Econometric Theory, **39**, 2023, 789–847.
doi:10.1017/S0266466622000184

HOW RELIABLE ARE BOOTSTRAP-BASED HETEROSKEDASTICITY ROBUST TESTS?

BENEDIKT M. PÖTSCHER
Department of Statistics, University of Vienna

DAVID PREINERSTORFER
SEPS-SEW, University of St. Gallen

We develop theoretical finite-sample results concerning the size of wild bootstrap-based heteroskedasticity robust tests in linear regression models. In particular, these results provide an efficient diagnostic check, which can be used to weed out tests that are unreliable for a given testing problem in the sense that they overreject substantially. This allows us to assess the reliability of a large variety of wild bootstrap-based tests in an extensive numerical study.

1. INTRODUCTION

Testing hypotheses on the parameters in a regression model with potentially heteroskedastic errors is a time-honored problem in econometrics and statistics. As the classical t-statistic (F-statistic, respectively) is not pivotal, or asymptotically pivotal, in such a case in general, even under Gaussianity of the errors, so-called heteroskedasticity robust (aka heteroskedasticity consistent) modifications of these test statistics have been proposed. These statistics are asymptotically standard normally (chi-square, respectively) distributed under the null. The first generation of such procedures is rooted in the results of Eicker (1963, 1967), see also Hinkley (1977), and has been popularized in econometrics by White (1980). It soon transpired that tests obtained from these heteroskedasticity robust test statistics by relying on critical values obtained from the respective asymptotic distributions are prone to overrejecting the null hypothesis in finite samples, especially so if the design matrix contains leverage points; see, e.g., MacKinnon and White (1985), Davidson and MacKinnon (1985), and Chesher and Jewitt (1987). One factor contributing to this tendency to overreject is a downward bias present in the covariance matrix estimators used in these test statistics, see

Financial support of the second author by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles is gratefully acknowledged. We thank two referees, the Co-Editor, and the Editor for helpful comments. Address correspondence to Benedikt Pötscher, Department of Statistics, University of Vienna, A-1090 Oskar-Morgenstern Platz 1, Vienna, Austria; e-mail: benedikt.poetscher@univie.ac.at.

Chesher and Jewitt (1987). Attempts at remedying the overrejection problem have led to the development of second-generation heteroskedasticity robust test statistics (often denoted by HC1 through HC4, with HC0 denoting the first generation test statistic). These statistics use various ways of rescaling the least-squares residuals before computing the covariance matrix estimator employed in the construction of the test statistic; see Hinkley (1977), MacKinnon and White (1985), and Cribari-Neto (2004). Simulation studies reported in, e.g., Davidson and MacKinnon (1985) and Cribari-Neto (2004) show that these modifications, especially HC3 and HC4, ameliorate the overrejection problem to some extent, but do not eliminate it. Further numerical results are provided in Chesher and Austin (1991), see also Chesher (1989). Davidson and MacKinnon (1985) also consider variants of HC0–HC3, denoted by HC0R–HC3R, obtained by using restricted instead of unrestricted least-squares residuals in the computation of the covariance matrix estimators employed by the various test statistics (the restriction alluded to being the restriction defining the null hypothesis). In their simulation experiments, this typically leads to tests that do not overreject (but that may underreject); see also the simulation results in Godfrey (2006), who additionally also considers HC4R. Of course, these simulation results do not rule out that the tests based on HC0R–HC4R (relying on critical values suggested by asymptotic theory) may overreject in some situations outside of the scope of the simulation studies; in fact, Pötscher and Preinerstorfer (2021) provide numerical proof that also these tests can suffer from considerable overrejection. Note that under the typical assumptions used in the literature all of the modifications of HC0 discussed so far have the same asymptotic distribution under the null, and thus use the same critical value.

An alternative approach is to use bootstrap methods to compute critical values for the test statistics HC0–HC4 or HC0R–HC4R, with the intention to improve upon the critical values derived from the asymptotic null distributions.[1] Inspired by earlier work in the statistics literature (e.g., Wu (1986) and the discussion in Beran (1986), Liu (1988), Mammen (1993)), Horowitz (1997) used the wild bootstrap to obtain critical values for HC0. This was followed up by Flachaire (1999, 2005a) and Davidson and Flachaire (2008), who considered the test statistics HC0–HC3 as well as HC0R–HC3R, and who stressed the version of the wild bootstrap that imposes the null restriction on the bootstrap data generating process; see also Godfrey and Orme (2004) and the more recent survey MacKinnon (2013). Further simulation studies, some of which also include HC4 and HC4R, can be found in Cribari-Neto (2004), Godfrey (2006), Cribari-Neto and Lima (2009), and Richard (2017). Once one has turned to bootstrap methods, one can also think of reverting

[1] Another possibility is to use Edgeworth expansions to find better critical values, see Rothenberg (1988) for the case of the HC0 test statistic and Davidson and MacKinnon (1985) for the HC0R test statistic. Simulation results in MacKinnon and White (1985) and Davidson and MacKinnon (1985) indicate that this does not work too well in practice. Of course, such expansions could also be worked out for the other versions of the test statistics mentioned, but this does not seem to have been pursued in the literature. Other adjustments are discussed in Imbens and Kolesár (2016); however, as shown in Pötscher and Preinerstorfer (2021), these adjustments do also not resolve the overrejection problem in general.

to the classical (i.e., uncorrected) t-statistic (F-statistic, respectively) and to apply the bootstrap methods to determine appropriate critical values. This has been considered in Mammen (1993); see also Godfrey (2006) for some Monte Carlo results. Since the majority of the literature on bootstrap-based heteroskedasticity robust tests favors the wild bootstrap over other bootstrap methods, we shall concentrate on the wild bootstrap in the sequel.

While the before mentioned bootstrap procedures have their merits and overrejection is ameliorated in many of the cases considered in the simulation studies cited, it is unclear whether these observations generalize beyond the situations studied in these simulation experiments. In particular, it is unclear if—and under which conditions—these bootstrap procedures (and which of the many variants thereof) are immune to overrejection in finite samples.[2,3] In the present paper, we set out to study this question theoretically and numerically. On the theoretical side we show the following finite-sample result: For any test statistic T from a large class of test statistics (including HC0–HC4, HC0R–HC4R, the classical F-statistic and a variant thereof that uses restricted residuals) and for any bootstrap method from a large class of wild bootstrap methods (including virtually all wild bootstrap methods considered in the literature) there is a computable number ϑ (depending only on observables like the design matrix, the restriction to be tested, etc.), such that the size of the corresponding bootstrap-based test is 1 for nominal significance levels α satisfying $\alpha > \vartheta$.[4] That is, for $\alpha > \vartheta$ the bootstrap-based test fails miserably in that it has null rejection probabilities arbitrarily close to 1 for some forms of heteroskedasticity.[5] We note that our results also provide information concerning the infimal coverage probabilities of confidence sets obtained by "inverting" the bootstrap-based tests under consideration. We discuss this in more detail in Remark 5.15.

In practice, our theoretical finite-sample result can be used as a diagnostic tool to weed out procedures in the following sense: as mentioned before, there is a large menu of heteroskedasticity robust test statistics and wild bootstrap methods available in the literature from which the applied researcher has to choose. As it is unlikely that simulation results in the literature are available that precisely fit the problem the researcher is interested in (i.e., use the same design matrix and the same restriction to be tested), the researcher is typically left with little guidance

[2] We are not interested in asymptotic justifications here.

[3] In the quite special case where the number of restrictions tested equals the number of regression parameters, Davidson and Flachaire (2008) have a result which implies that certain wild bootstrap-based tests have size equal to the nominal significance level (and hence do not overreject) in finite samples. We note that this result in Davidson and Flachaire (2008) is not entirely correct as stated, but needs some amendments and corrections; see also Footnote 34.

[4] The size is the maximal (i.e., worst-case) null rejection probability, where one maximizes over all possible forms of heteroskedasticity, reflecting agnosticism about the form of the heteroskedasticity; see (4) for a formal definition. It is also assumed that the normalized regression error vector $(\mathbf{u}_1/\mathrm{Var}^{1/2}(\mathbf{u}_1), \dots, \mathbf{u}_n/\mathrm{Var}^{1/2}(\mathbf{u}_n))'$ follows a given (fixed) distribution (e.g., a normal distribution). Note that the theoretical result mentioned in the main text does *not* depend on the choice of this distribution (as long as it is absolutely continuous), see Section 7.

[5] By construction $\vartheta \leq 1$ always holds. If $\vartheta < 1$ (which will often be the case) we can then conclude that the bootstrap-based test has size 1 at least for some values of α.

on which of the many bootstrap-based test procedures to choose for the problem at hand. Based on our theoretical results, the applied researcher can now eliminate procedures that break down in the researchers problem, by computing—for any initially selected procedure—the corresponding ϑ for the given design matrix and restriction to be tested. If it turns out that $\vartheta < \alpha$ holds, this procedure should not be used, because these tests have size equal to one according to our theoretical results. Numerical routines for computing ϑ are provided in the associated R-package **wbsd** by Preinerstorfer (2020).[6]

The before mentioned theoretical result will typically have practical consequences only in testing problems for which ϑ is sufficiently small so that standard choices of α like 0.05 or 0.1 satisfy the condition $\alpha > \vartheta$. We hence investigate this numerically for the test statistics HC0–HC4, HC0R–HC4R, for the classical F-statistic, and for a variant thereof that uses restricted residuals, each combined with a large variety of wild bootstrap methods.[7] We now summarize the results of our numerical experiments for $n = 10$ (the results for $n = 20, 30$ being similar): For each combination of test statistic and wild bootstrap method (960 combinations) we compute ϑ for a range of design matrices and null hypotheses (i.e., restrictions to be tested) and report $\vartheta_{\min}$, the smallest of these values of ϑ.[8,9] We find that for 826 of the 960 combinations $\vartheta_{\min}$ is less than 0.05, and for 936 combinations $\vartheta_{\min}$ is less than 0.1. As a consequence, for the bootstrap-based tests corresponding to these 826 (936, respectively) combinations our theoretical results imply that size is equal to 1 for some design matrices and null hypotheses, if a nominal significance level of 0.05 (0.1, respectively) is being used. Thus these bootstrap-based tests are found not to be reliable in general, in that they suffer from severe overrejection for some design matrices and null hypotheses. Furthermore, for each combination of test statistic/wild bootstrap method we also compute a lower bound for the size of the bootstrap-based test conducted at nominal significance level $\alpha = 0.05$ (as well as at $\alpha = 0.1$).[10] We find that for 95 out of the remaining 134 combinations (11 out of the remaining 24 combinations, respectively) the (lower bound for the) size exceeds 3α for some of the design matrices and null hypotheses, sometimes by a considerable margin. Thus also these combinations do not lead to reliable bootstrap-based tests. This leaves us with 39 (13, respectively) combinations. Exploiting that some of these combinations left are in fact equivalent to some of the above mentioned *unreliable* procedures (see Sections 8.1 and 8.3 for an

[6] As discussed in Section 8.2 and Sections E.2 and E.3 of Appendix E, computing ϑ is a nontrivial numerical problem. Supplementing the calculation of ϑ by numerically evaluating null rejection probabilities for strategically chosen heteroskedasticity structures as discussed in Section E.2 of Appendix E may be advisable.

[7] In the wild bootstrap methods we vary the following elements: (i) centering the bootstrap sample at the unrestricted versus at the restricted least squares estimator, (ii) bootstrapping from unrestricted versus restricted residuals, (iii) the distribution of the bootstrap noise, and (iv) various bootstrap multiplicator weights. See Section 8 for more details.

[8] For reasons of numerical stability we actually compute an upper bound for ϑ, see Section E.3 in Appendix E for more information.

[9] Some of the 960 combinations actually give rise to one and the same bootstrap-based test. The reasons for nevertheless considering all 960 combinations are discussed in Sections 8.1 and 8.3.

[10] For the size computations we assume the errors to be normally distributed.

explanation), allows us to even conclude that in the end only 16 (4, respectively) bootstrap-based heteroskedasticity robust test procedures do not exhibit severe overrejection within the range of our numerical study when $n = 10$.

Combining the just-described results with similar findings for the other sample sizes $n = 20, 30$ leads to the sobering conclusion that *none* of the bootstrap-based tests considered is reliable for all sample sizes and for $\alpha = 0.05$ as well as $\alpha = 0.1$. That is, for every combination of test statistic and bootstrap method considered, there is a sample size $n \in \{10, 20, 30\}$, a significance level $\alpha \in \{0.05, 0.1\}$, a testing problem and a design matrix, such that the size of the corresponding bootstrap-based test equals 1 by our theoretical results or is numerically found to exceed 3α. We must hence conclude that *none* of the bootstrap-based tests considered is guaranteed to be immune to overrejection, and thus such tests are no reliable panacea for heteroskedasticity robust testing.

If one considers a fixed α, the situation is somewhat more encouraging. While there is no bootstrap-based test that is reliable for all sample sizes for the significance level $\alpha = 0.1$, for $\alpha = 0.05$ there are two bootstrap-based tests that are found not to break down in the above sense for any of the sample sizes considered in the numerical study. Both of these tests use a heteroskedasticity robust test statistic based on a HC3R covariance estimator, a wild bootstrap method based on the Mammen distribution, and impose the null restriction on the bootstrap data generating process. For more details see Section 8. It is interesting to note that these findings call into question the recommendation in Davidson and Flachaire (2008) to base the wild bootstrap on the Rademacher distribution.

Of course, the above are worst-case results in spirit and do not preclude a given bootstrap-based test to be reasonably sized for certain instances of design matrix and null hypothesis. Therefore, in a given application, one could in principle imagine the following strategy: Numerically evaluate the size of the given bootstrap-based test (this will require to commit to a distributional assumption on the errors) and use the test only if the so-evaluated size does not exceed the nominal level α (by much).[11] Otherwise, switch to another one of the many other bootstrap-based tests, repeat, and stop upon finding an acceptable test. As mentioned before, a partial shortcut for this strategy could be to compute ϑ first and to check if $\alpha > \vartheta$, as we then know from our theoretical results that size must be equal to 1. Of course, such a strategy would be computationally expensive and moreover would only be a stab into the dark, as there is no guarantee that one would end up with a bootstrap-based test that performs well in the sense of delivering size less than or equal to α. It seems that a better and more direct strategy is to forgo the bootstrap idea and rather to construct size-controlling critical values for the original test statistics, e.g., for HC0–HC4 or HC0R–HC4R. This is pursued in the companion paper Pötscher and Preinerstorfer (2021). Certainly, this also leads to a computationally intensive method, but one that comes with guaranteed size control.

[11] Of course, this could also be pursued with nonbootstrap-based tests.

All test statistics mentioned so far are based on the ordinary least squares estimator. An alternative is to start from a feasible generalized least squares estimator, computed from a (potentially misspecified) model for heteroskedasticity. Again heteroskedasticity robust test procedures can then be developed in a similar manner, see, e.g., Cragg (1983, 1992), Flachaire (2005b), Romano and Wolf (2017), Lin and Chou (2018), DiCiccio, Romano, and Wolf (2019). While results similar to the ones given in the present paper can probably also be developed for this alternative class of heteroskedasticity robust test procedures, we do not pursue this avenue here.

2. FRAMEWORK

Consider the linear regression model

$$\mathbf{Y} = X\beta + \mathbf{U}, \tag{1}$$

where X is a (real) nonstochastic regressor (design) matrix of dimension $n \times k$ and where $\beta \in \mathbb{R}^k$ denotes the unknown regression parameter vector. We always assume $\mathrm{rank}(X) = k$ and $1 \leq k < n$. We furthermore assume that the $n \times 1$ disturbance vector $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)'$ has mean zero and unknown covariance matrix $\sigma^2\Sigma$, where Σ varies in a user-specified (nonempty) set $\mathfrak{C}$ describing the allowed forms of heteroskedasticity, with $\mathfrak{C}$ satisfying $\mathfrak{C} \subseteq \mathfrak{C}_{Het}$, and where $0 < \sigma^2 < \infty$ holds (σ always denoting the positive square root).[12] The set $\mathfrak{C}$ will be referred to as the "heteroskedasticity model." Here

$$\mathfrak{C}_{Het} = \left\{ \mathrm{diag}(\tau_1^2, \ldots, \tau_n^2) : \tau_i^2 > 0 \text{ for all } i, \sum_{i=1}^{n} \tau_i^2 = 1 \right\},$$

where $\mathrm{diag}(\tau_1^2, \ldots, \tau_n^2)$ denotes the $n \times n$ matrix with diagonal elements given by τ_i^2. That is, the errors in the regression model are uncorrelated but can be heteroskedastic. In particular, if $\mathfrak{C}$ is chosen to be $\mathfrak{C}_{Het}$, one allows for heteroskedasticity of completely unknown form. The normalization condition $\sum_{i=1}^{n} \tau_i^2 = 1$ is included here only in order to guarantee identifiability of σ^2 and Σ, and could be replaced by any other normalization condition such as, e.g., $\max \tau_i^2 = 1$, or $\tau_1^2 = 1$, without affecting the final results (because any of these normalizations leads to the same overall set of covariance matrices $\sigma^2\Sigma$ when σ^2 varies through the positive real line).

Although of no real significance for the results of this paper as explained in Section 7, we shall, for ease of exposition, maintain in the sequel that the disturbance vector $\mathbf{U}$ *is normally distributed.* The linear model described in (1), together with the just made Gaussianity assumption on $\mathbf{U}$ and with the given heteroskedasticity model $\mathfrak{C}$, then induces a collection of distributions on the

[12] Since we are concerned with finite-sample results only, the elements of $\mathbf{Y}$, X, and $\mathbf{U}$ (and even the probability space supporting $\mathbf{Y}$ and $\mathbf{U}$) may depend on sample size n, but this will not be expressed in the notation. Furthermore, the obvious dependence of $\mathfrak{C}$ on n will also not be shown in the notation.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Borel-sets of $\mathbb{R}^n$, the sample space of $\mathbf{Y}$. Denoting a Gaussian probability measure with mean $\mu \in \mathbb{R}^n$ and (possibly singular) covariance matrix A by $P_{\mu,A}$, the induced collection of distributions is then given by

$$\left\{P_{\mu,\sigma^2\Sigma} : \mu \in \mathrm{span}(X), 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C}\right\}. \tag{2}$$

Since every $\Sigma \in \mathfrak{C}$ is positive definite by assumption, each element of the set in the previous display is absolutely continuous with respect to (w.r.t.) Lebesgue measure on $\mathbb{R}^n$.

We shall consider the problem of testing a linear (better: affine) hypothesis on the parameter vector $\beta \in \mathbb{R}^k$, i.e., the problem of testing the null $R\beta = r$ against the alternative $R\beta \neq r$, where R is a $q \times k$ matrix always of rank $q \geq 1$ and $r \in \mathbb{R}^q$. Set $\mathfrak{M} = \mathrm{span}(X)$. Define the affine space

$$\mathfrak{M}_0 = \{\mu \in \mathfrak{M} : \mu = X\beta \text{ and } R\beta = r\}$$

and let

$$\mathfrak{M}_1 = \{\mu \in \mathfrak{M} : \mu = X\beta \text{ and } R\beta \neq r\}.$$

Adopting these definitions, the above testing problem can then be written more precisely as

$$H_0 : \mu \in \mathfrak{M}_0,\ 0 < \sigma^2 < \infty,\ \Sigma \in \mathfrak{C} \quad \text{vs.} \quad H_1 : \mu \in \mathfrak{M}_1,\ 0 < \sigma^2 < \infty,\ \Sigma \in \mathfrak{C}. \tag{3}$$

With $\mathfrak{M}_0^{lin}$ we shall denote the linear space parallel to $\mathfrak{M}_0$, i.e., $\mathfrak{M}_0^{lin} = \mathfrak{M}_0 - \mu_0 = \{X\beta : R\beta = 0\}$ where $\mu_0 \in \mathfrak{M}_0$. Of course, $\mathfrak{M}_0^{lin}$ does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$.

As already mentioned, the assumption of Gaussianity is made for the sake of exposition only and does not really restrict the scope of the results in the paper as is discussed in Section 7. The assumption of nonstochastic regressors entails little loss of generality either: For example, if X is random and $\mathbf{U}$ is conditionally on X distributed as $N(0,\sigma^2\Sigma)$, with $\sigma^2 = \sigma^2(X)$ and $\Sigma = \Sigma(X) \in \mathfrak{C}_{Het}$, the results of the paper can be applied after one conditions on X (and a similar statement applies to the generalizations to non-Gaussianity discussed in Section 7). See Section 7 for more discussion. For arguments supporting conditional inference see, e.g., Robinson (1979). Note that such a "strict exogeneity" assumption is quite natural in the situation considered here.

We next collect some further terminology and notation used throughout the paper. A (nonrandomized) *test* is the indicator function of a Borel-set W in $\mathbb{R}^n$, with W called the corresponding *rejection region*. The *size* of such a test (rejection region) is—as usual—defined as the supremum over all rejection probabilities under the null hypothesis H_0 given in (3), i.e.,

$$\sup_{\mu \in \mathfrak{M}_0} \sup_{0<\sigma^2<\infty} \sup_{\Sigma \in \mathfrak{C}} P_{\mu,\sigma^2\Sigma}(W). \tag{4}$$

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

In slight abuse of terminology, we shall sometimes refer to this quantity as "the size of W over $\mathfrak{C}$" when we want to emphasize the rôle of $\mathfrak{C}$. Throughout the paper, we let $\hat{\beta}(y) = \left(X'X\right)^{-1} X'y$, where X is the design matrix appearing in (1) and $y \in \mathbb{R}^n$. The corresponding ordinary least squares (OLS) residual vector is denoted by $\hat{u}(y) = y - X\hat{\beta}(y)$ and its elements are denoted by $\hat{u}_t(y)$. The elements of X are denoted by x_{ti}, while $x_{t\cdot}$ and $x_{\cdot i}$ denote the tth row and ith column of X, respectively. For $\mathcal{A}$ an affine subspace of $\mathbb{R}^n$ satisfying $\mathcal{A} \subseteq \operatorname{span}(X)$ let $\tilde{\beta}_{\mathcal{A}}(y)$ denote the restricted least-squares estimator, i.e., $X\tilde{\beta}_{\mathcal{A}}(y)$ solves

$$\min_{z \in \mathcal{A}} (y-z)'(y-z).$$

Lebesgue measure on the Borel-sets of $\mathbb{R}^n$ will be denoted by $\lambda_{\mathbb{R}^n}$. The set of real matrices of dimension $l \times m$ is denoted by $\mathbb{R}^{l \times m}$ (all matrices in the paper will be real matrices). The Euclidean norm is denoted by $\|\cdot\|$. Let B' denote the transpose of a matrix $B \in \mathbb{R}^{l \times m}$ and let $\operatorname{span}(B)$ denote the subspace in $\mathbb{R}^l$ spanned by its columns. For a symmetric and nonnegative definite matrix B we denote the unique symmetric and nonnegative definite square root by $B^{1/2}$. For a linear subspace $\mathcal{L}$ of $\mathbb{R}^n$ we let $\mathcal{L}^{\perp}$ denote its orthogonal complement and we let $\Pi_{\mathcal{L}}$ denote the orthogonal projection onto $\mathcal{L}$. The jth standard basis vector in $\mathbb{R}^n$ is written as $e_j(n)$. Furthermore, we let $\mathbb{N}$ denote the set of all positive integers. A sum (product, respectively) over an empty index set is to be interpreted as 0 (1, respectively). For a subset A of a topological space we denote by $\operatorname{int}(A)$ the interior of A (w.r.t. the ambient space). Finally, for $\mathcal{A}$ an affine subspace of $\mathbb{R}^n$, let $G(\mathcal{A})$ denote the group of all affine transformations $y \mapsto \delta(y-a) + a^*$ where $\delta \in \mathbb{R}$, $\delta \neq 0$, and a as well as a^* are elements of $\mathcal{A}$; for more information see Section 5.1 of Preinerstorfer and Pötscher (2016).

3. HETEROSKEDASTICITY ROBUST TEST STATISTICS USING UNRESTRICTED RESIDUALS

We next introduce two test statistics that will feature prominently. Variants of these statistics using restricted residuals are discussed in Section 5.2. For a result pertaining to a more general class of test statistics see Theorem A.1 in Appendix A. The test statistic we shall consider first is a standard heteroskedasticity robust test statistic frequently considered in the literature and is given by

$$T_{Het}(y) = \begin{cases} (R\hat{\beta}(y) - r)' \hat{\Omega}_{Het}^{-1}(y) (R\hat{\beta}(y) - r), & \text{if } \det \hat{\Omega}_{Het}(y) \neq 0, \\ 0, & \text{if } \det \hat{\Omega}_{Het}(y) = 0, \end{cases} \tag{5}$$

where $\hat{\Omega}_{Het} = R\hat{\Psi}_{Het}R'$ and where $\hat{\Psi}_{Het}$ is a heteroskedasticity robust estimator as considered in Eicker (1963, 1967), which later on has found its way into the econometrics literature (e.g., White, 1980). It is of the form

$$\hat{\Psi}_{Het}(y) = (X'X)^{-1} X' \operatorname{diag}\left(d_1 \hat{u}_1^2(y), \ldots, d_n \hat{u}_n^2(y)\right) X (X'X)^{-1},$$

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

where the constants $d_i > 0$ sometimes depend on the design matrix. Typical choices for d_i suggested in the literature are $d_i = 1$, $d_i = n/(n-k)$, $d_i = (1-h_{ii})^{-1}$, or $d_i = (1-h_{ii})^{-2}$, where h_{ii} denotes the ith diagonal element of the projection matrix $X(X'X)^{-1}X'$, see Long and Ervin (2000) for an overview. Another suggestion is $d_i = (1-h_{ii})^{-\delta_i}$ for $\delta_i = \min(nh_{ii}/k, 4)$, see Cribari-Neto (2004). For the last three choices of d_i just given, we use the convention that we set $d_i = 1$ in case $h_{ii} = 1$. Note that $h_{ii} = 1$ implies $\hat{u}_i(y) = 0$ for every y, and hence it is irrelevant which real value is assigned to d_i in case $h_{ii} = 1$.[13] The five examples for the weights d_i just given correspond to what is often called HC0–HC4 weights in the literature.

In conjunction with the test statistic T_{Het}, we shall consider the following mild assumption, which is Assumption 3 in Preinerstorfer and Pötscher (2016). As discussed further below, this assumption is in a certain sense unavoidable when using T_{Het}. It furthermore also entails that our choice of assigning $T_{Het}(y)$ the value zero in case $\hat{\Omega}_{Het}(y)$ is singular has no import on the rejection probabilities of the (nonbootstrap-based) tests obtained from T_{Het} (because of Lemma 3.1(c) and absolute continuity of the measures $P_{\mu,\sigma^2\Sigma}$). As will be seen later, our results for the corresponding bootstrap-based tests do also not depend on this choice.

Assumption 1. Let $1 \le i_1 < \cdots < i_s \le n$ denote all the indices for which $e_{i_j}(n) \in \operatorname{span}(X)$ holds where $e_j(n)$ denotes the jth standard basis vector in $\mathbb{R}^n$. If no such index exists, set $s = 0$. Let $X'(\neg(i_1, \ldots i_s))$ denote the matrix which is obtained from X' by deleting all columns with indices i_j, $1 \le i_1 < \cdots < i_s \le n$ (if $s = 0$ no column is deleted). Then $\operatorname{rank}\left(R(X'X)^{-1}X'(\neg(i_1, \ldots i_s))\right) = q$ holds.

Observe that this assumption only depends on X and R and hence can be checked. Obviously, a simple sufficient condition for Assumption 1 to hold is that $s = 0$ (i.e., that $e_j(n) \notin \operatorname{span}(X)$ for all j), a generically satisfied condition. Furthermore, we introduce the matrix

$$
\begin{aligned}
B(y) &= R(X'X)^{-1}X' \operatorname{diag}\left(\hat{u}_1(y), \ldots, \hat{u}_n(y)\right) \\
&= R(X'X)^{-1}X' \operatorname{diag}\left(e_1'(n)\Pi_{\operatorname{span}(X)^\perp} y, \ldots, e_n'(n)\Pi_{\operatorname{span}(X)^\perp} y\right). \qquad (6)
\end{aligned}
$$

The facts collected in the subsequent lemma will be used in the sequel. Parts (a)–(c) have been shown in Lemma 4.1 in Preinerstorfer and Pötscher (2016), while Part (d) is taken from Lemma 5.18 of Pötscher and Preinerstorfer (2018). Part (e) is obvious (observe that $B(y)$ depends only on $\hat{u}(y)$ and that $\hat{u}(\gamma(y-\mu)+\mu^\bullet) = \gamma\hat{u}(y)$ for every $\gamma \in \mathbb{R}$, every $\mu \in \operatorname{span}(X)$, and every $\mu^\bullet \in \operatorname{span}(X)$).

LEMMA 3.1. *(a)* $\hat{\Omega}_{Het}(y)$ *is nonnegative definite for every* $y \in \mathbb{R}^n$.

(b) $\hat{\Omega}_{Het}(y)$ *is singular (zero, respectively) if and only if* $\operatorname{rank}(B(y)) < q$ $(B(y) = 0$*, respectively).*

(c) The set B *given by* $\{y \in \mathbb{R}^n : \operatorname{rank}(B(y)) < q\}$ *(or in view of (b) equivalently given by* $\{y \in \mathbb{R}^n : \det(\hat{\Omega}_{Het}(y)) = 0\}$*) is either a* $\lambda_{\mathbb{R}^n}$*-null set or the entire sample*

[13] In fact, $h_{ii} = 1$ is equivalent to $\hat{u}_i(y) = 0$ for every y, each of which in turn is equivalent to $e_i(n) \in \operatorname{span}(X)$.

space $\mathbb{R}^n$*. The latter occurs if and only if Assumption 1 is violated (in which case the test based on* T_{Het} *becomes trivial, as then* T_{Het} *is identically zero).*

(d) Under Assumption 1, the set B *is a finite union of proper linear subspaces of* $\mathbb{R}^n$*; in case* $q = 1$*,* B *is even a proper linear subspace itself.*[14]

(e) B *is a closed set and contains* $\operatorname{span}(X)$*. Furthermore,* B *is* $G(\mathfrak{M})$*-invariant and, in particular,* $\mathsf{B} + \operatorname{span}(X) = \mathsf{B}$ *holds.*

In light of Part (c) of the lemma, we see that Assumption 1 is a natural and unavoidable condition if one wants to obtain a sensible test from T_{Het}.[15] Furthermore, note that, if $\mathsf{B} = \operatorname{span}(X)$ is true, then Assumption 1 must be satisfied (since $\operatorname{span}(X)$ is a $\lambda_{\mathbb{R}^n}$-null set due to the maintained assumption $k < n$). As shown in Lemma A.3 in Pötscher and Preinerstorfer (2018), for any given restriction matrix R, the relation $\mathsf{B} = \operatorname{span}(X)$ holds generically in various universes of design matrices. For later use we also mention that under Assumption 1 the test statistic T_{Het} is continuous at every $y \in \mathbb{R}^n \backslash \mathsf{B}$.[16]

Next, we also consider the classical (i.e., uncorrected) F-test statistic, i.e.,

$$T_{uc}(y) = \begin{cases} (R\hat{\beta}(y) - r)' \left(\hat{\sigma}^2(y) R \left(X'X\right)^{-1} R'\right)^{-1} (R\hat{\beta}(y) - r), & \text{if } y \notin \operatorname{span}(X), \\ 0, & \text{if } y \in \operatorname{span}(X), \end{cases} \tag{7}$$

where $\hat{\sigma}^2(y) = \hat{u}(y)'\hat{u}(y)/(n-k) \geq 0$ (which vanishes if and only if $y \in \operatorname{span}(X)$). Our choice to set $T_{uc}(y) = 0$ for $y \in \operatorname{span}(X)$ has no import on the rejection probabilities of the (nonbootstrap-based) tests obtained from T_{uc}, since $\operatorname{span}(X)$ is a $\lambda_{\mathbb{R}^n}$-null set as a consequence of the maintained assumption that $k < n$ (and since the measures $P_{\mu,\sigma^2\Sigma}$ are absolutely continuous). It will turn out also not to affect our results for bootstrap-based tests obtained from T_{uc}. For reasons of comparability with (5), we have chosen not to normalize the numerator in (7) by q, the number of restrictions to be tested, as is often done in the definition of the classical F-test statistic. This also has no import on the results as the bootstrap automatically adapts to scaling. For later use we also mention that the test statistic T_{uc} is continuous at every $y \in \mathbb{R}^n \backslash \operatorname{span}(X)$.

Remark 3.2. (i) The test statistics T_{Het} as well as T_{uc} are $G(\mathfrak{M}_0)$-invariant as is easily seen (with the respective exceptional sets B and $\operatorname{span}(X)$ being $G(\mathfrak{M})$ -invariant).

(ii) Both statistics actually belong to the class of nonsphericity-corrected F-type test statistics in the sense of Section 5.4 in Preinerstorfer and Pötscher (2016) (terminology being somewhat unfortunate in the case of T_{uc} as no correction for the nonsphericity is applied in this case). See Remark B.1 in Appendix B for more discussion.

[14] If Assumption 1 is violated, B equals $\mathbb{R}^n$ by Part (c).

[15] If this assumption is violated then T_{Het} is identically zero, an uninteresting trivial case.

[16] If Assumption 1 is violated, then T_{Het} is constant equal to zero, and hence is trivially continuous everywhere.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

4. SOME INTUITION FOR THE SIZE ONE RESULTS

The mechanism leading to the size one results put forward formally in the next section is a concentration effect in the distribution generating the data $\mathbf{Y} = (y_1, \ldots, y_n)'$, entailing a similar effect in the distribution of $T(\mathbf{Y})$, where we denote by T any of the test statistics considered in the paper. This concentration effect emerges when the data-generating process (DGP) is "strongly heteroskedastic." For simplicity, in this section we call a DGP strongly heteroskedastic, if a single observation has a (relatively) high variance, whereas all other observations have (relatively) low variance. We shall denote the index corresponding to the highly varying observation by i^*. Denote the expectation of the data vector $\mathbf{Y}$ by μ_0, where we assume for the discussion in this section that $\mu_0 \in \mathfrak{M}_0$, i.e., that the null hypothesis is satisfied.

We now provide a nonrigorous explanation of the above-mentioned concentration effect and how it leads to the size one results:

1. If the DGP is strongly heteroskedastic, only the single highly varying observation y_{i^*} will substantially deviate from its expectation $\mu_0^{(i^*)}$, whereas all other observations will be very close to their expectations. That is, under such a DGP we approximately have

   $$\mathbf{Y} \approx \mu_0 + (y_{i^*} - \mu_0^{(i^*)}) e_{i^*}(n),$$

   where we recall that $e_{i^*}(n)$ is the i^*th $n \times 1$ standard basis vector. That is, essentially, the data are concentrated on a one-dimensional affine subspace of the sample space $\mathbb{R}^n$.
2. Invariance properties of T common to all test statistics used in this paper (and in practice) imply that

   $$T(\mu_0 + c e_{i^*}(n)) = T(\mu_0 + e_{i^*}(n)) \text{ for every } c \neq 0.$$

   That is, the test statistic under consideration is essentially constant on the one-dimensional affine subspace just obtained in the previous item.
3. Combining the two previous observations (and ignoring the case where $y_{i^*} = \mu_0^{(i^*)}$), suggests that for strongly heteroskedastic DGPs we have

   $$T(\mathbf{Y}) \approx T(\mu_0 + e_{i^*}(n)).$$

   That is, essentially, the distribution of the test statistic collapses at the value $T(\mu_0 + e_{i^*}(n))$.

Now, recall that a wild bootstrap-based test rejects if the test statistic evaluated at the data $T(\mathbf{Y})$ exceeds the bootstrap critical value. This bootstrap critical value is a $1-\alpha$ quantile of the distribution of the test statistic, but now induced by the distribution that corresponds to a bootstrap scheme $\mathbf{Y}^*$, say. In general, the distribution of $T(\mathbf{Y}^*)$ depends on two sources of randomness: first, the DGP itself, and second, the randomization mechanism used to generate the bootstrap scheme $\mathbf{Y}^*$. Making use of the concentration mechanism outlined above, one can, for

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

the class of bootstrap schemes considered, however, show that the dependence on the DGP essentially vanishes for strongly heteroskedastic DGPs. That is, the distribution function of $T(\mathbf{Y}^*)$ approximately equals a distribution function $\digamma_{i^*}$, say, which depends on i^*, but on no other aspect of the DGP. The bootstrap critical value is then a $1-\alpha$ quantile of $\digamma_{i^*}$. Recalling from 3. that for strongly heteroskedastic DGPs $T(\mathbf{Y}) \approx T(\mu_0 + e_{i^*}(n))$, it follows that for such DGPs the event that the wild bootstrap based test rejects the null hypothesis essentially coincides with the event that $T(\mu_0 + e_{i^*}(n))$ exceeds the $1-\alpha$ quantile of $\digamma_{i^*}$. Both $T(\mu_0 + e_{i^*}(n))$ and $\digamma_{i^*}$ are nonrandom. Hence (recall that we are operating under the null hypothesis) for strongly heteroskedastic DGPs the test will have a rejection probability close to one if $\digamma_{i^*}(T(\mu_0 + e_{i^*}(n))) > 1-\alpha$.[17] In the above argument i^* was fixed. Varying $i^* \in \{1,\dots,n\}$, we finally come to the conclusion that the maximal rejection probability under the null will be close to 1 in case

$$\max_{i^*=1,\dots,n} \digamma_{i^*}(T(\mu_0 + e_{i^*}(n))) > 1-\alpha.$$

In other words, the bootstrap-based test under consideration will have rejection probabilities close to one for all levels of significance satisfying

$$\alpha > 1 - \max_{i^*=1,\dots,n} \digamma_{i^*}\left(T(\mu_0 + e_{i^*}(n))\right).$$

The quantity to the right is closely related to our constants ϑ. Note that the above reasoning is nonrigorous and, in particular, does not take into consideration some technical subtleties that arise in the just given approximation arguments and that we have tacitly ignored in the preceding discussion. Therefore, the expressions for the constants ϑ we arrive at in the theorems in the subsequent section are somewhat more complicated, albeit the underlying intuition is the same.

As transpires from the preceding heuristic discussion, the method for establishing the size one results given in the next section relies on the assumption that the heteroskedasticity model employed is rich enough to approximate extreme cases of strongly heteroskedastic DGPs, namely the ones where all but one observation have zero variance, arbitrarily well. This is certainly so for the leading case of the heteroskedasticity model $\mathfrak{C}_{Het}$, which describes agnosticism about the form of heteroskedasticity. Therefore, the results in the next section are presented for this case, and a discussion to which other heteroskedasticity models these results generalize is given in Section 7.

If one maintains a heteroskedasticity model that does not allow one to approximate any of the above mentioned extreme cases of strongly heteroskedastic DGPs (such as, e.g., the heteroskedasticity model $\mathfrak{C}_{Het}(a)$ which consists of all error-covariance matrices in $\mathfrak{C}_{Het}$ with diagonal elements bounded from below by $a > 0$), then the method of proof underlying our size one results no longer is applicable. However, this does *not* imply that the size of a bootstrap-based test over $\mathfrak{C}_{Het}(a)$

[17] More precisely, if the left hand side limit $\digamma_{i^*}(T(\mu_0 + e_{i^*}(n))-)$ exceeds $1-\alpha$. We ignore this technical detail here for the sake of simplicity.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

is about right: Since the rejection probabilities are continuous in the parameters (in particular, in Σ), the size over $\mathfrak{C}_{Het}(a)$ will be much larger than the nominal significance level at least for small a (in fact, it will be close to one if a is sufficiently small) in any situation where the size over $\mathfrak{C}_{Het}$ equals one (e.g., in the situations described in the theorems further below). [The actual size over $\mathfrak{C}_{Het}(a)$ depends on the chosen bound a and on the design matrix, the hypothesis to be tested, the test statistic, and also on the bootstrap scheme used.] Furthermore, the bound a has to be decided on prior to the data analysis and is part of modeling the form of heteroskedasticity. It is difficult to see how one would come up with a reasonable bound a in practice: if a is chosen to be very small, this may result in a heteroskedasticity model under which the tests are still severely oversized as just discussed, while choosing a large will typically not be defendable as it presumes considerable knowledge about the admissible forms of heteroskedasticity.

5. SIZE ONE RESULTS

In this section, we provide sufficient conditions for the size of bootstrap-based heteroskedasticity robust tests to be equal to one when the heteroskedasticity model is $\mathfrak{C}_{Het}$, which is the largest possible heteroskedasticity model and which reflects agnosticism regarding the form of heteroskedasticity. For extensions to other heteroskedasticity models see Section 7. We next discuss the bootstrap schemes that will be considered and which all are based on the wild bootstrap idea. The first bootstrap scheme is given by

$$y^*(y,\xi) = X\tilde{\beta}_{\mathfrak{M}_0}(y) + \operatorname{diag}(\xi)(y - X\tilde{\beta}_{\mathcal{A}}(y)) \tag{8}$$

for every $y \in \mathbb{R}^n$, where $\mathcal{A}$ will always be an affine subspace of $\mathbb{R}^n$ satisfying $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \operatorname{span}(X)$, and where ξ is a draw from Ξ, a given (Borel) probability measure on $\mathbb{R}^n$. Typical choices in the literature are $\mathcal{A} = \mathfrak{M}_0$, i.e., one uses restricted residuals in the wild bootstrap, or $\mathcal{A} = \operatorname{span}(X)$, in which case unrestricted residuals are used.[18] In practice only these two choices will typically arise, but the theory given below covers the more general case where $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \operatorname{span}(X)$ at no extra cost. The measure Ξ may depend on observable quantities like, e.g., X, R, or $\mathcal{A}$, but not on y. For example, Ξ could be the n-fold product of Mammen or Rademacher distributions, but other choices (e.g., ones obtained by modifying the aforementioned distributions by weights, or nondiscrete distributions) are also covered. See Section 8 for some examples. For the theoretical results in this section, there is no need to specify a particular form of Ξ.[19]

[18] Because all test statistics (and associated exceptional sets) considered below are at least $G(\mathfrak{M}_0)$-invariant (see Remarks 3.2 and 5.6), the bootstrap scheme (8) can be replaced by $y^{**}(y,\xi) = \mu_0 + \operatorname{diag}(\xi)(y - X\tilde{\beta}_{\mathcal{A}}(y))$ for an arbitrary value $\mu_0 \in \mathfrak{M}_0$ without affecting the bootstrapped test statistic.

[19] Suppose Ξ is the empirical distribution of B draws (possibly modified by weights) from an underlying distribution Ξ_0, which will often be the case if n is large and the ideal bootstrap using Ξ_0 is infeasible. In this case Ξ is strictly speaking a random probability measure (depending on the particular sample of size B drawn from Ξ_0) and the

The second bootstrap scheme differs from the first one only insofar as centering is at the unrestricted estimator $X\hat{\beta}(y)$ rather than at the restricted estimator $X\tilde{\beta}_{\mathfrak{M}_0}(y)$. That is, the second bootstrap scheme is given by

$$y^{\maltese}(y,\xi) = X\hat{\beta}(y) + \operatorname{diag}(\xi)(y - X\tilde{\beta}_{\mathcal{A}}(y)) \tag{9}$$

for every $y \in \mathbb{R}^n$. Note that $y^*(y,\xi)$ as well as $y^{\maltese}(y,\xi)$ depend also on the choice of $\mathcal{A}$, but we shall not show this dependence in the notation.

5.1. Bootstrap-Based Tests Derived from T_{Het} and T_{uc}

In the subsequent theorems, Ξ is always a (Borel) probability measure on $\mathbb{R}^n$, and $\mathcal{A}$ is an affine subspace of $\mathbb{R}^n$ satisfying $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \operatorname{span}(X)$. If we use the first bootstrap scheme, i.e., (8), the bootstrapped test statistic corresponding to T_{Het} is given by T^*_{Het}, where $T^*_{Het} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined via

$$T^*_{Het}(y,\xi) = T_{Het}\left(y^*(y,\xi)\right).$$

Furthermore, for every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $F_{Het,y}$, i.e., $F_{Het,y}(t) = \Xi(T^*_{Het}(y,\xi) \leq t)$ for $t \in \mathbb{R}$. For reasons that are discussed further below, we also need to consider a modification of T_{Het} defined by $T^{\blacktriangle}_{Het}(y) = T_{Het}(y)$ if $y \notin \mathsf{B}$ and $T^{\blacktriangle}_{Het}(y) = \infty$ otherwise. Its bootstrapped version is then given by $T^{\blacktriangle,*}_{Het}(y,\xi) = T^{\blacktriangle}_{Het}(y^*(y,\xi))$. Similarly as before, for every $y \in \mathbb{R}^n$ we denote its distribution function under Ξ by $F^{\blacktriangle}_{Het,y}$, i.e., $F^{\blacktriangle}_{Het,y}(t) = \Xi(T^{\blacktriangle,*}_{Het}(y,\xi) \leq t)$ for $t \in \mathbb{R} \cup \{\infty\}$.

THEOREM 5.1. *Suppose Assumption 1 holds.*

(a) For every $\alpha \in (0,1)$, let $f_{Het,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $F_{Het,y}$. Define $\vartheta_{Het} = 1 - \max(\vartheta_{1,Het}, \vartheta_{2,Het})$, where

$$\vartheta_{1,Het} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\notin \mathsf{B}}} \Xi\left(\left\{\xi : T^*_{Het}(\mu_0 + e_i(n),\xi) < T_{Het}(\mu_0 + e_i(n)), y^*(\mu_0 + e_i(n),\xi) \notin \mathsf{B}\right\}\right) \tag{10}$$

and

$$\vartheta_{2,Het} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\in \operatorname{span}(X),\, R\hat{\beta}(e_i(n))\neq 0}} \Xi\left(\left\{\xi : y^*(\mu_0 + e_i(n),\xi) \notin \mathsf{B}\right\}\right) \tag{11}$$

for some $\mu_0 \in \mathfrak{M}_0$, with the convention that $\vartheta_{1,Het} = 0$ ($\vartheta_{2,Het} = 0$, respectively) if the index set in the maximum operator in (10) ((11), respectively) is empty. Then

bootstrap-based tests also depend on this sample. However, working conditionally on this sample brings us back into the current framework.

neither $\vartheta_{1,Het}$ *nor* $\vartheta_{2,Het}$ *depend on the choice of* $\mu_0 \in \mathfrak{M}_0$. *Furthermore, for every* $\alpha \in (0,1)$ *such that* $\alpha > \vartheta_{Het}$ *holds, we have*

$$\sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{Het} \geq f_{Het,1-\alpha}\right) \geq \sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{Het} > f_{Het,1-\alpha}\right) = 1 \tag{12}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (12) are to be interpreted as inner probabilities*[20]*).*

(b) For every $\alpha \in (0,1)$, *let* $f^{\blacktriangle}_{Het,1-\alpha}(y)$ *denote a* $(1-\alpha)$*-quantile of* $F^{\blacktriangle}_{Het,y}$. *Then, with* ϑ_{Het} *defined in Part (a), for every* $\alpha \in (0,1)$ *such that* $\alpha > \vartheta_{Het}$ *holds, we have*

$$\sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{Het} \geq f^{\blacktriangle}_{Het,1-\alpha}\right) \geq \sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{Het} > f^{\blacktriangle}_{Het,1-\alpha}\right) = 1 \tag{13}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (13) are to be interpreted as inner probabilities).*

Part (a) of the preceding theorem implies that for every nominal significance level $\alpha > \vartheta_{Het}$ the size (over $\mathfrak{C}_{Het}$) of the bootstrap-based test derived from T_{Het} is equal to 1 and thus is inflated (and this is true whether the bootstrap-based test uses the rejection region $\{y : T_{Het}(y) \geq f_{Het,1-\alpha}(y)\}$ or $\{y : T_{Het}(y) > f_{Het,1-\alpha}(y)\}$).[21] Note that the lower bound ϑ_{Het} is observable and can be computed, see Section 5.3 for some more detail. As we shall see from the numerical results in Section 8, the lower bound ϑ_{Het} can be quite small, the results in the theorem thus covering standard choices for α such as $\alpha = 0.05$. A consequence of Theorem 5.1 thus is, in particular, that there is in general no guarantee for bootstrap-based tests derived from T_{Het} (or from the other statistics considered in the theorems further below), conducted at a nominal significance level α, to be truly level α tests. Although trivial, we note that Theorem 5.1 provides only a *sufficient* condition for size being equal to one and thus, in case $\alpha \leq \vartheta_{Het}$ holds, the size of the bootstrap-based test may nevertheless be much larger than α (and may perhaps even be equal to 1).

The significance of Part (b) of the theorem is as follows: Recall from Lemma 3.1 that under Assumption 1 the way T_{Het} is defined on B is immaterial for the rejection probabilities of (nonbootstrap-based) tests obtained from this test statistic since the set B is a Lebesgue null set and since the probability measures in (2) are all absolutely continuous w.r.t. Lebesgue measure; in particular, the (nonbootstrap-based) tests derived from T_{Het} and $T^{\blacktriangle}_{Het}$ have the same rejection probabilities. However, when it comes to the *bootstrapped* test statistics, the situation becomes more complicated as Ξ often will be a discrete measure. That

[20] This allows one to ignore measurability issues regarding $f_{Het,1-\alpha}$.

[21] We note that in principle it is conceivable that these two rejection regions have different probabilities under $P_{\mu_0,\sigma^2\Sigma}$.

is, it is a priori conceivable that the value we assign to the test statistic on the set B may have an effect on the *bootstrapped* test statistic and thus on the $(1-\alpha)$-quantile computed from it; in particular, it might be that an assignment of a value different from zero on the set B may lead to a larger $(1-\alpha)$-quantile. This then raises the question, whether a bootstrap-based test that uses such a (potentially) larger $(1-\alpha)$-quantile may have a smaller size than when the quantile $f_{Het,1-\alpha}$ is being used. Within the context of the theorem, Part (b) answers this in the negative by showing that, even if one defines the *bootstrapped* test statistic as ∞ on the event where the bootstrap sample $y^*(y,\xi)$ falls into the exceptional set B and uses a resulting $(1-\alpha)$-quantile, the bootstrap-based test again has size 1 under the same condition on α. [As any other way of defining the bootstrapped test statistic on the event $y^*(y,\xi) \in \mathsf{B}$ obviously leads to $(1-\alpha)$-quantiles not larger than an (appropriately chosen) $(1-\alpha)$-quantile of $F^{\blacktriangle}_{Het,y}$, Part (b) covers also any such alternative definition of the bootstrapped test statistic.][22] For additional discussion see also Remark 5.12.

We also stress that the results in the preceding theorem hold for *any* choice $f_{Het,1-\alpha}$ ($f^{\blacktriangle}_{Het,1-\alpha}(y)$, respectively) from the set of $(1-\alpha)$-quantiles of $F_{Het,y}$ ($F^{\blacktriangle}_{Het,y}$, respectively).[23]

Furthermore, we note that the preceding theorem holds with the *same* lower bound ϑ_{Het} for a much larger class of error distributions than just Gaussian errors (an assumption we have made only for convenience), see Section 7. Hence, in this sense the lower bound ϑ_{Het} is "distribution free."

We next turn to the test statistic T_{uc}. Again using the first bootstrap scheme, the bootstrapped test statistic is then given by T^*_{uc} where $T^*_{uc} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined via

$$T^*_{uc}(y,\xi) = T_{uc}\left(y^*(y,\xi)\right).$$

Furthermore, for every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $F_{uc,y}$, i.e., $F_{uc,y}(t) = \Xi(T^*_{uc}(y,\xi) \le t)$ for $t \in \mathbb{R}$. As before, we also need to consider the modification of T_{uc} defined by $T^{\blacktriangle}_{uc}(y) = T_{uc}(y)$ if $y \notin \operatorname{span}(X)$ and $T^{\blacktriangle}_{uc}(y) = \infty$ otherwise. Its bootstrapped version is then given by $T^{\blacktriangle,*}_{uc}(y,\xi) = T^{\blacktriangle}_{uc}(y^*(y,\xi))$. Similarly as before, for every $y \in \mathbb{R}^n$ we denote its distribution function under Ξ by $F^{\blacktriangle}_{uc,y}$, i.e., $F^{\blacktriangle}_{uc,y}(t) = \Xi(T^{\blacktriangle,*}_{uc}(y,\xi) \le t)$ for $t \in \mathbb{R} \cup \{\infty\}$.

[22] An alternative approach, which—if successful—would make considering Part (b) obsolete, would be to try to show that the set of y's for which $T^{\blacktriangle,*}_{Het}$ and T^*_{Het} coincide Ξ-a.e., and thus their quantiles coincide, is the complement of a Lebesgue null set. While this alternative approach actually can be shown to work in some special cases, it does not so in general, as can be seen from examples.

[23] Suppose $0 < \delta < 1$ and F is a cdf defined on $\mathbb{R}$ ($\mathbb{R} \cup \{\infty\}$, respectively). An element $q \in \mathbb{R}$ ($q \in \mathbb{R} \cup \{\infty\}$, respectively) is said to be a δ-quantile of F iff it satisfies $F(q) \ge \delta \ge F(q-)$, where $F(q-)$ denotes the left-hand limit of F at q. Note that q need not be unique in general. There is always a smallest and a largest δ-quantile among all δ-quantiles. The smallest one is given by $F^{-1}(\delta)$, where F^{-1} is the "generalized" inverse of F. If δ does not belong to the range of F, then $F^{-1}(\delta)$ is also the largest δ-quantile. Otherwise, the largest δ-quantile is given by $\sup\{x \in \mathbb{R} : F(x) = \delta\}$ ($\sup\{x \in \mathbb{R} \cup \{\infty\} : F(x) = \delta\}$, respectively), which may or may not coincide with $F^{-1}(\delta)$.

THEOREM 5.2. *(a) For every* $\alpha \in (0,1)$*, let* $f_{uc,1-\alpha}(y)$ *denote a* $(1-\alpha)$*-quantile of* $F_{uc,y}$*. Define* $\vartheta_{uc} = 1 - \max(\vartheta_{1,uc}, \vartheta_{2,uc})$*, where*

$$\vartheta_{1,uc} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\notin \operatorname{span}(X)}} \Xi\left(\left\{\xi : T^*_{uc}(\mu_0 + e_i(n), \xi) < T_{uc}(\mu_0 + e_i(n)), y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)\right\}\right) \tag{14}$$

and

$$\vartheta_{2,uc} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\in \operatorname{span}(X), R\hat{\beta}(e_i(n))\neq 0}} \Xi\left(\left\{\xi : y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)\right\}\right) \tag{15}$$

for some $\mu_0 \in \mathfrak{M}_0$*, with the convention that* $\vartheta_{2,uc} = 0$ *if the index set in the maximum operator in (15) is empty.*[24] *Then neither* $\vartheta_{1,uc}$ *nor* $\vartheta_{2,uc}$ *depend on the choice of* $\mu_0 \in \mathfrak{M}_0$*. Furthermore, for every* $\alpha \in (0,1)$ *such that* $\alpha > \vartheta_{uc}$ *holds, we have*

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{uc} \geq f_{uc,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{uc} > f_{uc,1-\alpha}\right) = 1 \tag{16}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (16) are to be interpreted as inner probabilities).*

(b) For every $\alpha \in (0,1)$*, let* $f^{\blacktriangle}_{uc,1-\alpha}(y)$ *denote a* $(1-\alpha)$*-quantile of* $F^{\blacktriangle}_{uc,y}$*. Then, with* ϑ_{uc} *defined in Part (a), for every* $\alpha \in (0,1)$ *such that* $\alpha > \vartheta_{uc}$ *holds, we have*

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{uc} \geq f^{\blacktriangle}_{uc,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_{uc} > f^{\blacktriangle}_{uc,1-\alpha}\right) = 1 \tag{17}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (17) are to be interpreted as inner probabilities).*

Mutatis mutandis, a discussion similar to the one given subsequently to Theorem 5.1 also applies here.

So far we have only considered the bootstrap scheme (8). We now turn to the second bootstrap scheme given by (9). Here the bootstrapped version of T_{Het} is given by

$$T^{\maltese}_{Het}(y,\xi) = (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y))'\hat{\Omega}^{-1}_{Het}\left(y^{\maltese}(y,\xi)\right)(R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y)),$$

if $y^{\maltese}(y,\xi) \notin \mathsf{B}$, and by $T^{\maltese}_{Het}(y,\xi) = 0$ if $y^{\maltese}(y,\xi) \in \mathsf{B}$. And the bootstrapped version of T_{uc} is given by

$$T^{\maltese}_{uc}(y,\xi) = (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y))'\left(\hat{\sigma}^2(y^{\maltese}(y,\xi))R\left(X'X\right)^{-1}R'\right)^{-1} \times (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y)),$$

[24]Note that the index set in the maximum operator in (14) cannot be empty since we have assumed $k < n$.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

if $y^{\maltese}(y,\xi) \notin \operatorname{span}(X)$, and by $T_{uc}^{\maltese}(y,\xi) = 0$ if $y^{\maltese}(y,\xi) \in \operatorname{span}(X)$. Furthermore, $T_{Het}^{\blacktriangle,\maltese}(y,\xi)$ and $T_{uc}^{\blacktriangle,\maltese}(y,\xi)$ are defined in exactly the same way, except that $T_{Het}^{\blacktriangle,\maltese}(y,\xi) = \infty$ if $y^{\maltese}(y,\xi) \in \mathsf{B}$ and that $T_{uc}^{\blacktriangle,\maltese}(y,\xi) = \infty$ if $y^{\maltese}(y,\xi) \in \operatorname{span}(X)$.

We will show in the next lemma that $T_{Het}^{\maltese}(y,\xi)$ coincides with $T_{Het}^{*}(y,\xi)$, and that the same is true for $T_{uc}^{\maltese}(y,\xi)$ and $T_{uc}^{*}(y,\xi)$ (as well as for $T_{Het}^{\blacktriangle,\maltese}(y,\xi)$ and $T_{Het}^{\blacktriangle,*}(y,\xi)$, and $T_{uc}^{\blacktriangle,\maltese}(y,\xi)$ and $T_{uc}^{\blacktriangle,*}(y,\xi)$), provided the same affine space $\mathcal{A}$ is used in (8) and (9). As a consequence, this—together with Remark 5.4—shows that Theorems 5.1 and 5.2 also apply immediately to the bootstrap-based test when the second bootstrap scheme, i.e., (9), is used (with the same $\mathcal{A}$ and Ξ). The lemma is certainly not new and is a variant of a similar result given as Proposition 1 in van Giersbergen and Kiviet (2002).

LEMMA 5.3. *We have* $T_{Het}^{*}(y,\xi) = T_{Het}^{\maltese}(y,\xi)$, $T_{uc}^{*}(y,\xi) = T_{uc}^{\maltese}(y,\xi)$, $T_{Het}^{\blacktriangle,*}(y,\xi) = T_{Het}^{\blacktriangle,\maltese}(y,\xi)$, *and* $T_{uc}^{\blacktriangle,*}(y,\xi) = T_{uc}^{\blacktriangle,\maltese}(y,\xi)$ *for every* $y \in \mathbb{R}^n$ *and every* $\xi \in \mathbb{R}^n$. [*Here it is understood that both bootstrap schemes are based on the same affine space* $\mathcal{A}$.]

Remark 5.4. Define θ_{Het} exactly in the same way as ϑ_{Het}, except that T_{Het}^{*} and y^{*} are replaced by $T_{Het}^{\maltese}$ and $y^{\maltese}$. Similarly define θ_{uc}. Then $\theta_{Het} = \vartheta_{Het}$ and $\theta_{uc} = \vartheta_{uc}$ in view of Lemma 5.3 and the fact that $y^{*}(y,\xi) \notin \mathsf{B}$ iff $y^{\maltese}(y,\xi) \notin \mathsf{B}$ and $y^{*}(y,\xi) \notin \operatorname{span}(X)$ iff $y^{\maltese}(y,\xi) \notin \operatorname{span}(X)$ (note that $y^{*}(y,\xi) - y^{\maltese}(y,\xi) \in \operatorname{span}(X)$ and that $\mathsf{B} + \operatorname{span}(X) = \mathsf{B}$).

5.2. Bootstrap-Based Tests Derived from $\tilde{T}_{Het}$ and $\tilde{T}_{uc}$

Based on suggestions in the literature on bootstrapping heteroskedasticity robust tests, we next consider two further test statistics, which are versions of T_{Het} and T_{uc} with the only difference that the covariance matrix estimators used are computed from restricted—instead of unrestricted—residuals. We thus define

$$\tilde{T}_{Het}(y) = \begin{cases} (R\hat{\beta}(y) - r)' \tilde{\Omega}_{Het}^{-1}(y) (R\hat{\beta}(y) - r), & \text{if } \det \tilde{\Omega}_{Het}(y) \neq 0, \\ 0, & \text{if } \det \tilde{\Omega}_{Het}(y) = 0, \end{cases} \tag{18}$$

where $\tilde{\Omega}_{Het} = R\tilde{\Psi}_{Het}R'$ and where $\tilde{\Psi}_{Het}$ is given by

$$\tilde{\Psi}_{Het}(y) = (X'X)^{-1} X' \operatorname{diag}\left(\tilde{d}_1 \tilde{u}_1^2(y), \ldots, \tilde{d}_n \tilde{u}_n^2(y)\right) X(X'X)^{-1},$$

where the constants $\tilde{d}_i > 0$ sometimes depend on the design matrix and on the restriction matrix R. Here $\tilde{u}(y) = y - X\tilde{\beta}_{\mathfrak{M}_0}(y) = \Pi_{(\mathfrak{M}_0^{lin})^{\perp}}(y - \mu_0)$, where the last expression does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$, and where $\tilde{u}_t(y)$ denotes the tth component of $\tilde{u}(y)$. Typical choices for $\tilde{d}_i$ are $\tilde{d}_i = 1$, $\tilde{d}_i = n/(n-(k-q))$, $\tilde{d}_i = (1 - \tilde{h}_{ii})^{-1}$, or $\tilde{d}_i = (1 - \tilde{h}_{ii})^{-2}$ where $\tilde{h}_{ii}$ denotes the ith diagonal element of the projection matrix $\Pi_{\mathfrak{M}_0^{lin}}$, see, e.g., Davidson and MacKinnon (1985). Another suggestion is $\tilde{d}_i = (1 - \tilde{h}_{ii})^{-\tilde{\delta}_i}$ for $\tilde{\delta}_i = \min(n\tilde{h}_{ii}/(k-q), 4)$ with the convention that

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

$\tilde{\delta}_i = 0$ if $k = q$.[25] For the last three choices of $\tilde{d}_i$ just given we use the convention that we set $\tilde{d}_i = 1$ in case $\tilde{h}_{ii} = 1$. Note that $\tilde{h}_{ii} = 1$ implies $\tilde{u}_i(y) = 0$ for every y, and hence it is irrelevant which real value is assigned to $\tilde{d}_i$ in case $\tilde{h}_{ii} = 1$.[26] The five examples for the weights $\tilde{d}_i$ just given correspond to what is often called HC0R–HC4R weights in the literature.[27]

The subsequent assumption ensures that the set of y's for which $\tilde{\Omega}_{Het}(y)$ is singular is a Lebesgue null set, implying that our choice of assigning $\tilde{T}_{Het}(y)$ the value zero in case $\tilde{\Omega}_{Het}(y)$ is singular has no import on the rejection probabilities of the (nonbootstrap-based) tests obtained from $\tilde{T}_{Het}$ (as the measures $P_{\mu,\sigma^2\Sigma}$ are absolutely continuous). As will be seen later, our results for the corresponding bootstrap-based tests do also not depend on this choice. Also, as discussed further below, the assumption is in a certain sense unavoidable when using $\tilde{T}_{Het}$.

Assumption 2. Let $1 \le i_1 < \cdots < i_s \le n$ denote all the indices for which $e_{i_j}(n) \in \mathfrak{M}_0^{lin}$ holds where $e_j(n)$ denotes the jth standard basis vector in $\mathbb{R}^n$. If no such index exists, set $s = 0$. Let $X'(\neg(i_1, \dots i_s))$ denote the matrix which is obtained from X' by deleting all columns with indices i_j, $1 \le i_1 < \cdots < i_s \le n$ (if $s = 0$ no column is deleted). Then rank $\left(R(X'X)^{-1}X'(\neg(i_1, \dots i_s))\right) = q$ holds.

Observe that this assumption only depends on X and R and hence can be checked. Obviously, a simple sufficient condition for Assumption 2 to hold is that $s = 0$ (i.e., that $e_j(n) \notin \mathfrak{M}_0^{lin}$ for all j), a generically satisfied condition. Furthermore, we introduce the matrix

$$\begin{aligned}\tilde{B}(y) &= R(X'X)^{-1}X'\operatorname{diag}(\tilde{u}_1(y), \dots, \tilde{u}_n(y)) \\ &= R(X'X)^{-1}X'\operatorname{diag}\left(e_1'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0), \dots, e_n'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)\right).\end{aligned} \tag{19}$$

Note that this matrix does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. The following lemma collects some important properties of $\tilde{\Omega}_{Het}$ and $\tilde{\mathsf{B}}$ (defined in that lemma). Its proof is given in Appendix C.

LEMMA 5.5. *(a) $\tilde{\Omega}_{Het}(y)$ is nonnegative definite for every $y \in \mathbb{R}^n$.*

(b) $\tilde{\Omega}_{Het}(y)$ is singular (zero, respectively) if and only if $\operatorname{rank}(\tilde{B}(y)) < q$ *($\tilde{B}(y) = 0$, respectively).*

(c) The set $\tilde{\mathsf{B}}$ given by $\{y \in \mathbb{R}^n : \operatorname{rank}(\tilde{B}(y)) < q\}$ (or, in view of (b), equivalently given by $\{y \in \mathbb{R}^n : \det(\tilde{\Omega}_{Het}(y)) = 0\}$) is either a $\lambda_{\mathbb{R}^n}$-null set or the entire sample space $\mathbb{R}^n$. The latter occurs if and only if Assumption 2 is violated (in which case the test based on $\tilde{T}_{Het}$ becomes trivial, as then $\tilde{T}_{Het}$ is identically zero).

(d) Suppose Assumption 2 holds. Then for every $\mu_0 \in \mathfrak{M}_0$ the set $\tilde{\mathsf{B}} - \mu_0$ is a finite union of proper linear subspaces; in case $q = 1$, $\tilde{\mathsf{B}} - \mu_0$ is even a proper linear

[25]Note that in case $k = q$ we have $\tilde{h}_{ii} = 0$, and hence $\tilde{d}_i = 1$ regardless of our convention for $\tilde{\delta}_i$.

[26]In fact, $\tilde{h}_{ii} = 1$ is equivalent to $\tilde{u}_i(y) = 0$ for every y, each of which in turn is equivalent to $e_i(n) \in \mathfrak{M}_0^{lin}$.

[27]In the case $k = q$ the HC0R–HC4R weights all coincide ($\tilde{d}_i = 1$ for every i), and hence result in the same test statistic.

subspace itself.[28,29] [*Note that* $\tilde{\mathsf{B}} - \mu_0$ *does not depend on the choice of* $\mu_0 \in \mathfrak{M}_0$. *In particular, if* $r = 0$, *i.e., if* $\mathfrak{M}_0$ *is linear, we thus may set* $\mu_0 = 0$.]

(*e*) $\tilde{\mathsf{B}}$ *is a closed set and contains* $\mathfrak{M}_0$. *Also* $\tilde{\mathsf{B}}$ *is* $G(\mathfrak{M}_0)$*-invariant, and in particular* $\tilde{\mathsf{B}} + \mathfrak{M}_0^{lin} = \tilde{\mathsf{B}}$.

In light of Part (c) of the lemma, we see that Assumption 2 is a natural and unavoidable condition if one wants to obtain a sensible test from $\tilde{T}_{Het}$.[30] Furthermore, note that if $\tilde{\mathsf{B}} = \mathfrak{M}_0$ is true, then Assumption 2 must be satisfied (since $\mathfrak{M}_0$ is a $\lambda_{\mathbb{R}^n}$-null set as $k - q < n$ is always the case). For later use we also mention that under Assumption 2 the statistic $\tilde{T}_{Het}$ is continuous at every $y \in \mathbb{R}^n \backslash \tilde{\mathsf{B}}$.[31]

We finally consider in analogy with T_{uc}

$$\tilde{T}_{uc}(y) = \begin{cases} (R\hat{\beta}(y) - r)' \left(\tilde{\sigma}^2(y) R \left(X'X\right)^{-1} R'\right)^{-1} (R\hat{\beta}(y) - r), & \text{if } y \notin \mathfrak{M}_0, \\ 0, & \text{if } y \in \mathfrak{M}_0, \end{cases} \tag{20}$$

where $\tilde{\sigma}^2(y) = \tilde{u}(y)' \tilde{u}(y) / (n - (k - q)) \geq 0$ (which vanishes if and only if $y \in \mathfrak{M}_0$). Of course, our choice to set $\tilde{T}_{uc}(y) = 0$ for $y \in \mathfrak{M}_0$ has no import on the rejection probabilities of the (nonbootstrap-based) tests obtained from $\tilde{T}_{uc}$, since $\mathfrak{M}_0$ is a $\lambda_{\mathbb{R}^n}$-null set (and since the measures $P_{\mu,\sigma^2\Sigma}$ are absolutely continuous). It will turn out also not to affect our results for bootstrap-based tests obtained from $\tilde{T}_{uc}$. For later use we also mention that $\tilde{T}_{uc}$ is continuous at every $y \in \mathbb{R}^n \backslash \mathfrak{M}_0$.

Remark 5.6. The test statistics $\tilde{T}_{Het}$ as well as $\tilde{T}_{uc}$ are $G(\mathfrak{M}_0)$-invariant as is easily seen (with the respective exceptional sets $\tilde{\mathsf{B}}$ and $\mathfrak{M}_0$ also being $G(\mathfrak{M}_0)$-invariant), but typically they are *not* nonsphericity-corrected F-type tests in the sense of Section 5.4 in Preinerstorfer and Pötscher (2016).

In the theorems given in the next two subsections we use the same bootstrap schemes as before (i.e., (8) and (9)); in particular, recall that Ξ is a (Borel) probability measure on $\mathbb{R}^n$, and that $\mathcal{A}$ is an affine subspace of $\mathbb{R}^n$ satisfying $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \text{span}(X)$.

5.2.1. *The first bootstrap scheme.* We start with results where the first bootstrap scheme, i.e., (8), is being used. The bootstrapped test statistic corresponding to $\tilde{T}_{Het}$ is then given by $\tilde{T}^*_{Het}$, where $\tilde{T}^*_{Het} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined via

$$\tilde{T}^*_{Het}(y, \xi) = \tilde{T}_{Het}\left(y^*(y, \xi)\right).$$

[28] Consequently, $\tilde{\mathsf{B}}$ is a finite union of proper affine subspaces, and is a proper affine subspace itself in case $q = 1$.

[29] If Assumption 2 is violated, then $\tilde{\mathsf{B}} - \mu_0 = \tilde{\mathsf{B}} = \mathbb{R}^n$ in view of Part (c).

[30] If this assumption is violated then $\tilde{T}_{Het}$ is identically zero, an uninteresting trivial case.

[31] If Assumption 2 is violated, then $\tilde{T}_{Het}$ is constant equal to zero, and hence trivially continuous everywhere.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Furthermore, for every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $\tilde{F}_{Het,y}$, i.e., $\tilde{F}_{Het,y}(t) = \Xi(\tilde{T}^*_{Het}(y,\xi) \le t)$ for $t \in \mathbb{R}$. For similar reasons as in Section 5.1, we also consider the modification of $\tilde{T}_{Het}$ defined by $\tilde{T}^{\blacktriangle}_{Het}(y) = \tilde{T}_{Het}(y)$ if $y \notin \tilde{\mathsf{B}}$ and $\tilde{T}^{\blacktriangle}_{Het}(y) = \infty$ otherwise. Its bootstrapped version is then given by $\tilde{T}^{\blacktriangle,*}_{Het}(y,\xi) = \tilde{T}^{\blacktriangle}_{Het}(y^*(y,\xi))$. For every $y \in \mathbb{R}^n$ we denote its distribution function under Ξ by $\tilde{F}^{\blacktriangle}_{Het,y}$, i.e., $\tilde{F}^{\blacktriangle}_{Het,y}(t) = \Xi(\tilde{T}^{\blacktriangle,*}_{Het}(y,\xi) \le t)$ for $t \in \mathbb{R} \cup \{\infty\}$.

THEOREM 5.7. *Suppose Assumption 2 holds.*
(a) For every $\alpha \in (0,1)$, let $\tilde{f}_{Het,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{F}_{Het,y}$. Define

$$\tilde{\vartheta}_{Het} = 1 - \max_{\substack{i=1,\dots,n,\\ \mu_0+e_i(n)\notin\tilde{\mathsf{B}}}} \Xi\left(\left\{\xi : \tilde{T}^*_{Het}(\mu_0+e_i(n),\xi) < \tilde{T}_{Het}(\mu_0+e_i(n)), y^*(\mu_0+e_i(n),\xi) \notin \tilde{\mathsf{B}}\right\}\right) \tag{21}$$

for some $\mu_0 \in \mathfrak{M}_0$, with the convention that $\tilde{\vartheta}_{Het} = 1$ if the index set in the maximum operator in (21) is empty. Then $\tilde{\vartheta}_{Het}$ does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. Furthermore, for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\vartheta}_{Het}$ holds, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} \ge \tilde{f}_{Het,1-\alpha}\right) \ge \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} > \tilde{f}_{Het,1-\alpha}\right) = 1 \tag{22}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (22) are to be interpreted as inner probabilities).
(b) For every $\alpha \in (0,1)$, let $\tilde{f}^{\blacktriangle}_{Het,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{F}^{\blacktriangle}_{Het,y}$. Then, with $\tilde{\vartheta}_{Het}$ defined in Part (a), for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\vartheta}_{Het}$ holds, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} \ge \tilde{f}^{\blacktriangle}_{Het,1-\alpha}\right) \ge \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} > \tilde{f}^{\blacktriangle}_{Het,1-\alpha}\right) = 1 \tag{23}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (23) are to be interpreted as inner probabilities).

We next turn to the test statistic $\tilde{T}_{uc}$. Again using the first bootstrap scheme, the bootstrapped test statistic is then given by $\tilde{T}^*_{uc}$, where $\tilde{T}^*_{uc} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined via

$$\tilde{T}^*_{uc}(y,\xi) = \tilde{T}_{uc}\left(y^*(y,\xi)\right).$$

Furthermore, for every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $\tilde{F}_{uc,y}$, i.e., $\tilde{F}_{uc,y}(t) = \Xi(\tilde{T}^*_{uc}(y,\xi) \le t)$ for $t \in \mathbb{R}$. We also consider the modification of $\tilde{T}_{uc}$ defined by $\tilde{T}^{\blacktriangle}_{uc}(y) = \tilde{T}_{uc}(y)$ if $y \notin \mathfrak{M}_0$ and $\tilde{T}^{\blacktriangle}_{uc}(y) = \infty$ otherwise. Its bootstrapped version is then given by $\tilde{T}^{\blacktriangle,*}_{uc}(y,\xi) = \tilde{T}^{\blacktriangle}_{uc}(y^*(y,\xi))$. For every $y \in \mathbb{R}^n$ we denote its distribution function under Ξ by $\tilde{F}^{\blacktriangle}_{uc,y}$, i.e., $\tilde{F}^{\blacktriangle}_{uc,y}(t) = \Xi(\tilde{T}^{\blacktriangle,*}_{uc}(y,\xi) \le t)$ for $t \in \mathbb{R} \cup \{\infty\}$.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

THEOREM 5.8. *(a) For every $\alpha \in (0,1)$, let $\tilde{f}_{uc,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{F}_{uc,y}$. Define*

$$\tilde{\vartheta}_{uc} = 1 - \max_{\substack{i=1,\dots,n,\\ \mu_0+e_i(n)\notin\mathfrak{M}_0}} \Xi\left(\left\{\xi : \tilde{T}^*_{uc}(\mu_0+e_i(n),\xi) < \tilde{T}_{uc}(\mu_0+e_i(n)), y^*(\mu_0+e_i(n),\xi) \notin \mathfrak{M}_0\right\}\right) \quad \textbf{(24)}$$

for some $\mu_0 \in \mathfrak{M}_0$.[32] *Then $\tilde{\vartheta}_{uc}$ does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. Furthermore, for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\vartheta}_{uc}$ holds, we have*

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{uc} \geq \tilde{f}_{uc,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{uc} > \tilde{f}_{uc,1-\alpha}\right) = 1 \quad \textbf{(25)}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (25) are to be interpreted as inner probabilities).

(b) For every $\alpha \in (0,1)$, let $\tilde{f}^{\blacktriangle}_{uc,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{F}^{\blacktriangle}_{uc,y}$. Then, with $\tilde{\vartheta}_{uc}$ defined in Part (a), for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\vartheta}_{uc}$ holds, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{uc} \geq \tilde{f}^{\blacktriangle}_{uc,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{uc} > \tilde{f}^{\blacktriangle}_{uc,1-\alpha}\right) = 1 \quad \textbf{(26)}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (26) are to be interpreted as inner probabilities).

Mutatis mutandis, a discussion similar to the one given subsequently to Theorem 5.1 also applies to the preceding two theorems.

5.2.2. *The second bootstrap scheme.* For the test statistics $\tilde{T}_{Het}$ and $\tilde{T}_{uc}$ an analogon to Lemma 5.3 is not available. Hence, we need to provide separate theorems for the case where the second bootstrap scheme, i.e., (9), is being used. This is done next. With this bootstrap scheme, the bootstrapped test statistic corresponding to $\tilde{T}_{Het}$ is given by

$$\tilde{T}^{\maltese}_{Het}(y,\xi) = (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y))'\tilde{\Omega}^{-1}_{Het}\left(y^{\maltese}(y,\xi)\right)(R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y)),$$

if $y^{\maltese}(y,\xi) \notin \tilde{\mathsf{B}}$, and by $\tilde{T}^{\maltese}_{Het}(y,\xi) = 0$ if $y^{\maltese}(y,\xi) \in \tilde{\mathsf{B}}$. For every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $\tilde{H}_{Het,y}$, i.e., $\tilde{H}_{Het,y}(t) = \Xi(\tilde{T}^{\maltese}_{Het}(y,\xi) \leq t)$ for $t \in \mathbb{R}$. Furthermore, $\tilde{T}^{\blacktriangle,\maltese}_{Het}(y,\xi)$ is defined exactly as is $\tilde{T}^{\maltese}_{Het}(y,\xi)$, except that $\tilde{T}^{\blacktriangle,\maltese}_{Het}(y,\xi) = \infty$ if $y^{\maltese}(y,\xi) \in \tilde{\mathsf{B}}$. For every $y \in \mathbb{R}^n$ denote its distribution function under Ξ by $\tilde{H}^{\blacktriangle}_{Het,y}$, i.e., $\tilde{H}^{\blacktriangle}_{Het,y}(t) = \Xi(\tilde{T}^{\blacktriangle,\maltese}_{Het}(y,\xi) \leq t)$ for $t \in \mathbb{R} \cup \{\infty\}$.

THEOREM 5.9. *Suppose Assumption 2 holds.*

[32] Note that the index set in the maximum operator in (24) cannot be empty since $k - q \leq k$ and we have assumed $k < n$.

(a) For every $\alpha \in (0,1)$, let $\tilde{h}_{Het,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{H}_{Het,y}$. Define

$$\tilde{\theta}_{Het} = 1 - \max_{\substack{i=1,\dots,n,\\ \mu_0+e_i(n)\notin \tilde{\mathsf{B}}}} \Xi\left(\left\{\xi : \tilde{T}^{\maltese}_{Het}(\mu_0+e_i(n),\xi) < \tilde{T}_{Het}(\mu_0+e_i(n)), y^{\maltese}(\mu_0+e_i(n),\xi) \notin \tilde{\mathsf{B}}\right\}\right) \tag{27}$$

for some $\mu_0 \in \mathfrak{M}_0$, with the convention that $\tilde{\theta}_{Het} = 1$ if the index set in the maximum operator in (27) is empty. Then $\tilde{\theta}_{Het}$ does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. Furthermore, for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\theta}_{Het}$ holds, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} \geq \tilde{h}_{Het,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} > \tilde{h}_{Het,1-\alpha}\right) = 1 \tag{28}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (28) are to be interpreted as inner probabilities).

(b) For every $\alpha \in (0,1)$, let $\tilde{h}^{\blacktriangle}_{Het,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{H}^{\blacktriangle}_{Het,y}$. Then, with $\tilde{\theta}_{Het}$ defined in Part (a), for every $\alpha \in (0,1)$ such that $\alpha > \tilde{\theta}_{Het}$ holds, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} \geq \tilde{h}^{\blacktriangle}_{Het,1-\alpha}\right) \geq \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(\tilde{T}_{Het} > \tilde{h}^{\blacktriangle}_{Het,1-\alpha}\right) = 1 \tag{29}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (29) are to be interpreted as inner probabilities).

With the bootstrap scheme considered in this subsection, the bootstrapped test statistic corresponding to $\tilde{T}_{uc}$ is given by

$$\tilde{T}^{\maltese}_{uc}(y,\xi) = (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y))'\left(\tilde{\sigma}^2(y^{\maltese}(y,\xi))R\left(X'X\right)^{-1}R'\right)^{-1} \times (R\hat{\beta}\left(y^{\maltese}(y,\xi)\right) - R\hat{\beta}(y)),$$

if $y^{\maltese}(y,\xi) \notin \mathfrak{M}_0$, and by $\tilde{T}^{\maltese}_{uc}(y,\xi) = 0$ if $y^{\maltese}(y,\xi) \in \mathfrak{M}_0$. For every $y \in \mathbb{R}^n$ denote the distribution function of the bootstrapped test statistic under Ξ by $\tilde{H}_{uc,y}$, i.e., $\tilde{H}_{uc,y}(t) = \Xi(\tilde{T}^{\maltese}_{uc}(y,\xi) \leq t)$ for $t \in \mathbb{R}$. Furthermore, $\tilde{T}^{\blacktriangle,\maltese}_{uc}(y,\xi)$ is defined exactly as is $\tilde{T}^{\maltese}_{uc}(y,\xi)$, except that $\tilde{T}^{\blacktriangle,\maltese}_{uc}(y,\xi) = \infty$ if $y^{\maltese}(y,\xi) \in \mathfrak{M}_0$. For every $y \in \mathbb{R}^n$ denote its distribution function under Ξ by $\tilde{H}^{\blacktriangle}_{uc,y}$, i.e., $\tilde{H}^{\blacktriangle}_{uc,y}(t) = \Xi(\tilde{T}^{\blacktriangle,\maltese}_{uc}(y,\xi) \leq t)$ for $t \in \mathbb{R}\cup\{\infty\}$.

THEOREM 5.10. *(a) For every $\alpha \in (0,1)$, let $\tilde{h}_{uc,1-\alpha}(y)$ denote a $(1-\alpha)$-quantile of $\tilde{H}_{uc,y}$. Define*

$$\tilde{\theta}_{uc} = 1 - \max_{\substack{i=1,\dots,n,\\ \mu_0+e_i(n)\notin \mathfrak{M}_0}} \Xi\left(\left\{\xi : \tilde{T}^{\maltese}_{uc}(\mu_0+e_i(n),\xi) < \tilde{T}_{uc}(\mu_0+e_i(n)), y^{\maltese}(\mu_0+e_i(n),\xi) \notin \mathfrak{M}_0\right\}\right) \tag{30}$$

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

for some $\mu_0 \in \mathfrak{M}_0$.[33] *Then* $\tilde{\theta}_{uc}$ *does not depend on the choice of* $\mu_0 \in \mathfrak{M}_0$. *Furthermore, for every* $\alpha \in (0,1)$ *such that* $\alpha > \tilde{\theta}_{uc}$ *holds, we have*

$$\sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0, \sigma^2 \Sigma}\left(\tilde{T}_{uc} \geq \tilde{h}_{uc,1-\alpha}\right) \geq \sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0, \sigma^2 \Sigma}\left(\tilde{T}_{uc} > \tilde{h}_{uc,1-\alpha}\right) = 1 \tag{31}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (31) are to be interpreted as inner probabilities).*

(b) For every $\alpha \in (0,1)$, *let* $\tilde{h}^{\blacktriangle}_{uc,1-\alpha}(y)$ *denote a* $(1-\alpha)$*-quantile of* $\tilde{H}^{\blacktriangle}_{uc,y}$. *Then, with* $\tilde{\theta}_{uc}$ *defined in Part (a), for every* $\alpha \in (0,1)$ *such that* $\alpha > \tilde{\theta}_{uc}$ *holds, we have*

$$\sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0, \sigma^2 \Sigma}\left(\tilde{T}_{uc} \geq \tilde{h}^{\blacktriangle}_{uc,1-\alpha}\right) \geq \sup_{\Sigma \in \mathfrak{C}_{Het}} P_{\mu_0, \sigma^2 \Sigma}\left(\tilde{T}_{uc} > \tilde{h}^{\blacktriangle}_{uc,1-\alpha}\right) = 1 \tag{32}$$

for every $\mu_0 \in \mathfrak{M}_0$ *and every* $0 < \sigma^2 < \infty$ *(where the probabilities in (32) are to be interpreted as inner probabilities).*

Mutatis mutandis, a discussion similar to the one given subsequently to Theorem 5.1 also applies to the preceding two theorems.

5.3. Further Remarks

Remark 5.11. As already noted earlier, the various lower bounds for α given in the theorems depend only on observable quantities, and can thus be computed numerically. In particular, ϑ_{Het} in Theorem 5.1 depends only on X, R, r, $\mathcal{A}$, Ξ, and on the d_i's appearing in the definition of the test statistic, while ϑ_{uc} in Theorem 5.2 depends only on X, R, r, $\mathcal{A}$, and Ξ. Similarly, $\tilde{\vartheta}_{Het}$ in Theorem 5.7 and $\tilde{\theta}_{Het}$ in Theorem 5.9 depend only on X, R, r, $\mathcal{A}$, Ξ, and on the $\tilde{d}_i$'s appearing in the definition of the test statistic. Finally, $\tilde{\vartheta}_{uc}$ in Theorem 5.8 and $\tilde{\theta}_{uc}$ in Theorem 5.10 depend only on X, R, r, $\mathcal{A}$, and Ξ.

Remark 5.12. (i) Relations (12) and (13) continue to hold a fortiori if T_{Het} is replaced by $T^{\blacktriangle}_{Het}$, as $T^{\blacktriangle}_{Het}$ is never smaller than T_{Het} (in fact, $T^{\blacktriangle}_{Het}$ and T_{Het} coincide except on a Lebesgue null set under Assumption 1).

(ii) Relations (16) and (17) continue to hold a fortiori if T_{uc} is replaced by $T^{\blacktriangle}_{uc}$, as $T^{\blacktriangle}_{uc}$ is never smaller than T_{uc} (in fact, both coincide except on $\operatorname{span}(X)$, a Lebesgue null set).

(iii) Relations (22), (23), (28), and (29) continue to hold a fortiori if $\tilde{T}_{Het}$ is replaced by $\tilde{T}^{\blacktriangle}_{Het}$, as $\tilde{T}^{\blacktriangle}_{Het}$ is never smaller than $\tilde{T}_{Het}$ (in fact, both coincide except on a Lebesgue null set under Assumption 2).

(iv) Relations (25), (26), (31), and (32) continue to hold a fortiori if $\tilde{T}_{uc}$ is replaced by $\tilde{T}^{\blacktriangle}_{uc}$, as $\tilde{T}^{\blacktriangle}_{uc}$ is never smaller than $\tilde{T}_{uc}$ (in fact, both coincide except on $\mathfrak{M}_0$, a Lebesgue null set).

[33]Note that the index set in the maximum operator in (30) can not be empty since $k - q \leq k$ and we have assumed $k < n$.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Remark 5.13. In case $q = 1$, inspection of the proof of Theorem 5.1 shows that the bound ϑ_{Het} can be somewhat improved by allowing in the definition of $\vartheta_{2,Het}$ the index i to range over all indices such that $e_i(n) \in \mathsf{B}$ and $R\hat{\beta}(e_i(n)) \neq 0$. [This is so, since in case $q = 1$ singularity of $\hat{\Omega}_{Het}(y)$ is equivalent to $\hat{\Omega}_{Het}(y) = 0$.]

Remark 5.14. The test statistics T_{Het} using HC0 and HC1 weights, respectively, differ only by a multiplicative constant, and hence result in the same bootstrap-based test. For design matrices X with h_{ii} not depending on i, the same conclusion applies for all weights HC0–HC4. A similar remark applies to $\tilde{T}_{Het}$ (with $\tilde{h}_{ii}$ taking the rôle of h_{ii}).

Remark 5.15. The size one results for bootstrap-based tests given in the preceding theorems are easily seen to imply infimal coverage zero results for the corresponding confidence sets for $R\beta$ obtained by "inverting" the tests. The computation of such confidence sets is straightforward and leads to ellipsoids in the case where the bootstrap-based test is obtained from T_{Het} or T_{uc} and the bootstrap scheme (9) with $\mathcal{A} = \operatorname{span}(X)$ is used. This is so, since the covariance matrix estimator employed in T_{Het} (or T_{uc}, respectively) does not depend on r, and since the quantile of the bootstrap distribution is easily seen also not to depend on r in this case. By Lemma 5.3 the same is true for the bootstrap-based tests obtained from T_{Het} or T_{uc} and the bootstrap scheme (8) with $\mathcal{A} = \operatorname{span}(X)$. In all other combinations of test statistics and bootstrap schemes the "inversion" is typically more complicated and becomes numerically burdensome, as then the covariance matrix estimator employed in the test statistic and/or the quantile of the bootstrap distribution will typically depend on r.

6. SOME SPECIAL CASES

Here we consider the special case where the null hypothesis is simple (i.e., $q = k$). If restricted residuals are used in the bootstrap scheme (8) (i.e., if $\mathcal{A} = \mathfrak{M}_0$ holds), the following result shows that in these cases our theorems become vacuous, and thus do not allow us to draw any conclusion about the sizes of the corresponding bootstrap-based tests. [Of course, this by itself does not preclude the possibility that in these cases the size may be equal to one or may substantially exceed α.] The observations made in the theorem below are in line with a result in Davidson and Flachaire (2008) implying that—in the case corresponding to Part (c) of the subsequent theorem—the bootstrap-based test using the bootstrap scheme (8) with $\mathcal{A} = \mathfrak{M}_0$ indeed has size equal to the nominal significance level α, *provided* a particular choice of Ξ and particular values of α are used. [In fact, this result, which is Theorem 1 in Davidson and Flachaire (2008), is not entirely correct in the form given, but needs some amendments and corrections, which we shall not provide here.[34]]

[34] A simple counterexample to Theorem 1 in Davidson and Flachaire (2008) is provided by a regression model which has a standard basis vector as its only regressor. It is then easy to see that the test statistic is (almost surely) constant

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

THEOREM 6.1. *Suppose* $q = k$ *and* $\Xi(\{\xi \in \mathbb{R}^n : \xi_i \neq 0\}) = 1$ *for every* $i = 1, \ldots, n$. *Then:*

(a) $\vartheta_{Het} = 1$ *holds in Theorem 5.1, if* $\mathcal{A} = \mathfrak{M}_0$ *is used in the bootstrap scheme.*[35]

(b) $\vartheta_{uc} = 1$ *holds in Theorem 5.2, if* $\mathcal{A} = \mathfrak{M}_0$ *is used in the bootstrap scheme.*

(c) $\tilde{\vartheta}_{Het} = 1$ *holds in Theorem 5.7, if* $\mathcal{A} = \mathfrak{M}_0$ *is used in the bootstrap scheme.*[36]

(d) $\tilde{\vartheta}_{uc} = 1$ *holds in Theorem 5.8, if* $\mathcal{A} = \mathfrak{M}_0$ *is used in the bootstrap scheme.*

Remark 6.2. (i) Part (a) (Part (b), respectively) of the preceding theorem applies to the bootstrap-based test derived from T_{Het} (T_{uc}, respectively) when the bootstrap scheme (8) with $\mathcal{A} = \mathfrak{M}_0$ is employed. In view of Lemma 5.3 and Remark 5.4, these results also apply if the bootstrap scheme (9), again with $\mathcal{A} = \mathfrak{M}_0$, is used. However, as can be seen from simple examples, this is not so in the context of Parts (c) and (d) of the preceding theorem (i.e., $\tilde{\theta}_{Het} < 1$ and $\tilde{\theta}_{uc} < 1$ can occur in Theorems 5.9 and 5.10, respectively, even if $q = k$, $\mathcal{A} = \mathfrak{M}_0$, and Ξ is as in Theorem 6.1).

(ii) If bootstrap schemes (8) or (9) are used for the bootstrap-based tests derived from any of T_{Het}, T_{uc}, $\tilde{T}_{Het}$, and $\tilde{T}_{uc}$, but now with $\mathfrak{M}_0 \subsetneqq \mathcal{A}$, simple examples show that $\vartheta_{Het} < 1$, $\vartheta_{uc} < 1$, $\tilde{\vartheta}_{Het} < 1$, $\tilde{\vartheta}_{uc} < 1$, $\tilde{\theta}_{Het} < 1$, and $\tilde{\theta}_{uc} < 1$ can occur.

As a consequence of the preceding theorem, the function in the R-package **wbsd** for computing ϑ_{Het}, etc. first checks if the conditions of the theorem are satisfied, and if so, outputs 1 for ϑ_{Het}, etc.

7. EXTENSIONS AND GENERALIZATIONS

7.1. Other Covariance Models

The results given so far refer to the size of bootstrap-based tests when the covariance model $\mathfrak{C}_{Het}$ is maintained. Inspection of the proofs of the theorems in Sections 5 and 6 as well as of Theorem A.1 in Appendix A shows that they also hold with $\mathfrak{C}_{Het}$ replaced by any covariance model $\mathfrak{C} \subseteq \mathfrak{C}_{Het}$ other than $\mathfrak{C}_{Het}$, provided the closure of $\mathfrak{C}$ contains, for $i = 1, \ldots, n$, the matrices $e_i(n)e_i(n)'$. More generally, if the closure of $\mathfrak{C}$ contains the matrices $e_i(n)e_i(n)'$ only for $i \in I \subseteq \{1, \ldots, n\}$, then Theorem A.1 continues to hold with $\mathfrak{C}_{Het}$ replaced by $\mathfrak{C}$, provided the range of the maximum operator in (35) is intersected with I, and the other theorems mentioned before continue to hold with $\mathfrak{C}_{Het}$ replaced by $\mathfrak{C}$, provided the ranges

and coincides with the bootstrapped test statistic. Consequently, the bootstrap-based test becomes trivial. Its null rejection probabilities are equal to 0 if the p-value is defined as in Davidson and Flachaire (2008). [They are equal to 1 if an alternative definition of the p-value is used.]

[35] Theorem 5.1 maintains Assumption 1. If this assumption is violated, then T_{Het} is identically equal to zero, leading to a useless test. If one would formally apply bootstrap scheme (8), the bootstrapped test statistic T^*_{Het} would then also be identically zero, leading to a bootstrap critical value of zero. The rejection probability is then always equal to zero or always equal to one, depending on whether one uses a strict or weak inequality in the definition of the rejection region.

[36] Theorem 5.7 maintains Assumption 2. If this assumption is violated, then a similar comment as in Footnote 35 applies.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

of the maximum operators appearing in the definitions of the various quantities $\vartheta_{1,Het}$, $\vartheta_{2,Het}$, $\vartheta_{1,uc}$, $\vartheta_{2,uc}$, $\tilde{\vartheta}_{Het}$, $\tilde{\vartheta}_{uc}$, $\tilde{\theta}_{Het}$, and $\tilde{\theta}_{uc}$ are intersected with I.[37]

7.2. Non-Gaussian Errors

Consider now the regression model as in Section 2, except for the Gaussianity assumption.

(i) If we assume a (possibly semiparametric) model for the distribution of the errors such that the implied model for the distributions of $\mathbf{Y}$ contains all the Gaussian distributions shown in (2), then the results of the paper continue to hold a fortiori (with the same lower bounds for α), since the size of any test computed w.r.t. such a larger model for the distributions of $\mathbf{Y}$ is certainly not smaller than the size of the same test when computed w.r.t. to the Gaussian model (2).

(ii) Suppose next we assume that the standardized errors $\sigma^{-1}\Sigma^{-1/2}\mathbf{U}$ follow a (fixed) distribution G that does not depend on (μ,σ,Σ), and let $Q_{\mu,\sigma^2\Sigma,G}$ denote the implied distribution of $\mathbf{Y}$. If G is absolutely continuous w.r.t. $\lambda_{\mathbb{R}^n}$, then the results of the paper continue to hold with $P_{\mu,\sigma^2\Sigma}$ replaced by $Q_{\mu,\sigma^2\Sigma,G}$ (and with the same lower bounds for α). This is easily seen from an inspection of the proofs.[38]

(iii) Suppose we have the same framework as in (ii), except that now G varies in a set $\mathfrak{G}$ (independently of (μ,σ,Σ)), i.e., we have a semiparametric model. If at least one member $G \in \mathfrak{G}$ is absolutely continuous, then the results of the paper continue to hold a fortiori (with the same lower bounds for α) for reasons similar to the ones given in (i).[39]

(iv) Also note that the lower bounds for α in all the results do not involve the distribution of $\mathbf{Y}$ (and thus of $\mathbf{U}$), and, in particular, do not involve the Gaussianity assumption. Hence, in this sense the lower bounds are "distribution free."

7.3. Stochastic Regressors

The assumption of nonstochastic regressors can be easily relaxed as follows: Suppose X is random and $\mathbf{U}$ is conditionally on X distributed as $N(0,\sigma^2\Sigma)$, with $\sigma^2 = \sigma^2(X) > 0$ and $\Sigma = \Sigma(X) \in \mathfrak{C}_{Het}$, where $\sigma^2(\cdot)$ and $\Sigma(\cdot)$ vary in given classes of functions. Suppose further that $\sigma^2(X)$ and $\Sigma(X)$ vary independently through all of $(0,\infty)$ and $\mathfrak{C}_{Het}$, respectively, for (almost) every realization of X, when the functions $\sigma^2(\cdot)$ and $\Sigma(\cdot)$ vary in the before mentioned function classes.[40] Then the

[37] It is here understood that a maximum is interpreted as zero if it extends over an empty range.

[38] The assumption of absolute continuity of G can, in fact, be relaxed to the assumption that none of its one-dimensional marginals has positive mass at zero in case of Theorems A.1, 5.7–5.10, and of the weaker versions of Theorems 5.1 and 5.2 discussed in Remark B.2 in Appendix B. The proofs of Theorems 5.1 and 5.2 also extend to the situation discussed here under any additional condition that guarantees that the exceptional sets B and $\operatorname{span}(X)$, respectively, have probability zero under any $Q_{\mu_0,\sigma^2\Sigma,G}$.

[39] Again, the absolute continuity assumption can be weakened, cf. Footnote 38.

[40] This is certainly the case if no restrictions on the functions $\sigma^2(\cdot)$ and $\Sigma(\cdot)$ are imposed beyond $\sigma^2(\cdot)$ and $\Sigma(\cdot)$, respectively, taking values in $(0,\infty)$ and in the set of diagonal matrices with positive diagonal elements. Another

results of the paper obviously apply after one conditions on X provided (almost) all realizations of X satisfy the assumptions of our theorems, which will typically be the case (for brevity we do not provide a formal statement here). And again similar generalizations to non-Gaussianity as discussed in the preceding subsection are possible here.

8. NUMERICAL RESULTS

There is a considerable body of simulation studies investigating finite sample properties of bootstrap-based heteroskedasticity robust tests, see the references mentioned in Section 1. While these studies provide helpful information, there is—as always with simulation studies—an issue to what extent conclusions of such a study generalize. This is particularly so with *positive* findings (such as, e.g., that a particular bootstrap-based test has null rejection probabilities close to the nominal significance level) as it is less than clear that such a finding allows for generalization beyond the design matrices X, the restrictions (given by R, r), and the forms of heteroskedasticity considered in the simulation study. It is less of an issue with *negative* results (such as, e.g., that a particular bootstrap-based test has null rejection probabilities much larger than the nominal significance level), since they can be viewed as counterexamples disproving good behavior of the bootstrap-based test in general.

For these reasons we set out to study the *worst-case size performance* of a variety of bootstrap-based tests.[41] That is, for any given bootstrap-based test in a large class, we try to "break" the test by searching for a design matrix X, a restriction (given by R, r) to be tested, and a form of heteroskedasticity, such that the null rejection probability of the test is substantially larger than the nominal significance level α. Note that this is equivalent to finding X, R, and r such that the size of the bootstrap-based test computed over the heteroskedasticity model $\mathfrak{C}_{Het}$ is substantially larger than α. A bootstrap-based test that is "broken" in our study, should probably not be used by practitioners (at least not without first assessing its properties in the particular testing problem put before the practitioner, e.g., by attempting to determining the size of the test in that problem by Monte Carlo methods). A bootstrap-based test that "survives" the "stress test" imposed by our study may perhaps be considered to be a better choice, but note that our numerical results do *not* provide any guarantee for good performance in the practitioner's testing problem either (and thus again an assessment of its properties in the practitioner's testing problem may be called for).

Our theoretical results obtained in the previous sections play an important role in our study of the size performance of bootstrap-based tests, as these results allow

---

instance where this is seen to be satisfied is the case where $\sigma^2(X)\Sigma(X) = \mathrm{diag}(\varpi(x_{1\cdot}), \ldots, \varpi(x_{n\cdot}))$ with no further restrictions on the function ϖ (besides positivity) and with at least one regressor being absolutely continuous.

[41] The concept of the size of a test is by itself already a worst-case concept, as it is the supremum of the rejection probabilities over the null hypothesis (where X and the restrictions are being held fixed), cf. (4). The term "worst-case" in "worst-case size performance" here refers to varying X and the restrictions to be tested.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

us to deduce abysmal size behavior (i.e., size equal to 1) of a bootstrap-based test (for a given design matrix X and restriction (R, r) to be tested) by comparing the (numerically evaluated) ϑ with the nominal significance level α; in this section the symbol ϑ serves as a generic abbreviation for $\vartheta_{Het}(=\theta_{Het})$ (cf. Theorem 5.1), $\vartheta_{uc}(=\theta_{uc})$ (cf. Theorem 5.2), $\tilde{\vartheta}_{Het}$ (cf. Theorem 5.7), $\tilde{\theta}_{Het}$ (cf. Theorem 5.9), $\tilde{\vartheta}_{uc}$ (cf. Theorem 5.8), or $\tilde{\theta}_{uc}$ (cf. Theorem 5.10), depending on which of the theorems listed in parentheses (possibly after an appeal to Lemma 5.3 and Remark 5.4) applies to the bootstrap-based test under consideration. Recall that these theorems show that if $\alpha > \vartheta$ holds, then the corresponding bootstrap-based test has size 1 (over the heteroskedasticity model $\mathfrak{C}_{Het}$), and thus certainly breaks down (for the given testing problem, i.e., for the given X, R, r). Hence, we shall search for worst-case X, R, and r that lead to small values of ϑ.[42]

More precisely, we shall consider three *settings*, where *setting* refers to sample size $n = 10, 20$, and 30, and various *scenarios*, where *scenario* refers to a combination of k (number of regressors) and q (number of restrictions to be tested and where $R = (0 : I_q)$, $r = 0$). In every setting, we roughly do the following: we compute for every bootstrap-based test included in our study the value of ϑ for a variety of design matrices in a range of scenarios. We then determine the minimal value of ϑ over all design matrices considered. Then, we compare this minimum with two commonly used levels of significance ($\alpha = 0.05$ and $\alpha = 0.1$). In addition to studying the behavior of this minimal value of ϑ, we shall complement this by numerical size lower bound computations.

All computations were carried out in R (R Core Team, 2020) version 3.6.3 using version 1.0.0 of the R-package **wbsd** ("wild bootstrap size diagnostics") by Preinerstorfer (2020) generated with Rtools35. The package **wbsd** provides computationally efficient routines for determining the quantities $\vartheta_{Het}(=\theta_{Het})$, $\vartheta_{uc}(=\theta_{uc})$, $\tilde{\vartheta}_{Het}$, $\tilde{\theta}_{Het}$, $\tilde{\vartheta}_{uc}$, and $\tilde{\theta}_{uc}$, and for obtaining bootstrap p-values in order to obtain the numerical results reported here. The tools for computing ϑ provided in the R-package **wbsd** can be used by practitioners as a diagnostic device to check whether a bootstrap-based test is provably unreliable (in that $\alpha > \vartheta$, which implies size equal to 1) in a given testing problem. The package is available on CRAN.

In the following subsections, we describe the bootstrap-based tests studied, we explain the computations carried out for each test, and discuss the results obtained. Some of the details are deferred to Appendix E.

8.1. Description of the Bootstrap-Based Tests Studied

The number of bootstrap-based tests we cover in our study is vast: In total, we consider the 960 possible combinations of the 12 test statistics discussed in Section 3 and at the beginning of Section 5.2 around Equations (5), (7), (18), and (20) with

[42]Evaluating ϑ numerically is a nontrivial task as discussed in Section E.3 in Appendix E. In order to be on the safe side and to bias our results in favor of the tests (recall that we are after negative results), the ϑ we shall report will actually be a numerical obtained upper bound for the true ϑ.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

the bootstrap schemes discussed further below (80 in total), which are popular special cases of the two general bootstrap schemes discussed in Section 5.

To be precise, the 12 *test statistics studied* are:[43]

1. **Test statistics based on unrestricted residuals**: T_{uc}; T_{Het} with $d_i = 1$ (HC0); with $d_i = n/(n-k)$ (HC1); with $d_i = (1-h_{ii})^{-1}$ (HC2); with $d_i = (1-h_{ii})^{-2}$ (HC3); and with $d_i = (1-h_{ii})^{\delta_i}$ for $\delta_i = \min(nh_{ii}/k, 4)$ (HC4).
2. **Test statistics based on restricted residuals**: $\tilde{T}_{uc}$; $\tilde{T}_{Het}$ with $\tilde{d}_i = 1$ (HC0R); with $\tilde{d}_i = n/(n-(k-q))$ (HC1R); with $\tilde{d}_i = (1-\tilde{h}_{ii})^{-1}$ (HC2R); with $\tilde{d}_i = (1-\tilde{h}_{ii})^{-2}$ (HC3R); and with $\tilde{d}_i = (1-\tilde{h}_{ii})^{\tilde{\delta}_i}$ for $\tilde{\delta}_i = \min(n\tilde{h}_{ii}/(k-q), 4)$ (HC4R).

The *bootstrap schemes we study* are y^* as defined in (8), and $y^{\maltese}$ as defined in (9). Both bootstrap schemes are applied with $\mathcal{A} = \mathfrak{M}_0$ as well as with $\mathcal{A} = \operatorname{span}(X)$. In addition to choosing $\mathcal{A}$, both bootstrap schemes require a concrete choice of Ξ. All distributions Ξ we consider are constructed in the following way: first an auxiliary distribution $\Xi^\bullet$ on $\{-1,1\}^n$ has to be chosen. The way we choose this auxiliary distribution depends on the magnitude of n. We consider three cases: Setting A ($n = 10$), Setting B ($n = 20$), and Setting C ($n = 30$).

- In Setting A, we consider (i) $\Xi^\bullet$ equal to the n-fold Rademacher distribution (i.e., the n-fold product of the uniform distribution on $\{-1,1\}$), and (ii) $\Xi^\bullet$ equal to the n -fold Mammen distribution (i.e., the n-fold product of the distribution on $\{-(\sqrt{5}-1)/2, (\sqrt{5}+1)/2\}$ that assigns mass $(\sqrt{5}+1)/(2\sqrt{5})$ to $-(\sqrt{5}-1)/2$).
- In Settings B and C, we consider $\Xi^\bullet$ equal to an *empirical* distribution of a sample of size $10n-1$ from the n-fold Rademacher distribution and from the n-fold Mammen distribution, respectively.[44]

Given an auxiliary distribution $\Xi^\bullet$, the distribution Ξ actually used in the bootstrap scheme depends on a vector of weights w, itself typically depending on X or on X and R. Given a weights vector w, Ξ is then obtained as the distribution of $\operatorname{diag}(w)\xi^\bullet$ where $\xi^\bullet$ follows the distribution $\Xi^\bullet$. We consider the following choices for the vector of weights w:[45]

1. **Unrestricted HC0–HC4 weights** $w = (w_1, \ldots, w_n)$**:** $w_i = 1$ (HC0), $w_i = [n/(n-k)]^{1/2}$ (HC1), $w_i = (1-h_{ii})^{-1/2}$ (HC2), $w_i = (1-h_{ii})^{-1}$ (HC3), and $w_i = (1-h_{ii})^{\delta_i/2}$ for $\delta_i = \min(nh_{ii}/k, 4)$ (HC4).
2. **Null-restricted HC0R–HC4R weights** $w = (\tilde{w}_1, \ldots, \tilde{w}_n)$**:** $\tilde{w}_i = 1$ (HC0R), $\tilde{w}_i = [n/(n-(k-q))]^{1/2}$ (HC1R), $\tilde{w}_i = (1-\tilde{h}_{ii})^{-1/2}$ (HC2R), $\tilde{w}_i = (1-\tilde{h}_{ii})^{-1}$ (HC3R), and $\tilde{w}_i = (1-\tilde{h}_{ii})^{\tilde{\delta}_i/2}$ for $\tilde{\delta}_i = \min(n\tilde{h}_{ii}/(k-q), 4)$ (HC4R).

In total this gives 960 possible combinations of test statistics and bootstrap schemes. We emphasize that some of these combinations result in the same

[43] We use here the conventions for d_i and $\tilde{d}_i$ given below (5) and (18), respectively.

[44] The reason for treating Settings B and C differently from Setting A is that for values of n such as 20 or 30 an enumeration of all support points of the n-fold Rademacher or n-fold Mammen distribution is numerically too costly.

[45] We use here the same conventions as mentioned in Footnote 43.

bootstrap-based test: (i) For reasons discussed in Lemma 5.3, (ii) when changing HC0 weights to HC0R weights in the bootstrap scheme, and (iii) when changing HC0 (HC0R) weights to HC1 (HC1R) weights in the definition of T_{Het} ($\tilde{T}_{Het}$), cf. Remark 5.14. Concerning the run-time of the simulations, one could certainly argue that, for the computations, one should keep only one of the combinations that lead to the same bootstrap-based test. However, we have chosen not to, because we can then exploit the ensuing additional computations as a double-check for the methods that "survive" the worst-case analysis (as the design matrices are generated separately for each of the 960 combinations).

Given a test statistic, a bootstrap scheme, and a level of significance α, the corresponding bootstrap-based test is throughout taken as the test that, observing y, rejects the null hypothesis, if the bootstrap p-value computed for y is strictly smaller than α. Here, bootstrap p-value refers to the mass assigned by Ξ to the points ξ that give rise to elements in the bootstrap sample at which the test statistic is greater than or equal to the test statistic evaluated at y. To be precise, if, e.g., T_{Het} is used as a test statistic, we define the bootstrap p-value as $\Xi(\xi : T_{Het}^{\blacktriangle,*}(y,\xi) \geq T_{Het}(y))$ if a bootstrap scheme of the form y^* is used, and as $\Xi(\xi : T_{Het}^{\blacktriangle,\maltese}(y,\xi) \geq T_{Het}(y))$ if a bootstrap scheme of the form $y^{\maltese}$ is used; for the other test statistics T_{uc}, $\tilde{T}_{Het}$, and $\tilde{T}_{uc}$ we proceed similarly. Recall that $T_{Het}^{\blacktriangle}$ coincides with T_{Het}, except on the exceptional set B, on which $T_{Het}^{\blacktriangle}$ is set equal to ∞ (and a similar statement applies for the other test statistics). The reason for using $T_{Het}^{\blacktriangle,*}$ ($T_{Het}^{\blacktriangle,\maltese}$, respectively) rather than T_{Het}^{*} ($T_{Het}^{\maltese}$, respectively) in the definition of the p-value (and similarly for the other test statistics) is that this potentially gives a smaller rejection region, thus biasing the result in favor of the test (recall we are after negative results!); cf. the discussion relating to Part (b) of Theorem 5.1 given subsequent to this theorem. For the same reason we use $\geq$, and not $>$, in the definition of the p-value. It is easy to see that—in case a bootstrap scheme of the form y^* is used—the bootstrap-based test just defined via p-values can be rewritten as the test that rejects if $T_{Het}(y) > f_{Het,1-\alpha}^{\blacktriangle,upper}(y)$, where $f_{Het,1-\alpha}^{\blacktriangle,upper}(y)$ is the *upper* (i.e., largest) $(1-\alpha)$-quantile of $F_{Het,y}^{\blacktriangle}$; and a similar statement applies if a bootstrap scheme of the form $y^{\maltese}$ (or one of the other test statistics) is being used. [This also shows that the above defined rejection region is the smallest among all the rejection regions that can appear in the formulations of the theorems in Section 5.]

8.2. Computations Carried Out in Each Setting

In each setting ($n = 10, 20, 30$) and for each of the 960 combinations of test statistics and bootstrap schemes described above we perform a two-step procedure. A detailed description of the computations carried out can be found in Sections E.1 and E.2 in Appendix E. Here we only provide a brief summary of the

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

two-step procedure to the extent needed for an understanding of the results presented in Section 8.3.

1. The main goal of Step 1 is to find a scenario and a corresponding design matrix leading to a small value of ϑ.

   Essentially, this is done by randomly generating $n \times k$ design matrices (with first column the intercept, and the remaining coordinates i.i.d. log-(standard) normally distributed) and by computing the corresponding values of ϑ for the testing problems $R = (0 : I_q)$ and $r = 0$. In preparation for Step 2, for a suitably chosen subset of the design matrices generated, we also compute null rejection probabilities for strategically chosen variance parameters (assuming normality). All this is done for every pair (k, q) with $k = 2, \dots, 5$ and $q = 1, \dots, k-1$.
2. The goals of Step 2 are twofold: (a) to check the numerical reliability of the computation of ϑ in Step 1; and (b) to compute lower bounds on the size of the test, if necessary. We do the following for $\alpha \in \{0.05, 0.1\}$:

   If $\vartheta_{\min}$, the overall smallest ϑ identified in Step 1, turns out to be smaller than α, we further check the numerical reliability of $\vartheta_{\min}$ by making use of the null rejection probabilities computed in Step 1. If this numerical check, described in Section E.2 in Appendix E, is not passed, we update $\vartheta_{\min}$. Once this check is passed, we distinguish two cases: (i) If $\vartheta_{\min} < \alpha$ or if the maximum of the null rejection probabilities just referred to (maximized over the strategically chosen variance parameters) exceeds 3α, we stop. For these cases, we report the value of $\vartheta_{\min}$ together with the maximal rejection probability obtained for the design matrix pertaining to $\vartheta_{\min}$. (ii) For the exceptional set of tests for which $\vartheta_{\min} \geq \alpha$ and the maximum of the null rejection probabilities does not exceed 3α we perform a second search (again sampling as in Step 1) to find design matrices leading to high rejection probabilities under the null. For these cases we report the highest null rejection probability found, and the value of ϑ corresponding to the design matrix that led to the highest null rejection probability.

8.3. Results and Discussion

The results of the two-step procedure described in Section 8.2 are summarized in Figure 1 in the form of six plots corresponding to the six combinations of the three settings A, B, and C, and of the two values for α ($\alpha \in \{0.05, 0.1\}$). In each plot, the vertical dashed line intersects the axis at α, the lower (upper, respectively) horizontal dashed line intersects the axis at α (at 3α, respectively).

For every combination of the setting and the value of α, the plot is obtained as follows: for every bootstrap-based test procedure (i.e., combination of test statistic and bootstrap scheme), a null rejection probability is plotted against a corresponding ϑ indicated by a black or red circle. The black circles correspond to test procedures for which Step 2 terminated without starting a second set of searches (which was the case for the vast majority of procedures, see Footnote 55 in Appendix E). The red circles correspond to the remaining (exceptional) test procedures for which further null rejection probabilities were computed in Step 2.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

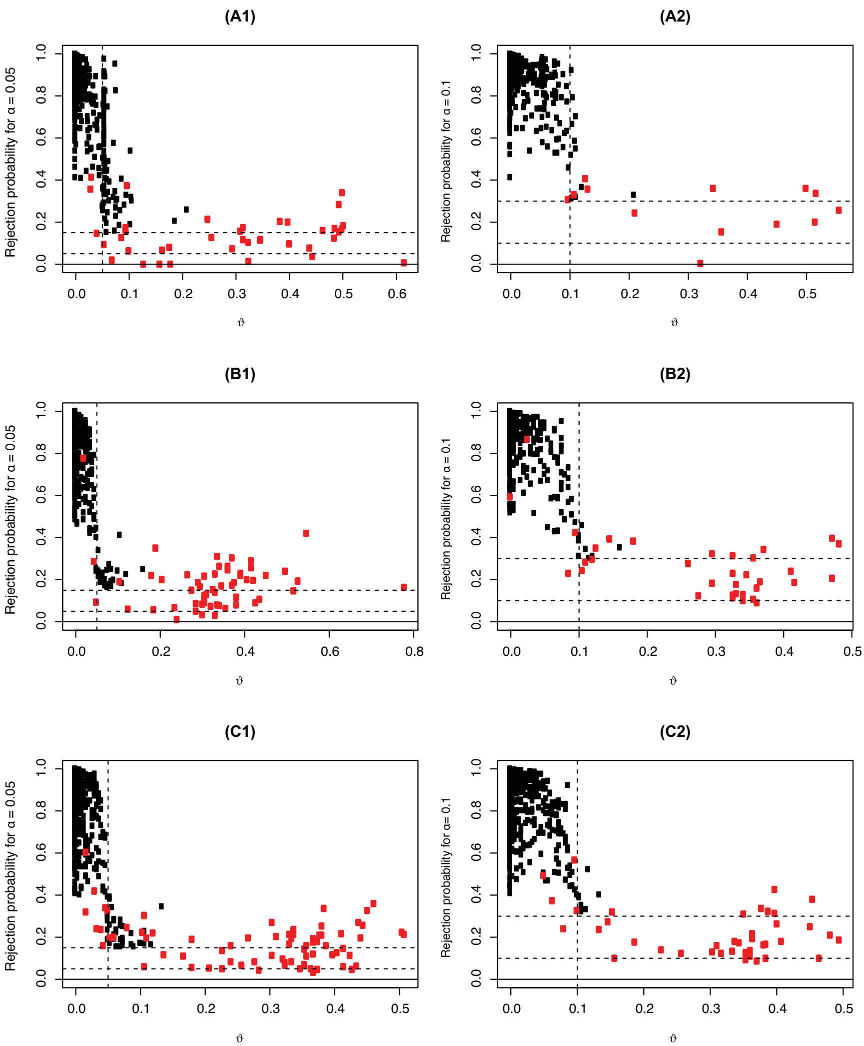

FIGURE 1. Results in Settings A, B, and C.

For all test procedures corresponding to black circles with $\vartheta < \alpha$, the reliability check applied in Step 1 guarantees that the corresponding null rejection probability found in Step 1 is greater than 0.4. Note that these null rejection probabilities do *not* coincide with the sizes of the respective tests, which actually all are equal to 1 by our theoretical results; they are only lower bounds for the sizes that are reported for completeness. For all black circles with $\vartheta \geq \alpha$, the null rejection probabilities are not less than 3α, which can be gathered from an inspection of the plots (and which is so by construction of the two-step procedure, see Section E.2 in Appendix

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

E); hence, they are much too large compared to the nominal significance level α. A bootstrap-based test procedure corresponding to a black circle hence "fails the worst-case check" (in the setting and for the α considered), because a scenario (i.e., k and q) and a corresponding design matrix has been found for which the size of the test is 1, or is exceedingly large, i.e., larger than or equal to 3α.

For the test procedures corresponding to red circles extra computations were carried out in Step 2. A test procedure resulting in a red circle is declared to "fail the worst-case check" (in the setting and for the α considered) if the null rejection probability plotted is not less than 3α (as then a scenario (i.e., k and q) and a corresponding design matrix have been found for which the size of the test is exceedingly large, namely larger than or equal to 3α); otherwise it is declared to "pass the worst-case check" (in the setting and for the α considered) *for the time being*. [There are a few instances of red circles for which the null rejection probability plotted is less than 3α, but $\vartheta < \alpha$ holds. While our theoretical results then tell us that the size of the test should be equal to 1, and we hence should classify the test procedure as "failing the worst-case check," we do not do so as we do not want to rely too much on the information provided by ϑ in such cases, since no reliability check for computing ϑ is included in the computation in Step 2.]

Bootstrap-based test procedures that fail the worst-case check (in at least one setting and for one of the values of α) should thus not be expected to be reliable in general. Therefore, such a test procedure should not be used in practice without first obtaining further guarantees concerning its size properties in the specific problem at hand, e.g., by running additional simulations geared towards the problem at hand.

Figure 1 shows that in all settings and for both significance levels considered, the vast majority of circles are black (see Footnote 55 in Appendix E), already leading to the conclusion that most bootstrap-based test procedures fail the worst-case check. Furthermore, most of these test procedures fail in such a way that the corresponding ϑ is smaller than the significance level α, and thus the test is known to have size equal to 1 (at least in one of the scenarios and for one of the design matrices considered) as a consequence of our theoretical results. Inspection of Figure 1 also shows that a good portion of the test procedures corresponding to red circles fail the worst-case check in that the red circles are above or on the 3α-line. Comparing the figures across different settings and values of α, we also see that the number of red circles decreases when passing from $\alpha = 0.05$ to $\alpha = 0.1$. The figure also shows that the number of red circles increases when increasing n. One reason could be that the randomized search for design matrices leading to low values of ϑ becomes more difficult as n increases. We used the same randomized search algorithm in all three scenarios, which could explain the difference. A conceptual difference between the methods used in Setting A and Settings B and C is that for Setting A ($n = 10$) exact computations (concerning $\Xi^\bullet$) are carried out, while for Settings B ($n = 20$) and C ($n = 30$) approximate computations (based on empirical distributions $\Xi^\bullet$) are done. This introduces an additional source of variation in the

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

computations in Settings B and C, which could also be responsible for the increase in the number of red circles.

Important questions now are whether (i) there is a bootstrap-based test procedure left that passes the worst-case check in all settings and for both significance levels considered; (ii) there is a test procedure that passes the worst-case check in all settings considered for a fixed α; and (iii) there is a pattern, in the sense that certain combinations of test statistics and bootstrap schemes often pass the worst-case check? To answer such questions, we shall next provide information on the procedures that pass the worst-case check in each setting and for each α considered.

The bootstrap-based test procedures which (for the time being) pass the worst-case check in Settings A–C, respectively, are summarized in Tables 1–3. In each table, the first row contains the test procedures that pass the worst-case check for $\alpha = 0.05$, the second row the ones that pass for $\alpha = 0.1$, and the third row the ones that pass at both nominal levels of significance.

In these tables, to facilitate the exposition, we use the following way of encoding a bootstrap-based test procedure: to each of the 960 possible procedures (i.e., combinations of test statistics and bootstrap schemes) considered we associate a 7 digit code: $x = x_1{:}x_2{:}x_3{:}x_4{:}x_5{:}x_6{:}x_7$. The encoding of the digits is as follows:

- $x_1 \ldots$ indicates which covariance matrix estimator is used in the test statistic ("-1" stands for the uncorrected estimator based on restricted or unrestricted residuals depending on whether x_2 is set to F or T; and $0,\ldots,4$ stand for HC0,...,HC4, respectively, or for HC0R,...,HC4R, respectively, depending on whether x_2 is set to F or T).
- $x_2 \ldots$ indicates whether the covariance matrix estimator used in the test statistic is based on null-restricted residuals (T) or not (F).
- $x_3 \ldots$ indicates the distribution underlying $\Xi^{\bullet}$ ("r" stands for Rademacher, "m" stands for Mammen).
- $x_4 \ldots$ indicates the weights w used in constructing Ξ from $\Xi^{\bullet}$ ($0,\ldots,4$ stands for HC0,...,HC4, respectively, or for HC0R,...,HC4R, respectively, depending on whether x_5 is set to F or T).
- $x_5 \ldots$ indicates whether the weights w are null-restricted (T) or not (F).
- $x_6 \ldots$ indicates whether the bootstrap-scheme was based on null-restricted residuals (T) (i.e., $\mathcal{A} = \mathfrak{M}_0$) or not (F) (i.e., $\mathcal{A} = \operatorname{span}(X)$).
- $x_7 \ldots$ indicates whether the bootstrap scheme y^* (T) or $y^{\maltese}$ (F) was used.

As an example, the code "-1:T:m:2:F:T:F" translates to the bootstrap-based test procedure which uses the test statistic $\tilde{T}_{uc}$ (determined by the first two digits of the code) and the following bootstrap scheme: $\Xi^{\bullet}$ based on the Mammen distribution, modified by HC2 weights (based on unrestricted residuals), and $y^{\maltese}$ with $\mathcal{A} = \mathfrak{M}_0$.

Before interpreting the results, we also need to explain why some test procedures are struck out in Tables 1–3. Recall (e.g., from the discussion in the last but one paragraph of Section 8.1) that some of the 960 procedures studied are in fact *equivalent*, meaning that they lead to exactly the same bootstrap-based test. Because the design matrices are generated anew for each of the 960 procedures, the

TABLE 1. Surviving Test Procedures in Setting A.

$\alpha = 0.05$	−1:T:m:0:F:T:F	−1:T:m:0:T:T:F	−1:T:m:1:F:T:F	−1:T:m:1:T:T:F
	−1:T:m:2:F:T:F	−1:T:m:2:T:T:F	−1:T:m:3:F:T:F	−1:T:m:3:T:T:F
	~~3:F:r:0:F:T:T~~	~~3:F:m:0:T:T:F~~	~~3:F:m:1:F:T:F~~	~~3:F:m:1:T:T:F~~
	3:F:m:2:F:T:F	3:F:m:2:F:T:T	3:T:r:2:T:T:T	3:T:r:3:F:T:F
	3:T:r:3:T:T:T	~~3:T:m:0:F:T:T~~	3:T:m:1:T:T:T	3:T:m:2:T:T:T
	3:T:m:3:F:F:F	~~4:T:m:0:T:T:T~~		
$\alpha = 0.1$	−1:T:m:2:F:T:F	−1:T:m:3:F:T:F	~~3:F:r:2:F:T:F~~	3:F:m:2:F:T:F
	3:F:m:2:F:T:T	~~4:F:m:2:F:T:F~~		
Both	−1:T:m:2:F:T:F	−1:T:m:3:F:T:F	3:F:m:2:F:T:F	3:F:m:2:F:T:T

TABLE 2. Surviving Test Procedures in Setting B.

$\alpha = 0.05$	−1:T:r:3:F:T:F	~~1:F:m:3:F:F:F~~	~~2:F:r:3:F:F:F~~	~~2:F:m:0:T:T:F~~
	~~3:F:m:0:F:T:F~~	~~3:F:m:2:F:T:F~~	~~3:F:m:3:F:F:F~~	~~3:T:r:0:T:F:T~~
	~~3:T:r:0:T:T:T~~	3:T:r:1:T:F:T	3:T:r:1:T:T:T	3:T:r:2:T:F:T
	3:T:r:3:F:T:F	3:T:m:0:F:F:T	3:T:m:0:F:T:T	3:T:m:0:T:F:T
	3:T:m:0:T:T:T	3:T:m:1:F:F:T	3:T:m:1:F:T:T	3:T:m:1:T:F:T
	3:T:m:1:T:T:T	3:T:m:2:F:F:T	3:T:m:2:T:F:T	3:T:m:2:T:T:T
	~~4:T:m:0:T:T:T~~			
$\alpha = 0.1$	−1:T:r:3:F:F:F	−1:T:r:3:F:T:F	~~1:F:r:3:F:F:F~~	~~1:F:m:3:F:F:F~~
	~~3:F:m:0:T:T:T~~	3:F:m:2:F:T:F	3:F:m:2:F:T:T	~~3:F:m:3:F:T:T~~
	~~3:T:r:0:T:T:T~~	3:T:r:2:F:F:T	3:T:r:2:F:T:T	3:T:r:2:T:F:T
	3:T:r:3:F:F:F	~~3:T:m:0:F:F:T~~	3:T:m:0:F:T:T	3:T:m:0:T:T:T
	3:T:m:1:F:F:T	3:T:m:1:F:T:T	3:T:m:1:T:F:T	3:T:m:1:T:T:T
	3:T:m:2:F:F:T	3:T:m:2:T:T:T		
Both	−1:T:r:3:F:T:F	~~1:F:m:3:F:F:F~~	~~3:F:m:2:F:T:F~~	~~3:T:r:0:T:T:T~~
	3:T:r:2:T:F:T	~~3:T:m:0:F:F:T~~	3:T:m:0:F:T:T	3:T:m:0:T:T:T
	3:T:m:1:F:F:T	3:T:m:1:F:T:T	3:T:m:1:T:F:T	3:T:m:1:T:T:T
	3:T:m:2:F:F:T	3:T:m:2:T:T:T		

Table 3. Surviving Test Procedures in Setting C.

$\alpha = 0.05$	~~0:F:m:3:F:F:F~~	~~1:F:m:3:F:F:T~~	~~2:F:m:0:T:T:F~~	~~2:F:m:3:T:T:T~~
	~~3:F:r:0:T:T:T~~	~~3:F:r:1:T:T:F~~	3:F:r:2:F:T:F	3:F:r:2:F:T:T
	~~3:F:m:0:F:T:T~~	~~3:F:m:1:F:T:T~~	3:F:m:2:F:T:F	3:F:m:2:F:T:T
	~~3:F:m:2:T:T:F~~	~~3:T:r:0:T:F:T~~	~~3:T:r:0:T:T:T~~	3:T:r:2:F:F:T
	3:T:r:2:T:T:T	3:T:m:0:F:F:T	3:T:m:0:F:T:T	3:T:m:0:T:F:T
	3:T:m:0:T:T:T	3:T:m:1:F:F:T	3:T:m:1:F:T:T	3:T:m:1:T:F:T
	3:T:m:1:T:T:T	3:T:m:2:F:F:T	3:T:m:2:T:F:T	3:T:m:2:T:T:T
	~~4:T:r:0:F:T:T~~	4:T:m:0:F:T:T	4:T:m:0:T:T:T	
$\alpha = 0.1$	−1:T:r:3:F:F:F	~~0:F:r:3:F:F:F~~	~~0:F:m:3:F:F:F~~	~~1:F:r:3:F:F:F~~
	~~1:F:m:3:F:T:F~~	~~2:F:m:3:F:T:T~~	~~2:F:m:3:T:T:T~~	~~3:F:r:2:F:T:F~~
	~~3:F:r:2:T:T:F~~	~~3:F:m:1:T:T:T~~	~~3:F:m:2:F:T:F~~	3:T:r:0:F:F:T
	3:T:r:0:T:F:T	~~3:T:r:0:T:T:T~~	3:T:r:1:F:F:T	3:T:r:1:T:T:T
	3:T:r:2:F:F:T	3:T:m:0:F:F:T	3:T:m:0:F:T:T	3:T:m:0:T:F:T
	3:T:m:0:T:T:T	3:T:m:1:F:F:T	3:T:m:1:T:F:T	3:T:m:1:T:T:T
	3:T:m:2:F:F:T	3:T:m:2:T:F:T	3:T:m:2:T:T:T	3:T:m:3:F:F:T
	3:T:m:3:T:F:T			
Both	~~0:F:m:3:F:F:F~~	~~2:F:m:3:T:T:T~~	~~3:F:r:2:F:T:F~~	~~3:F:m:2:F:T:F~~
	~~3:T:r:0:T:F:T~~	~~3:T:r:0:T:T:T~~	3:T:r:2:F:F:T	3:T:m:0:F:F:T
	3:T:m:0:F:T:T	3:T:m:0:T:F:T	3:T:m:0:T:T:T	3:T:m:1:F:F:T
	3:T:m:1:T:F:T	3:T:m:1:T:T:T	3:T:m:2:F:F:T	3:T:m:2:T:F:T
	3:T:m:2:T:T:T			

results found for equivalent procedures can be different. Therefore, it can happen that a procedure passes the worst-case check, but an equivalent version of this procedure does not. Now, a procedure is struck out in a given table, if—while this procedure passed the worst-case check underlying the table—an *equivalent* procedure did not (and thus does not appear in this table). Procedures that are struck out in a table are now *no longer* considered as having passed the worst-case check (although they appear in that table).

We exploit the following three reasons for equivalence between test procedures: (1) Lemma 5.3 shows that a bootstrap-based test using a covariance matrix estimator based on unrestricted residuals does not depend on whether y^* or $y^{\maltese}$ is used as a bootstrap scheme. Therefore, two procedures with codes x and x' are equivalent in case $x_2 = x_2' = F$ and $x_i = x_i'$ for $i = 1, \ldots, 6$. (2) Changing the weights vector from HC0 to HC0R (or vice versa) in the construction of Ξ does not change the bootstrap-based test. Therefore, two procedures with codes x and x' are equivalent in case $x_4 = x_4' = 0$, and $x_i = x_i'$ for all $i \neq 5$. (3) Changing HC0 (HC0R) weights to HC1 (HC1R) weights in the definition of T_{Het} ($\tilde{T}_{Het}$) also does not change the resulting bootstrap-based test, cf. Remark 5.14. Therefore, two procedures with codes x and x' are equivalent in case $x_1 \in \{0,1\}$, $x_1' \in \{0,1\}$, and $x_i = x_i'$ for all $i = 2, \ldots, 7$. Procedures x that are struck out by a slash, i.e., $\cancel{x}$, were eliminated based on reason (1); procedures that are struck out by a backslash, i.e., $\bcancel{x}$, were eliminated based on reason (2); and procedures that are struck out by a horizontal line, i.e., ~~x~~, were eliminated based on reason (3). Note that a procedure can be struck out, e.g., based on reasons (1) and (2), and thus is then crossed out, i.e., is marked by $\xcancel{x}$, etc.

Concerning the questions (i)–(iii) raised above, inspection of Tables 1–3 now delivers the following answers:

1. There is no bootstrap-based test procedure that passes the worst-case check in all settings and for both significance levels.
2. The tests 3:T:m:1:T:T:T and 3:T:m:2:T:T:T pass the worst-case checks in all three settings for the significance level $\alpha = 0.05$. The corresponding rejection probabilities shown in Figure 1 for the tests 3:T:m:1:T:T:T and 3:T:m:2:T:T:T are 0.077 and 0.067 (Setting A), 0.033 and 0.080 (Setting B), 0.050 and 0.033 (Setting C), respectively. For $\alpha = 0.1$ there is no test that passes the worst-case checks in all three settings.
3. The majority of test procedures appearing in the tables is based on the *HC3 or HC3R covariance estimator*, and uses a bootstrap scheme based on the *Mammen-distribution*. In Settings B and C, the tests passing the worst-case check are typically based on $\tilde{T}_{Het}$, i.e., they use *restricted residuals in the construction of the covariance estimator*.

On the one hand our results issue a distinct warning: overall *none* of the bootstrap-based tests considered comes with a guarantee that its size is (about) right. On the other hand, when restricting attention only to $\alpha = 0.05$, the tests 3:T:m:1:T:T:T and 3:T:m:2:T:T:T did not break down in our worst-case

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

analysis.[46] These two tests are based on $\tilde{T}_{Het}$ using a HC3R covariance estimator, and use the Mammen-distribution in the bootstrap scheme; properties that are common to many of the tests that pass the check (in one of the settings considered). While this obviously does not prove that 3:T:m:1:T:T:T and 3:T:m:2:T:T:T always will have perfect size properties for $\alpha = 0.05$, it shows that in the settings considered (and for $\alpha = 0.05$) they seem to have the best size performance among all bootstrap-based tests considered, and should therefore perhaps be preferred (over the other procedures) by practitioners who insist on applying a bootstrap-based test. However one should keep in mind that, while we have examined a considerable and reasonable range of scenarios and design matrices, the size-behavior of the bootstrap-based tests outside of the range studied can potentially be even worse.

In light of the findings above a better way forward seems to use heteroskedasticity robust test procedures that guarantee size-control as expounded in Pötscher and Preinerstorfer (2021).

9. CONCLUSION

Bootstrap-based heteroskedasticity robust tests have been suggested in the literature to ameliorate overrejection problems often arising with heteroskedasticity robust tests based on standard critical values (derived from asymptotic theory). While there is Monte Carlo evidence suggesting that the bootstrap can attenuate this overrejection problem, the question arises whether this observation generalizes beyond the specific Monte Carlo settings considered. In the present paper, we establish sufficient conditions under which bootstrap-based tests can be shown to "break down" in the sense that their size equals one. This theoretical insight can be used to check whether a given bootstrap-based test procedure should not be used in a given problem. Furthermore, the results allow us to conduct a numerical "stress test" on a wide variety of existing bootstrap-based test procedures, leading to the conclusion that none of these tests is immune to considerable overrejection. Thus any such bootstrap-based test is no reliable panacea for heteroskedasticity robust testing. An alternative to bootstrap-based procedures is to use smallest size-controlling critical values as studied in Pötscher and Preinerstorfer (2021).

APPENDIX

A. A Basic Theorem

THEOREM A.1. *Let* $T_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *be Borel-measurable and* $G(\mathfrak{M}_0)$*-invariant for* $i = 1, 2$. *Let* Ξ *be a (Borel) probability measure on* $\mathbb{R}^n$, *and let* $\mathcal{A}$ *be an affine subspace of* $\mathbb{R}^n$ *with* $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \operatorname{span}(X)$. *Define the function* $T_2^* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *via*

$$T_2^*(y, \xi) = T_2\left(y^*(y, \xi)\right), \tag{33}$$

[46] However, recall that we have only computed a lower bound for the size.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

where $y^*(y,\xi)$ has been defined in (8), and set

$K(y) = \{\xi \in \mathbb{R}^n : y^*(y,\xi)$ *is a continuity point of* $T_2\}$.

For every $y \in \mathbb{R}^n$ denote by F_y the distribution function of $\Xi \circ T_2^*(y,\cdot)$, i.e., $F_y(t) = \Xi(\{\xi : T_2^*(y,\xi) \le t\})$ for $t \in \mathbb{R} \cup \{\infty\}$. For every $\alpha \in (0,1)$, let $f_{1-\alpha}(y)$ denote a $(1-\alpha)$ -quantile of F_y. For every $i = 1,\ldots,n$ define

$$c_i = \sup_{\delta>0} \inf_{z\in B(e_i(n),\delta)} T_1(\mu_0 + z) \tag{34}$$

for some $\mu_0 \in \mathfrak{M}_0$, where $B(e_i(n),\delta) = \{z \in \mathbb{R}^n : \|z - e_i(n)\| < \delta\}$ (note that c_i does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$). Then, for every $\alpha \in (0,1)$ such that

$$\alpha > 1 - \max_{i=1,\ldots,n} \Xi\left(\left\{\xi : T_2^*(\mu_0 + e_i(n),\xi) < c_i, \xi \in K(\mu_0 + e_i(n))\right\}\right) \tag{35}$$

for some (and hence all) $\mu_0 \in \mathfrak{M}_0$, we have

$$\sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_1 \ge f_{1-\alpha}\right) \ge \sup_{\Sigma\in\mathfrak{C}_{Het}} P_{\mu_0,\sigma^2\Sigma}\left(T_1 > f_{1-\alpha}\right) = 1 \tag{36}$$

for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$ (where the probabilities in (36) are to be interpreted as inner probabilities[47]).[48]

Proof. By $G(\mathfrak{M}_0)$-invariance of T_i, for every $z \in \mathbb{R}^n$ the expression $T_i(\mu_0 + \gamma z)$ depends neither on the choice of $\mu_0 \in \mathfrak{M}_0$ nor on the value of $\gamma \in \mathbb{R}\backslash\{0\}$ (for $i = 1,2$).[49] This shows, in particular, that the c_i's do not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. Furthermore, since $\mathfrak{M}_0 \subseteq \mathcal{A}$ has been assumed, it is easy to see that $\mathcal{A}-\mu_0$ is a linear space containing $\mathfrak{M}_0^{lin}$ for every choice of $\mu_0 \in \mathfrak{M}_0$ (with $\mathcal{A}-\mu_0$ being the same space regardless of the choice of $\mu_0 \in \mathfrak{M}_0$), and that $y - X\tilde{\beta}_{\mathcal{A}}(y) = \Pi_{(\mathcal{A}-\mu_0)^\perp}(y-\mu_0)$ holds for every $y \in \mathbb{R}^n$ and for every choice of $\mu_0 \in \mathfrak{M}_0$. It now easily follows that $y^*(\mu_0 + y,\xi) = y^*(\mu_0' + y,\xi) - \mu_0' + \mu_0$ for every $\mu_0, \mu_0' \in \mathfrak{M}_0$. In view of $G(\mathfrak{M}_0)$-invariance of T_2 we can conclude that the right-hand side of (35) also does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. For later use we also make the following observation: Since $X\tilde{\beta}_{\mathfrak{M}_0}(y)$ obviously belongs to $\mathfrak{M}_0$, we have for every $y \in \mathbb{R}^n$, every $\xi \in \mathbb{R}^n$, and every $\mu_0 \in \mathfrak{M}_0$

$$\begin{aligned} T_2^*(y,\xi) &= T_2(X\tilde{\beta}_{\mathfrak{M}_0}(y) + \operatorname{diag}(\xi)(y - X\tilde{\beta}_{\mathcal{A}}(y)) = T_2(\mu_0 + \operatorname{diag}(\xi)(y - X\tilde{\beta}_{\mathcal{A}}(y)) \\ &= T_2(\mu_0 + \operatorname{diag}(\xi)\Pi_{(\mathcal{A}-\mu_0)^\perp}(y-\mu_0)), \end{aligned} \tag{37}$$

in view of what has been said at the beginning of the proof.

Now, denote the set of continuity points of T_2 by K_2, and define $\bar{T}_2 := T_2\mathbf{1}_{K_2} + \infty\mathbf{1}_{\mathbb{R}^n\backslash K_2}$ (with the convention $\infty\cdot 0 = 0$). Define $\bar{T}_2^*$ analogously to T_2^* (cf. (33)), but with T_2 replaced by $\bar{T}_2$. For every $y \in \mathbb{R}^n$, we let $\bar{F}_y$ denote the distribution function of $\Xi \circ \bar{T}_2^*(y,\cdot)$, i.e., $\bar{F}_y(t) = \Xi(\{\xi : \bar{T}_2^*(y,\xi) \le t\})$ for $t \in \mathbb{R}\cup\{\infty\}$. It easily follows that for every $y \in \mathbb{R}^n$ and every $t \in \mathbb{R}$ we have $\bar{F}_y(t) = \Xi(\{\xi : T_2^*(y,\xi) \le t, \xi \in K(y)\})$. Consequently, for every $y \in \mathbb{R}^n$

[47] This allows one to ignore measurability issues regarding $f_{1-\alpha}$.

[48] The set appearing in (35) is a Borel set since $T_2^*(y,\cdot)$ is Borel measurable and since $K(y)$ is a Borel set.

[49] Let $\mu_0' \in \mathfrak{M}_0$ and $\gamma' \in \mathbb{R}\backslash\{0\}$ be arbitrary. By $G(\mathfrak{M}_0)$-invariance, $T_i(\mu_0 + \gamma z) = T_i(h(\mu_0 + \gamma z))$ where $h(v) = \gamma'\gamma^{-1}(v - \mu_0) + \mu_0' \in G(\mathfrak{M}_0)$. This shows that $T_i(\mu_0 + \gamma z) = T_i(\mu_0' + \gamma' z)$.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

and every $t \in \mathbb{R} \cup \{\infty\}$, we have $\bar{F}_y(t-) = \Xi(\{\xi : T_2^*(y,\xi) < t, \xi \in K(y)\})$, where $\bar{F}_y(t-)$ denotes the left-hand side limit of $\bar{F}_y$ at t. In particular, (35) is equivalent to

$$\max_{i=1,\ldots,n} \bar{F}_{\mu_0+e_i(n)}(c_i-) > 1-\alpha$$

(with the convention that $\bar{F}_{\mu_0+e_i(n)}(c_i-) = 0$ in case $c_i = -\infty$). From now on, let i be an index that realizes the maximum in the previous display. If $\bar{F}_{\mu_0+e_i(n)}(c_i-) = 0$, there is nothing to prove and we are done. Hence, it remains to consider the case where $\bar{F}_{\mu_0+e_i(n)}(c_i-) > 0$. Note that this implies $c_i > -\infty$. In this case, let $\alpha \in (0,1)$ now be such that $\bar{F}_{\mu_0+e_i(n)}(c_i-) > 1-\alpha$ (where $\mu_0 \in \mathfrak{M}_0$ can be chosen arbitrarily). From $\bar{F}_{\mu_0+e_i(n)}(c_i-) > 1-\alpha$ we can then conclude existence of a real number x_i smaller than c_i (recall $c_i > -\infty$ must hold) such that $\bar{F}_{\mu_0+e_i(n)}(x_i) > 1-\alpha$ holds and such that x_i is a continuity point of $\bar{F}_{\mu_0+e_i(n)}$. In view of (34), there exists a $\delta > 0$ such that every $z \in B(e_i(n),\delta)$ satisfies $T_1(\mu_0+z) > x_i$ (and the same is true if we replace δ by a smaller positive number). We claim that for every sequence $z_m \to e_i(n)$ ($z_m \in \mathbb{R}^n$) we have

$$\liminf_{m\to\infty} F_{\mu_0+z_m}(x_i) \geq \bar{F}_{\mu_0+e_i(n)}(x_i).$$

Define $V_m = V_m(\xi) = y^*(\mu_0+z_m,\xi)$ and $V = V(\xi) = y^*(\mu_0+e_i(n),\xi)$, which can be viewed as random vectors defined on $\mathbb{R}^n$ (equipped with the Borel σ-field) and where the probability measure is given by Ξ. Note that V_m converges to V everywhere as $m \to \infty$ (as $y^*(y,\xi)$ is continuous w.r.t. y). Furthermore, $F_{\mu_0+z_m}(x_i) = \Xi(T_2(V_m) \leq x_i)$ and $\bar{F}_{\mu_0+e_i(n)}(x_i) = \Xi(\bar{T}_2(V) \leq x_i)$ hold (recall that x_i is a real number). The statement in the previous display now follows from Lemma A.3, recalling that we have chosen x_i as a continuity point of $\bar{F}_{\mu_0+e_i(n)}$, which implies $\Xi(\bar{T}_2(V) = x_i) = 0$. Summarizing, we hence arrive, replacing δ by another element of $(0,\delta)$ if necessary, at

$$T_1(\mu_0+z) > x_i \text{ and } F_{\mu_0+z}(x_i) > 1-\alpha \text{ for every } z \in B(e_i(n),\delta). \tag{38}$$

From (37) and the observation in the first sentence in this proof it readily follows that for every $z \in \mathbb{R}^n$ and every $\gamma \neq 0$ we have

$$\begin{aligned} F_{\mu_0+\gamma z}(x_i) &= \Xi(\xi : T_2^*(\mu_0+\gamma z,\xi) \leq x_i) \\ &= \Xi(\xi : T_2(\mu_0+\operatorname{diag}(\xi)\Pi_{(\mathcal{A}-\mu_0)^\perp}(\mu_0+\gamma z-\mu_0)) \leq x_i) \\ &= \Xi(\xi : T_2(\mu_0+\gamma\operatorname{diag}(\xi)\Pi_{(\mathcal{A}-\mu_0)^\perp}(z)) \leq x_i) \\ &= \Xi(\xi : T_2(\mu_0+\operatorname{diag}(\xi)\Pi_{(\mathcal{A}-\mu_0)^\perp}(z)) \leq x_i) = F_{\mu_0+z}(x_i). \end{aligned}$$

Using the observation in the first sentence of the proof again, the preceding display and (38) thus give

$$T_1(\mu_0+y) > x_i \text{ and } F_{\mu_0+y}(x_i) > 1-\alpha \text{ for every } y \in \{\gamma z : \gamma \neq 0, z \in B(e_i(n),\delta)\} =: \Delta,$$

or equivalently

$$T_1(y) > x_i \text{ and } F_y(x_i) > 1-\alpha \text{ for every } y \in \mu_0+\Delta. \tag{39}$$

By Lemma A.4, we see that $F_y(x_i) > 1-\alpha$ implies $x_i \geq f_{1-\alpha}(y)$. Hence,

$$\left\{y \in \mathbb{R}^n : T_1(y) > x_i,\ F_y(x_i) > 1-\alpha\right\} \subseteq \left\{y \in \mathbb{R}^n : T_1(y) > f_{1-\alpha}(y)\right\}.$$

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

This, together with (39), gives

$$\left\{y \in \mathbb{R}^n : T_1(y) > f_{1-\alpha}(y)\right\} \supseteq \mu_0 + \Delta \supseteq \mu_0 + \operatorname{int}(\Delta) \supseteq \mu_0 + (\operatorname{span}(e_i(n)) \backslash \{0\}), \tag{40}$$

where $\operatorname{int}(\Delta)$ denotes the interior of Δ. Finally, to complete the proof of (36), let Σ_m be a sequence in $\mathfrak{C}_{Het}$ that converges to $e_i(n)e_i(n)'$ and let $0 < \sigma^2 < \infty$. Lemma E.1 in Preinerstorfer and Pötscher (2016) then shows that $P_{\mu_0, \sigma^2 \Sigma_m}$ converges weakly to $P_{\mu_0, \sigma^2 e_i(n)e_i(n)'}$. The previous display implies that for every $m \in \mathbb{N}$

$$P_{\mu_0, \sigma^2 \Sigma_m}\left(\left\{y \in \mathbb{R}^n : T_1(y) > f_{1-\alpha}(y)\right\}\right) \geq P_{\mu_0, \sigma^2 \Sigma_m}(\mu_0 + \operatorname{int}(\Delta)),$$

where the left-hand side is to be interpreted as an inner probability. The Portmanteau theorem delivers

$$\begin{aligned}\liminf_{m\to\infty} P_{\mu_0, \sigma^2 \Sigma_m}(\mu_0 + \operatorname{int}(\Delta)) &\geq P_{\mu_0, \sigma^2 e_i(n)e_i(n)'}(\mu_0 + \operatorname{int}(\Delta)) \\ &\geq P_{\mu_0, \sigma^2 e_i(n)e_i(n)'}(\mu_0 + (\operatorname{span}(e_i(n)) \backslash \{0\})) = 1,\end{aligned}$$

where the second inequality follows from (40), and the final equality from $P_{\mu_0, \sigma^2 e_i(n)e_i(n)'}$ being supported by $\mu_0 + \operatorname{span}(e_i(n))$ and assigning probability zero to $\{\mu_0\}$. This establishes (36). $\square$

Remark A.2. (i) If T_1 is lower semi-continuous at $\mu_0 + e_i(n)$ for some (and hence all) $\mu_0 \in \mathfrak{M}_0$, then $c_i = T_1(\mu_0 + e_i(n))$.

(ii) If the set $K(\mu_0 + e_i(n))$ has Ξ-measure 1, then $\Xi\left(\xi : T_2^*(\mu_0 + e_i(n), \xi) < c_i, \xi \in K(\mu_0 + e_i(n))\right)$ reduces to $F_{\mu_0 + e_i(n)}(c_i-)$.

(iii) The lower bound for α in (35) depends only on observable quantities and thus can be computed.

LEMMA A.3. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space, and let* V_m *be a sequence of* $\mathbb{R}^p$*-valued random vectors* ($p \in \mathbb{N}$) *defined on that space that converges almost everywhere to the random vector* V. *Let* $T : \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ *be Borel measurable, denote the set of continuity points of* T *by* K_T, *and define* $\bar{T} := T\mathbf{1}_{K_T} + \infty \mathbf{1}_{\mathbb{R}^p \backslash K_T}$ *(with the convention* $\infty \cdot 0 = 0$*). Then, for every* $t \in \mathbb{R}$ *such that* $\mathbb{P}(\bar{T}(V) = t) = 0$ *as well as for* $t = \infty$*, it holds that*

$$\mathbb{P}\left(\bar{T}(V) \leq t\right) \leq \liminf_{m\to\infty} \mathbb{P}(T(V_m) \leq t). \tag{41}$$

Proof. It is well known that K_T is a countable intersection of open sets, and hence is a Borel set. As a consequence $\bar{T} : \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ is Borel measurable. For $t = \infty$ the probabilities in (41) are all equal to 1, and the inequality therefore trivially holds in this case. Next, let $t \in \mathbb{R}$ be such that $\mathbb{P}(\bar{T}(V) = t) = 0$. Set $A_m := \{\omega \in \Omega : T(V_m(\omega)) \leq t\}$, $A := \{\omega \in \Omega : \bar{T}(V(\omega)) \leq t\}$ and $\Omega' = \{\omega \in \Omega : \bar{T}(V(\omega)) \neq t, V_m(\omega) \to V(\omega)\}$. From the definition of $\bar{T}$ we obtain

$$\limsup_{m\to\infty} T(V_m(\omega)) \leq \bar{T}(V(\omega)) \text{ for every } \omega \in \Omega \text{ such that } V_m(\omega) \to V(\omega). \tag{42}$$

Furthermore, if $\omega \in A \cap \Omega'$, we have that $\bar{T}(V(\omega)) < t$ must hold and hence $T(V_m(\omega)) < t$ must hold eventually in view of (42), i.e., that $\omega \in A_m$ eventually must be true. This implies

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

that $\liminf_{m\to\infty} \mathbf{1}_{A_m}(\omega) \geq \mathbf{1}_A(\omega)$ for every $\omega \in \Omega'$ (the inequality being trivial for $\omega \notin A$), and thus almost everywhere. Fatou's lemma now yields

$$\liminf_{m\to\infty} \mathbb{P}(T(V_m) \leq t) = \liminf_{m\to\infty} \mathbb{E}(\mathbf{1}_{A_m}) \geq \mathbb{E}(\liminf_{m\to\infty} \mathbf{1}_{A_m}) \geq \mathbb{E}(\mathbf{1}_A) = \mathbb{P}(\bar{T}(V) \leq t),$$

where $\mathbb{E}$ denotes the expectation operator associated with $\mathbb{P}$. □

LEMMA A.4. *Let F be the distribution function of a random variable taking values in $\mathbb{R} \cup \{\infty\}$ and let $\delta \in (0,1)$. If $s \in \mathbb{R} \cup \{\infty\}$ satisfies $F(s) > \delta$, then s is larger than or equal to any δ-quantile of F.*

Proof. If $r \in \mathbb{R} \cup \{\infty\}$ satisfies $r > s$, it follows that $F(s) \leq F(r-)$ must hold. Together with the hypothesis, we conclude that $F(r-) > \delta$. But this shows that r can not be a δ-quantile of F, cf. Footnote 23. □

B. Proofs for Section 5.1

The facts collected in the subsequent remark will be used in the proofs further below.

Remark B.1. (i) Suppose Assumption 1 holds. Then the test statistic T_{Het} is a nonsphericity corrected F-type test statistic in the sense of Section 5.4 in Preinerstorfer and Pötscher (2016). More precisely, T_{Het} is of the form (28) in Preinerstorfer and Pötscher (2016) and Assumption 5 in the same reference is satisfied with $\check{\beta} = \hat{\beta}$, $\check{\Omega} = \hat{\Omega}_{Het}$, and $N = \emptyset$. Furthermore, the set N^* defined in (27) of Preinerstorfer and Pötscher (2016) satisfies $N^* = \mathsf{B}$. And also Assumptions 6 and 7 of Preinerstorfer and Pötscher (2016) are satisfied. All these claims follow easily in view of Lemma 4.1 in Preinerstorfer and Pötscher (2016), see also the proof of Theorem 4.2 in that reference.

(ii) The test statistic T_{uc} is also a nonsphericity corrected F-type test statistic in the sense of Section 5.4 in Preinerstorfer and Pötscher (2016) (terminology being somewhat unfortunate here as no correction for the nonsphericity is being attempted). More precisely, T_{uc} is of the form (28) in Preinerstorfer and Pötscher (2016) and Assumption 5 in the same reference is satisfied with $\check{\beta} = \hat{\beta}$, $\check{\Omega} = \hat{\sigma}^2 R\left(X'X\right)^{-1} R'$, and $N = \emptyset$. Furthermore, the set N^* defined in (27) of Preinerstorfer and Pötscher (2016) satisfies $N^* = \operatorname{span}(X)$. And also Assumptions 6 and 7 of Preinerstorfer and Pötscher (2016) are satisfied. All these claims are evident (and obviously do not rely on Assumption 1).

(iii) We note that any nonsphericity corrected F-type test statistic (for testing (3)) in the sense of Section 5.4 in Preinerstorfer and Pötscher (2016), i.e., any test statistic T of the form (28) in Preinerstorfer and Pötscher (2016) that also satisfies Assumption 5 in that reference, is invariant under the group $G(\mathfrak{M}_0)$. Furthermore, the associated set N^* defined in (27) of Preinerstorfer and Pötscher (2016) is even invariant under the larger group $G(\mathfrak{M})$. See Sections 5.1 and 5.4 of Preinerstorfer and Pötscher (2016) as well as Lemma 5.16 in Pötscher and Preinerstorfer (2018) for more information.

Proof of Theorem 5.1. We first prove Part (b) and apply Theorem A.1 in Appendix A with $T_1 = T_2 = T^{\blacktriangle}_{Het}$. Borel-measurability and $G(\mathfrak{M}_0)$-invariance of T_1 and T_2 follow from the corresponding properties of T_{Het}, see the discussion in Section 3 and, in particular, Remark 3.2. Furthermore, we see that F_y in Theorem A.1 coincides with $F^{\blacktriangle}_{Het,y}$, implying that we may set $f_{1-\alpha}(y) = f^{\blacktriangle}_{Het,1-\alpha}(y)$. For indices i such that $e_i(n) \notin \mathsf{B}$, we have also

$\mu_0 + e_i(n) \notin \mathsf{B}$ and thus continuity of T_1 at $\mu_0 + e_i(n)$ (as T_1 coincides with T_{Het} on the complement of the closed set B); cf. Remarks 3.2, B.1, and Lemma 5.15 in Preinerstorfer and Pötscher (2016). In particular, $c_i = T_1(\mu_0 + e_i(n)) = T_{Het}(\mu_0 + e_i(n))$ follows for such indices i. Also note that $y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}$ implies that $y^*(\mu_0 + e_i(n), \xi)$ is a continuity point of T_2. This shows that $\vartheta^{\blacktriangle}_{1,Het}$ defined by

$$\vartheta^{\blacktriangle}_{1,Het} = \max_{\substack{i=1,\ldots,n,\\ e_i(n)\notin \mathsf{B}}} \Xi\left(\left\{\xi : T^{\blacktriangle,*}_{Het}(\mu_0 + e_i(n), \xi) < T_{Het}(\mu_0 + e_i(n)), y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}\right\}\right)$$

is less than or equal to the maximum appearing on the right-hand side of (35) (where we set $\vartheta^{\blacktriangle}_{1,Het} = 0$ if the maximum operator is taken over an empty index set). It is also obvious that $\vartheta^{\blacktriangle}_{1,Het} = \vartheta_{1,Het}$ holds, since $T^{\blacktriangle,*}_{Het}(\mu_0 + e_i(n), \xi)$ coincides with $T^{*}_{Het}(\mu_0 + e_i(n), \xi)$ when $y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}$. Next, consider an index i such that $e_i(n) \in \operatorname{span}(X)$ and $R\hat{\beta}(e_i(n)) \neq 0$ hold. We show that then $c_i = \infty$: For this it suffices to show that $T_1(\mu_0 + v_m) \to \infty$ for any sequence v_m with $v_m \to e_i(n)$ for $m \to \infty$. Now, for all m such that $\mu_0 + v_m \in \mathsf{B}$ holds we have $T_1(\mu_0 + v_m) = T^{\blacktriangle}_{Het}(\mu_0 + v_m) = \infty$; and for all m with $\mu_0 + v_m \notin \mathsf{B}$ we obtain

$$T_1(\mu_0 + v_m) = T_{Het}(\mu_0 + v_m) \geq \left\| R\hat{\beta}(\mu_0 + v_m) - r \right\|^2 \lambda^{-1}_{\max}(\hat{\Omega}_{Het}(\mu_0 + v_m)),$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix indicated. Note that $R\hat{\beta}(\mu_0 + v_m) - r \to R\hat{\beta}(\mu_0 + e_i(n)) - r = R\hat{\beta}(e_i(n)) \neq 0$ and that $\hat{\Omega}_{Het}(\mu_0 + v_m) \to \hat{\Omega}_{Het}(\mu_0 + e_i(n)) = 0$, the last equality following from $\mu_0 + e_i(n) \in \operatorname{span}(X)$, which in turn is a consequence of $e_i(n) \in \operatorname{span}(X)$. Obviously, $T_1(\mu_0 + v_m) \to \infty$ for $m \to \infty$ now follows. Now define

$$\vartheta^{\blacktriangle}_{2,Het} = \max_{\substack{i=1,\ldots,n,\\ e_i(n)\in \operatorname{span}(X), R\hat{\beta}(e_i(n))\neq 0}} \Xi\left(\left\{\xi : T^{\blacktriangle,*}_{Het}(\mu_0 + e_i(n), \xi) < \infty, y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}\right\}\right),$$

where we set $\vartheta^{\blacktriangle}_{2,Het} = 0$ if the maximum operator is taken over an empty index set. Then $\vartheta^{\blacktriangle}_{2,Het}$ is obviously less than or equal to the maximum appearing on the right-hand side of (35), again using that $y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}$ implies $y^*(\mu_0 + e_i(n), \xi)$ being a continuity point of T_2. Observe now that the condition $T^{\blacktriangle,*}_{Het}(\mu_0 + e_i(n), \xi) < \infty$ in the above set is always satisfied because of $y^*(\mu_0 + e_i(n), \xi) \notin \mathsf{B}$. Hence, $\vartheta^{\blacktriangle}_{2,Het} = \vartheta_{2,Het}$ holds (of course, also in the trivial case where the index set in question is empty). Consequently, any α satisfying $\alpha > \vartheta_{Het} = 1 - \max(\vartheta_{1,Het}, \vartheta_{2,Het})$ also satisfies (35). An application of Theorem A.1 now yields (13) but with T_{Het} replaced by $T^{\blacktriangle}_{Het}$. Since both the latter functions agree outside of B, a $\lambda_{\mathbb{R}^n}$-null set, and since the measures $P_{\mu_0,\sigma^2\Sigma}$ are absolutely continuous with respect to $\lambda_{\mathbb{R}^n}$, also (13) as given in the theorem follows. Independence of $\vartheta_{1,Het}$ and $\vartheta_{2,Het}$ from the choice of $\mu_0 \in \mathfrak{M}_0$ follows since $y^*(\mu_0 + y, \xi) = y^*(\mu'_0 + y, \xi) - \mu'_0 + \mu_0$ for every $\mu'_0 \in \mathfrak{M}_0$ as shown in the proof of Theorem A.1 and since T_{Het} and B are $G(\mathfrak{M}_0)$-invariant, see Remark 3.2.

Part (a) now follows from the already established Part (b), noting that we may choose $f^{\blacktriangle}_{Het,1-\alpha}$ such that $f^{\blacktriangle}_{Het,1-\alpha}(y) \geq f_{Het,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$.[50] □

[50]For example, by choosing $f^{\blacktriangle}_{Het,1-\alpha}(y)$ as the largest $(1-\alpha)$-quantile of $F^{\blacktriangle}_{Het,y}$.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Proof of Theorem 5.2. We first prove Part (b) and apply Theorem A.1 in Appendix A with $T_1 = T_2 = T_{uc}^{\blacktriangle}$. Borel-measurability and $G(\mathfrak{M}_0)$-invariance of T_1 and T_2 follow from the corresponding properties of T_{uc}, see the discussion in Section 3 and, in particular, Remark 3.2. Furthermore, we see that F_y in Theorem A.1 coincides with $F_{uc,y}^{\blacktriangle}$, implying that we may set $f_{1-\alpha}(y) = f_{uc,1-\alpha}^{\blacktriangle}(y)$. For indices i such that $e_i(n) \notin \operatorname{span}(X)$, we have also $\mu_0 + e_i(n) \notin \operatorname{span}(X)$ and thus continuity of T_1 at $\mu_0 + e_i(n)$ (as T_1 coincides with T_{uc} on the complement of the closed set $\operatorname{span}(X)$); cf. Remarks 3.2, B.1, and Lemma 5.15 in Preinerstorfer and Pötscher (2016). In particular, $c_i = T_1(\mu_0 + e_i(n)) = T_{uc}(\mu_0 + e_i(n))$ follows for such indices i. Also note that $y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)$ implies that $y^*(\mu_0 + e_i(n), \xi)$ is a continuity point of T_2. This shows that $\vartheta_{1,uc}^{\blacktriangle}$ defined by

$$\vartheta_{1,uc}^{\blacktriangle} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\notin \operatorname{span}(X)}} \Xi\left(\left\{\xi : T_{uc}^{\blacktriangle,*}(\mu_0 + e_i(n), \xi) < T_{uc}(\mu_0 + e_i(n)), y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)\right\}\right)$$

is less than or equal to the maximum appearing on the r.h.s. of (35). It is also obvious that $\vartheta_{1,uc}^{\blacktriangle} = \vartheta_{1,uc}$ holds, since $T_{uc}^{\blacktriangle,*}(\mu_0 + e_i(n), \xi)$ coincides with $T_{uc}^{*}(\mu_0 + e_i(n), \xi)$ when $y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)$. Next, consider an index i such that $e_i(n) \in \operatorname{span}(X)$ and $R\hat{\beta}(e_i(n)) \neq 0$ hold. We show that then $c_i = \infty$: For this it suffices to show that $T_1(\mu_0 + v_m) \to \infty$ for any sequence v_m with $v_m \to e_i(n)$ for $m \to \infty$. Now, for all m such that $\mu_0 + v_m \in \operatorname{span}(X)$ holds we have $T_1(\mu_0 + v_m) = T_{uc}^{\blacktriangle}(\mu_0 + v_m) = \infty$; and for all m with $\mu_0 + v_m \notin \operatorname{span}(X)$ we obtain

$$T_1(\mu_0 + v_m) = T_{uc}(\mu_0 + v_m) \geq \left\| R\hat{\beta}(\mu_0 + v_m) - r \right\|^2 \lambda_{\max}^{-1}(R\left(X'X\right)^{-1} R')\hat{\sigma}^{-2}(\mu_0 + v_m),$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix indicated. Note that $R\hat{\beta}(\mu_0 + v_m) - r \to R\hat{\beta}(\mu_0 + e_i(n)) - r = R\hat{\beta}(e_i(n)) \neq 0$ and that $\hat{\sigma}^2(\mu_0 + v_m) \to \hat{\sigma}^2(\mu_0 + e_i(n)) = 0$, the last equality following from $\mu_0 + e_i(n) \in \operatorname{span}(X)$, which in turn is a consequence of $e_i(n) \in \operatorname{span}(X)$. Obviously, $T_1(\mu_0 + v_m) \to \infty$ for $m \to \infty$ now follows. Now define

$$\vartheta_{2,uc}^{\blacktriangle} = \max_{\substack{i=1,\dots,n,\\ e_i(n)\in \operatorname{span}(X), R\hat{\beta}(e_i(n))\neq 0}} \Xi\left(\left\{\xi : T_{uc}^{\blacktriangle,*}(\mu_0 + e_i(n), \xi) < \infty, y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)\right\}\right),$$

where we set $\vartheta_{2,uc}^{\blacktriangle} = 0$ if the maximum operator is taken over an empty index set. Then $\vartheta_{2,uc}^{\blacktriangle}$ is obviously less than or equal to the maximum appearing on the right-hand side of (35), again using that $y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)$ implies $y^*(\mu_0 + e_i(n), \xi)$ being a continuity point of T_2. Observe now that the condition $T_{uc}^{\blacktriangle,*}(\mu_0 + e_i(n), \xi) < \infty$ in the above set is always satisfied because of $y^*(\mu_0 + e_i(n), \xi) \notin \operatorname{span}(X)$. Hence, $\vartheta_{2,uc}^{\blacktriangle} = \vartheta_{2,uc}$ holds (of course, also in the trivial case where the index set in question is empty). Consequently, any α satisfying $\alpha > \vartheta_{uc} = 1 - \max(\vartheta_{1,uc}, \vartheta_{2,uc})$ also satisfies (35). An application of Theorem A.1 now yields (17) but with T_{uc} replaced by $T_{uc}^{\blacktriangle}$. Since both the latter functions agree outside of $\operatorname{span}(X)$, a $\lambda_{\mathbb{R}^n}$-null set, and since the measures $P_{\mu_0,\sigma^2\Sigma}$ are absolutely continuous with respect to $\lambda_{\mathbb{R}^n}$, also (17) as given follows. Independence of $\vartheta_{1,uc}$ and $\vartheta_{2,uc}$ from the choice of $\mu_0 \in \mathfrak{M}_0$ follows since $y^*(\mu_0 + y, \xi) = y^*(\mu_0' + y, \xi) - \mu_0' + \mu_0$ for every $\mu_0' \in \mathfrak{M}_0$ as shown in the proof of Theorem A.1 and since T_{uc} and $\operatorname{span}(X)$ are $G(\mathfrak{M}_0)$-invariant, see Remark 3.2.

Part (a) now follows from the already established Part (b), noting that we may choose $f^{\blacktriangle}_{uc,1-\alpha}$ such that $f^{\blacktriangle}_{uc,1-\alpha}(y) \geq f_{uc,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$. □

Remark B.2. A weaker version of Theorem 5.1 (Theorem 5.2, respectively) is obtained if ϑ_{Het} is replaced by $1-\vartheta_{1,Het}$ (ϑ_{uc} is replaced by $1-\vartheta_{1,uc}$, respectively). These weaker versions not only have simpler proofs in that one does not need to deal with the quantities $\vartheta_{2,Het}$ ($\vartheta_{2,uc}$, respectively), but can also be derived from Theorem A.1 more directly by setting $T_1 = T_{Het}$ and $T_2 = T^{\blacktriangle}_{Het}$ ($T_1 = T_{uc}$ and $T_2 = T^{\blacktriangle}_{uc}$, respectively), leading to a somewhat simpler proof that directly establishes (13) ((17), respectively) instead of establishing these relations for $T^{\blacktriangle}_{Het}$ ($T^{\blacktriangle}_{uc}$, respectively) first. Cf. the proofs of Theorems 5.7 and 5.8 further below, which have a similar structure.

Proof of Lemma 5.3. Let $y \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^n$ be arbitrary. Observe that $y^*(y,\xi) - y^{\maltese}(y,\xi) = X\tilde{\beta}_{\mathfrak{M}_0}(y) - X\hat{\beta}(y)$. It is then elementary to see that $R\hat{\beta}(y^*(y,\xi)) - r = R\hat{\beta}(y^{\maltese}(y,\xi)) - R\hat{\beta}(y)$ since $R\tilde{\beta}_{\mathfrak{M}_0}(y) = r$ certainly holds. Another immediate consequence is that $\hat{u}(y^*(y,\xi)) = \hat{u}(y^{\maltese}(y,\xi))$ holds, which implies $\hat{\Omega}_{Het}(y^*(y,\xi)) = \hat{\Omega}_{Het}(y^{\maltese}(y,\xi))$ as well as $\hat{\sigma}^2(y^*(y,\xi)) = \hat{\sigma}^2(y^{\maltese}(y,\xi))$. From the first observation we see that $y^*(y,\xi)$ and $y^{\maltese}(y,\xi)$ differ only by an element of span(X). Hence, $y^*(y,\xi) \in \mathsf{B}$ if and only if $y^{\maltese}(y,\xi) \in \mathsf{B}$, since $\mathsf{B} + \text{span}(X) = \mathsf{B}$ holds as noted in Lemma 3.1. And similarly, $y^*(y,\xi) \in \text{span}(X)$ if and only if $y^{\maltese}(y,\xi) \in \text{span}(X)$. This proves all the claims. □

C. Proofs for Section 5.2

Proof of Lemma 5.5. Observe that $\tilde{\Omega}_{Het}(y) = \tilde{B}(y)\,\text{diag}\left(\tilde{d}_1,\ldots,\tilde{d}_n\right)\tilde{B}'(y)$. Given that $\tilde{d}_i > 0$, this immediately establishes Parts (a) and (b) of the lemma. We next prove Part (c). Let s be as in Assumption 2 and consider first the case where this assumption is satisfied. If now $y \in \tilde{\mathsf{B}}$ it follows that $\tilde{u}_l(y) = 0$ must hold at least for some $l \notin \{i_1,\ldots i_s\}$ where l may depend on y. But this means that y satisfies $e_l'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0) = 0$. Since $e_l'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp} \neq 0$ by construction of l, it follows that $\tilde{\mathsf{B}}$ is contained in a finite union of proper affine subspaces, and hence is a $\lambda_{\mathbb{R}^n}$-null set. Next consider the case where Assumption 2 is not satisfied. Observe that then $s > 0$ must hold. Note that $\tilde{u}_i(y) = 0$ holds for all $y \in \mathbb{R}^n$ and all $i \in \{i_1,\ldots i_s\}$ by construction of $\{i_1,\ldots i_s\}$. But then for every $y \in \mathbb{R}^n$

$$\begin{aligned}\text{rank}(\tilde{B}(y)) &= \text{rank}\left(R(X'X)^{-1}X'(\neg(i_1,\ldots i_s))A(y)\right)\\ &\leq \text{rank}\left(R(X'X)^{-1}X'(\neg(i_1,\ldots i_s))\right) < q\end{aligned}$$

is satisfied where $A(y)$ is obtained from diag$(\tilde{u}_1(y),\ldots,\tilde{u}_n(y))$ by deleting rows and columns i with $i \in \{i_1,\ldots i_s\}$. This completes the proof of Part (c). To prove Part (d) recall that $\tilde{\mathsf{B}}$ is the set where

$$\begin{aligned}\tilde{B}(y) &= R(X'X)^{-1}X'\text{diag}\left(e_1'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0),\ldots,e_n'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)\right)\\ &= R(X'X)^{-1}X'\left[e_1(n)e_1'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0),\ldots,e_n(n)e_n'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)\right]\end{aligned}$$

has rank less than q. Define the set

$$D=\{(j_1,\ldots,j_s):1\le s\le n, 1\le j_1<\cdots<j_s \le n, \operatorname{rank}\left(R(X'X)^{-1}X'\left[e_{j_1}(n),\ldots,e_{j_s}(n)\right]\right)<q\right\},$$

which may be empty in case $q=1$. Consider first the case where D is nonempty: Since $R(X'X)^{-1}X'$ has rank q, it is then easy to see that we have $y\in\tilde{\mathsf{B}}$ if and only if there exists $(j_1,\ldots,j_s)\in D$ such that $e_j'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)=0$ for $j\neq j_i$ for $i=1,\ldots,s$. This shows, that $\tilde{\mathsf{B}}-\mu_0$ is a finite union of (not necessarily distinct) linear subspaces. In case D is empty, $\operatorname{rank}(R(X'X)^{-1}X')=q$ implies that $y\in\tilde{\mathsf{B}}$ if and only if $e_j'(n)\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)=0$ for all $1\le j\le n$, i.e., if and only if $y\in\mathfrak{M}_0$. That is, in this case $\tilde{\mathsf{B}}-\mu_0=\mathfrak{M}_0^{lin}$, a linear subspace. That the linear subspaces making up $\tilde{\mathsf{B}}-\mu_0$ are proper, follows since otherwise $\tilde{\mathsf{B}}-\mu_0$, and thus $\tilde{\mathsf{B}}$, would be all of $\mathbb{R}^n$, which is impossible under Assumption 2 as shown in Part (c). To prove the second claim, observe that in case $q=1$ the condition that $\operatorname{rank}(\tilde{B}(y))$ is less than q is equivalent to $\tilde{B}(y)=0$. Since the expressions $R(X'X)^{-1}X'e_j(n)$ are now scalar, we may thus write the condition $\tilde{B}(y)=0$ equivalently as

$$\left[R(X'X)^{-1}X'e_1(n)e_1(n),\ldots,R(X'X)^{-1}X'e_n(n)e_n(n)\right]\Pi_{(\mathfrak{M}_0^{lin})^\perp}(y-\mu_0)=0.$$

But this shows that $\mathsf{B}-\mu_0$ is a linear subspace, namely the kernel of the matrix appearing on the left-hand side of the preceding display. It remains to prove Part (e). Closedness of $\tilde{\mathsf{B}}$ follows from Parts (c) and (d), and the remaining claims are trivial (note that $\tilde{B}(y)$ only depends on $\tilde{u}(y)$, which obviously is $G(\mathfrak{M}_0)$-invariant). □

Proof of Theorem 5.7. We first prove Part (b) and apply Theorem A.1 in Appendix A with $T_1=\tilde{T}_{Het}$ and $T_2=\tilde{T}_{Het}^{\blacktriangle}$. Borel-measurability and $G(\mathfrak{M}_0)$-invariance of T_1 and T_2 follow from the corresponding properties of $\tilde{T}_{Het}$, see Remark 5.6. Furthermore, we see that F_y in Theorem A.1 coincides with $\tilde{F}_{Het,y}^{\blacktriangle}$, implying that we may set $f_{1-\alpha}(y)=\tilde{f}_{Het,1-\alpha}^{\blacktriangle}(y)$. For indices i such that $\mu_0+e_i(n)\notin\tilde{\mathsf{B}}$, the function T_1 is continuous at $\mu_0+e_i(n)$. In particular, $c_i=T_1(\mu_0+e_i(n))=\tilde{T}_{Het}(\mu_0+e_i(n))$ follows for such indices i. Also note that $y^*(\mu_0+e_i(n),\xi)\notin\tilde{\mathsf{B}}$ implies that $y^*(\mu_0+e_i(n),\xi)$ is a continuity point of T_2. This shows that $\tilde{\vartheta}_{Het}^{\blacktriangle}$ defined by

$$\tilde{\vartheta}_{Het}^{\blacktriangle}=1-\max_{\substack{i=1,\ldots,n,\\ \mu_0+e_i(n)\notin\tilde{\mathsf{B}}}}\Xi\left(\left\{\xi:\tilde{T}_{Het}^{\blacktriangle,*}(\mu_0+e_i(n),\xi) < \tilde{T}_{Het}(\mu_0+e_i(n)), y^*(\mu_0+e_i(n),\xi)\notin\tilde{\mathsf{B}}\right\}\right)$$

is not less than the right-hand side of (35) (where we set $\tilde{\vartheta}_{Het}^{\blacktriangle}=1$ if the maximum operator is taken over an empty index set). It is also obvious that $\tilde{\vartheta}_{Het}^{\blacktriangle}=\tilde{\vartheta}_{Het}$ holds, since $\tilde{T}_{Het}^{\blacktriangle,*}(\mu_0+e_i(n),\xi)$ coincides with $\tilde{T}_{Het}^{*}(\mu_0+e_i(n),\xi)$ when $y^*(\mu_0+e_i(n),\xi)\notin\tilde{\mathsf{B}}$. Consequently, any α satisfying $\alpha>\tilde{\vartheta}_{Het}$ also satisfies (35). An application of Theorem A.1 now yields (23). Independence of $\tilde{\vartheta}_{Het}$ from the choice of $\mu_0\in\mathfrak{M}_0$ follows since $y^*(\mu_0+y,\xi)=y^*(\mu_0'+y,\xi)-\mu_0'+\mu_0$ for every $\mu_0'\in\mathfrak{M}_0$ as shown in the proof of Theorem A.1 and since $\tilde{T}_{Het}$ and $\tilde{\mathsf{B}}$ are $G(\mathfrak{M}_0)$-invariant, see Remark 5.6.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Part (a) now follows from the already established Part (b), noting that we may choose $\tilde{f}^{\blacktriangle}_{Het,1-\alpha}$ such that $\tilde{f}^{\blacktriangle}_{Het,1-\alpha}(y) \geq \tilde{f}_{Het,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$. □

Proof of Theorem 5.8. [51] We first prove Part (b) and apply Theorem A.1 in Appendix A with $T_1 = \tilde{T}_{uc}$ and $T_2 = \tilde{T}^{\blacktriangle}_{uc}$. Borel-measurability and $G(\mathfrak{M}_0)$-invariance of T_1 and T_2 follow from the corresponding properties of $\tilde{T}_{uc}$, see Remark 5.6. Furthermore, we see that F_y in Theorem A.1 coincides with $\tilde{F}^{\blacktriangle}_{uc,y}$, implying that we may set $f_{1-\alpha}(y) = \tilde{f}^{\blacktriangle}_{uc,1-\alpha}(y)$. For indices i such that $\mu_0 + e_i(n) \notin \mathfrak{M}_0$, the function T_1 is continuous at $\mu_0 + e_i(n)$. In particular, $c_i = T_1(\mu_0 + e_i(n)) = \tilde{T}_{uc}(\mu_0 + e_i(n))$ follows for such indices i. Also note that $y^*(\mu_0 + e_i(n), \xi) \notin \mathfrak{M}_0$ implies that $y^*(\mu_0 + e_i(n), \xi)$ is a continuity point of T_2. This shows that $\tilde{\vartheta}^{\blacktriangle}_{uc}$ defined by

$$\tilde{\vartheta}^{\blacktriangle}_{uc} = 1 - \max_{\substack{i=1,\ldots,n,\\ \mu_0+e_i(n)\notin\mathfrak{M}_0}} \Xi\left(\left\{\xi : \tilde{T}^{\blacktriangle,*}_{uc}(\mu_0 + e_i(n), \xi) \right.\right.$$
$$\left.\left. < \tilde{T}_{uc}(\mu_0 + e_i(n)), y^*(\mu_0 + e_i(n), \xi) \notin \mathfrak{M}_0 \right\}\right)$$

is not less than the right-hand side of (35). It is also obvious that $\tilde{\vartheta}^{\blacktriangle}_{uc} = \tilde{\vartheta}_{uc}$ holds, since $\tilde{T}^{\blacktriangle,*}_{uc}(\mu_0 + e_i(n), \xi)$ coincides with $\tilde{T}^{*}_{uc}(\mu_0 + e_i(n), \xi)$ when $y^*(\mu_0 + e_i(n), \xi) \notin \mathfrak{M}_0$. Consequently, any α satisfying $\alpha > \tilde{\vartheta}_{uc}$ also satisfies (35). An application of Theorem A.1 now yields (26). Independence of $\tilde{\vartheta}_{uc}$ from the choice of $\mu_0 \in \mathfrak{M}_0$ follows since $y^*(\mu_0 + y, \xi) = y^*(\mu'_0 + y, \xi) - \mu'_0 + \mu_0$ for every $\mu'_0 \in \mathfrak{M}_0$ as shown in the proof of Theorem A.1 and since $\tilde{T}_{uc}$ and $\mathfrak{M}_0$ are $G(\mathfrak{M}_0)$-invariant, see Remark 5.6.

Part (a) now follows from the already established Part (b), noting that we may choose $\tilde{f}^{\blacktriangle}_{uc,1-\alpha}$ such that $\tilde{f}^{\blacktriangle}_{uc,1-\alpha}(y) \geq \tilde{f}_{uc,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$. □

LEMMA C.1. *Let $\mathcal{A}$ be an affine subspace of $\mathbb{R}^n$ satisfying $\mathfrak{M}_0 \subseteq \mathcal{A} \subseteq \operatorname{span}(X)$, and let $y^{\maltese}(y, \xi)$ be as defined in (9).*

(a) For every $\gamma \in \mathbb{R}$, every $\mu'_0 \in \mathfrak{M}_0$, every $\mu''_0 \in \mathfrak{M}_0$, every $y \in \mathbb{R}$, and every $\xi \in \mathbb{R}^n$ we have

$$y^{\maltese}(\gamma(y - \mu'_0) + \mu''_0, \xi) = \gamma\left(y^{\maltese}(y, \xi) - \mu'_0\right) + \mu''_0,$$
$$R\hat{\beta}(y^{\maltese}(\gamma(y - \mu'_0) + \mu''_0, \xi)) - R\hat{\beta}(\gamma(y - \mu'_0) + \mu''_0) = \gamma\left[R\hat{\beta}(y^{\maltese}(y, \xi)) - R\hat{\beta}(y)\right],$$
$$\tilde{u}(y^{\maltese}(\gamma(y - \mu'_0) + \mu''_0, \xi)) = \gamma\tilde{u}(y^{\maltese}(y, \xi)),$$
$$\tilde{\Omega}_{Het}\left(y^{\maltese}(\gamma(y - \mu'_0) + \mu''_0, \xi)\right) = \gamma^2\tilde{\Omega}_{Het}\left(y^{\maltese}(y, \xi)\right)$$

where $\tilde{u}(y) = y - X\tilde{\beta}_{\mathfrak{M}_0}(y)$ denotes the vector of restricted residuals corresponding to the restricted estimator $\tilde{\beta}_{\mathfrak{M}_0}$.

(b) The set $\{y \in \mathbb{R}^n : \det\tilde{\Omega}_{Het}\left(y^{\maltese}(y, \xi)\right) = 0\}$ as well as $\tilde{T}^{\maltese}_{Het}(\cdot, \xi)$ are $G(\mathfrak{M}_0)$-invariant for every $\xi \in \mathbb{R}^n$.

(c) The set $\{y \in \mathbb{R}^n : y^{\maltese}(y, \xi) \in \mathfrak{M}_0\}$ as well as $\tilde{T}^{\maltese}_{uc}(\cdot, \xi)$ are $G(\mathfrak{M}_0)$-invariant for every $\xi \in \mathbb{R}^n$.

[51] In view of the relationship between $\tilde{T}_{uc}$ and T_{uc} discussed in Section 5.1.1 in Pötscher and Preinerstorfer (2021), one could attempt to use this relationship to derive Theorem 5.8 from Theorem 5.2. However, it is not obvious how this can actually be executed since the relationship mentioned only holds outside of a Lebesgue null-set, but $y^*(y, \xi)$ typically follows a discrete distribution.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Proof. (a) Using elementary properties of the least squares estimator $\hat{\beta}$, using that $y - X\tilde{\beta}_{\mathcal{A}}(y) = \Pi_{(\mathcal{A}-\mu_0)^{\perp}}(y-\mu_0)$ as noted in the proof of Theorem A.1 in Appendix A, and noting that $\Pi_{(\mathcal{A}-\mu_0)^{\perp}}(\mu_0''-\mu_0) = 0 = \Pi_{(\mathcal{A}-\mu_0)^{\perp}}(\mu_0-\mu_0')$ (since $\mu_0''-\mu_0$ as well as $\mu_0-\mu_0'$ belong to $\mathfrak{M}_0^{lin}$ and since $(\mathcal{A}-\mu_0)^{\perp} \subseteq (\mathfrak{M}_0^{lin})^{\perp}$) we get

$$\begin{aligned} y^{\maltese}(\gamma(y-\mu_0')+\mu_0'',\xi) &= X\hat{\beta}(\gamma(y-\mu_0')+\mu_0'') + \operatorname{diag}(\xi) \\ &\quad \times \Pi_{(\mathcal{A}-\mu_0)^{\perp}}(\gamma(y-\mu_0')+\mu_0''-\mu_0) \\ &= \gamma(X\hat{\beta}(y)-\mu_0') + \mu_0'' + \gamma\operatorname{diag}(\xi)\Pi_{(\mathcal{A}-\mu_0)^{\perp}}(y-\mu_0) \\ &= \gamma\left(y^{\maltese}(y,\xi)-\mu_0'\right)+\mu_0''. \end{aligned}$$

The second displayed equation is then an immediate consequence. Using what has already been established

$$\begin{aligned} \tilde{u}(y^{\maltese}(\gamma(y-\mu_0')+\mu_0'',\xi)) &= \tilde{u}(\gamma\left(y^{\maltese}(y,\xi)-\mu_0'\right)+\mu_0'') \\ &= \Pi_{(\mathfrak{M}_0^{lin})^{\perp}}(\gamma\left(y^{\maltese}(y,\xi)-\mu_0'\right)+\mu_0''-\mu_0) \\ &= \gamma\Pi_{(\mathfrak{M}_0^{lin})^{\perp}}\left(y^{\maltese}(y,\xi)-\mu_0'\right) \\ &= \gamma\Pi_{(\mathfrak{M}_0^{lin})^{\perp}}\left(y^{\maltese}(y,\xi)-\mu_0\right) = \gamma\tilde{u}(y^{\maltese}(y,\xi)), \end{aligned}$$

which then also immediately gives the fourth displayed equation.

(b) and (c). Follows from (a). □

Proof of Theorem 5.9. We prove Part (b) first. By $G(\mathfrak{M}_0)$-invariance of $\tilde{T}_{Het}^{\maltese}(\cdot,\xi)$ and of the set $\{y\in\mathbb{R}^n : y^{\maltese}(y,\xi)\in\tilde{\mathsf{B}}\}$ (for every ξ) established in Lemma C.1 we can conclude that for every $z\in\mathbb{R}^n$ the expression $\tilde{T}_{Het}^{\maltese}(\mu_0+\gamma z,\xi)$ depends neither on the choice of $\mu_0\in\mathfrak{M}_0$ nor on the value of $\gamma\in\mathbb{R}\backslash\{0\}$, cf. Footnote 49. By $G(\mathfrak{M}_0)$-invariance (see Remark 5.6), the same is true also for $\tilde{T}_{Het}(\mu_0+\gamma z)$. By Lemma C.1 and by $G(\mathfrak{M}_0)$-invariance of $\tilde{\mathsf{B}}$, the truth-value of the statement $y^{\maltese}(\mu_0+\gamma z,\xi)\notin\tilde{\mathsf{B}}$ depends neither on the choice of $\mu_0\in\mathfrak{M}_0$ nor on the value of $\gamma\in\mathbb{R}\backslash\{0\}$. A similar comment applies to the statement $\mu_0+e_i(n)\notin\tilde{\mathsf{B}}$. This shows, in particular, that $\tilde{\theta}_{Het}$ does not depend on the choice of $\mu_0\in\mathfrak{M}_0$. Next, observe that for every $y\in\mathbb{R}^n$ and every $t\in\mathbb{R}$ we have $\tilde{H}_{Het,y}^{\blacktriangle}(t) = \Xi(\{\xi : \tilde{T}_{Het}\left(y^{\maltese}(y,\xi)\right)\leq t, y^{\maltese}(y,\xi)\notin\tilde{\mathsf{B}}\})$. In particular, the condition $\alpha>\tilde{\theta}_{Het}$ is equivalent to

$$\max_{\substack{i=1,\dots,n,\\ \mu_0+e_i(n)\notin\tilde{\mathsf{B}}}} \tilde{H}_{Het,\mu_0+e_i(n)}^{\blacktriangle}(\tilde{T}_{Het}(\mu_0+e_i(n))-) > 1-\alpha,$$

with the convention that the left-hand side is zero if the maximum operator extends over an empty index set, in which case there is then nothing to prove. Otherwise, let from now on i be an index that realizes the maximum in the previous display. In the case where $\tilde{H}_{Het,\mu_0+e_i(n)}^{\blacktriangle}(\tilde{T}_{Het}(\mu_0+e_i(n))-)=0$, there is again nothing to prove and we are done. Hence, it remains to consider the case where $\tilde{H}_{Het,\mu_0+e_i(n)}^{\blacktriangle}(\tilde{T}_{Het}(\mu_0+e_i(n))-)>0$. In this case then, let $\alpha\in(0,1)$ be such that $\tilde{H}_{Het,\mu_0+e_i(n)}^{\blacktriangle}(\tilde{T}_{Het}(\mu_0+e_i(n))-)>1-\alpha$ (where $\mu_0\in\mathfrak{M}_0$ can be chosen arbitrarily). From this inequality we can conclude existence of a real number x_i smaller than $\tilde{T}_{Het}(\mu_0+e_i(n))$ such that $\tilde{H}_{Het,\mu_0+e_i(n)}^{\blacktriangle}(x_i)>1-\alpha$ holds

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

and such that x_i is a continuity point of $\tilde{H}^{\blacktriangle}_{Het,\,\mu_0+e_i(n)}$. Since $\mu_0 + e_i(n) \notin \tilde{\mathsf{B}}$, it is obvious that $\tilde{T}_{Het}$ is continuous at $\mu_0 + e_i(n)$. In view of this, there exists a $\delta > 0$ such that every $z \in B(e_i(n), \delta)$ satisfies $\tilde{T}_{Het}(\mu_0 + z) > x_i$ (and the same is true if we replace δ by a smaller positive number). We claim that for every sequence $z_m \to e_i(n)$ ($z_m \in \mathbb{R}^n$) we have

$$\liminf_{m\to\infty} \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+z_m}(x_i) \geq \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+e_i(n)}(x_i). \tag{43}$$

Define $V_m = V_m(\xi) = (\mu_0 + z_m, y^{\maltese}(\mu_0 + z_m, \xi))$ and $V = V(\xi) = (\mu_0 + e_i(n), y^{\maltese}(\mu_0 + e_i(n), \xi))$, which can be viewed as random vectors defined on $\mathbb{R}^n$ (equipped with the Borel σ-field) and where the probability measure is given by Ξ. Note that V_m converges to V everywhere as $m \to \infty$ (since $y^{\maltese}(y, \xi)$ is continuous w.r.t. y). Define $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ via

$$S(z,x) = \begin{cases} (R\hat{\beta}(x) - R\hat{\beta}(z))' \tilde{\Omega}^{-1}_{Het}(x)(R\hat{\beta}(x)) - R\hat{\beta}(z)), & \text{if } x \notin \tilde{\mathsf{B}}, \\ \infty, & \text{if } x \in \tilde{\mathsf{B}}, \end{cases}$$

and note that $\tilde{H}^{\blacktriangle}_{Het,\,\mu_0+z_m}(t) = \Xi(S(V_m) \leq t)$ and $\tilde{H}^{\blacktriangle}_{Het,\,\mu_0+e_i(n)}(t) = \Xi(S(V) \leq t)$ holds for $t \in \mathbb{R}$ and thus also for $t = x_i$ (recall that x_i is a real number). The statement in (43) now follows from Lemma A.3 in Appendix A, recalling that we have chosen x_i as a continuity point of $\tilde{H}^{\blacktriangle}_{Het,\,\mu_0+e_i(n)}$, which implies $\Xi(S(V) = x_i) = 0$, and noting that $\bar{S} = S$ here holds. Summarizing, we hence arrive, replacing δ by another element of $(0, \delta)$ if necessary, at

$$\tilde{T}_{Het}(\mu_0 + z) > x_i \text{ and } \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+z}(x_i) > 1 - \alpha \text{ for every } z \in B(e_i(n), \delta). \tag{44}$$

Now, for every $z \in \mathbb{R}^n$ and every $\gamma \neq 0$ we have

$$\begin{aligned} \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+\gamma z}(x_i) &= \Xi\left(\left\{\xi : \tilde{T}^{\blacktriangle,\maltese}_{Het}(\mu_0 + \gamma z, \xi) \leq x_i\right\}\right) \\ &= \Xi\left(\left\{\xi : \tilde{T}^{\blacktriangle,\maltese}_{Het}(\mu_0 + z, \xi) \leq x_i\right\}\right) = \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+z}(x_i) \end{aligned}$$

by what has been shown at the very beginning of the proof. Using also the observation for $\tilde{T}_{Het}(\mu_0 + \gamma z)$ made at the beginning of the proof, the preceding display and (44) thus give

$$\begin{aligned} &\tilde{T}_{Het}(\mu_0 + y) > x_i \text{ and } \tilde{H}^{\blacktriangle}_{Het,\,\mu_0+y}(x_i) \\ &\qquad > 1 - \alpha \text{ for every } y \in \{\gamma z : \gamma \neq 0, z \in B(e_i(n), \delta)\} =: \Delta, \end{aligned}$$

or equivalently

$$\tilde{T}_{Het}(y) > x_i \text{ and } \tilde{H}^{\blacktriangle}_{Het,\,y}(x_i) > 1 - \alpha \text{ for every } y \in \mu_0 + \Delta. \tag{45}$$

By Lemma A.4, we see that $\tilde{H}^{\blacktriangle}_{Het,\,y}(x_i) > 1 - \alpha$ implies $x_i \geq \tilde{h}^{\blacktriangle}_{Het,\,1-\alpha}(y)$. Hence,

$$\left\{y \in \mathbb{R}^n : \tilde{T}_{Het}(y) > x_i,\ \tilde{H}^{\blacktriangle}_{Het,\,y}(x_i) > 1 - \alpha\right\} \subseteq \left\{y \in \mathbb{R}^n : \tilde{T}_{Het}(y) > \tilde{h}^{\blacktriangle}_{Het,\,1-\alpha}(y)\right\}.$$

This, together with (45), gives

$$\left\{y \in \mathbb{R}^n : \tilde{T}_{Het}(y) > \tilde{h}^{\blacktriangle}_{Het,\,1-\alpha}(y)\right\} \supseteq \mu_0 + \Delta \supseteq \mu_0 + \operatorname{int}(\Delta) \supseteq \mu_0 + (\operatorname{span}(e_i(n)) \setminus \{0\}),$$

where $\operatorname{int}(\Delta)$ denotes the interior of Δ. The proof is now completed by following the steps after (40) in the proof of Theorem A.1 in Appendix A.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Part (a) now follows from the already established Part (b), noting that we may choose $\tilde{h}^{\blacktriangle}_{Het,1-\alpha}$ such that $\tilde{h}^{\blacktriangle}_{Het,1-\alpha}(y) \geq \tilde{h}_{Het,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$. □

Proof of Theorem 5.10. We prove Part (b) first. By $G(\mathfrak{M}_0)$-invariance of $\tilde{T}^{\maltese}_{uc}(\cdot,\xi)$ and of the set $\{y \in \mathbb{R}^n : y^{\maltese}(y,\xi) \in \mathfrak{M}_0\}$ (for every ξ) established in Lemma C.1 we can conclude that for every $z \in \mathbb{R}^n$ the expression $\tilde{T}^{\maltese}_{uc}(\mu_0+\gamma z,\xi)$ depends neither on the choice of $\mu_0 \in \mathfrak{M}_0$ nor on the value of $\gamma \in \mathbb{R}\backslash\{0\}$, cf. Footnote 49. By $G(\mathfrak{M}_0)$-invariance (see Remark 5.6), the same is true also for $\tilde{T}_{uc}(\mu_0+\gamma z)$. By Lemma C.1, the truth-value of the statement $y^{\maltese}(\mu_0+\gamma z,\xi) \notin \mathfrak{M}_0$ depends neither on the choice of $\mu_0 \in \mathfrak{M}_0$ nor on the value of $\gamma \in \mathbb{R}\backslash\{0\}$. A similar comment applies to the statement $\mu_0 + e_i(n) \notin \mathfrak{M}_0$. This shows, in particular, that $\tilde{\theta}_{uc}$ does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$. Next, observe that for every $y \in \mathbb{R}^n$ and every $t \in \mathbb{R}$ we have $\tilde{H}^{\blacktriangle}_{uc,y}(t) = \Xi(\{\xi : \tilde{T}_{uc}\left(y^{\maltese}(y,\xi)\right) \leq t, y^{\maltese}(y,\xi) \notin \mathfrak{M}_0\})$. In particular, the condition $\alpha > \tilde{\theta}_{uc}$ is equivalent to

$$\max_{\substack{i=1,\ldots,n,\\ \mu_0+e_i(n)\notin\mathfrak{M}_0}} \tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(\tilde{T}_{uc}(\mu_0+e_i(n))-) > 1-\alpha.$$

From now on let i be an index that realizes the maximum in the previous display (note that the index set is not empty since $k-q \leq k < n$ by assumption). In the case where $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(\tilde{T}_{uc}(\mu_0+e_i(n))-) = 0$, there is nothing to prove and we are done. Hence, it remains to consider the case where $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(\tilde{T}_{uc}(\mu_0+e_i(n))-) > 0$. In this case then, let $\alpha \in (0,1)$ be such that $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(\tilde{T}_{uc}(\mu_0+e_i(n))-) > 1-\alpha$ (where $\mu_0 \in \mathfrak{M}_0$ can be chosen arbitrarily). From this inequality, we can conclude existence of a real number x_i smaller than $\tilde{T}_{uc}(\mu_0+e_i(n))$ such that $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(x_i) > 1-\alpha$ holds and such that x_i is a continuity point of $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}$. Since $\mu_0+e_i(n) \notin \mathfrak{M}_0$, it is obvious that $\tilde{T}_{uc}$ is continuous at $\mu_0+e_i(n)$. In view of this, there exists a $\delta > 0$ such that every $z \in B(e_i(n),\delta)$ satisfies $\tilde{T}_{uc}(\mu_0+z) > x_i$ (and the same is true if we replace δ by a smaller positive number). We claim that for every sequence $z_m \to e_i(n)$ ($z_m \in \mathbb{R}^n$) we have

$$\liminf_{m\to\infty} \tilde{H}^{\blacktriangle}_{uc,\mu_0+z_m}(x_i) \geq \tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(x_i). \tag{46}$$

Define $V_m = V_m(\xi) = (\mu_0+z_m, y^{\maltese}(\mu_0+z_m,\xi))$ and $V = V(\xi) = (\mu_0+e_i(n), y^{\maltese}(\mu_0+e_i(n),\xi))$, which can be viewed as random vectors defined on $\mathbb{R}^n$ (equipped with the Borel σ-field) and where the probability measure is given by Ξ. Note that V_m converges to V everywhere as $m \to \infty$ (since $y^{\maltese}(y,\xi)$ is continuous w.r.t. y). Define $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ via

$$S(z,x) = \begin{cases} (R\hat{\beta}(x) - R\hat{\beta}(z))'\left(\tilde{\sigma}^2(x)R\left(X'X\right)^{-1}R'\right)^{-1}(R\hat{\beta}(x)) - R\hat{\beta}(z)), & \text{if } x \notin \mathfrak{M}_0, \\ \infty, & \text{if } x \in \mathfrak{M}_0, \end{cases}$$

and note that $\tilde{H}^{\blacktriangle}_{uc,\mu_0+z_m}(t) = \Xi(S(V_m) \leq t)$ and $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}(t) = \Xi(S(V) \leq t)$ holds for $t \in \mathbb{R}$ and thus also for $t = x_i$ (recall that x_i is a real number). The statement in (46) now follows from Lemma A.3 in Appendix A, recalling that we have chosen x_i as a continuity point of $\tilde{H}^{\blacktriangle}_{uc,\mu_0+e_i(n)}$, which implies $\Xi(S(V) = x_i) = 0$, and noting that $\bar{S} = S$ here holds. Summarizing, we hence arrive, replacing δ by another element of $(0,\delta)$ if necessary, at

$$\tilde{T}_{uc}(\mu_0+z) > x_i \text{ and } \tilde{H}^{\blacktriangle}_{uc,\mu_0+z}(x_i) > 1-\alpha \text{ for every } z \in B(e_i(n),\delta). \tag{47}$$

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

Now, for every $z \in \mathbb{R}^n$ and every $\gamma \neq 0$ we have

$$\begin{aligned}\tilde{H}^{\blacktriangle}_{uc,\mu_0+\gamma z}(x_i) &= \Xi\left(\left\{\xi : \tilde{T}^{\blacktriangle,\maltese}_{uc}(\mu_0+\gamma z,\xi) \le x_i\right\}\right)\\ &= \Xi\left(\left\{\xi : \tilde{T}^{\blacktriangle,\maltese}_{uc}(\mu_0+z,\xi) \le x_i\right\}\right) = \tilde{H}^{\blacktriangle}_{uc,\mu_0+z}(x_i)\end{aligned}$$

by what has been shown at the very beginning of the proof. Using also the observation for $\tilde{T}_{uc}(\mu_0+\gamma z)$ made at the beginning of the proof, the preceding display and (47) thus give

$$\tilde{T}_{uc}(\mu_0+y)>x_i \text{ and } \tilde{H}^{\blacktriangle}_{uc,\mu_0+y}(x_i)>1-\alpha \text{ for every } y\in\{\gamma z : \gamma \neq 0, z \in B(e_i(n),\delta)\} =: \Delta,$$

or equivalently

$$\tilde{T}_{uc}(y) > x_i \text{ and } \tilde{H}^{\blacktriangle}_{uc,y}(x_i) > 1-\alpha \text{ for every } y \in \mu_0+\Delta. \tag{48}$$

By Lemma A.4, we see that $\tilde{H}^{\blacktriangle}_{uc,y}(x_i) > 1-\alpha$ implies $x_i \geq \tilde{h}^{\blacktriangle}_{uc,1-\alpha}(y)$. Hence,

$$\left\{y \in \mathbb{R}^n : \tilde{T}_{uc}(y) > x_i,\ \tilde{H}^{\blacktriangle}_{uc,y}(x_i) > 1-\alpha\right\} \subseteq \left\{y \in \mathbb{R}^n : \tilde{T}_{uc}(y) > \tilde{h}^{\blacktriangle}_{uc,1-\alpha}(y)\right\}.$$

This, together with (48), gives

$$\left\{y \in \mathbb{R}^n : \tilde{T}_{uc}(y) > \tilde{h}^{\blacktriangle}_{uc,1-\alpha}(y)\right\} \supseteq \mu_0+\Delta \supseteq \mu_0+\operatorname{int}(\Delta) \supseteq \mu_0+(\operatorname{span}(e_i(n))\backslash\{0\}),$$

where $\operatorname{int}(\Delta)$ denotes the interior of Δ. The proof is now completed by following the steps after (40) in the proof of Theorem A.1 in Appendix A.

Part (a) now follows from the already established Part (b), noting that we may choose $\tilde{h}^{\blacktriangle}_{uc,1-\alpha}$ such that $\tilde{h}^{\blacktriangle}_{uc,1-\alpha}(y) \geq \tilde{h}_{uc,1-\alpha}(y)$ holds for every $y \in \mathbb{R}^n$. □

D. Proofs for Section 6

Proof of Theorem 6.1. Because of the assumption $q = k$, we have $\mathcal{A} = \mathfrak{M}_0 = \{\mu_0\}$, where $\mu_0 = X\beta_0$ and β_0 is defined as $R^{-1}r$. Observe that $y^*(y,\xi) = X\beta_0+\operatorname{diag}(\xi)(y-X\beta_0)$, and hence $y^*(\mu_0+e_i(n),\xi) = \mu_0+\xi_i e_i(n)$ for every $i = 1,\dots,n$. Furthermore, note that our assumption on Ξ is equivalent to $\Xi(\{\xi \in \mathbb{R}^n : \prod_{i=1}^n \xi_i \neq 0\}) = 1$.

(a) Because Assumption 1 is assumed in Theorem 5.1, we may conclude that $e_i(n) \notin \operatorname{span}(X)$ holds for every $i = 1,\dots,n$.[52] This shows that $\vartheta_{2,Het} = 0$. We now turn to $\vartheta_{1,Het}$: Consider $i = 1,\dots,n$, such that $e_i(n) \notin \mathsf{B}$ (if no such i exists $\vartheta_{1,Het} = 0$ follows immediately from our convention). Observe that $T^*_{Het}(\mu_0+e_i(n),\xi) = T_{Het}(\mu_0+\xi_i e_i(n))$ by definition and that, in view of $G(\mathfrak{M}_0)$-invariance of T_{Het} (Remark 3.2), this equals $T_{Het}(\mu_0+e_i(n))$ provided that $\xi_i \neq 0$ (cf. the argument at the beginning of the proof of Theorem A.1). By our assumption on Ξ, we can conclude that $\vartheta_{1,Het} = 0$. This now gives $\vartheta_{Het} = 1$.

(b) Consider first $i = 1,\dots,n$, such that $e_i(n) \in \operatorname{span}(X)$. Then $y^*(\mu_0+e_i(n),\xi) = \mu_0+\xi_i e_i(n) \in \operatorname{span}(X)$ also holds for every ξ, since $\mu_0 \in \mathfrak{M}_0 \subseteq \operatorname{span}(X)$ and since $\operatorname{span}(X)$ is a linear space. This shows that $\vartheta_{2,uc} = 0$. We turn now to $\vartheta_{1,uc}$: Consider $i = 1,\dots,n$, such that $e_i(n) \notin \operatorname{span}(X)$ (which has to be the case for at least one i in view of the assumption $k < n$). Observe that $T^*_{uc}(\mu_0+e_i(n),\xi) = T_{uc}(\mu_0+\xi_i e_i(n))$ by definition and that, in view of $G(\mathfrak{M}_0)$-invariance of T_{uc} (Remark 3.2), this equals $T_{uc}(\mu_0+e_i(n))$ provided that $\xi_i \neq 0$

[52] If $e_i(n) \in \operatorname{span}(X)$ would hold for some i, then there would exist a $k \times 1$ vector a, such that $Xa = e_i(n)$. It would follow that $(X'(\neg(i_1,\dots i_s)))'a = 0$, where $X'(\neg(i_1,\dots i_s))$ is the matrix appearing in Assumption 1. As a consequence, this assumption would be violated, a contradiction.

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

(cf. the argument at the beginning of the proof of Theorem A.1). By our assumption on Ξ, we can conclude that $\vartheta_{1,uc} = 0$. This now gives $\vartheta_{uc} = 1$.

(c) Consider $i = 1, \ldots, n$, such that $\mu_0 + e_i(n) \notin \tilde{\mathsf{B}}$ (if no such i exists $\tilde{\vartheta}_{Het} = 0$ follows immediately from our convention). Observe that $\tilde{T}^*_{Het}(\mu_0 + e_i(n), \xi) = \tilde{T}_{Het}(\mu_0 + \xi_i e_i(n))$ by definition and that, in view of $G(\mathfrak{M}_0)$-invariance of $\tilde{T}_{Het}$ (Remark 5.6), this equals $\tilde{T}_{Het}(\mu_0 + e_i(n))$ provided that $\xi_i \neq 0$ (cf. the argument at the beginning of the proof of Theorem A.1). By our assumption on Ξ, we can conclude that $\tilde{\vartheta}_{Het} = 1$.

(d) Consider $i = 1, \ldots, n$ such that $\mu_0 + e_i(n) \notin \mathfrak{M}_0$ (which is actually the case here for every i since $\mathfrak{M}_0 = \{\mu_0\}$). Observe that $\tilde{T}^*_{uc}(\mu_0 + e_i(n), \xi) = \tilde{T}_{uc}(\mu_0 + \xi_i e_i(n))$ by definition and that, in view of $G(\mathfrak{M}_0)$-invariance of $\tilde{T}_{uc}$ (Remark 5.6), this equals $\tilde{T}_{uc}(\mu_0 + e_i(n))$ provided that $\xi_i \neq 0$ (cf. the argument at the beginning of the proof of Theorem A.1). By our assumption on Ξ, we can conclude that $\tilde{\vartheta}_{uc} = 1$. □

E. Computational Details

For every Setting A, B, and C and every bootstrap-based test procedure (i.e., combination of test statistic and bootstrap scheme) the two-step procedure of computations outlined in Section 8.2 proceeds as detailed in the following two subsections.

E.1. *Step 1: Search for Design Matrices Leading to a Small ϑ.* In every scenario, i.e., for every combination of $k = 2, \ldots, 5$ and $q = 1, \ldots, k-1$, we do the following: In a loop, we generate $n \times k$ -dimensional design matrices X with first column the intercept, and the remaining coordinates drawn independently from a log-(standard) normal distribution. For every X generated in this way, we determine the corresponding value of ϑ for testing the restriction $R = (0 : I_q)$ and $r = 0$ (after checking whether X has full rank and satisfies the assumption in the theorem corresponding to ϑ [if applicable]).[53] If a matrix X is found such that the corresponding ϑ satisfies $\vartheta < 0.01$, or if 150 design matrices have been generated in total we stop the loop.

Then, we determine a matrix $X_*(k,q)$, say, that realizes the minimal value of ϑ among the (at most 150) matrices considered. In preparation for Step 2, for $X_*(k,q)$ (and the restriction given by the associated q) we also compute, by Monte Carlo based on 300 replications for computing each probability (and assuming normality), the null rejection probabilities $\pi_{\alpha,\rho}$, say, of the bootstrap-based test under consideration for $\alpha = 0.05$ and $\alpha = 0.1$ under the parameters $\beta = 0$, $\sigma^2 = 1$, and $\Sigma(\rho, i^*) = \operatorname{diag}(\tau_1^2(\rho, i^*), \ldots, \tau_n^2(\rho, i^*))$ given by

$$\tau_{i^*}^2(\rho, i^*) = \rho, \text{ and } \tau_i^2(\rho, i^*) = (1-\rho)/(n-1) \text{ for } i \neq i^*, \tag{49}$$

for $\rho \in \{n^{-1}, 0.9, 0.99, 0.999, 0.9999\}$ and where i^* denotes the first index which realizes the optimum in the optimization problem defining $\vartheta = \vartheta(X_*(k,q))$ in the appropriate theorem that applies to the bootstrap-based test under consideration. The rationale for computing $\pi_{\alpha,\rho}$ is the following: Inspection of the proofs of the just mentioned theorems shows that if $\vartheta < \alpha$ holds, then the null rejection probabilities of the bootstrap-based test converge to 1 along a sequence of $\Sigma_m \in \mathfrak{C}_{Het}$ converging to $e_{i^*}(n)e_{i^*}(n)'$. Thus if $\vartheta < \alpha$, the null

[53] That is, Assumption 1 (Assumption 2, respectively) is checked if the test under consideration falls under the regime of Theorem 5.1 (Theorem 5.7 or 5.9, respectively). All these checks, including the full rank check, were always passed (which should not come as a surprise as these conditions are generically satisfied).

rejection probability $\pi_{\alpha,\rho}$ should be large for the variance specification considered in (49) and ρ close to 1. This will be exploited as a numerical check in Step 2.

Note also that $\rho = n^{-1}$ corresponds to homoskedastic errors.

For $n \in \{20, 30\}$ (i.e., in Settings B and C) the *empirical* distribution $\Xi^{\bullet}$ used in the bootstrap scheme is generated once for each combination of k and q and then held fixed in the computations of ϑ and the rejection probabilities described before.

E.2. *Step 2: Numerical Check and Computation of Additional Size Lower Bounds if Necessary.* First, we determine the matrix X_{**}, say, that corresponds to the minimal value of ϑ among the matrices $\{X_*(k,q) : k = 2,\dots,5,\ q = 1,\dots,k-1\}$ determined in Step 1. We then conduct the following numerical check: if, for $\alpha = 0.05$ or $\alpha = 0.1$, the value of ϑ corresponding to X_{**} (and the associated q) was smaller than α, but $\max_\rho \pi_{\alpha,\rho} < 0.4$, we took this as an indication of numerical unreliability of ϑ (cf. the discussion after (49) for an explanation, and see the discussion in Section E.3 below concerning some numerical issues that make determining ϑ nontrivial). If the value of ϑ corresponding to X_{**} was deemed unreliable, we replaced X_{**} by the matrix that led to the minimal value of ϑ among all matrices in $\{X_*(k,q) : k = 2,\dots,5,\ q = 1,\dots,k-1\} \backslash \{X_{**}\}$, and iterated this procedure until the check was passed (which eventually was always the case before the set of all matrices $X_*(k,q)$ was exhausted). Typically, the check was passed right away.[54] Denote by $\vartheta_{\min}$ the value of ϑ corresponding to the so-obtained matrix X_{**}.

Once the check was passed, we proceeded as follows, separately for $\alpha = 0.05$ and $\alpha = 0.1$: On the one hand, if $\vartheta_{\min} < \alpha$ (and the maximal rejection probability computed thus was guaranteed to be at least 0.4 by the numerical check), or if $\max_\rho \pi_{\alpha,\rho} \geq 3\alpha$, no additional computations were carried out, which was the case for the vast majority of test procedures.[55] In this case, the values of $\vartheta_{\min}$ and of $\max_\rho \pi_{\alpha,\rho}$ are reported. Note that if $\vartheta_{\min} < \alpha$ our theoretical results indicate that the test has size 1 (in the scenario corresponding to X_{**}); and if $\max_\rho \pi_{\alpha,\rho} \geq 3\alpha$, the null rejection probability is exceedingly large (even if $\vartheta_{\min} \not< \alpha$) and thus the test is also not reliable. On the other hand, if $\vartheta_{\min} \geq \alpha$ and $\max_\rho \pi_{\alpha,\rho} < 3\alpha$, we started a second set of computations to determine null rejection probabilities of the test over a range of additional design matrices.

In this second set of computations, we focus exclusively on the scenario k and q pertaining to X_{**}. In this scenario we first generate in a loop further design matrices in the same way as in Step 1, and compute for each matrix the value of ϑ (after checking whether X has full rank and satisfies the assumption in the theorem corresponding to ϑ [if applicable]).[56] If one of these new design matrices leads to a ϑ smaller than α, or once 150 matrices have been considered, we stop the loop. The matrix corresponding to the smallest value of ϑ is then used as the starting value in the following routine, in addition to (at most) 29 newly generated design matrices (of dimension $n \times k$) that are generated in the same way as the design matrices in Step 1:

[54] More specifically, in Settings A–C no replacements were necessary for 99.38%, 98.96%, and 98.33%, respectively, of the 960 bootstrap based test procedures considered.

[55] For $\alpha = 0.05$ no additional computations were necessary for 921, 904, and 887 of the 960 bootstrap based test procedures in Settings A–C, respectively; for $\alpha = 0.1$ no additional computations were necessary for 947, 926, and 920 of the 960 bootstrap based test procedures in Settings A–C, respectively

[56] Again all these checks, including the full rank check, were always passed.

(I) For every X considered we compute via a Monte Carlo method (again based on 300 replications and normality) the maximal null rejection probability for $\beta = 0, \sigma^2 = 1$ and $\Sigma = \operatorname{diag}(\tau_1^2, \dots, \tau_n^2)$ where the vector $(\tau_1^2, \dots, \tau_n^2)'$ varies in the set of vectors that is obtained from

$$\{(\tau_1^2(0.99, i), \dots, \tau_n^2(0.99, i))' + s|G| : i = 1, \dots, n,\ s = 0, 0.1, 0.5\},$$

after dividing each vector by its l_1-norm (cf. (49)); here the vector $|G|$ denotes coordinate-wise absolute values of G, and the latter is obtained as a draw from an n-variate standard normal distribution (here G was re-drawn for each X, i, and s). Once an X is found such that the maximal null rejection probability exceeds 4α, or once all 30 design matrices have been considered, we stop these computations.

(II) Denote by X_{***} the matrix for which the largest rejection probability was obtained in (I). All of the following computations are done on X_{***}.

If the largest rejection probability in (I) is greater than 4α, we re-compute this rejection probability on a new Monte Carlo sample (of size 300 and under normality) and stop.

Otherwise, we run (at most) 20 iterations of a Nelder–Mead optimization algorithm initialized at the vector $(\tau_1^2, \dots, \tau_n^2)'$ that leads to the maximal rejection probability in (I) to maximize the rejection probabilities further (all probabilities are again determined by Monte Carlo with 300 replications assuming normality each time a rejection probability is required during the Nelder–Mead optimization). After these iterations, the rejection probability for the "optimal" $(\tau_1^2, \dots, \tau_n^2)'$ so obtained is re-computed on a new Monte Carlo sample (of size 300 and assuming normality).

In both cases we report the last null rejection probability computed. Furthermore, we compute the value of ϑ corresponding to X_{***} (and the restrictions implied by the value of q under consideration) and report this value.[57,58]

In the second set of computations in Step 2, for $n \in \{20, 30\}$, and in contrast to Step 1, whenever a new design matrix X was considered, the empirical distribution $\Xi^{\bullet}$ used in the bootstrap scheme was newly initialized, and computations of ϑ and of rejection probabilities for this X were done based on this corresponding empirical distribution $\Xi^{\bullet}$.

As a point of interest we note that for every bootstrap-based test we also performed the following check: For every design matrix X (with associated scenario (k, q)) that underlies a result shown in Figure 1 or Tables 1–3 we have checked that all the relevant assumptions for size-controllability of the corresponding *non*-bootstrap-based test given in Pötscher and Preinerstorfer (2021) are satisfied.[59] Hence, for all these design matrices X (and their associated scenarios (k, q)), that we have shown to result in oversized bootstrap-based tests in the vast majority of cases, the testing procedures established in Pötscher and Preinerstorfer (2021), however, work and deliver valid tests, i.e., tests that are *not* oversized.

[57] The matrix X_{***} was also checked to have full rank and to satisfy the assumption in the theorem corresponding to ϑ (if applicable). These checks were always passed.

[58] The values of ϑ computed in the second set of computations are not subject to a numerical reliability check. While we report these (unvetted) values of ϑ as auxiliary information also in cases where this second set of computations are needed, we base our classification of the corresponding bootstrap-based test procedure as reliable or unreliable only on the magnitude of the computed null rejection probabilities and not on the value of such a ϑ. See Section 8.3.

[59] We have also checked that all these design matrices X satisfy $h_{ii} < 1$ for all $i = 1, \dots, n$ (which implies some, but not all, of the before mentioned conditions in Pötscher and Preinerstorfer, 2021).

https://doi.org/10.1017/S0266466622000184 Published online by Cambridge University Press

E.3. *Numerical Details Concerning the Computation of* ϑ. Evaluating ϑ numerically is a nontrivial task. In order to be on the safe side and to bias our results in favor of the bootstrap-based tests (recall we are after negative results), the ϑ we shall obtain will actually be a numerical obtained upper bound for the true ϑ (as can be seen from the subsequent discussion). In the following we discuss how the routines implemented in the R-package **wbsd** address the numerical challenges encountered in determining ϑ. We give the discussion for the case where ϑ_{Het} is to be computed. The routines proceed similarly for the other cases.

As can be seen from Theorem 5.1, to obtain $\vartheta_{1,Het}$ one repeatedly needs to compute $T_{Het}(z)$, after checking that $z \notin \mathsf{B}$, for various vectors z, say. Because this check is needed anyway for computing $T_{Het}(z)$ (see (5)), it is carried out in **wbsd** within the sub-routine computing $T_{Het}(z)$. To this end, the subroutine uses the function “isInvertible” associated to the rank-revealing LU decomposition (with complete pivoting) from the Eigen-library for linear algebra in C++ (recall that $z \notin \mathsf{B}$ is equivalent to $\hat{\Omega}_{Het}(z)$ being invertible). The package **wbsd** makes use of this library through the package **RcppEigen** by Bates and Eddelbuettel (2013). The function “isInvertible” requires the specification of a tolerance parameter, which we set to 10^{-6}. Note that the function “isInvertible” works in such a way that the larger the tolerance parameter, the more likely it is that the input matrix is categorized as numerically noninvertible (i.e., that z is categorized as satisfying $z \in \mathsf{B}$). As a consequence, the larger the tolerance parameter, the smaller the numerically determined value of $\vartheta_{1,Het}$, because violated rank conditions decrease $\vartheta_{1,Het}$.[60]

In addition to rank computations, the definition of $\vartheta_{1,Het}$ requires to numerically check the inequality in the events defining $\vartheta_{1,Het}$, see (10). To this end, we introduce another tolerance parameter and instead of checking “$\cdots < \cdots$” directly, we checked “$\cdots + 10^{-5} < \cdots$.” Note that also here, increasing the tolerance parameter decreases the numerically determined value of $\vartheta_{1,Het}$.

In the computation of $\vartheta_{2,Het}$, checks of the form $z \notin \mathsf{B}$ were done exactly in the same way as in the computation of $\vartheta_{1,Het}$ described above.[61] Again, the larger the tolerance parameter the more likely it is that the input matrix becomes numerically rank deficient. Thus, increasing the tolerance parameter leads to a possible decrease in the numerically determined value of $\vartheta_{2,Het}$. The computation of $\vartheta_{2,Het}$ also requires checking whether $R\hat{\beta}(e_i(n)) \neq 0$. Numerically, we checked this via determining whether $\|R\hat{\beta}(e_i(n))\|_\infty > 10^{-6}$ where $\|\cdot\|_\infty$ denotes the maximum norm (again the larger the tolerance parameter, the smaller the numerically determined value of $\vartheta_{2,Het}$). To compute $\vartheta_{2,Het}$ one also needs to check if $\operatorname{rank}((X : e_i(n))) < k+1$. These rank computations are implemented in **wbsd** based on the rank-revealing LU decomposition (with complete pivoting) and the corresponding function “rank” obtained from the Eigen-library mentioned above. Checking rank conditions via the function “rank” requires the specification of a tolerance parameter. To check $\operatorname{rank}((X : e_i(n))) < k+1$, we chose the tolerance parameter equal to the machine epsilon 2.220446×10^{-16}, because—in contrast to the previous checks—decreasing the tolerance parameter used in this check decreases the numerically determined value of $\vartheta_{2,Het}$. [Note that decreasing $\vartheta_{1,Het}$ or $\vartheta_{2,Het}$ increases ϑ_{Het}.]

[60] For the tests T_{uc} ($\tilde{T}_{uc}$, respectively) we checked whether the residual variance estimate $\hat{\sigma}^2(z)$ ($\tilde{\sigma}^2(z)$, respectively) exceeds 10^{-6} instead of using the function “isInvertible.”

[61] When computing $\vartheta_{2,uc}$, we checked whether the residual variance estimate $\hat{\sigma}^2(z)$ exceeded 10^{-6} instead of using the function “isInvertible.”

Theorem 5.1 requires Assumption 1 to hold, which (as noted above) was checked throughout. This condition can be verified by a series of rank computations, which was done based on the function “rank” as discussed above, but with a tolerance parameter of 10^{-7}. A similar remark applies to Theorems 5.7 and 5.9 with regard to Assumption 2.

REFERENCES

Bates, D. & D. Eddelbuettel (2013) Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software* 52, 1–24.

Beran, R. (1986). Discussion: Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1295–1298.

Chesher, A. & I. Jewitt (1987) The bias of a heteroskedasticity consistent covariance matrix estimator. *Econometrica* 55, 1217–1222.

Chesher, A.D. (1989) Hájek inequalities, measures of leverage, and the size of heteroskedasticity robust Wald tests. *Econometrica* 57, 971–977.

Chesher, A.D. & G. Austin (1991) The finite-sample distributions of heteroskedasticity robust Wald statistics. *Journal of Econometrics* 47, 153–173.

Cragg, J.G. (1983) More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* 51, 751–763.

Cragg, J.G. (1992) Quasi-Aitken estimation for heteroscedasticity of unknown form. *Journal of Econometrics* 54, 179–201.

Cribari-Neto, F. (2004) Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* 45, 215–233.

Cribari-Neto, F. & M.d.G.A. Lima (2009) Heteroskedasticity-consistent interval estimators. *Journal of Statistical Computation and Simulation* 79, 787–803.

Davidson, R. & E. Flachaire (2008) The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162–169.

Davidson, R. & J.G. MacKinnon (1985) Heteroskedasticity-robust tests in regressions directions. *Annales de l’INSÉÉ* 59/60, 183–218.

DiCiccio, C.J., J.P. Romano, & M. Wolf (2019) Improving weighted least squares inference. *Econometrics and Statistics* 10, 96–119.

Eicker, F. (1963) Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics* 34, 447–456.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, vol. I: Statistics, pp. 59–82. University of California Press.

Flachaire, E. (1999) A better way to bootstrap pairs. *Economics Letters* 64, 257–262.

Flachaire, E. (2005a) Bootstrapping heteroskedastic regression models: Wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis* 49, 361–376.

Flachaire, E. (2005b) More efficient tests robust to heteroskedasticity of unknown form. *Econometric Reviews* 24, 219–241.

Godfrey, L.G. (2006) Tests for regression models with heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* 50, 2715–2733.

Godfrey, L.G. & C.D. Orme (2004) Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients. *Economics Letters* 82, 281–287.

Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics* 19, 285–292.

Horowitz, J.L. (1997). Bootstrap methods in econometrics: Theory and numerical performance. In D.M. Kreps & K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications: Seventh World Congress*, Econometric Society Monographs, 3, pp. 188–222. Cambridge University Press.

Imbens, G.W. & M. Kolesár (2016) Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics* 98, 701–712.

Lin, E.S. & T.-S. Chou (2018) Finite-sample refinement of GMM approach to nonlinear models under heteroskedasticity of unknown form. *Econometric Reviews* 37, 1–28.

Liu, R.Y. (1988) Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics* 16, 1696–1708.

Long, J.S. & L.H. Ervin (2000) Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician* 54, 217–224.

MacKinnon, J.G. (2013) Thirty years of heteroskedasticity-robust inference. In X. Chen & N.R.E. Swanson (eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 437–462. Springer.

MacKinnon, J.G. & H. White (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29, 305–325.

Mammen, E. (1993) Bootstrap and wild bootstrap for high-dimensional linear models. *Annals of Statistics* 21, 255–285.

Pötscher, B.M. & D. Preinerstorfer (2018) Controlling the size of autocorrelation robust tests. *Journal of Econometrics* 207, 406–431.

Pötscher, B.M. & D. Preinerstorfer (2021) Valid Heteroskedasticity Robust Testing. Working Paper.

Preinerstorfer, D. (2020) wbsd: Wild Bootstrap Size Diagnostics, version 1.0.0.

Preinerstorfer, D. & B.M. Pötscher (2016) On size and power of heteroskedasticity and autocorrelation robust tests. *Econometric Theory* 32, 261–358.

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Richard, P. (2017) Robust heteroskedasticity-robust tests. *Economics Letters* 159, 28–32.

Robinson, G. (1979) Conditional properties of statistical procedures. *Annals of Statistics* 7, 742–755.

Romano, J.P. & M. Wolf (2017) Resurrecting weighted least squares. *Journal of Econometrics* 197, 1–19.

Rothenberg, T.J. (1988) Approximate power functions for some robust tests of regression coefficients. *Econometrica* 56, 997–1019.

van Giersbergen, N.P.A. & J.F. Kiviet (2002) How to implement the bootstrap in static or stable dynamic regression models: Test statistic versus confidence region approach. *Journal of Econometrics* 108, 133–156.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

Wu, C.-F.J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261–1350. With discussion and a rejoinder by the author.