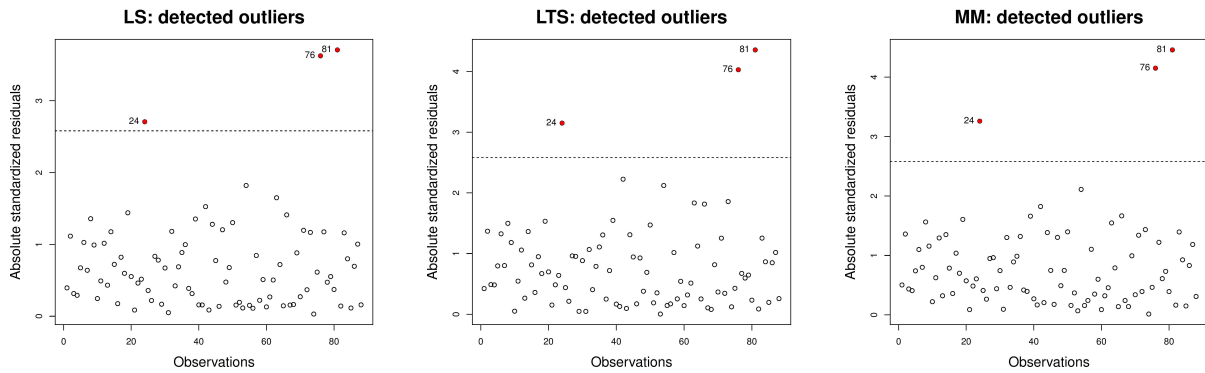


Figure 4: Detected outliers (indicated in red) in regression model for $\log(\text{price})$.

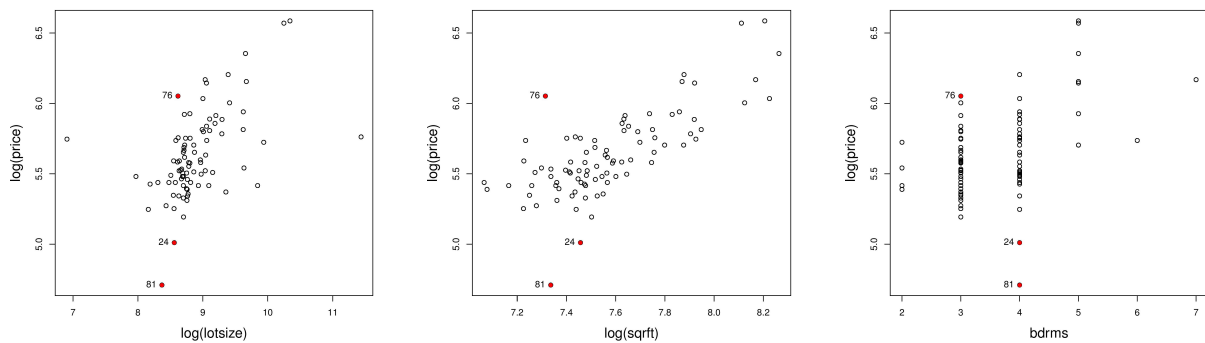


Figure 5: Detected outliers (indicated in red) in regression model for $\log(\text{price})$.

E References

- Akhtaruzzaman, M., C. Hajzler and P.D. Owen (2018) Does institutional quality resolve the Lucas paradox? *Applied Economics* 50(5), 455-474.
- Ali, M.M. and C. Giaccotto (1984) A study of several new and existing tests for heteroscedasticity in the general linear model. *Journal of Econometrics* 26(3), 355-373.
- Alih, E. and C.O. Ong (2015) An outlier-resistant test for heteroscedasticity in linear models. *Journal of Applied Statistics* 42(8), 1617-1634.
- Atalla, T., F. Joutz and A. Pierru (2017) Does disagreement among oil price forecasters reflect volatility? Evidence from ECB surveys. *International Journal of Forecasting* 32(4), 1178-1192.
- Bauer, C. and M. Neuenkirch (2017) Forecast uncertainty and the Taylor rule. *Journal of International Money and Finance* 77, 99-116.
- Bercu, B. and A.Touati (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18(5), 1848-1869.
- Berenguer-Rico, V., S. Johansen and B. Nielsen (2019a) Uniform consistency of marked and weighted empirical distribution functions of residuals. *Department of Economics, Discussion Paper Series, University of Oxford*.
- Berenguer-Rico, V., S. Johansen and B. Nielsen (2019b) The analysis of marked and weighted empirical processes of estimated residuals. *Department of Economics, Discussion Paper Series, University of Oxford*.
- Berenguer-Rico, V. and B. Nielsen (2018) Marked and weighted empirical processes of residuals with applications to robust regressions. *Department of Economics, Discussion Paper Series, University of Oxford*.
- Bold, T., K.C. Kaizzi, J. Svensson and D. Yanagizawa-Drott (2017) Lemon technologies and adoption: Measurement, theory, and evidence from agricultural markets in Uganda. *Quarterly Journal of Economics* 132(3), 1055-1100.
- Breusch, T.S. and A.R. Pagan (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47(5), 1287-1294.
- Carrera, N., T. Sohail and S. Carmona (2017) Audit committees' social capital and financial reporting quality. *Accounting and Business Research* 47(6), 633-672.
- Chen, S. and J. Bien (2019) Valid inference corrected for outlier removal. *Journal of Computational and Graphical Statistics*, DOI: 10.1080/10618600.2019.1660180.
- Chen, B., J.H. Choi and J.C. Escanciano (2017) Testing for fundamental vector moving average representations. *Quantitative Economics* 8(1), 149-180.
- Goldfeld, S.M. and R.E. Quandt (1965) Some tests for homoscedasticity. *Journal of the American Statistical Association* 61(310), 539-547.
- Hendry, D.F. and B. Nielsen (2007) *Econometric modeling: A likelihood approach*. Princeton University Press.
- Hendry, D.F. and C. Santos (2005) Regression models with data-based indicator variables. *Oxford Bulletin of Economics and Statistics* 67(5), 571-595.
- Hendry, D.F. and C. Santos (2010) An automatic test of super exogeneity. *Volatility and time series econometrics: essays in honor of Robert F. Engle* (eds Bollerslev, T., J.R. Russell and M.W. Watson), 164-193. Oxford University Press, Oxford.
- Hirvonen, K., B. Minten and D. Stifel (2017) Children's diets, nutrition knowledge, and access to markets. *World Development* 95, 303-315.
- Hoover, K.D. and Perez, S.J. (1999) Data mining reconsidered: encompassing and the general-to-specific approach to specification search (with discussion). *Econometrics Journal* 2, 167-191.
- Im, K.S. (2000) Robustifying Glejser test of heteroskedasticity. *Journal of Econometrics* 97(1), 179-188.
- Johansen, S. and B. Nielsen (2016a) Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli* 22(2), 1131-1183.
- Johansen, S. and B. Nielsen (2016b) Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics* 43(2), 321-348.

- Johansen, S. and B. Nielsen (2009) An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator. *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford University Press.
- Kalina, J. (2011) Testing heteroscedasticity in robust regression. *Research Journal of Economics Business and ICT* 4, 25-28.
- Koul, H. L. and M. Ossiander (1994) Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22(1), 540-562.
- Leeb, H. and B.M. Pötscher (2005) Model selection and inference: Facts and fiction. *Econometric Theory* 21(1), 21-59.
- Ling, T.W. and N.S.A. Wahab (2018) Roles of tax planning in market valuation of corporate social responsibility. *Cogent Business & Management* 5, Article number 1482595.
- Lyon, J.D. and C. Tsai (1996) A comparison of tests for heteroscedasticity. *The Statistician* 45(3), 337-349.
- Maronna, R.A., R.D. Martin and V.J. Yohai (2006) *Robust Statistics: Theory and Methods*. John Wiley & Sons, Hoboken, New Jersey.
- Montagnoli, A., A. Napolitano and B. Siliverstovs (2016) Regional interest rate pass-through in Italy. *Regional Studies* 50(8), 1404-1419.
- Pötscher, B.M. (1991) Effects of model selection on inference. *Econometric Theory* 7(2), 163-185.
- Rojas, M.G.A., E.R.R. Solis, R. Edgar and J.J. Zhu (2018) Innovation and network multiplexity: R&D and the concurrent effects of two collaboration networks in an emerging economy. *Research Policy* 47(6), 1111-1124.
- Romano, J.P. and M. Wolf (2017) Resurrecting weighted least squares. *Journal of Econometrics* 197(1), 1-19.
- Rousseeuw, P.J. (1984) Least median of squares regression. *Journal of the American Statistical Association* 79(388), 871-880.
- Shea, J.M. and K.H. Brown (2018) *wooldridge : 111 Data Sets from "Introductory Econometrics: A Modern Approach, 6e" by Jeffrey M. Wooldridge*. URL <https://CRAN.R-project.org/package=wooldridge>, R package version 1.3.1.
- Stanko, M.A. and D.H. Henard (2017) Toward a better understanding of crowdfunding, openness and the consequences for innovation. *Research Policy* 46(4), 784-798.
- Van Dijk, D., P.H. Franses and A. Lucas (1999) Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* 14(5), 539-562.
- Víšek, J.Á. (1999) The least trimmed squares – random carriers. *Bulletin of the Czech Econometric Society* 10, 1-30.
- White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4), 817-838.
- 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(1-2), 287-310.
- Wooldridge, J.M. (2013) *Introductory econometrics: A modern approach*. Fifth Edition. South-Western, Cengage Learning.
- Yohai, V.J. (1987) High breakdown-point and high efficiency estimates for regression. *Annals of Statistics* 15(2), 642-656.