On Size and Power of Heteroscedasticity and Autocorrelation Robust Tests

Preinerstorfer, David and Pötscher, Benedikt M.

Department of Statistics, University of Vienna

January 2013

On Size and Power of Heteroscedasticity and Autocorrelation Robust Tests [*]

David Preinerstorfer and Benedikt M. Pötscher[†]

Department of Statistics, University of Vienna

Abstract

Testing restrictions on regression coefficients in linear models often requires correcting the conventional F-test for potential heteroscedasticity or autocorrelation amongst the disturbances, leading to so-called heteroskedasticity and autocorrelation robust test procedures. These procedures have been developed with the purpose of attenuating size distortions and power deficiencies present for the uncorrected F-test. We develop a general theory to establish positive as well as negative finite-sample results concerning the size and power properties of a large class of heteroskedasticity and autocorrelation robust tests. Using these results we show that nonparametrically as well as parametrically corrected F-type tests in time series regression models with stationary disturbances have either size equal to one or nuisance-infimal power equal to zero under very weak assumptions on the covariance model and under generic conditions on the design matrix. In addition we suggest an adjustment procedure based on artificial regressors. This adjustment resolves the problem in many cases in that the so-adjusted tests do not suffer from size distortions. At the same time their power function is bounded away from zero. As a second application we discuss the case of heteroscedastic disturbances.

AMS Mathematics Subject Classification 2010: 62F03, 62J05, 62F35, 62M10, 62M15

Keywords: Size distortion, power deficiency, invariance, robustness, autocorrelation, heteroscedasticity, HAC, fixed-bandwidth, long-run-variance, feasible GLS

1 Introduction

So-called autocorrelation robust tests have received considerable attention in the econometrics literature in the last two and a half decades. These tests are Wald-type tests which make use of an appropriate nonparametric variance estimator that tries to take into account the autocorrelation in the data. The early papers on such nonparametric variance estimators in econometrics date from the late 1980s and early 1990s (see, e.g., Newey and West (1987, 1994), Andrews (1991), and Andrews and Monahan (1992)) and typically consider consistent variance estimators. The ideas

---

and techniques underlying this literature derive from the much earlier literature on spectral estimation and can be traced back to work by Bartlett (1950), Jowett (1955), Hannan (1957), and Grenander and Rosenblatt (1957), the latter explicitly discussing what would now be called autocorrelation robust tests and confidence intervals (Section 7.9 of Grenander and Rosenblatt (1957)). For book-length treatments of spectral estimation see the classics Hannan (1970) or Anderson (1971). Autocorrelation robust tests for the location parameter also play an important rôle in the field of simulation, see, e.g., Heidelberger and Welch (1981) or Flegal and Jones (2010). In a similar vein, so-called heteroskedasticity robust variance estimators and associated tests have been invented by Eicker (1963, 1967) and have later been introduced into the econometrics literature. As mentioned before, the autocorrelation robust test statistics considered in the above cited econometrics literature employ consistent variance estimators leading to an asymptotic chi-square distribution under the null. It soon transpired from Monte Carlo studies that these tests (using as critical values the quantiles of the asymptotic chi-square distribution) are often severely oversized in finite samples. This has led to the proposal to use a test statistic of the same form, but to obtain the critical values from another (nuisance parameter-free) distribution which arises as the limiting distribution in an alternative asymptotic framework ("fixed bandwidth asymptotics") in which the variance estimator is no longer consistent, see Kiefer et al. (2000), Kiefer and Vogelsang (2002a,b, 2005). The idea of using "fixed bandwidth asymptotics" can be traced back to earlier work by Neave (1970). Monte Carlo studies have shown that these tests typically are also oversized, albeit less so than the tests mentioned earlier. This improvement, however, is often achieved at the expense of some loss of power. In an attempt to better understand size and power properties of autocorrelation robust tests, higher-order asymptotic properties of these tests have been studied (Velasco and Robinson (2001), Jansson (2004), Sun et al. (2008, 2011), Zhang and Shao (2013)).

The first-order as well as the higher-order asymptotic results in the literature cited above are all *pointwise* asymptotic results in the sense that they are derived under the assumption of a *fixed* underlying data-generating process (DGP). Therefore, while these results tell us something about the limit of the rejection probability, or the rate of convergence to this limit, for a *fixed* underlying DGP, they do not necessarily inform us about the *size* of the test or its asymptotic behavior (e.g., limit of the size as sample size increases) nor about the *power function* or its asymptotic behavior. The reason is that the asymptotic results do not hold uniformly in the underlying DGP under the typical assumptions on the feasible set of DGPs in this literature. Of course, one could restrict the set of feasible DGPs in such a way that the asymptotic results hold uniformly, but this would require the imposition of unnatural and untenable assumptions on the set of feasible DGPs as will transpire from the subsequent discussion; cf. also Subsection 3.1.2.

In Section 3 of the present paper we provide a theoretical finite-sample analysis of the size and power properties of autocorrelation robust tests for linear restrictions on the parameters in a linear regression model with autocorrelated errors. Being finite-sample results, the findings of the paper apply equally well regardless of whether we fancy that the variance estimator being used would be consistent or not would sample size go to infinity. Under a mild assumption on the richness of the set of allowed autocorrelation structures in the maintained model, the results in Section 3 imply that in most cases the size of common autocorrelation robust tests is 1 or that the worst case power is 0 (or both). The richness assumption just mentioned only amounts to requiring that all correlation structures corresponding to stationary Gaussian autoregressive processes of order 1 are allowed for in the model. Compared to the much wider assumptions on the DGP appearing in the literature on autocorrelation robust tests cited above, this certainly is a very mild assumption. [Not including all stationary Gaussian autoregressive models of order 1 into the set of feasible disturbance

2

processes appears to be an unnatural restriction in a theory of autocorrelation robust tests, cf. also the discussion in Subsection 3.1.2.] A similar negative result is derived for tests that do not use a nonparametric variance estimator but use a variance estimator derived from a parametric model as well as for tests based on a feasible generalized least squares estimator (Subsection 3.2). We also show that the just mentioned negative results hold generically in the sense that, given the linear restrictions to be tested, the set of design matrices such that the negative results do *not* apply is a negligible set (Propositions 3.5 and 3.15). Furthermore, we provide a positive result in that we isolate conditions (on the design matrix and on the restrictions to be tested) such that the size of the test can be controlled. While this result is obtained under the strong assumption that the set of feasible correlation structures coincides with the correlation structures of all stationary autoregressive process of order 1, it should be noted that the negative results equally well hold under this parametric correlation model. The positive result just mentioned is then used to show how for the majority of testing problems autocorrelation robust tests can be adjusted in such a way that they do not suffer from the "size equals 1" and the "worst case power equals 0" problem. In Section 4 we provide an analogous negative result for heteroskedasticity robust tests and discuss why a (nontrivial) positive result is not possible.

We next discuss some related literature. Problems with tests and confidence sets for the intercept in a linear regression model with autoregressive disturbances have been pointed out in Section 5.3 of Dufour (1997) (in a somewhat different setup). These results are specific to testing the intercept and do not apply to other linear restrictions. This is, in particular witnessed by our positive results for certain testing problems. Furthermore, there is a considerable body of literature concerned with the properties of the *standard* F-test (i.e., the F-test constructed without any correction for autocorrelation) in the presence of autocorrelation, see the references cited in Krämer et al. (1990) and Banerjee and Magnus (2000). Much of this literature concentrates on the case where the errors follow a stationary autoregressive process of order 1. As the correlation in the errors is not accounted for when considering the standard F-test, it is not too surprising that the standard F-test typically shows deplorable performance for large values of the autocorrelation coefficient ρ, see Krämer (1989), Krämer et al. (1990), Banerjee and Magnus (2000), and Subsection 3.3 for more discussion. Section 3 of the present paper shows that autocorrelation robust tests, which despite having built into them a correction for autocorrelation, exhibit a similarly bad behavior. Finally, in a different testing problem (the leading case being testing the correlation of the errors in a spatial regression model) Martellosio (2010) has studied the power of a class of invariant tests including standard tests like the Cliff-Ord test and observed somewhat similar results in that the power of the tests considered typically approaches (as the strength of the correlation increases) either 0 or 1. While his results are similar in spirit to some of our results, his arguments are unfortunately fraught with a host of problems. See Preinerstorfer and Pötscher (2013) for discussion, corrections, and extensions.

The results in Section 3 for autocorrelation robust tests and in Section 4 for heteroskedasticity robust tests are derived as special cases of a more general theory for size and power properties of a larger class of tests that are invariant under a particular group of affine transformations. This theory is provided in Section 5. One of the mechanisms behind the negative results in the present paper is a concentration mechanism explained subsequent to Theorem 3.3 and in more detail in Subsection 5.2, cf. also Corollary 5.17. A second mechanism generating negative results is described in Theorem 5.19. The theory underlying the positive results mentioned above is provided in Subsection 5.3 and in Theorem 5.21 as well as Proposition 5.23. Furthermore, the results in Section 5 allow for covariance structures more general than the ones discussed in Sections 3 and 4.

For example, from the results in Section 5 results similar to the ones in Section 3 could be derived for heteroskedasticity/autocorrelation robust tests of regression coefficients in spatial regression models or in panel data models; for an overview of heteroskedasticity/autocorrelation robust tests in these models see Kelejian and Prucha (2007, 2010), and Vogelsang (2012). We do not provide any such results for lack of space. We note that for the uncorrected standard F-test in this setting negative results have been derived in Krämer (2003) and Krämer and Hanck (2009).

2   The Hypothesis Testing 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 $\beta \in \mathbb{R}^k$ denotes the unknown regression parameter vector. We assume rank$(X) = k$ and $1 \leq k < n$. The $n \times 1$ disturbance vector $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)'$ is normally distributed with mean zero and unknown covariance matrix $\sigma^2 \Sigma$, where $0 < \sigma^2 < \infty$ holds (and σ always denotes the positive square root). The matrix Σ varies in a prescribed (nonempty) set $\mathfrak{C}$ of symmetric and positive definite $n \times n$ matrices.[1] Throughout the paper we make the assumption that $\mathfrak{C}$ is such that σ^2 and $\Sigma \in \mathfrak{C}$ can be uniquely determined from $\sigma^2 \Sigma$. [For example, if the first diagonal element of each $\Sigma \in \mathfrak{C}$ equals 1 this is satisfied; alternatively, if the largest diagonal element or the trace of each $\Sigma \in \mathfrak{C}$ is normalized to a fixed constant, $\mathfrak{C}$ has this property.] Of course, this assumption entails little loss of generality and can, if necessary, always be achieved by a suitable reparameterization of $\sigma^2 \Sigma$.

The linear model described above induces a collection of distributions on $\mathbb{R}^n$, the sample space of $\mathbf{Y}$. Denoting a Gaussian probability measure with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\sigma^2 \Sigma$ by $P_{\mu,\sigma^2\Sigma}$ and setting $\mathfrak{M} = \mathrm{span}(X)$, the induced collection of distributions is given by

$$\left\{ P_{\mu,\sigma^2\Sigma} : \mu \in \mathfrak{M}, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C} \right\}. \tag{2}$$

Note that each $P_{\mu,\sigma^2\Sigma}$ is absolutely continuous with respect to (w.r.t.) Lebesgue measure on $\mathbb{R}^n$, since every $\Sigma \in \mathfrak{C}$ is positive definite by assumption. We consider the problem of testing a linear (better: affine) restriction on the parameter vector $\beta \in \mathbb{R}^k$, namely the problem of testing the null $R\beta = r$ versus the alternative $R\beta \neq r$, where R is a $q \times k$ matrix of rank q and $r \in \mathbb{R}^q$. To be more precise and to emphasize that the testing problem is in fact a compound one, the testing problem needs to be written as

$$H_0 : R\beta = r, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C} \quad \text{vs.} \quad H_1 : R\beta \neq r, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C}. \tag{3}$$

This is important to stress because size and power properties of tests critically depend on nuisance parameters and, in particular, on the complexity of $\mathfrak{C}$. Define the affine space

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

and let

$$\mathfrak{M}_1 = \mathfrak{M} \backslash \mathfrak{M}_0 = \{ \mu \in \mathfrak{M} : \mu = X\beta \text{ and } R\beta \neq r \}.$$

---

[1] Although not expressed in the notation, 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. Furthermore, the obvious dependence of C on n will also not be shown in the notation. [Note that C depends on n even if it is induced by a covariance model for the entire process $(\mathbf{u}_t)_{t \in \mathbb{N}}$ that does not depend on n.]

Adopting these definitions, the above hypothesis can also be written as

$$H_0 : \mu \in \mathfrak{M}_0, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C} \quad \text{vs.} \quad H_1 : \mu \in \mathfrak{M}_1, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C}. \tag{4}$$

Two remarks are in order: First, the Gaussiantiy assumption is not really a restriction for the negative results in the paper, since they hold a fortiori in any enlarged model that allows not only for Gaussian but also for non-Gaussian disturbances. Furthermore, a large portion of the results in the paper (positive or negative) continue to hold for certain classes of non-Gaussian distributions such as, e.g., elliptical distributions, see Subsection 5.5. Second, if X were allowed to be stochastic but independent of $\mathbf{U}$, the results of the paper apply to size and power conditional on X. Because X is observable, one could then argue in the spirit of conditional inference (see, e.g., Robinson (1979)) that conditional size and power and not their unconditional counterparts are the more relevant characteristics of a test.

Recall that a (randomized) test is a Borel-measurable function φ from the sample space $\mathbb{R}^n$ to $[0,1]$. If $\varphi = \mathbf{1}_W$, the set W is called the rejection region of the test. As usual, the size of a test φ is the supremum over all rejection probabilities under the null hypothesis H_0 and thus is given by $\sup_{\mu \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{\Sigma \in \mathfrak{C}} E_{\mu, \sigma^2 \Sigma} \varphi$ where $E_{\mu, \sigma^2 \Sigma}$ refers to expectation under the probability measure $P_{\mu, \sigma^2 \Sigma}$.

Throughout the paper we shall always reserve the symbol $\hat{\beta}(y)$ for $(X'X)^{-1} X'y$, where X is the design matrix appearing in (1) and $y \in \mathbb{R}^n$. Furthermore, random vectors and random variables are always written in bold capital and bold lower case letters, respectively. Lebesgue measure on $\mathbb{R}^n$ will be denoted by $\lambda_{\mathbb{R}^n}$, whereas Lebesgue measure on an affine subspace $\mathcal{A}$ of $\mathbb{R}^n$ (but viewed as a measure on the Borel-sets of $\mathbb{R}^n$) will be denoted by $\lambda_{\mathcal{A}}$, with zero-dimensional Lebesgue measure being interpreted as point mass. We shall write $\text{int}(A)$, $\text{cl}(A)$, and $\text{bd}(A)$ for the interior, closure and boundary of a set $A \subseteq \mathbb{R}^n$, respectively, taken with respect to the Euclidean topology. The Euclidean norm is denoted by $\|\cdot\|$, while $d(x, A)$ denotes the Euclidean distance of the point $x \in \mathbb{R}^n$ to the set $A \subseteq \mathbb{R}^n$. Let B' denote the transpose of a matrix B and let $\text{span}(B)$ denote the space spanned by the columns of B. 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 on $\mathcal{L}$. For a vector x in Euclidean space we define the symbol $\langle x \rangle$ to denote $\pm x$ for $x \neq 0$, the sign being chosen in such a way that the first nonzero component of $\langle x \rangle$ is positive, and we set $\langle 0 \rangle = 0$. The j-th standard basis vector in $\mathbb{R}^n$ is denoted by $e_j(n)$. The set of real matrices of dimension $m \times n$ is denoted by $\mathbb{R}^{m \times n}$. We also introduce the following terminology.

Definition 2.1. Let $\mathfrak{C}$ be a set of symmetric and positive definite $n \times n$ matrices. An l-dimensional linear subspace $\mathcal{Z}$ of $\mathbb{R}^n$ with $0 \leq l < n$ is called a *concentration space* of $\mathfrak{C}$, if there exists a sequence $(\Sigma_m)_{m \in \mathbb{N}}$ in $\mathfrak{C}$, such that $\Sigma_m \to \bar{\Sigma}$ and $\text{span}(\bar{\Sigma}) = \mathcal{Z}$.

While we shall in the sequel often refer to $\mathfrak{C}$ as the covariance model, one should keep in mind that the set of all feasible covariance matrices corresponding to (2) is given by $\{\sigma^2 \Sigma : 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C}\}$. In this context we note that two covariance models $\mathfrak{C}$ and $\mathfrak{C}^*$ can be equivalent in the sense of giving rise to the same set of feasible covariance matrices, but need not have the same concentration spaces.[2]

---

[2] In applying the general results in Section 5.2 or Corollary 5.17 to a particular problem some skill in choosing between equivalent C and C* may thus be required as one choice for C may lead to more interesting results than does another choice.

3 Size and Power of Tests of Linear Restrictions in Regression Models with Autocorrelated Disturbances

In this section we investigate size and power properties of autocorrelation robust tests that have been designed for use in case of stationary disturbances. Studies of the properties of such tests in the literature (Newey and West (1987, 1994), Andrews (1991), Andrews and Monahan (1992), Kiefer et al. (2000), Kiefer and Vogelsang (2002a,b, 2005), Jansson (2002, 2004), Sun et al. (2008, 2011)) maintain assumptions that allow for nonparametric models for the spectral distribution of the disturbances. For example, a typical nonparametric model results from assuming that the disturbance vector consists of n consecutive elements of a weakly stationary process with spectral density equal to

$$f(\omega) = (2\pi)^{-1} \left| \sum_{j=0}^{\infty} c_j \exp(-\iota j \omega) \right|^2,$$

where, for $\xi \geq 0$ a given number, the coefficients c_j satisfy the summability condition $\sum_{j=0}^{\infty} j^\xi |c_j| < \infty$. Here ι denotes the imaginary unit. Let $\mathfrak{F}_\xi$ denote the collection of all such spectral densities f. The corresponding covariance model $\mathfrak{C}_\xi$ is then given by $\{\Sigma(f) : f \in \mathfrak{F}_\xi\}$ where $\Sigma(f)$ is the $n \times n$ correlation matrix

$$\Sigma(f) = \left(\int_{-\pi}^{\pi} \exp(-\iota \omega (i-j)) f(\omega) d\omega \bigg/ \int_{-\pi}^{\pi} f(\omega) d\omega \right)_{i,j=1}^{n}.$$

Certainly, $\mathfrak{F}_\xi$ contains all spectral densities of stationary autoregressive moving average models of arbitrary large order. Hence, the following assumption on the covariance model $\mathfrak{C}$ that we shall impose for most results in this section is very mild and is satisfied by the typical nonparametric model allowed for in the above mentioned literature. It certainly covers the case where $\mathfrak{C} = \mathfrak{C}_\xi$ or where $\mathfrak{C}$ corresponds to an autoregressive model of order $p \geq 1$.

Assumption 1. $\mathfrak{C}_{AR(1)} \subseteq \mathfrak{C}$.

Here $\mathfrak{C}_{AR(1)}$ denotes the set of correlation matrices corresponding to n successive elements of a stationary autoregressive processes of order 1, i.e., $\mathfrak{C}_{AR(1)} = \{\Lambda(\rho) : \rho \in (-1,1)\}$ where the (i,j)-th entry in the $n \times n$ matrix $\Lambda(\rho)$ is given by $\rho^{|i-j|}$. As hinted at in the introduction, parameter values (μ, σ^2, Σ) with $\Sigma = \Lambda(\rho)$ where ρ gets close to ± 1 and σ^2 is constant will play an important rôle as they will be instrumental for establishing the bad size and power properties of the tests presented below. We want to stress here that, as $\rho \to \pm 1$, the corresponding stationary process does *not* converge to an integrated process but rather to a harmonic process. This is so because the variance of the disturbances σ^2 is held constant. [If we parameterized in terms of ρ and the innovation variance $\sigma_\varepsilon^2 = \sigma^2 (1 - \rho^2)$, this would correspond to $\sigma_\varepsilon^2 \to 0$ at the appropriate rate.]

For later use we note that under Assumption 1 the matrices $e_+ e_+'$ and $e_- e_-'$ are limiting points of the covariance model $\mathfrak{C}$ where $e_+ = (1, \ldots, 1)'$ and $e_- = (-1, 1, \ldots, (-1)^n)'$ are $n \times 1$ vectors (since $\Lambda(\rho_m)$ converges to $e_+ e_+'$ ($e_- e_-'$, respectively) if $\rho_m \to 1$ ($\rho_m \to -1$, respectively)). Other singular limiting points of $\mathfrak{C}$ are possible, but $e_+ e_+'$ and $e_- e_-'$ are the only singular limiting points of $\mathfrak{C}_{AR(1)}$.

The next subsection presents the results for common nonparametrically based autocorrelation robust tests whereas parametrically based tests are the subject of Subsection 3.2.

3.1 Nonparametrically based autocorrelation robust tests

Commonly used autocorrelation robust tests for the null hypothesis H_0 given by (3) are based on test statistics of the form $(R\hat{\beta}(y) - r)'\hat{\Omega}^{-1}(y)(R\hat{\beta}(y) - r)$, with the statistic typically being undefined if $\hat{\Omega}(y)$ is singular. Here

$$\hat{\Omega}(y) = nR(X'X)^{-1}\hat{\Psi}(y)(X'X)^{-1}R' \tag{5}$$

and $\hat{\Psi}$ is a nonparametric estimator for $n^{-1}\mathbb{E}(X'\mathbf{U}\mathbf{U}'X)$. The type of estimator $\hat{\Psi}$ we consider in this subsection is obtained as a weighted sum of sample autocovariances of $\hat{v}_t(y) = \hat{u}_t(y)x'_{t\cdot}$, where $\hat{u}_t(y)$ is the t-th coordinate of the least squares residual vector $\hat{u}(y) = y - X\hat{\beta}(y)$ and $x_{t\cdot}$ denotes the t-th row vector of X. That is

$$\hat{\Psi}(y) = \hat{\Psi}_w(y) = \sum_{j=-(n-1)}^{n-1} w(j,n)\hat{\Gamma}_j(y) \tag{6}$$

for every $y \in \mathbb{R}^n$ with $\hat{\Gamma}_j(y) = n^{-1}\sum_{t=j+1}^n \hat{v}_t(y)\hat{v}_{t-j}(y)'$ if $j \geq 0$ and $\hat{\Gamma}_j(y) = \hat{\Gamma}_{-j}(y)'$ else. The associated estimator $\hat{\Omega}$ will be denoted by $\hat{\Omega}_w$. We make the following assumption on the weights.

Assumption 2. *The weights $w(j,n)$ for $j = -(n-1), \ldots, n-1$ are data-independent and satisfy $w(0,n) = 1$ as well as $w(-j,n) = w(j,n)$. Furthermore, the symmetric $n \times n$ Toeplitz matrix $\mathcal{W}_n$ with elements $w(i-j,n)$ is positive definite.*[3]

The positive definiteness assumption on $\mathcal{W}_n$ is weaker than the frequently employed assumption that the Fourier transform $w\dagger(\omega)$ of the weights is nonnegative for all $\omega \in [-\pi, \pi]$.[4] It certainly implies that $\hat{\Psi}_w(y)$, and hence $\hat{\Omega}_w(y)$, is always nonnegative definite, but it will allow us to show more, see Lemma 3.1 below. In many applications the weights take the form $w(j,n) = w_0(|j|/M_n)$, where the lag-window w_0 is an even function with $w_0(0) = 1$ and where $M_n > 0$ is a truncation lag (bandwidth) parameter. In this case the first part of the above assumption means that we are considering deterministic bandwidths only (as is the case, e.g., in Newey and West (1987), Sections 3-5 of Andrews (1991), Hansen (1992), Kiefer and Vogelsang (2002b, 2005), and Jansson (2002, 2004)). Extensions of the results in this subsection to data-dependent bandwidth choices and prewhitening will be discussed in Preinerstorfer (2013). Assumption 2 is known to be satisfied, e.g., for the (modified) Bartlett, Parzen, or the Quadratic Spectral lag-window, but is not satisfied, e.g., for the rectangular lag-window (with $M_n > 1$).[5] See Anderson (1971) or Hannan (1970) for more discussion. It is also satisfied for many exponentiated lag-windows as used in Phillips et al. (2006, 2007) and Sun et al. (2011).

In the typical asymptotic analysis of this sort of tests in the literature the event where the estimator $\hat{\Omega}_w$ is singular is asymptotically negligible (as $\hat{\Omega}_w$ converges to a positive definite or almost surely positive definite matrix), and hence there is no need to be specific about the definition of the test statistic on this event. However, if one is concerned with finite-sample properties, one has

---

[3]For the case where $\mathcal{W}_n$ is only nonnegative definite see Subsection 3.1.1.

[4]Note that the quadratic form $\alpha'\mathcal{W}_n\alpha$ can be represented as $\int_{-\pi}^{\pi} \left|\sum_{j=1}^n \alpha_j \exp(\iota j\omega)\right|^2 w\dagger(\omega)\,d\omega$. If $w\dagger(\omega) \geq 0$ for all $\omega \in [-\pi, \pi]$ is assumed, the integrand is nonnegative; and if $\alpha \neq 0$ it is positive almost everywhere (since it is then a product of two nontrivial trigonometric polynomials).

[5]The estimator in Keener et al. (1991) coincides with (n times) the estimator given by (5) if the rectangular lag-window is used and $R = I_k$.

to think about the definition of the test statistic also in the case where $\hat{\Omega}_w(y)$ is singular. We thus define the test statistic as follows:[6]

$$T(y) = \begin{cases} (R\hat{\beta}(y) - r)'\hat{\Omega}_w^{-1}(y)(R\hat{\beta}(y) - r) & \text{if } \det\hat{\Omega}_w(y) \neq 0, \\ 0 & \text{if } \det\hat{\Omega}_w(y) = 0. \end{cases} \tag{7}$$

Of course, assigning the test statistic T the value zero on the set where $\hat{\Omega}_w(y)$ is singular is arbitrary. However, it will be irrelevant for size and power properties of the test *provided* we can ensure that the set of $y \in \mathbb{R}^n$ for which $\det\hat{\Omega}_w(y) = 0$ holds is a $\lambda_{\mathbb{R}^n}$-null set (since all relevant distributions $P_{\mu,\sigma^2\Sigma}$ are absolutely continuous w.r.t. $\lambda_{\mathbb{R}^n}$ due to the fact that every element of $\Sigma \in \mathfrak{C}$ is positive definite by assumption). We thus need to study under which circumstances this is ensured. This will be done in the subsequent lemma. It will prove useful to introduce the following matrix for every $y \in \mathbb{R}^n$

$$\begin{aligned} B(y) &= R(X'X)^{-1}X'\operatorname{diag}(\hat{u}_1(y), \ldots, \hat{u}_n(y)) \\ &= R(X'X)^{-1}X'\operatorname{diag}(e_1'(n)\Pi_{\operatorname{span}(X)^\perp}y, \ldots, e_n'(n)\Pi_{\operatorname{span}(X)^\perp}y), \end{aligned} \tag{8}$$

as well as the following assumption on the design matrix X (and on the restriction matrix R):

Assumption 3. *Let* $1 \leq i_1 < \ldots < i_s \leq n$ *denote all the indices for which* $e_{i_j}(n) \in \operatorname{span}(X)$ *holds where* $e_j(n)$ *denotes the j-th 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,* $1 \leq i_1 < \ldots < i_s \leq n$ *(if $s = 0$ no column is deleted). Then* $\operatorname{rank}(R(X'X)^{-1}X'(\neg(i_1, \ldots i_s))) = q$ *holds.*

The lemma is now as follows. Note that the matrix $B(y)$ does *not* depend on the weights $w(j, n)$.

Lemma 3.1. *Suppose Assumption 2 is satisfied. Then the following holds:*

1. $\hat{\Omega}_w(y)$ *is nonnegative definite for every* $y \in \mathbb{R}^n$.

2. $\hat{\Omega}_w(y)$ *is singular if and only if* $\operatorname{rank}(B(y)) < q$.

3. $\hat{\Omega}_w(y) = 0$ *if and only if* $B(y) = 0$.

4. *The set of all* $y \in \mathbb{R}^n$ *for which* $\hat{\Omega}_w(y)$ *is singular (or, equivalently, for which* $\operatorname{rank}(B(y)) < q)$ *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 3 is violated.*

Remark 3.2. (i) Setting $R = X'X$ and $q = k$ shows that a necessary and sufficient condition for $\hat{\Psi}_w$ to be $\lambda_{\mathbb{R}^n}$-almost everywhere nonsingular is that $e_i(n) \notin \operatorname{span}(X)$ for all $i = 1, \ldots, n$. [If this condition is not satisfied $\hat{\Psi}_w(y)$ is singular for every $y \in \mathbb{R}^n$.] In particular, it follows that under this simple condition $\hat{\Omega}_w(y)$ is nonsingular $\lambda_{\mathbb{R}^n}$-almost everywhere for *every* choice of the restriction matrix R.

(ii) In the case $q = 1$ Assumption 3 is easily seen to be violated if and only if

$$R(X'X)^{-1}X'e_i(n) = 0 \quad \text{or} \quad e_i(n) \in \operatorname{span}(X) \text{ holds for every } i = 1, \ldots, n.$$

---

[6]Some authors (e.g., Kiefer and Vogelsang (2002b, 2005)) choose to normalize also by q, the number of restrictions to be tested. This is of course immaterial as long as one accordingly adjusts the critical vlaue.

We learn from the preceding lemma that, provided Assumption 3 is satisfied (which only depends on X and R and hence can be verified by the user), our choice of defining the test statistic T to be zero on the set where $\hat{\Omega}_w$ is singular is immaterial and has no effect on the size and power properties of the test. We also learn from that lemma that, in case Assumption 3 is violated, the commonly used autocorrelation robust tests break down completely in a trivial way as $\hat{\Omega}_w(y)$ is then singular for *every* data point y. We are therefore forced to impose Assumption 3 on the design matrix X if we want commonly used autocorrelation robust tests to make any sense at all. We shall thus impose Assumption 3 in the following development. We also note that, given a restriction matrix R, the set of design matrices that lead to a violation of Assumption 3 is a "thin" subset in the set of all $n \times k$ matrices of full rank.

As usual, the test based on T rejects H_0 if $T(y) \geq C$ where $C > 0$ is an appropriate critical value. In applications the critical value is usually taken from the asymptotic distribution of T (obtained either under assumptions that guarantee consistency of $\hat{\Omega}_w$ or under the assumption of a "fixed bandwidth", i.e., $M_n/n > 0$ independent of n). In the subsequent theorem, which discusses size and power properties of autocorrelation robust tests based on T, we allow for *arbitrary* (nonrandom) critical values $C > 0$.[7] Because of this, and since the theorem is a finite-sample result, it applies equally well to standard autocorrelation robust tests (for which one fancies that $M_n \to \infty$ and $M_n/n \to 0$ if n would increase to infinity) and to so-called "fixed-bandwidth" tests (which assume $M_n/n > 0$ independent of n).

Theorem 3.3. *Suppose Assumptions 1, 2, and 3 are satisfied. Let T be the test statistic defined in (7) with $\hat{\Psi}_w$ as in (6). Let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ be the rejection region where C is a real number satisfying $0 < C < \infty$. Then the following holds:*

1. *Suppose $\operatorname{rank}(B(e_+)) = q$ and $T(e_+ + \mu_0^*) > C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$, or $\operatorname{rank}(B(e_-)) = q$ and $T(e_- + \mu_0^*) > C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C)) = 1$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

2. *Suppose $\operatorname{rank}(B(e_+)) = q$ and $T(e_+ + \mu_0^*) < C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$, or $\operatorname{rank}(B(e_-)) = q$ and $T(e_- + \mu_0^*) < C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C)) = 0$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$, and hence

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W(C)) = 0$$

   holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased. Furthermore, the nuisance-infimal rejection probability at every point $\mu_1 \in \mathfrak{M}_1$ is zero, i.e.,

$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W(C)) = 0.$$

   In particular, the infimal power of the test is equal to zero.

---

[7]Because the theorem is a finite-sample result, we are free to imagine that C depends on sample size n. In fact, there is nothing in the theory that prohibits us from imagining that C depends even on the design matrix X, on the restriction given by (R, r), or on the weights $w(j, n)$.

3. *Suppose $B(e_+) = 0$ and $R\hat{\beta}(e_+) \neq 0$ hold, or $B(e_-) = 0$ and $R\hat{\beta}(e_-) \neq 0$ hold. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 1$$

holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

Remark 3.4. (i) As a point of interest we note that the rejection probabilities $P_{\mu, \sigma^2 \Sigma}(W(C))$ can be shown to depend on $\left(\mu, \sigma^2, \Sigma\right)$ only through $\left(\left(R\beta - r\right)/\sigma, \Sigma\right)$ (in fact, only through $\left(\langle\left(R\beta - r\right)/\sigma\rangle, \Sigma\right)$), see Lemma A.1 in Appendix A.

(ii) Although trivial, it is useful to note that the conclusions of the preceding theorem also apply to any rejection region $W^* \in \mathcal{B}(\mathbb{R}^n)$ which differs from $W(C)$ by a $\lambda_{\mathbb{R}^n}$-null set.

(iii) By the way T is defined in (7), the condition $T(e_+ + \mu_0^*) > C$ $(T(e_- + \mu_0^*) > C$, respectively) in Part 1 of the preceding theorem already implies $\operatorname{rank}(B(e_+)) = q$ $(\operatorname{rank}(B(e_-)) = q$, respectively). For reasons of comparability with Part 2 we have nevertheless included this rank condition into the formulation of Part 1.

(iv) Inspection of the proof of Theorem 3.3 shows that Assumption 1 can obviously be weakened to the assumption that $\mathfrak{C}$ contains AR(1) correlation matrices $\Lambda(\rho_m^{(1)})$ and $\Lambda(\rho_m^{(2)})$ for two sequences $\rho_m^{(i)} \in (-1, 1)$ with $\rho_m^{(1)} \to 1$ and $\rho_m^{(2)} \to -1$. In fact, this can be further weakened to the assumption that there exist $\Sigma_m^{(i)} \in \mathfrak{C}$ with $\Sigma_m^{(1)} \to e_+ e_+'$ and $\Sigma_m^{(2)} \to e_- e_-'$ for $m \to \infty$.

The conditions in Parts 1-3 of the theorem only depend on the design matrix X, the restriction (R, r), the vector e_+ (e_-, respectively), the critical value C, and the weights $w(j, n)$ (via $T(e_+ + \mu_0^*)$ or $T(e_- + \mu_0^*)$, respectively). Hence, in any particular application it can be decided whether (and which of) these conditions are satisfied. Furthermore, as will become transparent from the examples to follow and from Proposition 3.5 below, in the majority of applications at least one of these conditions will be satisfied, implying that common autocorrelation robust tests have size 1 and/or have power arbitrarily close to 0 in certain parts of the alternative hypothesis. Before we turn to these examples, we want to provide some intuition for Theorem 3.3: Consider a sequence $\rho_m \in (-1, 1)$ with $\rho_m \to 1$ ($\rho_m \to -1$, respectively) as $m \to \infty$. Then $\Sigma_m = \Lambda(\rho_m) \in \mathfrak{C}$ by Assumption 1 and $\Lambda(\rho_m) \to e_+ e_+'$ ($e_- e_-'$) holds. Consequently, $P_{\mu_0, \sigma^2 \Sigma_m}$ concentrates more and more around the one-dimensional subspace $\operatorname{span}(e_+)$ ($\operatorname{span}(e_-)$, respectively) in the sense that it converges weakly to the singular Gaussian distribution $P_{\mu_0, \sigma^2 e_+ e_+'}$ ($P_{\mu_0, \sigma^2 e_- e_-'}$, respectively). The conditions in Part 1 (or Part 3) of the preceding theorem then essentially allow one to show that (i) the measure $P_{\mu_0, \sigma^2 e_+ e_+'}$ ($P_{\mu_0, \sigma^2 e_- e_-'}$, respectively) is supported by $W(C)$ (more precisely, after $W(C)$ has been modified by a suitable $\lambda_{\mathbb{R}^n}$-null set), and (ii) that $P_{\mu_0, \sigma^2 e_+ e_+'}$ ($P_{\mu_0, \sigma^2 e_- e_-'}$, respectively) puts no mass on the boundary of (the modified) set $W(C)$. By the Portmanteau theorem we can then conclude that the sequence of measures $P_{\mu_0, \sigma^2 \Sigma_m}$ concentrates more and more on $W(C)$ in the sense that $P_{\mu_0, \sigma^2 \Sigma_m}(W(C)) \to 1$ as $m \to \infty$, which establishes the conclusion of Part 1 of the theorem. The proof of the first claim in Part 2 works along similar lines but where concentration is now on the complement of the rejection region $W(C)$. For more discussion see Subsection 5.2. The remaining results in Part 2 are obtained from the first claim in Part 2 exploiting invariance and continuity properties of the rejection probabilities. While concentration of the probability measures $P_{\mu_0, \sigma^2 \Sigma_m}$ constitutes an important ingredient in the proof of Theorem 3.3, it should, however, be stressed that there are also other cases (cf. Theorems 3.6 and 3.7), where

despite concentration of $P_{\mu_0, \sigma^2 \Sigma_m}$ as above, the conditions for an application of the Portmanteau theorem are *not* satisfied; in fact, in some of these cases size < 1 and infimal power > 0 can be shown.

We now consider a few examples that illustrate the implications of the preceding theorem. As in most applications the regression model contains an intercept, we concentrate on this case in the examples.

Example 3.1. *(Testing a restriction involving the intercept)* Suppose that Assumptions 1, 2, and 3 hold. For definiteness assume that the first column of X corresponds to the intercept (i.e., the first column of X is e_+). Assume also that the restriction involves the intercept, i.e., the first column of R is nonzero. Then it is easy to see that $B(e_+) = 0$ and $R\hat\beta(e_+) \neq 0$ holds (the latter since $\hat\beta(e_+) = e_1(k)$). Consequently, Part 3 of Theorem 3.3 applies and shows that the size of the test T is always 1. Additionally, the power deficiency results in Part 2 of the theorem will apply whenever rank$(B(e_-)) = q$ and $T(e_- + \mu_0^*) < C$ hold. [Whether or not this is the case will depend on C, X, R, and the weights.]

Example 3.2. *(Location model)* Suppose that Assumptions 1 and 2 hold. Suppose $X = e_+$ and the hypothesis is $\beta = \beta_0$ (hence $k = q = 1$). As just noted in Example 3.1, the size of the test T is then always 1 (as Assumption 3 is certainly satisfied). In this simple model the conditions for the power deficiencies to arise can be made more explicit: Note that $B(e_-) \neq 0$ clearly always holds, and hence rank $B(e_-) = 1 = q$. If n is even, it is also easy to see that $T(e_- + \beta_0 e_+) = 0 < C$ always holds. Consequently, Part 2 of Theorem 3.3 applies and shows that the power of the test gets arbitrarily close to zero in certain parts of the parameter space as described in the theorem. If n is odd, then $T(e_- + \beta_0 e_+) = n^{-1}\hat\Psi_w^{-1}(e_-)$ and the same conclusion applies provided this quantity is less than C. For example, for the (modified) Bartlett lag-window numerical computations show that $n^{-1}\hat\Psi_w^{-1}(e_-)$ is less than 1.563 for every odd n in the range $1 < n < 1000$ and every choice of $M_n/n \in (0,1]$; hence, if C has been chosen to be larger than or equal to 1.563, which is typically the case at conventional nominal significance levels, the power deficiencies are also guaranteed to arise. We note here that this simple location model is often used in Monte Carlo studies that try to assess finite-sample properties of autocorrelation robust tests. Furthermore, autocorrelation robust testing of the location parameter plays an important rôle in the field of simulation, see, e.g., Heidelberger and Welch (1981), Flegal and Jones (2010).

Example 3.3. *(Testing a zero restriction on a slope parameter)* Consider the same regression model as in Example 3.1 with the same assumptions, but now suppose that the hypothesis is $\beta_i = 0$ for some $i > 1$, i.e., we are interested in testing a slope parameter. Since in this case $B(e_+) = 0$ and $R\hat\beta(e_+) = 0$ obviously hold, where $R = e_i'(k)$, we need to investigate the behavior of $B(e_-)$ in order to be able to apply Theorem 3.3. If rank $B(e_-) = 1$ holds (which will generically be the case) then size equals 1 in case $T(e_-) > C$ and the power deficiencies arise in case $T(e_-) < C$.

Example 3.4. *(Testing for a change in mean)* A special case of the preceding example is the case where $k = 2$, the first column of X is e_+ and the second column has entries $x_{t2} = 0$ for $1 \leq t \leq t_*$ and $x_{t2} = 1$ else. We assume t_* to be known and to satisfy $1 < t_* < n$. The hypothesis to be tested is $\beta_2 = 0$. It is then easy to see that Assumption 3 is satisfied. Furthermore, some simple computations show that rank $B(e_-) = q = 1$ always holds. Hence, the test T has size 1 if $T(e_-) > C$ and the power deficiencies arise if $T(e_-) < C$. In case n as well as $n - t_0$ are even, the latter case always arises since $T(e_-) = 0$ holds. [If n or $n - t_0$ is odd, $T(e_-)$ can of course be computed and depends only on n, t_0, and $\hat\Psi_w^{-1}(e_-)$. We omit the details.]

The cases in Theorem 3.3 leading to size 1 or to power deficiencies of the test based on T, while not being exhaustive, are often satisfied in applications. We make this formal in the subsequent proposition in that we prove that, for given restriction (R, r) and critical value C, the conditions in Theorem 3.3 involving X are generically satisfied. The first part of the proposition shows that these conditions are generically satisfied in the universe of all possible $n \times k$ design matrices of rank k. Parts 2 and 3 show that the same is true if we impose that the regression model has to contain an intercept. In the subsequent proposition the dependence of $B(y)$, of $T(y)$, as well as of $\hat{\Omega}_w(y)$ on X will be important and thus we shall write $B_X(y)$, $T_X(y)$, and $\hat{\Omega}_{w,X}(y)$ for these quantities in the result to follow.

Proposition 3.5. *Suppose Assumption 1 holds. Fix (R, r) with $\operatorname{rank}(R) = q$, fix $0 < C < \infty$, and fix the weights $w(j, n)$ which are assumed to satisfy Assumption 2. Let T be the test statistic defined in (7) with $\hat{\Psi}_w$ as in (6) and let $\mu_0^* \in \mathfrak{M}_0$ be arbitrary.*

1. *Define*

$$\mathfrak{X}_0 = \left\{ X \in \mathbb{R}^{n \times k} : \operatorname{rank}(X) = k \right\}, \quad \mathfrak{X}_1(e_+) = \left\{ X \in \mathfrak{X}_0 : \operatorname{rank}(B_X(e_+)) < q \right\},$$

$$\mathfrak{X}_2(e_+) = \left\{ X \in \mathfrak{X}_0 \backslash \mathfrak{X}_1(e_+) : T_X(e_+ + \mu_0^*) = C \right\},$$

and similarly define $\mathfrak{X}_1(e_-)$, $\mathfrak{X}_2(e_-)$. [Note that $\mathfrak{X}_2(e_+)$ and $\mathfrak{X}_2(e_-)$ do not depend on the choice of μ_0^*.] Then $\mathfrak{X}_1(e_+)$, $\mathfrak{X}_2(e_+)$, $\mathfrak{X}_1(e_-)$, and $\mathfrak{X}_2(e_-)$ are $\lambda_{\mathbb{R}^{n \times k}}$-null sets. The set of all design matrices $X \in \mathfrak{X}_0$ for which Theorem 3.3 does not apply is a subset of $(\mathfrak{X}_1(e_+) \cup \mathfrak{X}_2(e_+)) \cap (\mathfrak{X}_1(e_-) \cup \mathfrak{X}_2(e_-))$ and hence is a $\lambda_{\mathbb{R}^{n \times k}}$-null set. It thus is a "negligible" subset of $\mathfrak{X}_0$ in view of the fact that $\mathfrak{X}_0$ differs from $\mathbb{R}^{n \times k}$ only by a $\lambda_{\mathbb{R}^{n \times k}}$-null set.

2. *Suppose $k \geq 2$, X has e_+ as its first column, i.e., $X = \left(e_+, \tilde{X} \right)$, and suppose the first column of R consists of zeros only. Define*

$$
\begin{aligned}
\tilde{\mathfrak{X}}_0 &= \left\{ \tilde{X} \in \mathbb{R}^{n \times (k-1)} : \operatorname{rank}\left(\left(e_+, \tilde{X} \right) \right) = k \right\}, \\
\tilde{\mathfrak{X}}_1(e_-) &= \left\{ \tilde{X} \in \tilde{\mathfrak{X}}_0 : \operatorname{rank}\left(B_{(e_+, \tilde{X})}(e_-) \right) < q \right\}, \\
\tilde{\mathfrak{X}}_2(e_-) &= \left\{ \tilde{X} \in \tilde{\mathfrak{X}}_0 \backslash \tilde{\mathfrak{X}}_1(e_-) : T_{(e_+, \tilde{X})}(e_- + \mu_0^*) = C \right\},
\end{aligned}
$$

and note that $\tilde{\mathfrak{X}}_2(e_-)$ does not depend on the choice of μ_0^*. Then $\tilde{\mathfrak{X}}_1(e_-)$ and $\tilde{\mathfrak{X}}_2(e_-)$ are $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null sets (with the analogously defined sets $\tilde{\mathfrak{X}}_1(e_+)$ and $\tilde{\mathfrak{X}}_2(e_+)$ satisfying $\tilde{\mathfrak{X}}_1(e_+) = \tilde{\mathfrak{X}}_0$ and $\tilde{\mathfrak{X}}_2(e_+) = \emptyset$.). The set of all matrices $\tilde{X} \in \tilde{\mathfrak{X}}_0$ such that Theorem 3.3 does not apply to the design matrix $X = \left(e_+, \tilde{X} \right)$ is a subset of $\tilde{\mathfrak{X}}_1(e_-) \cup \tilde{\mathfrak{X}}_2(e_-)$ and hence is a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set. It thus is a "negligible" subset of $\tilde{\mathfrak{X}}_0$ in view of the fact that $\tilde{\mathfrak{X}}_0$ differs from $\mathbb{R}^{n \times (k-1)}$ only by a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set.

3. *Suppose $k \geq 2$, $X = \left(e_+, \tilde{X} \right)$, and suppose the first column of R is nonzero. Then Theorem 3.3 applies to the design matrix $X = \left(e_+, \tilde{X} \right)$ for every $\tilde{X} \in \tilde{\mathfrak{X}}_0$ (provided X satisfies Assumption 3).[8]*

---

[8] If X does not satisfy Assumption 3, then the test breaks down in a trivial way as already discussed.

The proof of the proposition actually shows more, namely that the set of design matrices for which Theorem 3.3 does not apply is contained in an algebraic set. We also remark that if the regressor matrix X is viewed as randomly drawn from a distribution that is absolutely continuous w.r.t. $\lambda_{\mathbb{R}^{n \times k}}$, Proposition 3.5 implies that then the conditions of Theorem 3.3 are almost surely satisfied; if X is also independent of $\mathbf{U}$, Theorem 3.3 then establishes negative results for the conditional rejection probabilities for almost all realizations of X.

We next discuss an exceptional case to which Theorem 3.3 does not apply and which is interesting in that a positive result can be established, at least if the covariance model $\mathfrak{C}$ is assumed to be $\mathfrak{C}_{AR(1)}$ or is approximated by $\mathfrak{C}_{AR(1)}$ near the singular points in the sense of Remark 3.9 below. This positive result will then guide us to an improved version of the test statistic T.

Theorem 3.6. *Suppose $\mathfrak{C} = \mathfrak{C}_{AR(1)}$ and suppose Assumptions 2 and 3 are satisfied. Let T be the test statistic defined in (7) with $\hat{\Psi}_w$ as in (6). Let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ be the rejection region where C is a real number satisfying $0 < C < \infty$. If $e_+, e_- \in \mathfrak{M}$ and $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$ is satisfied, then the following holds:*

1. *The size of the rejection region $W(C)$ is strictly less than 1, i.e.,*

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)} \left(W(C) \right) < 1.$$

   Furthermore,

$$\inf_{\mu_0 \in \mathfrak{M}_0} \inf_{0 < \sigma^2 < \infty} \inf_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)} \left(W(C) \right) > 0.$$

2. *The infimal power is bounded away from zero, i.e.,*

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{0 < \sigma^2 < \infty} \inf_{-1 < \rho < 1} P_{\mu_1, \sigma^2 \Lambda(\rho)}(W(C)) > 0.$$

3. *For every $0 < c < \infty$*

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, 0 < \sigma^2 < \infty \\ d(\mu_1, \mathfrak{M}_0)/\sigma \geq c}} P_{\mu_1, \sigma^2 \Lambda(\rho_m)}(W(C)) \to 1$$

   holds for $m \to \infty$ and for any sequence $\rho_m \in (-1, 1)$ satisfying $|\rho_m| \to 1$. Furthermore, for every sequence $0 < c_m < \infty$ and every $0 < \varepsilon < 1$

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, \\ d(\mu_1, \mathfrak{M}_0) \geq c_m}} \inf_{-1 + \varepsilon \leq \rho \leq 1 - \varepsilon} P_{\mu_1, \sigma_m^2 \Lambda(\rho)}(W(C)) \to 1$$

   holds for $m \to \infty$ whenever $0 < \sigma_m^2 < \infty$ and $c_m/\sigma_m \to \infty$. [The very last statement holds even without the conditions $e_+, e_- \in \mathfrak{M}$ and $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$.]

4. *For every δ, $0 < \delta < 1$, there exists a $C(\delta)$, $0 < C(\delta) < \infty$, such that*

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)}(W(C(\delta))) \leq \delta.$$

13

The first statement of the theorem says that, in contrast to the cases considered in Theorem 3.3, the size of the test T is now bounded away from 1 for any choice of the critical value C. Moreover, the last part of the theorem shows that the size can be controlled to be less than or equal to any prespecified significance level δ by a suitable choice of the critical value $C(\delta)$. Because $P_{\mu_0,\sigma^2\Lambda(\rho)}(W(C))$ does not depend on μ_0 and σ^2 but only on ρ (see Proposition 5.4) and because this probability can be computed via simulation, the supremum of this probability over μ_0, σ^2, and ρ can be easily found by a grid search; exploiting monotonicity of the probability with respect to C, the value of $C(\delta)$ can then be found by a simple search algorithm. The theorem furthermore shows that, again in contrast to the scenario considered in Theorem 3.3, the infimal power of the test is at least bounded away from zero. The power even approaches 1 if either $\left\|\left(R\beta^{(1)} - r\right)/\sigma\right\|$ is bounded away from zero and $|\rho| \to 1$, or if $\left\|\left(R\beta^{(1)} - r\right)/\sigma\right\| \to \infty$ and $|\rho|$ is bounded away from 1. [Here $\beta^{(1)}$ is the parameter vector corresponding to μ_1. Note that $d(\mu_1, \mathfrak{M}_0)$ is bounded from above as well as from below by multiples of $\left\|R\beta^{(1)} - r\right\|$, where the constants involved are positive and depend only on X, R, and r.]

The preceding theorem required $e_+, e_- \in \mathfrak{M}$ and $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$. To illustrate, these conditions are, e.g., satisfied if e_+ and e_- constitute the first two columns of the matrix X and the hypothesis tested only involves coefficients β_i with $i \geq 3$ (i.e., the first two columns of R are zero). While an intercept will typically be present in a regression model and thus e_+ appears as one of the regressors (and hence satisfies $e_+ \in \mathfrak{M}$), e_- will not necessarily be an element of $\mathfrak{M}$, and hence the preceding theorem will not apply. However, the following theorem shows how we can nevertheless extend the same positive results to this case if we apply a simple adjustment to the test statistic T.

Theorem 3.7. *Suppose $\mathfrak{C} = \mathfrak{C}_{AR(1)}$ and suppose Assumption 2 is satisfied. Suppose one of the following scenarios applies:*

1. *$e_+ \in \mathfrak{M}$ with $R\hat{\beta}(e_+) = 0$ and $e_- \notin \mathfrak{M}$. Furthermore, $k+1 < n$ holds and the $n \times (k+1)$ matrix $\bar{X} = (X, e_-)$ (which necessarily has rank $k+1$) satisfies Assumption 3 relative to the $q \times (k+1)$ restriction matrix $\bar{R} = (R, 0)$. Define $\bar{\beta}(y) = (I_k, 0)(\bar{X}'\bar{X})^{-1}\bar{X}'y$.*

2. *$e_+ \notin \mathfrak{M}$ and $e_- \in \mathfrak{M}$ with $R\hat{\beta}(e_-) = 0$. Furthermore, $k+1 < n$ holds and the $n \times (k+1)$ matrix $\bar{X} = (X, e_+)$ (which necessarily has rank $k+1$) satisfies Assumption 3 relative to the $q \times (k+1)$ restriction matrix $\bar{R} = (R, 0)$. Define $\bar{\beta}(y) = (I_k, 0)(\bar{X}'\bar{X})^{-1}\bar{X}'y$.*

3. *$e_+ \notin \mathfrak{M}$ and $e_- \notin \mathfrak{M}$ with $\text{rank}(X, e_+, e_-) = k+2$. Furthermore, $k+2 < n$ holds and the $n \times (k+2)$ matrix $\bar{X} = (X, e_+, e_-)$ (which necessarily has rank $k+2$) satisfies Assumption 3 relative to the $q \times (k+2)$ restriction matrix $\bar{R} = (R, 0, 0)$. Define $\bar{\beta}(y) = (I_k, 0, 0)(\bar{X}'\bar{X})^{-1}\bar{X}'y$.*

4. *$e_+ \notin \mathfrak{M}$ and $e_- \notin \mathfrak{M}$ with $\text{rank}(X, e_+, e_-) = k+1$. Furthermore, $k+1 < n$ holds and the $n \times (k+1)$ matrix $\bar{X} = (X, e_+)$ (which necessarily has rank $k+1$) satisfies Assumption 3 relative to the $q \times (k+1)$ restriction matrix $\bar{R} = (R, 0)$. Suppose further that $\bar{R}(\bar{X}'\bar{X})^{-1}\bar{X}'e_- = 0$ holds. Define $\bar{\beta}(y) = (I_k, 0)(\bar{X}'\bar{X})^{-1}\bar{X}'y$.*

5. *$e_+ \notin \mathfrak{M}$ and $e_- \notin \mathfrak{M}$ with $\text{rank}(X, e_+, e_-) = k+1$. Furthermore, $k+1 < n$ holds and the $n \times (k+1)$ matrix $\bar{X} = (X, e_-)$ (which necessarily has rank $k+1$) satisfies Assumption 3 relative*

to the $q \times (k+1)$ restriction matrix $\bar{R} = (R, 0)$. Suppose further that $\bar{R} \left(\bar{X}' \bar{X} \right)^{-1} \bar{X}' e_+ = 0$ holds. Define $\bar{\bar{\beta}} (y) = (I_k, 0) \left(\bar{X}' \bar{X} \right)^{-1} \bar{X}' y$.

In all five scenarios define

$$\bar{T}(y) = \begin{cases} (R\bar{\beta}(y) - r)' \bar{\Omega}_w^{-1}(y)(R\bar{\beta}(y) - r) & \text{if } \det \bar{\Omega}_w(y) \neq 0, \\ 0 & \text{if } \det \bar{\Omega}_w(y) = 0, \end{cases}$$

where $\bar{\Omega}_w(y) = n\bar{R}(\bar{X}'\bar{X})^{-1}\bar{\Psi}_w(y)(\bar{X}'\bar{X})^{-1}\bar{R}'$, and $\bar{\Psi}_w(y)$ is computed from (6) based on $\bar{v}_t(y) = \bar{u}_t(y)\bar{x}_t'$. instead of $\hat{v}_t(y)$. Here $\bar{u}_t(y)$ are the residuals from the regression of y on $\bar{X}$, and $\bar{x}_t$. are the rows of $\bar{X}$. Let $\bar{W}(C) = \{ y \in \mathbb{R}^n : \bar{T}(y) \geq C \}$ be the rejection region where C is a real number satisfying $0 < C < \infty$. Then for each of the five scenarios the conclusions of Theorem 3.6 hold with $W(C)$ replaced by $\bar{W}(C)$.

Theorem 3.3 together with Proposition 3.5 has shown that generically the commonly used test based on the statistic T has severe size or power deficiencies even for $\mathfrak{C} = \mathfrak{C}_{AR(1)}$, while Theorem 3.6 has isolated a special case where this is not so. Theorem 3.7 now shows that in many of the cases falling under the wrath of Theorem 3.3 the ensuing problems can be circumvented (if $\mathfrak{C} = \mathfrak{C}_{AR(1)}$) by making use of the adjusted version $\bar{T}$ of the test statistic. The adjustment mechanism is simple and amounts to basing the test statistic on estimators $\bar{\beta}$ and $\bar{\Omega}_w$ that are obtained from a "working model" that always adds the regressors e_+ and/or e_- to the design matrix. Note that these regressors effect a purging of the residuals from harmonic components of angular frequency 0 and π. This purging effect together with the fact that the restrictions to be tested do not involve the coefficients of the "purging" regressors e_+ and e_- lies at the heart of the positive results expressed in Theorems 3.6 and 3.7. Numerical results that will be presented elsewhere support the theoretical result and show that the adjusted test based on $\bar{T}$ considerably improves over the unadjusted one based on T.

Remark 3.8. (i) Suppose the scenario in Part 1 of the above theorem applies except that $k+1 = n$ holds or $\bar{X} = (X, e_-)$ does not satisfy Assumption 3. Then the test statistic $\bar{T}$ is identically zero and the adjustment procedure does not work. A similar remark applies to Parts 2-5.

(ii) Suppose the scenario of Part 4 of the above theorem applies except that $\bar{R} \left(\bar{X}' \bar{X} \right)^{-1} \bar{X}' e_- \neq 0$ holds. Applying Part 3 of Theorem 3.3 to $\bar{T}$ shows that this test has size 1 and hence the adjustment procedure fails. A similar comment applies to the scenario of Part 5.

Remark 3.9. The results in Theorems 3.6 and 3.7 have assumed $\mathfrak{C} = \mathfrak{C}_{AR(1)}$. The results immediately extend to other covariance models $\mathfrak{C}$ as long as $\mathfrak{C}$ is norm-bounded, the only singular accumulation points of $\mathfrak{C}$ are e_+e_+' and e_-e_-', and for every $\Sigma_m \in \mathfrak{C}$ converging to one of these limit points there exists a sequence $(\rho_m)_{m \in \mathbb{N}}$ in $(-1, 1)$ such that $\Lambda(\rho_m)^{-1/2}\Sigma_m\Lambda(\rho_m)^{-1/2} \to I_n$ for $m \to \infty$ (that is, near the "singular boundary" the covariance model $\mathfrak{C}$ behaves similar to $\mathfrak{C}_{AR(1)}$). This can be seen from an inspection of the proof. An extension of Theorems 3.6 and 3.7 to even more general covariance models will be discussed elsewhere.

3.1.1 Alternative nonparametric estimators for the variance covariance matrix

We next discuss test statistics of the form (7) that use estimators other than $\hat{\Psi}_w$.

A. *(General quadratic estimators based on $\hat{v}_t$)* The estimator $\hat{\Psi}_w$ given by (6) is a special case of general quadratic estimators $\hat{\Psi}_{GQ}(y)$ of the form

$$\hat{\Psi}_{GQ}(y) = \sum_{t,s=1}^{n} w(t,s;n)\,\hat{v}_t(y)\hat{v}_s(y)'$$

for every $y \in \mathbb{R}^n$, where the $n \times n$ weighting matrix $\mathcal{W}_n^* = (w(t,s;n))_{t,s}$ is symmetric and data-independent. While estimators of this more general form have been studied in the early literature on spectral estimation, much of the literature has focused on the special case of weighted autocovariance estimators of the form $\hat{\Psi}_w$ (partly as a consequence of a result in Grenander and Rosenblatt (1957) that the restriction to the smaller class of estimators does not lead to inferior estimators in an asymptotic sense). However, if the data are preprocessed by tapering before an estimator like $\hat{\Psi}_w$ is computed from the tapered data, the final estimator belongs to the class of general quadratic estimators. Also, many modern spectral estimators studied in the engineering literature fall into this class (see Thomson (1982)), but not into the more narrow class of weighted autocovariance estimators. Another example are the estimators proposed in Phillips (2005) and Sun (2012). We now distinguish two cases:

Case 1: The weighting matrix $\mathcal{W}_n^* = (w(t,s;n))_{t,s}$ is positive definite. Inspection of the proofs then shows that all results given above for the tests T based on $\hat{\Psi}_w$ remain valid as they stand if $\hat{\Psi}_w$ is replaced by $\hat{\Psi}_{GQ}$ in the definition of the test statistic.

Case 2: The weighting matrix $\mathcal{W}_n^* = (w(t,s;n))_{t,s}$ is only assumed to be nonnegative definite (as is the case for the estimators considered in Phillips (2005) and Sun (2012)). Arguing similar as in the proof of Lemma 3.1 one can show the following:

Lemma 3.10. *Suppose $\mathcal{W}_n^* = (w(t,s;n))_{t,s}$ is nonnegative definite and define*

$$\hat{\Omega}_{GQ}(y) = nR(X'X)^{-1}\hat{\Psi}_{GQ}(y)(X'X)^{-1}R'.$$

Then the following hold:

1. $\hat{\Omega}_{GQ}(y)$ *is nonnegative definite for every* $y \in \mathbb{R}^n$.

2. $\hat{\Omega}_{GQ}(y)$ *is singular if and only if* $\operatorname{rank}(B(y)\mathcal{W}_n^*) < q$ *(or, equivalently, if* $\operatorname{rank}\left(B(y)\mathcal{W}_n^{*1/2}\right) < q$*).*

3. $\hat{\Omega}_{GQ}(y) = 0$ *if and only if* $B(y)\mathcal{W}_n^* = 0$ *(or, equivalently, if* $B(y)\mathcal{W}_n^{*1/2} = 0$*).*

4. *The set of all* $y \in \mathbb{R}^n$ *for which* $\hat{\Omega}_{GQ}(y)$ *is singular (or, equivalently, for which* $\operatorname{rank}(B(y)\mathcal{W}_n^*) < q$*) is either a* $\lambda_{\mathbb{R}^n}$*-null set or the entire sample space* $\mathbb{R}^n$.

As a consequence we see that two cases can arise: In the first case $\hat{\Omega}_{GQ}(y)$ is singular for all $y \in \mathbb{R}^n$, in which case the test statistic T breaks down in a trivial way. Note that this arises precisely if and only if $\operatorname{rank}(B(y)\mathcal{W}_n^*) < q$ for all $y \in \mathbb{R}^n$, which is a condition solely on the design matrix X, the restriction matrix R, and the weighting matrix $\mathcal{W}_n^*$, and thus can be verified in any particular application. Now suppose that the second case arises, i.e., $\operatorname{rank}(B(y)\mathcal{W}_n^*) = q$ for $\lambda_{\mathbb{R}^n}$-almost all y. Then inspection of the proofs shows that Theorems 3.3 and 3.6 continue to hold for the test statistic T based on $\hat{\Psi}_{GQ}$ provided Assumption 3 is replaced by the just mentioned condition

16

rank $(B(y)\mathcal{W}_n^*) = q$ for $\lambda_{\mathbb{R}^n}$-almost all y, and the matrix $B(y)$ in those theorems is replaced by $B(y)\mathcal{W}_n^*$. Also Theorem 3.7 generalizes with the obvious changes.

B. *(An estimator based on $\hat{u}$)* Because $n^{-1}\mathbb{E}(X'\mathbf{U}\mathbf{U}'X) = n^{-1}X'\mathbb{E}(\mathbf{U}\mathbf{U}')X$, a natural estimator is

$$\hat{\Psi}_E(y) = n^{-1}X'\hat{K}(y)X$$

for every $y \in \mathbb{R}^n$, where $\hat{K}(y)$ is the symmetric $n \times n$ Toeplitz matrix with block elements $n^{-1}\sum_{l=j+1}^{n} \hat{u}_l(y)\hat{u}_{l-j}(y)$ in the j-th diagonal above the main diagonal. This estimator has already been discussed in Eicker (1967), but does not seem to have been used much in the econometrics literature. It is not difficult to see that $\hat{\Psi}_E(y)$ is always nonnegative definite. It is positive definite if and only if $y \notin \text{span}(X)$; and it is equal to zero for $y \in \text{span}(X)$. Define the statistic T_E via (7) with $\hat{\Omega}_w(y)$ replaced by $\hat{\Omega}_E(y)$ where the latter is obtained from (5) by replacing $\hat{\Psi}_w(y)$ by $\hat{\Psi}_E(y)$. It is then easy to see that Theorems 3.3, 3.6, and 3.7 carry over to the test based on T_E provided Assumption 3 is deleted from the formulation, the condition rank $(B(e_+)) = q$ (rank $(B(e_-)) = q$, respectively) is replaced by $e_+ \notin \text{span}(X)$ ($e_- \notin \text{span}(X)$), and the condition $B(e_+) = 0$ ($B(e_-) = 0$, respectively) is replaced by $e_+ \in \text{span}(X)$ ($e_- \in \text{span}(X)$). [While Eicker (1967) provided conditions on the regressors under which consistency of $\hat{\Psi}_E(y)$ results, it may not be consistent for some common forms of regressors (as noted in Eicker (1967)). Therefore one may want to replace $\hat{K}(y)$ by a variant where the empirical second moments are downweighted (or more generally are obtained from an estimate of the spectral density of the errors $\mathbf{u}_t$). Similar results can then be obtained for this variant of the test. We omit the details.]

C. *(Data-driven bandwidth, prewhitening, flat-top kernels, autoregressive estimates, random critical values)* Tests based on weighted autocovariance estimators $\hat{\Psi}$ but where the weights are allowed to depend on the data (e.g., lag-window estimators with data-driven bandwidth choice or exponentiated lag-window estimators with data-driven choice of the exponent), or where prewhitening is used, or where the estimators $\hat{\Psi}$ may not always be nonnegative definite (as is the case with so-called flat-top kernels, see Politis (2011)), are discussed in detail in Preinerstorfer (2013). Like the results given above, they are obtained as special cases of very general results provided in Subsection 5.4. The results in Subsection 5.4 essentially rely only on a certain equivariance property of the estimator $\hat{\Omega}$ and allow for cases where $\hat{\Omega}$ is not always nonnegative definite. These results also accommodate situations where the estimator $\hat{\Omega}(y)$ is only well-defined for $\lambda_{\mathbb{R}^n}$-almost all y, a case that arises often when data-driven bandwidth or prewithening are employed. Furthermore, certain cases where the critical value is allowed to be random are covered by these results, see Remark 5.16(ii) in Section 5.4 (and this is also true for Subsection 3.2 and Section 4). Finally, tests based on estimators $\hat{\Psi}$ obtained from parametric models like vector autoregressions (see, e.g., den Haan and Levin (1997) or Sun and Kaplan (2012) and references therein) also fall into the domain of the results in Subsection 5.4, but we abstain from a detailed analysis. See, however, Subsection 3.2 for a related analysis.

3.1.2 Some discussion

A. In the subsequent discussion we concentrate for definiteness on the test statistic T that is based on the estimator $\hat{\Omega}_w$. However, the discussion carries over mutatis mutandis to the case where the alternative estimators $\hat{\Omega}$ discussed in Subsection 3.1.1 are used.

The negative results in Theorem 3.3 (i.e., size equal to 1 and/or nuisance-infimal rejection probability equal to 0) are driven by the fact that, due to Assumption 1, the covariance model $\mathfrak{C}$

17

has $e_+e'_+$ and $e_-e'_-$ as limiting points; cf. also Remark 3.4(iv). Suppose now that one would be willing to assume that $\mathfrak{C}$ does not have any singular limiting point (and is norm-bounded which is not really a restriction here). Then the negative results in Theorem 3.3 do not apply. In fact, an application of Theorem 5.21 shows that size is now strictly less than 1 and the infimal power is larger than 0. Does such an assumption on $\mathfrak{C}$ now solve the problem? We do not think so for at least two reasons: First, making the assumption that the covariance model $\mathfrak{C}$ does not have singular limit points (like $e_+e'_+$ and $e_-e'_-$) is highly questionable, especially in view of the fact that the main motivation for the development of autocorrelation robust tests has been the desire to avoid strong assumptions on $\mathfrak{C}$ which could lead to misspecification issues. In particular, in the not unreasonable case where $\mathfrak{C}$ contains AR(1) correlation matrices $\Lambda(\rho)$, such an assumption would require to restrict ρ to an interval $(-1+\varepsilon, 1-\varepsilon)$ for some positive ε. Given the emphasis on unit root and near unit root processes in econometrics, such an assumption seems untenable. Second, even if one is willing to make such an heroic assumption, size or power problems can be present. To see this assume for definiteness of the discussion that $\mathfrak{C} = \mathfrak{C}_{AR(1)}(\varepsilon, \varepsilon) = \{\Lambda(\rho) : \rho \in (-1+\varepsilon, 1-\varepsilon)\}$ for some small $\varepsilon > 0$. As mentioned above, the size of the test based on T will be less than 1 and the infimal power will be larger than 0. However, an upshot of Theorem 3.3 still is that the size will be close to 1 and/or the infimal power will be close to 0 for generic design matrices X, provided ε is small (more precisely, for given sample size n this will happen for sufficiently small ε).[9] Hence, even under such an assumption, size/power problems will disappear only if one is willing to assume a relatively large ε (in relation to sample size n), making the assumption look even more heroic.

In the important special case where an intercept is in the regression and the hypothesis tested does not involve the intercept, a positive result (i.e., size < 1 and infimal power > 0) can be obtained from Theorem 5.21 provided $\mathfrak{C}$ has $e_+e'_+$ as its only singular limit point (or none at all) and is norm-bounded. [To be precise Theorem 5.21 requires matrices in $\mathfrak{C}$ that approach $e_+e'_+$ to do so in a particular manner. This is, e.g., automatically satisfied for $\mathfrak{C} = \mathfrak{C}_{AR(1)}(\varepsilon, 0)$ defined subsequently, cf. Lemma G.1 in Appendix G.] While perhaps a bit more palatable, this is still a restrictive assumption on the covariance model $\mathfrak{C}$. In the AR(1)-case this amounts to $\mathfrak{C} = \mathfrak{C}_{AR(1)}(\varepsilon, 0) = \{\Lambda(\rho) : \rho \in (-1+\varepsilon, 1)\}$, and size or power problems will generically still be present if ε is small as explained in the preceding paragraph.

In the context of the preceding discussion one should recall that in case $\mathfrak{C} = \mathfrak{C}_{AR(1)}$ Theorems 3.6 and 3.7 show how tests, which have size less than 1 and infimal power larger than 0, can easily be obtained without any need of bounding $|\rho|$ away from unity. Therefore, it would be desirable to free Theorems 3.6 and 3.7 from the assumption $\mathfrak{C} = \mathfrak{C}_{AR(1)}$. To what extent this can be achieved without introducing implausible assumptions like the ones discussed in the preceding paragraphs will be discussed elsewhere.

B. In a recent paper Perron and Ren (2011) argue that the impossibility results in Pötscher (2002) for estimating the value of the spectral density at frequency zero are irrelevant in the context of autocorrelation robust testing: In the framework of a Gaussian location model they compare the behavior of common autocorrelation robust tests t_{Robust}, which are standardized with the help of a spectral density estimate $\hat{f}_n(0)$, with a benchmark given by the infeasible test statistic $t_{f(0)}$ that uses the value of the unknown spectral density at frequency zero for standardization. They find that common autocorrelation robust tests beat the infeasible test statistic along a sequence of DGPs similar to the ones that have been used in Pötscher (2002) to establish ill-posedness of

---

[9] Of course, size could be reduced by increasing the critical value, but this would come at the price of even further reduced power.

the spectral density estimation problem. This is certainly true and in fact easy to understand: Consider as another benchmark the infeasible test statistic t_{ideal}, say, which uses the (unknown) finite-sample variance s_n of the arithmetic mean for standardization rather than the asymptotic variance $2\pi f(0)$, and observe that this statistic is exactly $N(0,1)$ distributed (under the null) and has well-behaved size and power properties. Because s_n does in general not converge uniformly to the asymptotic variance $2\pi f(0)$ (for the very same reasons that underlie the impossibility result in Pötscher (2002)) $t_{f(0)}$ is *not* uniformly close to the ideal test t_{ideal}. The fact that $\hat{f}_n(0)$ is also not uniformly close to $f(0)$ (due to the ill-posedness results in Pötscher (2002)) is now "helpful" in the sense that it in principle allows for the possibility that $2\pi\hat{f}_n(0)$ might be closer to the ideal standardization factor s_n than is $2\pi f(0)$, thus allowing for the possibility that t_{Robust} might be closer to the ideal test t_{ideal} than to $t_{f(0)}$. [Observe that $2\pi\hat{f}_n(0)$ as well as s_n each not being uniformly close to $2\pi f(0)$ does in principle not preclude (uniform) closeness between $2\pi\hat{f}_n(0)$ and s_n.] In other words, "aiming" at $f(0)$ in standardizing the test statistic is simply the wrong thing to do. In that sense, the ill-posedness of estimating $f(0)$ is then indeed irrelevant for autocorrelation robust testing (simply because the benchmark $t_{f(0)}$ is irrelevant). As a matter of fact, there is no statement to the contrary in Pötscher (2002): Note that Pötscher (2002) only discusses ill-posedness of the problem of estimating $f(0)$ (considered to be the parameter of interest), and does not make any statements regarding consequences of this ill-posedness for autocorrelation robust tests that use $2\pi\hat{f}_n(0)$ as an estimate of the variance nuisance parameter. The claim opening the last but one paragraph on p.1 in Perron and Ren (2011) is thus simply false. Finally, the preceding discussion begs the question whether or not uniform closeness of $2\pi\hat{f}_n(0)$ and s_n can indeed be established under sufficiently general assumptions on the underlying correlation structure. If possible, this would then immediately transfer the good size and power properties of t_{ideal} to t_{Robust}. However, unfortunately this is not possible: Recall from Example 3.2 that in the location model considered in Perron and Ren (2011) the size of common autocorrelation robust tests like t_{Robust} is always equal to 1.

3.1.3 Further obstructions to favorable size and power properties

The negative results given in Theorem 3.3 rest on Assumption 1, i.e., $\mathfrak{C} \supseteq \mathfrak{C}_{AR(1)}$, and the fact that there exist sequences $\Sigma_m \in \mathfrak{C}_{AR(1)}$ that converge to the singular matrices e_+e_+' or $e_-'e_-$ leading to a concentration phenomenon as discussed in the wake of Theorem 3.3. The commonly used nonparametric covariance models like $\mathfrak{C}_\xi$ discussed at the beginning of Section 3 of course also satisfy $\mathfrak{C}_\xi \supseteq \mathfrak{C}_{AR(p)}$ for every p, where $\mathfrak{C}_{AR(p)}$ is the set of all $n \times n$ correlation matrices arising from stationary autoregressive process of order not larger than p. In this case additional singular limit matrices arise which lead to additional conditions under which size equals 1 or infimal power equals 0. We illustrate this shortly for the case where $\mathfrak{C} \supseteq \mathfrak{C}_{AR(2)}$. To this end define for $\nu \in (0,\pi)$ the matrix $E(\nu)$ as the $n \times 2$ matrix with t-th row equal to $(\cos(t\nu),\sin(t\nu))$. Furthermore set $E(0) = e_+$ and $E(\pi) = e_-$. In Lemma G.2 in Appendix G we show that the matrices $E(\nu)E(\nu)'$ for $\nu \in [0,\pi]$ arise as limits of sequences of matrices in $\mathfrak{C}_{AR(2)}$. Obviously, $E(\nu)E(\nu)'$ is singular whenever $n \geq 3$. Restricting ν to the set $\{0,\pi\}$ in the subsequent theorem reproduces the conditions appearing in Theorem 3.3 (albeit under the stronger assumptions that $\mathfrak{C} \supseteq \mathfrak{C}_{AR(2)}$ and $n \geq 3$).

Theorem 3.11. *Suppose $\mathfrak{C} \supseteq \mathfrak{C}_{AR(2)}$, Assumptions 2 and 3 are satisfied, and $n \geq 3$ holds. Let T be the test statistic defined in (7) with $\hat{\Psi}_w$ as in (6). Let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ be the rejection region where C is a real number satisfying $0 < C < \infty$. Then the following holds:*

1. *Suppose there exists a $\nu \in [0, \pi]$ such that $\mathrm{rank}\left(B(z)\right) = q$ and $T(z + \mu_0^*) > C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$ and for $\lambda_{\mathrm{span}(E(\nu))}$-almost all $z \in \mathrm{span}\left(E(\nu)\right)$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 1$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

2. *Suppose there exists a $\nu \in [0, \pi]$ such that $\mathrm{rank}\left(B(z)\right) = q$ and $T(z + \mu_0^*) < C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$ and for $\lambda_{\mathrm{span}(E(\nu))}$-almost all $z \in \mathrm{span}\left(E(\nu)\right)$. Then*

$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 0$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$, and hence

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 0$$

   holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased. Furthermore, the nuisance-infimal rejection probability at every point $\mu_1 \in \mathfrak{M}_1$ is zero, i.e.,

$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 0.$$

   In particular, the infimal power of the test is equal to zero.

3. *Suppose there exists a $\nu \in [0, \pi]$ such that $B(z) = 0$ and $R\hat{\beta}(z) \neq 0$ hold for $\lambda_{\mathrm{span}(E(\nu))}$-almost all $z \in \mathrm{span}\left(E(\nu)\right)$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}\left(W\left(C\right)\right) = 1$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

To illustrate the value added of the preceding theorem when compared to Theorem 3.3 consider the following example: Assume that e_+ and e_- are both elements of $\mathfrak{M}$ and $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$. Then none of the conditions in Theorem 3.3 are satisfied and thus this theorem is not applicable. Suppose now that the design matrix X contains $E(\nu)$ for some $\nu \in (0, \pi)$ as a submatrix, i.e., seasonal regressors are included. Without loss of generality assume that $X = (E(\nu), X^{(2)})$. If we want to test for absence of seasonality at angular frequency ν, this corresponds to $R = (I_2, 0)$ and $r = 0$. In case Assumption 3 holds, the conditions in Case 3 of the preceding theorem are then obviously satisfied and we conclude that the size of the test for absence of seasonality is equal to one. [In case Assumption 3 is violated, the test breaks down in a trivial way as noted earlier.]

We finally ask what happens if we allow for covariance structures deriving from even higher-order autoregressive models, i.e., $\mathfrak{C} \supseteq \mathfrak{C}_{AR(p)}$ with $p > 2$. While additional concentration spaces arise and theorems like the one above can be easily obtained form Corollary 5.17, these theorems do not generate new obstructions to good size and power properties. The reason for this is that any of the newly arising concentration spaces already contains one of the concentration spaces $\mathrm{span}\left(E(\nu)\right)$ for $\nu \in [0, \pi]$ as a subset.

3.2 Parametrically based autocorrelation robust tests

The results in Subsection 3.1 were given for autocorrelation robust tests that make use of a non-parametric estimator $\hat{\Omega}$. In this subsection we show that the phenomena encountered in Subsection 3.1 (size distortions and power deficiencies) are not a consequence of the nonparametric nature of the estimator, but can equally arise if a parametric estimator is being used (and even if the parametric model employed correctly describes the covariance structure of the errors). We illustrate this for the case where the test statistic is obtained from a feasible generalized least squares (GLS) estimator predicated on an AR(1) covariance structure, as well as for the case where the test statistic is obtained from the ordinary least squares (OLS) estimator combined with an estimator for the variance covariance matrix again predicated on the same covariance structure. The theoretical results derived below are in line with Monte Carlo results provided in Park and Mitchell (1980) and Magee (1989).

We start with the estimator $\hat{\rho}$ that will be used in the feasible GLS procedure as well as in the estimator for the variance covariance matrix of the OLS estimator.

Assumption 4. *For $a_1 \in \{1, 2\}$ and $a_2 \in \{n - 1, n\}$ with $a_1 \leq a_2$ the estimator $\hat{\rho}$ is of the form*

$$\hat{\rho}(y) = \sum_{t=2}^{n} \hat{u}_t(y)\hat{u}_{t-1}(y) \left/ \sum_{t=a_1}^{a_2} \hat{u}_t^2(y) \right.$$

for all $y \in \mathbb{R}^n \backslash N_0(a_1, a_2)$ and it is undefined for $y \in N_0(a_1, a_2) = \left\{ y \in \mathbb{R}^n : \sum_{t=a_1}^{a_2} \hat{u}_t^2(y) = 0 \right\}$.

The Yule-Walker estimator, which we shall abbreviate by $\hat{\rho}_{YW}$, corresponds to $a_1 = 1$, $a_2 = n$, while the least squares estimator $\hat{\rho}_{LS}$ corresponds to $a_1 = 1$, $a_2 = n - 1$. The estimators which use $a_1 = 2$, $a_2 = n - 1$ or $a_1 = 2$, $a_2 = n$ have also been considered in the literature (see, e.g., Park and Mitchell (1980), Magee (1989)).

Remark 3.12. *(Some properties of $\hat{\rho}$)* (i) For the Yule-Walker estimator $\hat{\rho}_{YW}$ we have $N_0(1, n) = \mathfrak{M}$, i.e., $\hat{\rho}_{YW}$ is well-defined for every $y \in \mathbb{R}^n \backslash \mathfrak{M}$. Furthermore, $\hat{\rho}_{YW}$ is bounded away from 1 in modulus uniformly over its domain of definition, i.e., $\sup_{y \in \mathbb{R}^n \backslash \mathfrak{M}} |\hat{\rho}_{YW}(y)| < 1$ holds. This follows easily from the well-known fact that $|\hat{\rho}_{YW}(y)| < 1$, that the supremum in question does not change its value if the range for y is replaced by the compact set $\left\{ y \in \mathfrak{M}^\perp : \|y\| = 1 \right\}$, and the fact that $\hat{\rho}_{YW}$ is continuous on this set. [It can also be derived from the discussion in Section 3.5 in Grenander and Rosenblatt (1957).]

(ii) The least squares estimator $\hat{\rho}_{LS}$ exhibits a somewhat different behavior: First, $\hat{\rho}_{LS}$ is well defined only on $\mathbb{R}^n \backslash N_0(1, n - 1)$, with $N_0(1, n - 1)$ given by $\{ y \in \mathbb{R}^n : \hat{u}(y) \in \text{span}(e_n(n)) \}$. Note that $\mathbb{R}^n \backslash N_0(1, n - 1)$ is contained in $\mathbb{R}^n \backslash \mathfrak{M}$, but is strictly smaller in case $e_n(n)$ is orthogonal to each column of X. Second, $\hat{\rho}_{LS}$ is not bounded away from one in modulus, in fact $|\hat{\rho}_{LS}| \geq 1$ can occur.[10]

(iii) The behavior of the remaining two estimators $\hat{\rho}$ is similar to the behavior of $\hat{\rho}_{LS}$.

(iv) The set $N_0(a_1, a_2)$ is always a closed subset of $\mathbb{R}^n$. It is guaranteed to be a $\lambda_{\mathbb{R}^n}$-null set provided $k \leq a_2 - a_1$ holds, cf. Lemma 3.13 below. This condition on k is no restriction in the case of the Yule-Walker estimator (since we have assumed $k < n$ from the beginning), and is a very mild condition in the other cases (requiring $k \leq n - 2$ or $k \leq n - 3$ at most).

---

[10] There are even cases where $\hat{\rho}_{LS}$ is unbounded.

The definition of the test statistics below will require inversion of $\Lambda(\hat{\rho})$. While $\Lambda(\hat{\rho})$ is nonsingular if $|\hat{\rho}| \neq 1$, $\Lambda(\hat{\rho})$ is singular if $|\hat{\rho}| = 1$, and hence we need to study the set of y where $|\hat{\rho}(y)| = 1$ (or $\hat{\rho}(y)$ is undefined).

Lemma 3.13. *Let $\hat{\rho}$ satisfy Assumption 4. Then $\mathfrak{M} \subseteq N_0(a_1, a_2) \subseteq N_1(a_1, a_2)$ where*

$$N_1(a_1, a_2) = \left\{ y \in \mathbb{R}^n : \left| \sum_{t=2}^{n} \hat{u}_t(y)\hat{u}_{t-1}(y) \right| = \sum_{t=a_1}^{a_2} \hat{u}_t^2(y) \right\}.$$

The set $N_1(a_1, a_2)$ is a closed subset of $\mathbb{R}^n$ and is precisely the set where the estimator $\hat{\rho}$ is either not well-defined or is equal to 1 in modulus. The estimator $\hat{\rho}$ is continuous on $\mathbb{R}^n \backslash N_0(a_1, a_2) \supseteq \mathbb{R}^n \backslash N_1(a_1, a_2)$. If $k \leq a_2 - a_1$ holds, the set $N_1(a_1, a_2)$ is a $\lambda_{\mathbb{R}^n}$-null set.

While for the Yule-Walker estimator $N_1(1, n) = N_0(1, n)$ holds as a consequence of Remark 3.12(i), for the other estimators $\hat{\rho}$ the corresponding set $N_1(a_1, a_2)$ can be a proper superset of $N_0(a_1, a_2)$.

Given an estimator $\hat{\rho}$ satisfying Assumption 4 we now introduce the test statistic

$$T_{FGLS}(y) = \begin{cases} (R\tilde{\beta}(y) - r)'\tilde{\Omega}^{-1}(y)(R\tilde{\beta}(y) - r) & \text{if } y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2), \\ 0 & \text{else.} \end{cases}$$

where

$$\tilde{\beta}(y) = (X'\Lambda^{-1}(\hat{\rho}(y))X)^{-1}X'\Lambda^{-1}(\hat{\rho}(y))y,$$
$$\tilde{\sigma}^2(y) = (n-k)^{-1}(y - X\tilde{\beta}(y))'\Lambda^{-1}(\hat{\rho}(y))(y - X\tilde{\beta}(y)),$$
$$\tilde{\Omega}(y) = \tilde{\sigma}^2(y)R(X'\Lambda^{-1}(\hat{\rho}(y))X)^{-1}R'.$$

Here $N_2^*(a_1, a_2)$ is defined via

$$\mathbb{R}^n \backslash N_2^*(a_1, a_2) = \left\{ y \in \mathbb{R}^n \backslash N_2(a_1, a_2) : \tilde{\sigma}^2(y) \neq 0, \det\left(R(X'\Lambda^{-1}(\hat{\rho}(y))X)^{-1}R'\right) \neq 0 \right\},$$

where $N_2(a_1, a_2)$ is given by

$$\mathbb{R}^n \backslash N_2(a_1, a_2) = \left\{ y \in \mathbb{R}^n \backslash N_1(a_1, a_2) : \det\left(X'\Lambda^{-1}(\hat{\rho}(y))X\right) \neq 0 \right\}.$$

Note that $\tilde{\beta}$, $\tilde{\sigma}^2$, and $\tilde{\Omega}$ are well-defined on $\mathbb{R}^n \backslash N_2(a_1, a_2)$, with $\tilde{\Omega}(y)$ being nonsingular if and only if $y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$, see Lemma B.1 in Appendix B. Furthermore, define

$$T_{OLS}(y) = \begin{cases} (R\hat{\beta}(y) - r)'\hat{\Omega}^{-1}(y)(R\hat{\beta}(y) - r) & \text{if } y \in \mathbb{R}^n \backslash N_0^*(a_1, a_2), \\ 0 & \text{else,} \end{cases}$$

where $\hat{\beta}(y)$ is the OLS-estimator, $\hat{\sigma}^2(y) = (n-k)^{-1}\hat{u}'(y)\hat{u}(y)$, and

$$\hat{\Omega}(y) = \hat{\sigma}^2(y)R(X'X)^{-1}X'\Lambda(\hat{\rho}(y))X(X'X)^{-1}R'.$$

Here $N_0^*(a_1, a_2)$ is defined via

$$\mathbb{R}^n \backslash N_0^*(a_1, a_2) = \left\{ y \in \mathbb{R}^n \backslash N_0(a_1, a_2) : \det\left(R(X'X)^{-1}X'\Lambda(\hat{\rho}(y))X(X'X)^{-1}R'\right) \neq 0 \right\}.$$

22

Of course, $\hat{\beta}$ and $\hat{\sigma}^2$ are well-defined on all of $\mathbb{R}^n$, while $\hat{\Omega}$ is well-defined on $\mathbb{R}^n \backslash N_0(a_1, a_2) \supseteq$ $\mathbb{R}^n \backslash N_0^*(a_1, a_2)$. Furthermore, $\hat{\Omega}(y)$ is nonsingular for $y \in \mathbb{R}^n \backslash N_0^*(a_1, a_2)$, see Lemma B.1 in Appendix B. We note that the exceptional sets $N_0^*(a_1, a_2)$ and $N_2^*(a_1, a_2)$, respectively, appearing in the definition of the test statistics are $\lambda_{\mathbb{R}^n}$-null sets provided $k \leq a_2 - a_1$ holds, see Lemma B.1. [For the case of the Yule-Walker estimator actually $N_2^*(1, n) = N_2(1, n) = N_1(1, n) = N_0^*(1, n) = N_0(1, n) = \mathfrak{M}$ holds, because $\Lambda(\hat{\rho}_{YW}(y))$ is positive definite for every $y \notin N_0(1, n) = \mathfrak{M}$ in view of $|\hat{\rho}_{YW}(y)| < 1$, cf. Remark 3.12(i).]

As already noted in Remark 3.12, except for the Yule-Walker estimator we can not rule out that $\hat{\rho}(y)$ is larger than one in absolute value. For such values of y the matrix $\Lambda(\hat{\rho}(y))$, although being nonsingular, is indefinite. [To see this, note that $\det \Lambda(\hat{\rho}(y)) = (1 - \hat{\rho}^2(y))^{n-1}$, which is negative for $|\hat{\rho}(y)| > 1$ if n is even. Hence there must exist a negative and a positive eigenvalue. For odd $n > 1$ the claim then follows from Cauchy's interlacing theorem.] In fact, if $|\hat{\rho}(y)| > 1$ occurs for some y, then it occurs on a set of positive $\lambda_{\mathbb{R}^n}$-measure in view of continuity of $\hat{\rho}$. As a consequence, $\tilde{\Omega}(y)$ and $\hat{\Omega}(y)$ are not guaranteed to be $\lambda_{\mathbb{R}^n}$-almost everywhere nonnegative definite (except if the Yule-Walker estimator is being used), although they are $\lambda_{\mathbb{R}^n}$-almost everywhere nonsingular in case $k \leq a_2 - a_1$. Of course, the probability of the event $|\hat{\rho}(y)| > 1$ will go to zero as sample size goes to infinity, but this is not relevant for the present finite-sample analysis and the complications ensuing from $|\hat{\rho}(y)| > 1$ have to be dealt with. Fortunately, the theory in Subsection 5.4 does not require the estimated variance covariance matrices to be nonnegative definite almost everywhere but only requires some weaker properties to be satisfied which are formalized in Assumptions 6 and 7 in Subsection 5.4. Lemma B.3 in Appendix B shows that $\tilde{\Omega}$ and $\hat{\Omega}$ satisfy these assumptions.

The subsequent theorem provides a negative result that is similar in spirit to Theorem 3.3.

Theorem 3.14. *Suppose Assumptions 1 and 4 are satisfied and $k \leq a_2 - a_1$ holds. Let $W_{FGLS}(C) = \{y \in \mathbb{R}^n : T_{FGLS}(y) \geq C\}$ and $W_{OLS}(C) = \{y \in \mathbb{R}^n : T_{OLS}(y) \geq C\}$ be the rejection regions corresponding to the test statistics T_{FGLS} and T_{OLS}, respectively, where C is a real number satisfying $0 < C < \infty$. Then the following holds:*

1. *Suppose $e_+ \notin N_2^*(a_1, a_2)$ and $T_{FGLS}(e_+ + \mu_0^*) > C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$, or $e_- \notin N_2^*(a_1, a_2)$ and $T_{FGLS}(e_- + \mu_0^*) > C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W_{FGLS}(C)) = 1$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

2. *Suppose $e_+ \notin N_2^*(a_1, a_2)$ and $T_{FGLS}(e_+ + \mu_0^*) < C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$, or $e_- \notin N_2^*(a_1, a_2)$ and $T_{FGLS}(e_- + \mu_0^*) < C$ hold for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W_{FGLS}(C)) = 0$$

   holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$, and hence

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W_{FGLS}(C)) = 0$$

   holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased. Furthermore, the nuisance-infimal rejection probability at every point $\mu_1 \in \mathfrak{M}_1$ is zero, i.e.,

$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W_{FGLS}(C)) = 0.$$

In particular, the infimal power of the test is equal to zero.

3. *Suppose that $e_+ \in \mathfrak{M}$ and $R\hat{\beta}(e_+) \neq 0$ hold. Then there exists a constant $K_{FGLS}(e_+)$, which depends only on e_+, R, X, and $\mathfrak{C}$, such that for every $\mu_0 \in \mathfrak{M}_0$, every σ with $0 < \sigma < \infty$, and every $M \geq 0$ we have*

$$\inf_{\gamma \in \mathbb{R}, |\gamma| \geq M} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_0 + \gamma e_+, \sigma^2 \Sigma} \left(W_{FGLS}(C) \right) \leq K_{FGLS}(e_+) \leq \sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma} \left(W_{FGLS}(C) \right);$$

Note that $\mu_0 + \gamma e_+ \in \mathfrak{M}_1$ for $\gamma \neq 0$. Furthermore, if $\hat{\rho} \equiv \hat{\rho}_{YW}$, then $K_{FGLS}(e_+) = 1$ and hence

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma} \left(W_{FGLS}(C) \right) = 1 \tag{9}$$

holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. If $e_- \in \mathfrak{M}$ and $R\hat{\beta}(e_-) \neq 0$ hold then the analogous statements hold with e_+ replaced by e_- where the constant $K_{FGLS}(e_-)$ now depends only on e_-, R, X, and $\mathfrak{C}$.

4. *Statements analogous to 1.-3. hold true if T_{FGLS} is replaced by T_{OLS}, $W_{FGLS}(C)$ is replaced by $W_{OLS}(C)$, the set $N_2^*(a_1, a_2)$ is replaced by $N_0^*(a_1, a_2)$, and the constants $K_{FGLS}(\cdot)$ are replaced by constants $K_{OLS}(\cdot)$.*

The meaning of Parts 1 and 2 of the preceding theorem is similar to the meaning of the corresponding parts of Theorem 3.3. Part 3 differs somewhat from the corresponding part of the earlier theorem, and tells us that, given the conditions in Part 3 are met, there exist points in the alternative, arbitrarily far away from the null hypothesis, at which power is not larger than the size of the test. The reason for the difference between Part 3 of Theorem 3.3 and Part 3 of the preceding theorem lies in the fact that the variance covariance matrix estimators $\tilde{\Omega}$ and $\hat{\Omega}$ used in the present subsection can be indefinite. This requires one in the proof of the preceding theorem to resort to Theorem 5.19 rather than to using Part 3 of Corollary 5.17. In the case where the Yule-Walker estimator $\hat{\rho}_{YW}$ is used, the matrices $\tilde{\Omega}$ and $\hat{\Omega}$ are always nonnegative definite, and Part 3 of Corollary 5.17 as well as Theorem 5.19 deliver the same conclusion (9) in this case. Recall also that in this case the exceptional null sets appearing in Parts 1 and 2 satisfy $N_2^*(1, n) = N_0^*(1, n) = \mathfrak{M}$. In view of the general results in Subsection 5.4 there is little doubt that similar negative results can also be obtained for FGLS or OLS based tests that are constructed on the basis of higher order autoregressive AR models or of other more profligate parametric models (as long as $\mathfrak{C} \supseteq \mathfrak{C}_{AR(1)}$ is assumed). Hence it is to be expected that autocorrelation robust tests based on autoregressive estimates (cf. Berk (1974), den Haan and Levin (1997), Sun and Kaplan (2012)) will also suffer from severe size and power problems.

The results given in the preceding theorem reveal serious size and power problems of the tests based on T_{FGLS} and T_{OLS}. Note that these problems arise even if $\mathfrak{C} = \mathfrak{C}_{AR(1)}$, i.e., even if the construction of the test statistics makes use of the correct covariance model. If $\mathfrak{C} = \mathfrak{C}_{AR(1)}$ holds, it is interesting to contrast the above results with the size and power properties of the corresponding infeasible tests based on T_{GLS}^* and T_{OLS}^* which are defined in a similar way as T_{FGLS} and T_{OLS} are, but with $\hat{\rho}$ replaced by the true value of ρ: These tests are standard F-tests (except for not being standardized by q), have well-known and reasonable size and power properties, and do not suffer from the size and power problems exhibited by their feasible counterparts.

Similar to the situation in Subsection 3.1, the conditions in Parts 1-3 of the preceding theorem only depend on a_1 and a_2 (i.e., on the choice of estimator $\hat{\rho}$), the design matrix X, the restriction

(R, r), the vector e_+ $(e_-,$ respectively), and the critical value C. Hence, in any particular application it can be decided whether or not (and which of) these conditions are satisfied. We furthermore note that a remark analogous to Remark 3.4 also applies mutatis mutandis to the preceding theorem. We also note that a result analogous to Theorem 3.11 could be given here, but we do not spell out the details.

We next show that the conditions of Theorem 3.14 involving the design matrix X are generically satisfied. The first part of the subsequent proposition shows that these conditions are generically satisfied in the class of all possible design matrices of rank k. Parts 2 and 3 show a corresponding result if we impose that the regression model has to contain an intercept. In the proposition the dependence of several quantities like T_{FGLS}, T_{OLS}, $N_2^*(a_1, a_2)$, etc on the design matrix X will be important and thus we shall write $T_{FGLS,X}$, $T_{OLS,X}$, $N_{2,X}^*(a_1, a_2)$, etc for these quantities in the result to follow.

Proposition 3.15. *Suppose Assumption 1 holds. Fix (R, r) with $\mathrm{rank}\,(R) = q$, fix $0 < C < \infty$, and fix $a_1 \in \{1, 2\}$ and $a_2 \in \{n-1, n\}$ in Assumption 4. Suppose $k \le a_2 - a_1$ holds. Let $T_{FGLS,X}$ and $T_{OLS,X}$ be the test statistic defined above and let $\mu_0^* \in \mathfrak{M}_0$ be arbitrary.*

1. *With $\mathfrak{X}_0$ defined in Proposition 3.5 define now*

$$
\begin{aligned}
\mathfrak{X}_{1,FGLS}(e_+) &= \left\{ X \in \mathfrak{X}_0 : e_+ \in N_{2,X}^*(a_1, a_2) \right\}, \\
\mathfrak{X}_{2,FGLS}(e_+) &= \left\{ X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+) : T_{FGLS,X}(e_+ + \mu_0^*) = C \right\},
\end{aligned}
$$

and similarly define $\mathfrak{X}_{1,FGLS}(e_-)$, $\mathfrak{X}_{2,FGLS}(e_-)$. [Note that $\mathfrak{X}_{2,FGLS}(e_+)$ and $\mathfrak{X}_{2,FGLS}(e_-)$ do not depend on the choice of μ_0^*.] Then $\mathfrak{X}_{1,FGLS}(e_+)$ and $\mathfrak{X}_{1,FGLS}(e_-)$ are $\lambda_{\mathbb{R}^{n \times k}}$-null sets. The same is true for $\mathfrak{X}_{2,FGLS}(e_+)$ ($\mathfrak{X}_{2,FGLS}(e_-)$, respectively) under the provision that it is a proper subset of $\mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ ($\mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_-)$, respectively). The set of all design matrices $X \in \mathfrak{X}_0$ for which Theorem 3.14 does not apply is a subset of

$$
\left(\mathfrak{X}_{1,FGLS}(e_+) \cup \mathfrak{X}_{2,FGLS}(e_+) \right) \cap \left(\mathfrak{X}_{1,FGLS}(e_-) \cup \mathfrak{X}_{2,FGLS}(e_-) \right),
$$

Hence it is a $\lambda_{\mathbb{R}^{n \times k}}$-null set provided the preceding provision holds for at least one of $\mathfrak{X}_{2,FGLS}(e_+)$ or $\mathfrak{X}_{2,FGLS}(e_-)$; it thus is a "negligible" subset of $\mathfrak{X}_0$ in view of the fact that $\mathfrak{X}_0$ differs from $\mathbb{R}^{n \times k}$ only by a $\lambda_{\mathbb{R}^{n \times k}}$-null set.

2. *Suppose $k \ge 2$ and $n \ge 4$ hold and suppose X has e_+ as its first column, i.e., $X = \left(e_+, \tilde{X} \right)$. With $\tilde{\mathfrak{X}}_0$ defined in Proposition 3.5 define*

$$
\begin{aligned}
\tilde{\mathfrak{X}}_{1,FGLS}(e_-) &= \left\{ \tilde{X} \in \tilde{\mathfrak{X}}_0 : e_- \in N_{2,(e_+,\tilde{X})}^*(a_1, a_2) \right\}, \\
\tilde{\mathfrak{X}}_{2,FGLS}(e_-) &= \left\{ \tilde{X} \in \tilde{\mathfrak{X}}_0 \backslash \tilde{\mathfrak{X}}_{1,FGLS}(e_-) : T_{FGLS,(e_+,\tilde{X})}(e_- + \mu_0^*) = C \right\},
\end{aligned}
$$

and note that $\tilde{\mathfrak{X}}_{2,FGLS}(e_-)$ does not depend on the choice of μ_0^*. Then $\tilde{\mathfrak{X}}_{1,FGLS}(e_-)$ is a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set. The set $\tilde{\mathfrak{X}}_{2,FGLS}(e_-)$ is a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set under the provision that it is a proper subset of $\tilde{\mathfrak{X}}_0 \backslash \tilde{\mathfrak{X}}_{1,FGLS}(e_-)$. [The analogously defined sets $\tilde{\mathfrak{X}}_{1,FGLS}(e_+)$ and $\tilde{\mathfrak{X}}_{2,FGLS}(e_+)$ satisfy $\tilde{\mathfrak{X}}_{1,FGLS}(e_+) = \tilde{\mathfrak{X}}_0$ and $\tilde{\mathfrak{X}}_{2,FGLS}(e_+) = \emptyset$.] The set of all matrices $\tilde{X} \in \tilde{\mathfrak{X}}_0$ such that Theorem 3.14 does not apply to the design matrix $X = \left(e_+, \tilde{X} \right)$ is a subset

of $\tilde{\mathfrak{X}}_{1,FGLS}(e_-) \cup \tilde{\mathfrak{X}}_{2,FGLS}(e_-)$ and hence is a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set under the preceding provision; it thus is a "negligible" subset of $\tilde{\mathfrak{X}}_0$ in view of the fact that $\tilde{\mathfrak{X}}_0$ differs from $\mathbb{R}^{n \times (k-1)}$ only by a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set.

3. Define $\mathfrak{X}_{1,OLS}(\cdot)$ and $\mathfrak{X}_{2,OLS}(\cdot)$ analogously, but with $N_{0,X}^*(a_1, a_2)$ replacing $N_{2,X}^*(a_1, a_2)$ and $T_{OLS,X}$ replacing $T_{FGLS,X}$. Similarly define $\tilde{\mathfrak{X}}_{1,OLS}(\cdot)$ and $\tilde{\mathfrak{X}}_{2,OLS}(\cdot)$. Then Part 1 (Part 2, respectively) holds analogously for $\mathfrak{X}_{1,OLS}(\cdot)$ and $\mathfrak{X}_{2,OLS}(\cdot)$ ($\tilde{\mathfrak{X}}_{1,OLS}(\cdot)$ and $\tilde{\mathfrak{X}}_{2,OLS}(\cdot)$, respectively) with obvious changes.

4. Suppose $X = \left(e_+, \tilde{X} \right)$, and suppose the first column of R is nonzero. Then Part 3 of Theorem 3.14 applies to the design matrix $X = \left(e_+, \tilde{X} \right)$ for every $\tilde{X} \in \tilde{\mathfrak{X}}_0$ (for the FGLS- as well as for the OLS-based test).

The preceding genericity result maintains in Part 1 the provision that $\mathfrak{X}_{2,FGLS}(e_+)$ is a proper subset of $\mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ or that $\mathfrak{X}_{2,FGLS}(e_-)$ is a proper subset of $\mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_-)$. Note that the provision depends on the critical value C. If the provision is satisfied for the given C, we can conclude from Part 1 that the set of all design matrices $X \in \mathfrak{X}_0$ for which Theorem 3.14 is not applicable to the test statistic T_{FGLS} is "negligible". If the provision is not satisfied, i.e., if $\mathfrak{X}_{2,FGLS}(e_+) = \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ and $\mathfrak{X}_{2,FGLS}(e_-) = \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_-)$ holds, and thus we cannot draw the desired conclusion for the given value of C, we immediately see that the provision must then be satisfied for any *other* choice C' of the critical value; hence, negligibility of the set of design matrices for which Theorem 3.14 is not applicable to the test statistic T_{FGLS} can then be concluded for any $C' \neq C$. Summarizing we see that the provision is always satisfied except possible for one particular choice of the critical value. A similar comment applies to Parts 2 and 3 of the proposition.[11]

Similar as in Subsection 3.1, we next discuss an exceptional case to which Theorem 3.14 does not apply and which allows for a positive result, at least if the covariance model $\mathfrak{C}$ is assumed to be $\mathfrak{C}_{AR(1)}$ or is approximated by $\mathfrak{C}_{AR(1)}$ near the singular points (in the sense of Remark 3.9).

Theorem 3.16. *Suppose* $\mathfrak{C} = \mathfrak{C}_{AR(1)}$, *Assumption 4 is satisfied, and* $k \leq a_2 - a_1$ *holds. Let* $W_{FGLS}(C) = \{y \in \mathbb{R}^n : T_{FGLS}(y) \geq C\}$ *and* $W_{OLS}(C) = \{y \in \mathbb{R}^n : T_{OLS}(y) \geq C\}$ *be the rejection regions corresponding to the test statistics* T_{FGLS} *and* T_{OLS}, *respectively, where* C *is a real number satisfying* $0 < C < \infty$. *If* $e_+, e_- \in \mathfrak{M}$ *and* $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$ *is satisfied, then the following holds for* $W(C) = W_{FGLS}(C)$ *as well as* $W(C) = W_{OLS}(C)$:

1. *The size of the rejection region* $W(C)$ *is strictly less than 1, i.e.,*

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)}(W(C)) < 1.$$

Furthermore,

$$\inf_{\mu_0 \in \mathfrak{M}_0} \inf_{0 < \sigma^2 < \infty} \inf_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)}(W(C)) > 0.$$

---

[11]For example, if T_{OLS} is used, $a_1 = 1$, $a_2 = n$ (Yule-Waker estimator), and X is not restricted to be of the form $\left(e_+, \tilde{X} \right)$, it is not difficult to show that the provision is in fact satisfied for *every* choice of C. This can also be shown for other choices of a_1 and a_2 and/or for the case where $X = \left(e_+, \tilde{X} \right)$ under additional assumptions on R. It may actually be true in general, but we do not want to pursue this.

2. *The infimal power is bounded away from zero, i.e.,*

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{0 < \sigma^2 < \infty} \inf_{-1 < \rho < 1} P_{\mu_1, \sigma^2 \Lambda(\rho)}(W(C)) > 0.$$

3. *Suppose that $a_1 = 1$ and $a_2 = n$. Then for every $0 < c < \infty$*

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, 0 < \sigma^2 < \infty \\ d(\mu_1, \mathfrak{M}_0)/\sigma \geq c}} P_{\mu_1, \sigma^2 \Lambda(\rho_m)}(W(C)) \to 1$$

holds for $m \to \infty$ and for any sequence $\rho_m \in (-1, 1)$ satisfying $|\rho_m| \to 1$. Furthermore, for every sequence $0 < c_m < \infty$ and every $0 < \varepsilon < 1$

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, \\ d(\mu_1, \mathfrak{M}_0) \geq c_m}} \inf_{-1 + \varepsilon \leq \rho \leq 1 - \varepsilon} P_{\mu_1, \sigma_m^2 \Lambda(\rho)}(W(C)) \to 1$$

holds for $m \to \infty$ whenever $0 < \sigma_m^2 < \infty$ and $c_m/\sigma_m \to \infty$. [The very last statement holds even without the conditions $e_+, e_- \in \mathfrak{M}$ and $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$.]

4. *For every δ, $0 < \delta < 1$, there exists a $C(\delta)$, $0 < C(\delta) < \infty$, such that*

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{-1 < \rho < 1} P_{\mu_0, \sigma^2 \Lambda(\rho)}(W(C(\delta))) \leq \delta.$$

A discussion similar to the one following Theorem 3.6 applies also here. Furthermore, a result paralleling Theorem 3.7 can again be obtained by a combined application of Theorem 5.21 and Proposition 5.23. The so-obtained result shows how adjusted test statistics $\bar{T}_{FGLS}$ and $\bar{T}_{OLS}$ can be constructed that have size/power properties as given in the preceding theorem also in many cases which fall under the wrath of Theorem 3.14 (and for which the tests based on T_{FGLS} and T_{OLS} suffer from extreme size or power deficiencies). The adjustment mechanism again amounts to using a "working model" that always adds the regressors e_+ and/or e_- to the design matrix. We abstain from providing details.

3.3 Some remarks on the F-test without correction for autocorrelation

As mentioned in the introduction, a considerable body of literature is concerned with the properties of the *standard F-test* (i.e., the F-test without correction for autocorrelation) in the presence of autocorrelation. Much of this literature concentrates on the case where the errors follow a stationary autoregressive process of order 1, i.e., $\mathfrak{C} = \mathfrak{C}_{AR(1)}$. As the correlation in the errors is not accounted for in the standard F-test, bad performance of the standard F-test for large values of the correlation ρ can be expected. This has been demonstrated formally in Krämer (1989), Krämer et al. (1990), and subsequently in Banerjee and Magnus (2000): These papers determine the limit as $\rho \to 1$ of the error of the first kind of the standard F-test and show that (i) this limit is 1 if the regression contains an intercept and the restrictions to be tested involve the intercept (i.e., the $n \times 1$ vector $e_+ = (1, \ldots, 1)'$ belongs to the span of the design matrix and $R\hat{\beta}(e_+) \neq 0$ holds) or if the regression does not contain an intercept (i.e., e_+ does not belong to the span of the design matrix) and a certain observable quantity, A say, is positive, (ii) it is 0 if the regression does not contain an intercept and the observable quantity A is negative, and (iii) it is a value between 0 and 1 if the regression

27

contains an intercept but the restrictions to be tested do not involve the intercept (i.e., e_+ belongs to the span of the design matrix and $R\hat{\beta}(e_+) = 0$ holds).[12] It perhaps comes as a surprise that autocorrelation robust tests, which have built into them a correction for autocorrelation, exhibit a similar behavior as shown in Section 3 of the present paper. We mention that, due to the relatively simple structure of the standard F-test statistic as a ratio of quadratic forms, the method of proof in Krämer (1989), Krämer et al. (1990), and Banerjee and Magnus (2000) is by direct computation of the limit (as $\rho \to 1$) of the test statistic. In contrast, the results for the much more complicated test statistics considered in the present paper rely on quite different methods which make use of invariance considerations and are of a more geometric flavor. Needless to say, the just mentioned results in Krämer (1989), Krämer et al. (1990), and Banerjee and Magnus (2000) can be rederived through a straightforward application of the general results in Subsection 5.4 to the standard F-test.

In light of the fact that the standard F-test makes no correction for autocorrelation at all, a perhaps surprising observation is that nevertheless an analogue to Theorems 3.6 and 3.16 can be established for the standard F-test by a simple application of Theorem 5.21. Even more, the adjustment procedure described in Proposition 5.23 can be applied to the standard F-test leading to a result analogous to Theorem 3.7. While these results show that the size and power of the so-adjusted standard F-test do not "break down" completely for extreme correlations, they do not tell us much about the performance of the adjusted test for moderate correlations.

4 Size and Power of Tests of Linear Restrictions in Regression Models with Heteroskedastic Disturbances

We next turn to size and power properties of commonly used heteroskedasticity robust tests. To this end we allow for heteroskedasticity of unknown form as is common in the literature and thus allow that the errors in the regression model have a variance covariance matrix $\sigma^2 \Sigma$ where Σ is an element of the covariance model given by

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

The normalization for Σ chosen is of course arbitrary and could equally well be replaced, e.g., by the normalization $\tau_1^2 = 1$. The heteroskedasticity robust test statistic considered is given by

$$T_{Het}(y) = \begin{cases} (R\hat{\beta}(y) - r)' \hat{\Omega}_{Het}^{-1}(y)(R\hat{\beta}(y) - r) & \text{if } \det \hat{\Omega}_{Het}(y) \neq 0, \\ 0 & \text{if } \det \hat{\Omega}_{Het}(y) = 0, \end{cases} \tag{10}$$

where $\hat{\Omega}_{Het} = R\hat{\Psi}_{Het}R'$ and $\hat{\Psi}_{Het}$ is a heteroskedasticity robust estimator. Such estimators were introduced in Eicker (1963, 1967) and have later found their way into the econometrics literature (e.g., White (1980)). They are of the form

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

where the constants $d_i > 0$ may depend on the design matrix. Typical choices for d_i 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 i-th diagonal element of

---

[12]Banerjee and Magnus (2000) claim in their Theorem 5 that the expression $\Pr(F(0) > \delta)$ converges to zero if $Mi \neq 0$ and $\bar{F}(0) \leq \delta$. In case $\bar{F}(0) = \delta$ the argument given there is, however, incorrect, because $F(0) \to \bar{F}(0) = \delta$ in probability does not imply $\Pr(F(0) > \delta) \to 0$ in general.

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 suitable choice of δ_i, see Cribari-Neto (2004). For the last three choices of d_i 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 $h_{ii} = 1$.

Similar as in Subsection 3.1 we need to ensure that $\hat{\Omega}_{Het}(y)$ is nonsingular $\lambda_{\mathbb{R}^n}$-everywhere. As shown in the subsequent lemma this is the case provided Assumption 3 introduced in Subsection 3.1 is satisfied. The lemma also shows that in case this assumption is violated the matrix $\hat{\Omega}_{Het}(y)$ is singular *everywhere*, leading to a complete and trivial breakdown of the test. Recall the definition of the matrix $B(y)$ given in (8) and note that it is independent of the constants d_i.

Lemma 4.1.   1. $\hat{\Omega}_{Het}(y)$ is nonnegative definite for every $y \in \mathbb{R}^n$.

2. $\hat{\Omega}_{Het}(y)$ is singular if and only if $\operatorname{rank}(B(y)) < q$.

3. $\hat{\Omega}_{Het}(y) = 0$ if and only if $B(y) = 0$.

4. The set of all $y \in \mathbb{R}^n$ for which $\hat{\Omega}_{Het}(y)$ is singular (or, equivalently, for which $\operatorname{rank}(B(y)) < q$) 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 3 is violated.

The proof of the preceding lemma is completely analogous to the proof of Lemma 3.1 and hence is omitted. We are now in the position to state the result on size and power of tests based on the statistic T_{Het} given in (10).

Theorem 4.2. *Suppose $\mathfrak{C} \supseteq \mathfrak{C}_{Het}$ holds and Assumption 3 is satisfied. Let T_{Het} be the test statistic defined in (10) and let $W_{Het}(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ be the rejection region where C is a real number satisfying $0 < C < \infty$. Then the following holds:*

1. *Suppose for some i, $1 \leq i \leq n$, we have $\operatorname{rank}(B(e_i(n))) = q$ and $T_{Het}(e_i(n) + \mu_0^*) > C$ for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W_{Het}(C)) = 1$$

holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

2. *Suppose for some i, $1 \leq i \leq n$, we have $\operatorname{rank}(B(e_i(n))) = q$ and $T_{Het}(e_i(n) + \mu_0^*) < C$ for some (and hence all) $\mu_0^* \in \mathfrak{M}_0$. Then*

$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W_{Het}(C)) = 0$$

holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$, and hence

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W_{Het}(C)) = 0$$

holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased. Furthermore, the nuisance-infimal rejection probability at every point $\mu_1 \in \mathfrak{M}_1$ is zero, i.e.,

$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W_{Het}(C)) = 0.$$

In particular, the infimal power of the test is equal to zero.

3. *Suppose for some i, $1 \leq i \leq n$, we have $B(e_i(n)) = 0$ and $R\hat{\beta}(e_i(n)) \neq 0$. Then*

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W_{Het}(C)) = 1$$

holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

We note that Remark 3.4(i)-(iii) as well as most of the discussion following Theorem 3.3 apply mutatis mutandis also here. Similar as in Subsection 3.1 it is also not difficult to show (for typical choices of d_i) that the set of design matrices X for which the conditions in Theorem 4.2 are not satisfied is a negligible set. We omit a formal statement. In contrast to the case considered in Subsection 3.1, however, no (nontrivial) analogues to the positive results given in Theorems 3.6 and 3.7 are possible due to the fact that in the present setting there are now too many concentration spaces (which together in fact span all of $\mathbb{R}^n$). Furthermore, the above theorem and its proof exploits only the one-dimensional concentration spaces $\mathcal{Z}_i = \text{span}(e_i(n))$. While every linear space of the form $\text{span}(e_{i_1}(n), \ldots, e_{i_p}(n))$ for $0 < p < n$ and $1 \leq i_1 < \ldots < i_p \leq n$ is a concentration space of the model $\mathfrak{C}$, using all these concentration spaces in conjunction with Corollary 5.17 does not deliver additional obstructions to good size or power properties, the reason being that each of these spaces already contains a concentration space $\mathcal{Z}_i$ as a subset. As a further point of interest we note that the assumptions imposed in Eicker (1963, 1967) require all variances $\sigma^2 \tau_i^2$ to be bounded away from zero in order to achieve uniformity in the convergence to the limiting distribution. Hence, Eicker's assumptions rule out the concentration effect that drives the above result. It appears that this insight in Eicker (1963, 1967) has not been fully appreciated in the ensuing econometrics literature. In connection with the preceding theorem, which points out size distortions and/or power deficiencies of heteroskedasticity robust tests even under a normality assumption, a result in Section 4.2 of Dufour (2003) needs to be mentioned which shows that the size of heteroskedasticity robust tests is always 1 if one allows for a sufficiently *large nonparametric* class of distributions for the errors **U**.

We shortly discuss the standard F-test statistic without any correction for heteroskedasticity. Let

$$T_{uncorr}(y) = \begin{cases} ((n-k)/q)(R\hat{\beta}(y) - r)' \left(R(X'X)^{-1}R'\right)^{-1} (R\hat{\beta}(y) - r)/(\hat{u}'(y)\hat{u}(y)) & \text{if } y \notin \mathfrak{M} \\ 0 & \text{if } y \in \mathfrak{M} \end{cases}$$

and define $W_{uncorr}(C)$ in the obvious way. It is then easy to see that a variant of Theorem 4.2 also holds with T_{uncorr} and $W_{uncorr}(C)$ replacing T_{Het} and $W_{Het}(C)$, respectively, if in this variant of the theorem Assumption 3 is dropped, the condition $\text{rank}(B(e_i(n))) = q$ is replaced by the condition $e_i(n) \notin \mathfrak{M}$, and the condition $B(e_i(n)) = 0$ is replaced by the condition $e_i(n) \in \mathfrak{M}$. In a recent paper Ibragimov and Müller (2010) consider the standard t-test for testing $\mu = 0$ versus $\mu \neq 0$ in a Gaussian location model and discuss a result by Bakirov and Székely (2005) to the effect that the size of this test under heteroskedasticity of unknown form equals the nominal significance level δ as long as $n \geq 2$ and $\delta \leq 0.08326$. It is not difficult to see that in this location problem $T_{uncorr}(e_i(n)) = 1$ holds for every i (note that $\mu_0^* = 0$) and thus the inequality $T_{uncorr}(e_i(n)) < C$ always holds whenever $C > 1$. Hence Case 1 of the variant of Theorem 4.2 just discussed does not arise whenever $C > 1$ which is in line with the results in Bakirov and Székely (2005). However, note that Case 2 of that theorem then always applies (since obviously $e_i(n) \notin \mathfrak{M} = \text{span}(e_+)$), showing that the standard t-test suffers from severe power deficiencies under heteroskedasticity of

unknown form in case $n \geq 2$ and $\delta \leq 0.08326$ (noting that the squared standard t-statistic is the standard F-statistic).

5 General Principles Underlying Size and Power Results for Tests of Linear Restrictions in Regression Models with Nonspherical Disturbances

The results on size and power properties given in the previous sections are obtained as special cases of a more general theory that applies to a large class of tests and to general covariance models $\mathfrak{C}$ (which thus are not restricted to covariance structures resulting from stationary disturbances or from heteroskedasticity). This theory is provided in the present section. We use the notation and assumptions of Section 2. Since invariance properties of tests will play an important rôle in some of the results to follow, the next subsection collects some relevant results related to invariance. In Subsection 5.2 we provide conditions under which the tests considered have highly unpleasant size or power properties. This result is based on a "concentration" effect. In contrast, Subsection 5.3 provides conditions under which tests do not suffer from the size and power problems just mentioned. Subsection 5.4 then specializes the results of the preceding subsections to a class of tests which can be described as nonsphericity-corrected F-type tests. This class of tests contains virtually all so-called heteroskedasticity and autocorrelation robust tests available in the literature as special cases. Furthermore, Subsection 5.4 also contains another negative result, the derivation of which exploits the particular structure of these tests.

5.1 Some preliminaries on groups and invariance

Let G be a group of Borel-measurable transformations of $\mathbb{R}^n$ into itself, the group operation being the composition of transformations. A function S defined on $\mathbb{R}^n$ is said to be invariant under the group G if $S(g(y)) = S(y)$ for all $y \in \mathbb{R}^n$ and all $g \in G$. A subset A of $\mathbb{R}^n$ is said to be invariant under G if $g(A) \subseteq A$ holds for every $g \in G$. Since with g also g^{-1} belongs to G, this is equivalent to $g(A) = A$ for every $g \in G$, and thus to invariance of the indicator function of A as defined before.[13] Clearly, invariance of $S : \mathbb{R}^n \to \overline{\mathbb{R}}$, the extended real line, under the group G implies invariance of the super-level sets $W = \{y : S(y) \geq C\}$. Furthermore, a function S defined on $\mathbb{R}^n$ is said to be almost invariant under the group G if $S(g(y)) = S(y)$ holds for all $g \in G$ and all $y \in \mathbb{R}^n \backslash N(g)$ with Borel-sets $N(g)$ satisfying $\lambda_{\mathbb{R}^n}(N(g)) = 0$ and also $\lambda_{\mathbb{R}^n}\left(g'^{-1}(N(g))\right) = 0$ for all $g' \in G$.[14] A subset A of $\mathbb{R}^n$ is said to be almost invariant if $g(A) \subseteq A \cup N(g)$ holds for every $g \in G$ with the Borel-sets $N(g)$ satisfying $\lambda_{\mathbb{R}^n}(N(g)) = 0$ and $\lambda_{\mathbb{R}^n}\left(g'^{-1}(N(g))\right) = 0$ for all $g' \in G$. It is easy to see that this is equivalent to $g(A) \triangle A \subseteq N^*(g)$ for every $g \in G$, with Borel-sets $N^*(g)$ satisfying $\lambda_{\mathbb{R}^n}(N^*(g)) = 0$ and $\lambda_{\mathbb{R}^n}\left(g'^{-1}(N^*(g))\right) = 0$ for all $g' \in G$; thus it is equivalent to almost invariance

---

[13]If G is only a collection of bijective transformations on $\mathbb{R}^n$ but is not a group, then invariance of A does not imply $g(A) = A$ in general, and in particular does not coincide with the notion of invariance of the indicator function of A.

[14]The additional requirement $\lambda_{\mathbb{R}^n}\left(g'^{-1}(N(g))\right) = 0$ for all $g' \in G$ of course implies $\lambda_{\mathbb{R}^n}(N(g)) = 0$ and may appear artificial at first sight. However, it arises naturally in the context of testing problems that are invariant under the group G and for which the relevant family of probability measures is equivalent to $\lambda_{\mathbb{R}^n}$, cf. Lehmann and Romano (2005), Section 6.5. Regardless of this, the additional requirement already follows from $\lambda_{\mathbb{R}^n}(N(g)) = 0$ in case the group G is a group of affine transformations on $\mathbb{R}^n$, which will be the groups we are interested in.

of the indicator function of A. Clearly, almost invariance of $S : \mathbb{R}^n \to \bar{\mathbb{R}}$ under the group G implies almost invariance of the super-level sets $W = \{y : S(y) \geq C\}$.

We are interested in some particular groups of affine transformations. For an affine subspace $\mathfrak{N}$ of $\mathbb{R}^n$ let

$$G(\mathfrak{N}) = \{g_{\alpha,\nu,\nu'} : \alpha \neq 0, \nu' \in \mathfrak{N}\}$$

for some fixed but arbitrary $\nu \in \mathfrak{N}$, where the affine map $g_{\alpha,\nu,\nu'}$ is given by $g_{\alpha,\nu,\nu'}(y) = \alpha(y - \nu) + \nu'$ with $\alpha \in \mathbb{R}$. Observe that $G(\mathfrak{N})$ does not depend on the choice of ν (in particular, if $\mathfrak{N}$ is a linear subspace, one may choose $\nu = 0$). Hence, $G(\mathfrak{N})$ can also be written in a redundant way as

$$G(\mathfrak{N}) = \{g_{\alpha,\nu,\nu'} : \alpha \neq 0, \nu \in \mathfrak{N}, \nu' \in \mathfrak{N}\}.$$

It is easy to see that $G(\mathfrak{N})$ is a group w.r.t. composition which is non-abelian except if $\mathfrak{N}$ is a singleton. For later use we also note that $\mathfrak{N}$ as well as $\mathbb{R}^n \backslash \mathfrak{N}$ are invariant under $G(\mathfrak{N})$, and that $G(\mathfrak{N})$ acts transitively on $\mathfrak{N}$ (but not on $\mathbb{R}^n \backslash \mathfrak{N}$ in general). Furthermore, note that the elements of $G(\mathfrak{N})$ can also be written as $g_{\alpha,\nu,\nu'}(y) = \alpha y + (1 - \alpha)\nu + (\nu' - \nu)$.

Remark 5.1. We make an observation on the structure of $G(\mathfrak{N})$. Let $G_1(\mathfrak{N})$ denote the collection of transformations $g_{\alpha,\nu,\nu}(y)$ for every $\alpha \neq 0$ and every $\nu \in \mathfrak{N}$, and let $G_2(\mathfrak{N})$ denote the collection of transformations $g_{1,\nu,\nu'}(y)$ for every pair $\nu, \nu' \in \mathfrak{N}$. Obviously, $G_1(\mathfrak{N})$ as well as $G_2(\mathfrak{N})$ are subsets of $G(\mathfrak{N})$, and every element of $G(\mathfrak{N})$ is the composition of an element in $G_2(\mathfrak{N})$ with an element of $G_1(\mathfrak{N})$. While $G_2(\mathfrak{N})$ is a subgroup, $G_1(\mathfrak{N})$ is not (as it is not closed under composition) except in the trivial case where $\mathfrak{N}$ is a singleton. However, the group generated by $G_1(\mathfrak{N})$ is precisely $G(\mathfrak{N})$. As a consequence, any function S which is invariant under the elements of $G_1(\mathfrak{N})$ (meaning that $S(g(y)) = S(y)$ for all $y \in \mathbb{R}^n$ and all $g \in G_1(\mathfrak{N})$) is already invariant under the entire group $G(\mathfrak{N})$, and a similar statement holds for almost invariance.

Proposition 5.2. *A maximal invariant for $G(\mathfrak{N})$ is given by*

$$h(y) = \left\langle \Pi_{(\mathfrak{N}-\nu_*)^\perp}(y - \nu_*) / \left\| \Pi_{(\mathfrak{N}-\nu_*)^\perp}(y - \nu_*) \right\| \right\rangle,$$

where ν_* is an arbitrary element of $\mathfrak{N}$. The maximal invariant h in fact does not depend on the choice of $\nu_* \in \mathfrak{N}$. [Here we use the convention $x/\|x\| = 0$ if $x = 0$.]

Remark 5.3. Specializing to the case $\mathfrak{N} = \mathfrak{M}_0$ it is obvious that $\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(y - \mu_0)$ can be computed as $y - X\hat{\beta}_{rest}(y)$, where $\hat{\beta}_{rest}$ denotes the restricted ordinary least squares estimator. It follows that any test that is invariant under $G(\mathfrak{M}_0)$ depends only on the normalized restricted least squares residuals, in fact only on $\left\langle y - X\hat{\beta}_{rest}(y) / \left\| y - X\hat{\beta}_{rest}(y) \right\| \right\rangle$. [For the tests considered in Subsection 5.4 one can obtain this result also directly from the definition of the tests.]

Consider now the problem of testing H_0 versus H_1 as defined in (4). First observe that the sets $\mathfrak{M}_0$ and $\mathfrak{M}_1$ are invariant under the transformations in $G(\mathfrak{M}_0)$. This implies that the parameter spaces $\overline{\mathfrak{M}_i} \times (0, \infty) \times \mathfrak{C}$ corresponding to H_i (for $i = 0, 1$) are each invariant under the associated group $\overline{G(\mathfrak{M}_0)}$, i.e., the group consisting of all transformations $\bar{g}_{\alpha,\mu_0,\mu_0'}$ defined on $\mathfrak{M} \times (0, \infty) \times \mathfrak{C}$ given by

$$\bar{g}_{\alpha,\mu_0,\mu_0'}(\mu, \sigma^2, \Sigma) = (\alpha(\mu - \mu_0) + \mu_0', \alpha^2\sigma^2, \Sigma)$$

where $\alpha \neq 0$, $\mu_0 \in \mathfrak{M}_0$, $\mu_0' \in \mathfrak{M}_0$. [Note that the associated group strictly speaking also depends on $\mathfrak{C}$, but we suppress this in the notation.] Second, the probability measures associated with H_0 and H_1 clearly satisfy

$$P_{\mu,\sigma^2\Sigma}(A) = P_{\alpha(\mu-\mu_0)+\mu_0',\alpha^2\sigma^2\Sigma}(\alpha(A-\mu_0)+\mu_0') \tag{11}$$

for every $(\mu, \sigma^2, \Sigma) \in \mathfrak{M} \times (0, \infty) \times \mathfrak{C}$ and every Borel set $A \subseteq \mathbb{R}^n$. This shows that the testing problem considered in (4) is invariant under the group $G(\mathfrak{M}_0)$ in the sense of Lehmann and Romano (2005), Chapters 6 and 8. While trivial, it will be useful to note that (11) continues to hold if $\Sigma \in \mathfrak{C}$ is replaced by an arbitrary nonnegative definite symmetric $n \times n$ matrix Φ. The next proposition discusses invariance properties of the rejection probabilities of an almost invariant test φ that will be needed in subsequent subsections. As will be seen later, it is useful to consider in that proposition the rejection probabilities $E_{\mu,\sigma^2\Phi}(\varphi)$ also for Φ a positive (or sometimes only nonnegative) definite symmetric $n \times n$ matrix not necessarily belonging to the assumed covariance model $\mathfrak{C}$.

Proposition 5.4. *Let $\varphi : \mathbb{R}^n \to [0,1]$ be a Borel-measurable function that is almost invariant under $G(\mathfrak{M}_0)$.*

1. *For every $(\mu, \sigma^2) \in \mathfrak{M} \times (0, \infty)$ and for every positive definite symmetric $n \times n$ matrix Φ the rejection probabilities satisfy*

$$E_{\mu,\sigma^2\Phi}(\varphi) = E_{\alpha(\mu-\mu_0)+\mu_0',\alpha^2\sigma^2\Phi}(\varphi) \tag{12}$$

for all $\alpha \neq 0$, $\mu_0 \in \mathfrak{M}_0$, $\mu_0' \in \mathfrak{M}_0$. [If φ is invariant under $G(\mathfrak{M}_0)$, then (12) holds even if Φ is only nonnegative definite and symmetric.]

2. *For every $(\mu, \sigma^2) \in \mathfrak{M} \times (0, \infty)$ and every positive definite symmetric $n \times n$ matrix Φ we have the representation*

$$E_{\mu,\sigma^2\Phi}(\varphi) = E_{\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)/\sigma+\mu_0,\Phi}(\varphi) = E_{\left\langle \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)/\sigma \right\rangle+\mu_0,\Phi}(\varphi) \tag{13}$$

where μ_0 is an arbitrary element of $\mathfrak{M}_0$. [Note that $\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)/\sigma$ actually does not depend on the choice of μ_0, and $\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)$ can be computed as $\mu - X\hat{\beta}_{rest}(\mu)$.]

3. *The rejection probability $E_{\mu,\sigma^2\Phi}(\varphi)$ depends on $(\mu, \sigma^2) \in \mathfrak{M} \times (0, \infty)$ and Φ only through $\left(\left\langle \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)/\sigma \right\rangle, \Phi \right)$. Furthermore, $\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu-\mu_0)/\sigma$ is in a bijective correspondence with $(R\beta - r)/\sigma$ where β denotes the coordinates of μ in the basis given by the columns of X. Thus the rejection probability $E_{\mu,\sigma^2\Phi}(\varphi)$ depends on $(\mu, \sigma^2) \in \mathfrak{M} \times (0, \infty)$ and Φ only through $(\langle (R\beta - r)/\sigma \rangle, \Phi)$.*

Remark 5.5. (i) For $\Phi = \Sigma \in \mathfrak{C}$ relation (12) expresses the fact that the rejection probability of the almost invariant test φ is invariant under the associated group $\overline{G(\mathfrak{M}_0)}$.

(ii) Setting $\alpha = 1$ in (12) and holding σ^2 and Φ fixed, we see that the rejection probability is, in particular, constant along that translation of $\mathfrak{M}_0$ which passes through μ.

(iii) If $\mu \in \mathfrak{M}_0$, choosing $\mu_0 = \mu$, $\alpha = \sigma^{-1}$, and fixing $\mu_0' \in \mathfrak{M}_0$, shows that $E_{\mu,\sigma^2\Phi}(\varphi) = E_{\mu_0',\Phi}(\varphi)$. Hence, for $\mu \in \mathfrak{M}_0$, the rejection probability is constant in (μ, σ^2) and only depends on Φ.

(iv) Occasionally we consider tests φ that are only required to be almost invariant under the subgroup of transformations $y \mapsto \alpha y + (1 - \alpha) \mu_0$ for a fixed $\mu_0 \in \mathfrak{M}_0$, i.e., under the group $G(\{\mu_0\})$. The results in the above propositions can be easily adapted to this case and we refrain from spelling out the details. We only note that the analogue to (12) in this case is given by

$$E_{\mu,\sigma^2\Phi}(\varphi) = E_{\alpha(\mu-\mu_0)+\mu_0,\alpha^2\sigma^2\Phi}(\varphi) \tag{14}$$

for all $\alpha \neq 0$.

Part 2 of the above proposition has shown that the rejection probability depends on the parameters only through $\left(\left\langle \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu - \mu_0)/\sigma \right\rangle, \Sigma\right)$. This quantity is recognized as a maximal invariant in the next result.

Proposition 5.6. *Let $\mu_0 \in \mathfrak{M}_0$ be arbitrary. Then $\left(\left\langle \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu - \mu_0)/\sigma \right\rangle, \Sigma\right)$ is a maximal invariant for the associated group $\overline{G(\mathfrak{M}_0)}$.*

5.2 Negative results

We next establish a negative result providing conditions under which (i) the size of a test is 1, and/or (ii) the power function of a test gets arbitrarily close to zero. The theorem is based on a "concentration effect" that we explain now: Suppose one can find a sequence $\Sigma_m \in \mathfrak{C}$ converging to a singular matrix $\bar{\Sigma}$ and let $\mathcal{Z}$ denote the span of the columns of $\bar{\Sigma}$. Let $\mu_0 \in \mathfrak{M}_0$. Since the probability measures $P_{\mu_0,\sigma^2\Sigma_m}$ converge weakly to $P_{\mu_0,\sigma^2\bar{\Sigma}}$, which has support $\mu_0 + \mathcal{Z}$, they concentrate their mass more and more around $\mu_0 + \mathcal{Z}$. Suppose first that one can show that $\mu_0 + \mathcal{Z}$ is essentially contained in the interior of the rejection region W in the sense that the set of points in $\mu_0 + \mathcal{Z}$ which are not interior points of W has $\lambda_{\mu_0+\mathcal{Z}}$-measure zero. It then follows that $P_{\mu_0,\sigma^2\Sigma_m}(W)$ converges to $P_{\mu_0,\sigma^2\bar{\Sigma}}(W) \geq P_{\mu_0,\sigma^2\bar{\Sigma}}(\mu_0 + \mathcal{Z}) = 1$, establishing that the size of the test is 1. Now, in some cases of interest it turns out that $\mu_0 + \mathcal{Z}$ fails to satisfy the just mentioned "interiority" condition with respect to the rejection region W, but it also turns out that it does satisfy the "interiority" condition with respect to an "equivalent" rejection region W', which is obtained by adjoining a $\lambda_{\mathbb{R}^n}$-null set to W (for example, for $W' = W \cup (\mu_0 + \mathcal{Z})$). Since the rejection probabilities corresponding to W and W' are identical (as any $\Sigma \in \mathfrak{C}$ is positive definite) and thus the two tests have the same size, the above reasoning can then be applied to W', again showing that the size of the test based on W is 1 for these cases. Part 1 of Theorem 5.7 below formalizes this reasoning. The same "concentration effect" reasoning applied to $\mathbb{R}^n \backslash W$ instead of W then gives (17). [The remaining claims in Part 2 as well as Part 3 are then consequences of (17) combined with continuity or invariance properties of the power function.] It should, however, be stressed that weak convergence of $P_{\mu_0,\sigma^2\Sigma_m}$ to $P_{\mu_0,\sigma^2\bar{\Sigma}}$ together with the inclusion $\mu_0 + \mathcal{Z} \subseteq W$ (except possibly for a $\lambda_{\mu_0+\mathcal{Z}}$-null set) alone is *not* sufficient to allow one to draw the conclusion – as tempting as it may be – that $P_{\mu_0,\sigma^2\Sigma_m}(W) \to 1$ although "in the limit" $P_{\mu_0,\sigma^2\bar{\Sigma}}(W) = 1$ holds. Counterexamples where $P_{\mu_0,\sigma^2\Sigma_m}$ converges weakly to $P_{\mu_0,\sigma^2\bar{\Sigma}}$ and $\mu_0 + \mathcal{Z} \subseteq W$ (and thus $P_{\mu_0,\sigma^2\bar{\Sigma}}(W) = 1$) holds, but where $P_{\mu_0,\sigma^2\Sigma_m}(W)$ converges to a positive number less than 1 are easily found with the help of Theorem 5.10. We furthermore note that in a different testing context Martellosio (2010) provides a result which also makes use of a "concentration effect", but his result is not correct as given. For a discussion of these issues and corrected results see Preinerstorfer and Pötscher (2013).

34

The "concentration effect" reasoning underlying Theorem 5.7 of course hinges crucially on the "interiority" condition (either w.r.t. W or w.r.t. $\mathbb{R}^n \backslash W$), raising the question why we should expect this to be satisfied in the applications we have in mind, rather than expect that $\mu_0 + \mathcal{Z}$ intersects with both W and $\mathbb{R}^n \backslash W$ is such a way that the "interiority" condition is neither satisfied w.r.t. W nor w.r.t. $\mathbb{R}^n \backslash W$. Consider the case where $\mathcal{Z}$ is one-dimensional, a case of paramount importance in the applications, and suppose also that W is invariant under the group $G(\mathfrak{M}_0)$. Then we have the dichotomy that $(\mu_0 + \mathcal{Z}) \backslash \{\mu_0\}$ either lies entirely in W or in $\mathbb{R}^n \backslash W$, showing that – except possibly for the point μ_0 – the set $\mu_0 + \mathcal{Z}$ never intersects both W and $\mathbb{R}^n \backslash W$. Moreover, if an element of $(\mu_0 + \mathcal{Z}) \backslash \{\mu_0\}$ belongs to the interior of W (of $\mathbb{R}^n \backslash W$, respectively), then $(\mu_0 + \mathcal{Z}) \backslash \{\mu_0\}$ in its entirety is a subset of the interior of W (of $\mathbb{R}^n \backslash W$, respectively). Hence, under the mentioned invariance and for one-dimensional $\mathcal{Z}$, one can expect the "interiority" conditions in the subsequent theorem to be satisfied not infrequently.

Theorem 5.7. *Let W be a Borel set in $\mathbb{R}^n$, the rejection region of a test. Furthermore, assume that $\mathcal{Z}$ is a concentration space of the covariance model $\mathfrak{C}$. Then the following holds:*

1. *If $\mu_0 \in \mathfrak{M}_0$ satisfies*
$$\lambda_{\mu_0 + \mathcal{Z}} \left(\mathrm{bd} \left(W \cup (\mu_0 + \mathcal{Z}) \right) \right) = 0, \tag{15}$$
   then for every $0 < \sigma^2 < \infty$
$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W) = 1$$
   holds; in particular, the size of the test equals 1. [In case W is of the form $\{y \in \mathbb{R}^n : T(y) \geq C\}$ for some Borel-measurable function $T : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and $0 < C < \infty$, a sufficient condition for (15) is that for $\lambda_{\mathcal{Z}}$-almost every $z \in \mathcal{Z}$ the test statistic T satisfies $T(\mu_0 + z) > C$ and is lower semicontinuous at $\mu_0 + z$.]

2. *If $\mu_0 \in \mathfrak{M}_0$ satisfies*
$$\lambda_{\mu_0 + \mathcal{Z}} \left(\mathrm{bd} \left((\mathbb{R}^n \backslash W) \cup (\mu_0 + \mathcal{Z}) \right) \right) = 0, \tag{16}$$
   then for every $0 < \sigma^2 < \infty$
$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W) = 0, \tag{17}$$
   and hence
$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W) = 0,$$
   holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased (except in the trivial case where its size is zero). [In case W is of the form $\{y \in \mathbb{R}^n : T(y) \geq C\}$ for some Borel-measurable function $T : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and $0 < C < \infty$, a sufficient condition for (16) is that for $\lambda_{\mathcal{Z}}$-almost every $z \in \mathcal{Z}$ the test statistic T satisfies $T(\mu_0 + z) < C$ and is upper semicontinuous at $\mu_0 + z$.]

3. *Suppose that condition (17) is satisfied for some $\mu_0 \in \mathfrak{M}_0$ and some $0 < \sigma^2 < \infty$. Furthermore, assume that W is almost invariant under the group $G(\{\mu_0\})$. Then for every $\mu_1 \in \mathfrak{M}_1$ we have*
$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W) = 0.$$

   [In case W is of the form $\{y \in \mathbb{R}^n : T(y) \geq C\}$ for some Borel-measurable function $T : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and $0 < C < \infty$, almost invariance of W under the group $G(\{\mu_0\})$ follows from almost invariance of T under $G(\{\mu_0\})$.]

Remark 5.8. (i) The conclusions of the above theorem immediately also apply to every test statistic T' that is $\lambda_{\mathbb{R}^n}$-almost surely equal to a test statistic T satisfying the assumptions of the theorem.

(ii) Let $\varphi : \mathbb{R}^n \mapsto [0,1]$ be Borel-measurable, i.e., a test. If the set $\{y : \varphi(y) = 1\}$ satisfies the assumptions on W in Part 1 of the above theorem, then for every $0 < \sigma^2 < \infty$

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \sigma^2 \Sigma} \varphi = 1$$

holds. If the set $\{y : \varphi(y) = 0\}$ satisfies the assumptions on $\mathbb{R}^n \backslash W$ in Part 2 of the above theorem then for every $0 < \sigma^2 < \infty$

$$\inf_{\Sigma \in \mathfrak{C}} E_{\mu_0, \sigma^2 \Sigma} \varphi = 0$$

holds. A similar remark applies to Part 3 of the theorem, provided φ is almost invariant under $G(\{\mu_0\})$.

Remark 5.9. If the covariance model $\mathfrak{C}$ contains AR(1) correlation matrices $\Lambda(\rho_m)$ for some sequence $\rho_m \in (-1,1)$ with $\rho_m \to 1$ ($\rho_m \to -1$, respectively), then $\mathrm{span}\,(e_+)$ ($\mathrm{span}\,(e_-)$, respectively) is a concentration space of $\mathfrak{C}$ (cf. Lemma G.1 in Appendix G). Hence Theorem 5.7 applies with $\mathcal{Z} = \mathrm{span}\,(e_+)$ ($\mathcal{Z} = \mathrm{span}\,(e_-)$, respectively). In particular, if $\mathfrak{C}$ contains $\mathfrak{C}_{AR(1)}$, then Theorem 5.7 applies with $\mathcal{Z} = \mathrm{span}\,(e_+)$ as well as with $\mathcal{Z} = \mathrm{span}\,(e_-)$.

5.3 Positive results

The next theorem isolates conditions under which a test does not suffer from the extreme size and power problems encountered in the preceding subsection. In particular, we provide conditions which guarantee that the size is bounded away from one and that the power function is bounded away from zero. The theorem assumes that the test φ – apart from being (almost) invariant under the group $G(\mathfrak{M}_0)$ – is also invariant under addition of elements of $J(\mathfrak{C})$ defined below. This additional invariance assumption will be automatically satisfied in the important special case where φ is invariant under the group $G(\mathfrak{M}_0)$ and where $J(\mathfrak{C}) \subseteq \mathfrak{M}_0 - \mu_0$ for some $\mu_0 \in \mathfrak{M}_0$ (and hence for all $\mu_0 \in \mathfrak{M}_0$) as then the maps $x \mapsto x + z$ for $z \in J(\mathfrak{C})$ are elements of $G(\mathfrak{M}_0)$; see also Proposition 5.23 and the attending discussion in Subsection 5.4. A second assumption of the subsequent theorem is that the covariance model $\mathfrak{C}$ is bounded which is typically a harmless assumption in applications as it is, e.g., always satisfied if the elements of $\mathfrak{C}$ are normalized such that the largest diagonal element is 1, or such that the trace is 1. The theorem also maintains a further assumption on the covariance model $\mathfrak{C}$ related to the way sequences of elements in $\mathfrak{C}$ approach singular matrices. This condition has to be verified for the covariance model $\mathfrak{C}$ in any particular application. A verification for $\mathfrak{C}_{AR(1)}$ is given in Appendix G, cf. also Remarks 5.14 and 5.20.

For a covariance model $\mathfrak{C}$ define now

$$J(\mathfrak{C}) = \bigcup \left\{ \mathrm{span}(\bar{\Sigma}) : \det \bar{\Sigma} = 0,\ \bar{\Sigma} = \lim_{m \to \infty} \Sigma_m \text{ for a sequence } \Sigma_m \in \mathfrak{C} \right\},$$

i.e., $J(\mathfrak{C})$ is the union of all concentration spaces of the covariance model $\mathfrak{C}$. [Note that the subsequent results remain valid in the case where $J(\mathfrak{C})$ is empty.]

Theorem 5.10. *Let $\varphi : \mathbb{R}^n \to [0,1]$ be a Borel-measurable function that is almost invariant under $G(\mathfrak{M}_0)$. Suppose that φ is neither $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 1 nor $\lambda_{\mathbb{R}^n}$-almost everywhere*

equal to 0. Suppose further that

$$\varphi(x+z) = \varphi(x) \qquad \text{for every } x \in \mathbb{R}^n \text{ and every } z \in J(\mathfrak{C}). \tag{18}$$

Assume that $\mathfrak{C}$ is bounded (as a subset of $\mathbb{R}^{n \times n}$). Assume also that for every sequence $\Sigma_m \in \mathfrak{C}$ converging to a singular $\bar{\Sigma}$ there exists a subsequence $(m_i)_{i \in \mathbb{N}}$ and a sequence of positive real numbers s_{m_i} such that the sequence of matrices $D_{m_i} = \Pi_{\mathrm{span}(\bar{\Sigma})^{\perp}} \Sigma_{m_i} \Pi_{\mathrm{span}(\bar{\Sigma})^{\perp}} / s_{m_i}$ converges to a matrix D which is regular on the orthogonal complement of $\mathrm{span}(\bar{\Sigma})$ (meaning that the linear map corresponding to D is injective when restricted to the orthogonal complement of $\mathrm{span}(\bar{\Sigma})$)[15]. Then the following holds:

1. *The size of the test φ is strictly less than 1, i.e.,*

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \sigma^2 \Sigma}(\varphi) < 1.$$

   Furthermore,

$$\inf_{\mu_0 \in \mathfrak{M}_0} \inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} E_{\mu_0, \sigma^2 \Sigma}(\varphi) > 0.$$

2. *Suppose additionally that for every sequence $\nu_m \in \Pi_{(\mathfrak{M}_0 - \mu_0)^{\perp}}(\mathfrak{M}_1 - \mu_0)$ with $\|\nu_m\| \to \infty$ and for every sequence Φ_m of positive definite symmetric $n \times n$ matrices with $\Phi_m \to \Phi$, Φ positive definite, we have*

$$\liminf_{m \to \infty} E_{\nu_m + \mu_0, \Phi_m}(\varphi) > 0, \tag{19}$$

   where μ_0 is an element of $\mathfrak{M}_0$. [This condition clearly does not depend on the particular choice of $\mu_0 \in \mathfrak{M}_0$.]. Then the infimal power is bounded away from zero, i.e.,

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} E_{\mu_1, \sigma^2 \Sigma}(\varphi) > 0.$$

3. *Suppose that the limit inferior in (19) is 1 for every sequence ν_m and Φ_m as specified above. Then for every $0 < c < \infty$*

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, 0 < \sigma^2 < \infty \\ d(\mu_1, \mathfrak{M}_0)/\sigma \geq c}} E_{\mu_1, \sigma^2 \Sigma_m}(\varphi) \to 1 \tag{20}$$

   holds for $m \to \infty$ and for any sequence $\Sigma_m \in \mathfrak{C}$ satisfying $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a singular matrix. Furthermore, for every sequence $0 < c_m < \infty$

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, \\ d(\mu_1, \mathfrak{M}_0) \geq c_m}} E_{\mu_1, \sigma_m^2 \Sigma_m}(\varphi) \to 1 \tag{21}$$

   holds for $m \to \infty$ whenever $0 < \sigma_m^2 < \infty$, $c_m/\sigma_m \to \infty$, and the sequence $\Sigma_m \in \mathfrak{C}$ satisfies $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a positive definite matrix. [The very last statement even holds without recourse to condition (18) and the condition on $\mathfrak{C}$ following (18).]

---

[15] Of course, D maps every element of $\mathrm{span}(\bar{\Sigma})$ into zero by construction.

The first two parts of the preceding theorem provide conditions under which the size is strictly less than 1 and the infimal power is strictly positive, while the third part provides conditions under which the power approaches 1 in certain parts of the parameter space, the parts being characterized by the property that either $\left\| \left(R\beta^{(1)} - r \right)/\sigma \right\|$ is bounded away from zero and Σ_m approaches a singular matrix, or that $\left\| \left(R\beta^{(1)} - r \right)/\sigma \right\| \to \infty$ and Σ_m approaches a positive definite matrix. Here $\beta^{(1)}$ is the parameter vector corresponding to μ_1. Note that $d\left(\mu_1, \mathfrak{M}_0\right)$ is bounded from above as well as from below by multiples of $\left\| R\beta^{(1)} - r \right\|$, where the constants involved are positive and depend only on X, R, and r.

Remark 5.11. (i) Because $J(\mathfrak{C})$ as a union of linear spaces is homogenous, condition (18) is equivalent to the condition that $\varphi(x + z) = \varphi(x)$ holds for every $x \in \mathbb{R}^n$ and every $z \in \mathrm{span}\,(J(\mathfrak{C}))$.
(ii) If condition (19) in Theorem 5.10 is replaced by the weaker condition

$$\liminf_{m \to \infty} E_{d_m(\mu_1 - \mu_0) + \mu_0, \Phi_m}(\varphi) > 0, \tag{22}$$

for every $\mu_1 \in \mathfrak{M}_1$, for every $d_m \to \infty$ and every sequence Φ_m of positive definite symmetric $n \times n$ matrices with $\Phi_m \to \Phi$, Φ a positive definite matrix, then we can only establish for every $\mu_1 \in \mathfrak{M}_1$ that

$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} E_{\mu_1, \sigma^2 \Sigma}(\varphi) > 0.$$

If the limes inferior in (22) is 1 for every μ_1, d_m, and Φ_m as specified above, then for every $\mu_1 \in \mathfrak{M}_1$ and every $0 < \sigma_*^2 < \infty$ we have

$$\inf_{0 < \sigma^2 \leq \sigma_*^2} E_{\mu_1, \sigma^2 \Sigma_m}(\varphi) \to 1$$

for any sequence $\Sigma_m \in \mathfrak{C}$ satisfying $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a singular matrix; and also $E_{\mu_1, \sigma_m^2 \Sigma_m}(\varphi) \to 1$ holds whenever $\sigma_m^2 \to 0$ and the sequence $\Sigma_m \in \mathfrak{C}$ satisfies $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a positive definite matrix. [The very last statement even holds without recourse to condition (18) and the condition on $\mathfrak{C}$ following (18).]

The subsequent theorem elaborates on Part 1 of Theorem 5.10 and shows that under the additional assumptions one can not only guarantee that the size of the test is smaller than 1, but one can, for any prescribed significance level δ $(0 < \delta < 1)$, construct the test in such a way that it has size not exceeding δ. The result applies in particular to the important case where the tests are of the form $\varphi_C = \mathbf{1}\,(T \geq C)$ for some test statistic T. Note that for any $C_k \uparrow \infty$ the sequence of tests φ_{C_k} clearly satisfies condition (23) in the subsequent theorem provided $\{y : T(y) = \infty\}$ is a $\lambda_{\mathbb{R}^n}$-null set. Thus in this case the theorem shows that for any given significance level δ, $0 < \delta < 1$, we can find a critical value $C(\delta)$ such that the test $\varphi_{C(\delta)}$ has a size not exceeding δ.

Theorem 5.12. *Let $\varphi_k : \mathbb{R}^n \to [0,1]$ for $k \geq 1$ be a sequence of Borel-measurable functions each of which satisfies the assumptions for Part 1 of Theorem 5.10, and let $\mathfrak{C}$ also satisfy the assumptions of that theorem. Furthermore assume that the sequence φ_k satisfies*

$$E_{\mu_0^*, \Phi}(\varphi_k) \downarrow 0 \tag{23}$$

as $k \uparrow \infty$ for some $\mu_0^* \in \mathfrak{M}_0$ and all positive definite symmetric $n \times n$ matrices Φ. Then for every δ, $0 < \delta < 1$, there exists a $k_0 = k_0(\delta)$ such that

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \sigma^2 \Sigma}(\varphi_{k_0}) \leq \delta.$$

Remark 5.13. (i) The assumption in Theorem 5.10 that φ_k is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 0 is of course irrelevant for the result in Theorem 5.12.

(ii) Of course, the second part of Part 1 of Theorem 5.10 immediately applies to φ_{k_0}; and Parts 2 and 3 of that theorem also apply to φ_{k_0} provided φ_{k_0} satisfies the respective additional conditions.

Remark 5.14. (i) In case the covariance model $\mathfrak{C}$ equals $\mathfrak{C}_{AR(1)}$, the boundedness condition in Theorems 5.10 and 5.12 is clearly satisfied and $J(\mathfrak{C})$ reduces to $\mathrm{span}\,(e_+) \cup \mathrm{span}\,(e_-)$. Furthermore, the condition on the covariance model $\mathfrak{C}$ in those theorems expressed in terms of the matrices D_m is then also satisfied as shown in Lemma G.1 in Appendix G. Also note that in this case the sequences Σ_m in Part 3 of Theorem 5.10 converging to a singular matrix are of the form $\Lambda\,(\rho_m)$ with $\rho_m \to 1$ or $\rho_m \to -1$.

(ii) More generally suppose that $\mathfrak{C}$ is norm-bounded, has $e_+e'_+$ and $e_-e'_-$ as the only singular accumulation points, and has the property that for every sequence $\Sigma_m \in \mathfrak{C}$ converging to one of these limit points there exists a sequence $(\rho_m)_{m \in \mathbb{N}}$ in $(-1, 1)$ such that $\Lambda(\rho_m)^{-1/2}\Sigma_m\Lambda(\rho_m)^{-1/2} \to I_n$ for $m \to \infty$ (that is, near the "singular boundary" the covariance model $\mathfrak{C}$ behaves similar to $\mathfrak{C}_{AR(1)}$). Then $J(\mathfrak{C})$ is as in (i) and again the conditions on the covariance model $\mathfrak{C}$ in Theorems 5.10 and 5.12 are satisfied.

5.4 Size and power properties of a common class of tests: Nonsphericity-corrected F-type tests

In this subsection we specialize the preceding results to a broad class of tests of linear restrictions in linear regression models with nonspherical errors and derive a further result specific to this class. The class considered in this subsection contains the vast majority of tests proposed in the literature for this testing problem. We start with a pair of estimators $\check{\beta}$ and $\check{\Omega}$, where $\check{\Omega}$ typically has the interpretation of an estimator of the variance covariance matrix of $R\check{\beta} - r$ under the null hypothesis. Similar as in previous sections, the estimators are viewed as functions of $y \in \mathbb{R}^n$, but it proves useful to allow for cases where the estimators are not defined for some exceptional values of y. We impose the following assumption on the estimators.

Assumption 5. *(i) The estimators $\check{\beta} : \mathbb{R}^n \backslash N \to \mathbb{R}^k$ and $\check{\Omega} : \mathbb{R}^n \backslash N \to \mathbb{R}^{q \times q}$ are well-defined and continuous on the complement of a closed $\lambda_{\mathbb{R}^n}$-null set N in the sample space $\mathbb{R}^n$, with $\check{\Omega}$ also being symmetric on $\mathbb{R}^n \backslash N$.*

(ii) The set $\mathbb{R}^n \backslash N$ is invariant under the group $G(\mathfrak{M})$, i.e., $y \in \mathbb{R}^n \backslash N$ implies $\alpha y + X\gamma \in \mathbb{R}^n \backslash N$ for every $\alpha \neq 0$ and every $\gamma \in \mathbb{R}^k$.

(iii) The estimators satisfy the equivariance properties $\check{\beta}(\alpha y + X\gamma) = \alpha\check{\beta}(y) + \gamma$ and $\check{\Omega}(\alpha y + X\gamma) = \alpha^2\check{\Omega}(y)$ for every $y \in \mathbb{R}^n \backslash N$, for every $\alpha \neq 0$, and for every $\gamma \in \mathbb{R}^k$.

(ii) $\check{\Omega}$ is $\lambda_{\mathbb{R}^n}$-almost surely nonsingular on $\mathbb{R}^n \backslash N$.

We make a few obvious observations: First, the invariance of $\mathbb{R}^n \backslash N$ under the group $G(\mathfrak{M})$ expressed in Assumption 5 is equivalent to the same invariance property of N itself. Second, since N is closed by Assumption 5, it follows that either N is empty or otherwise must at least contain $\mathfrak{M}$ (to see this note that $y \in N$ implies $\alpha y \in N$ for α arbitrarily close to zero which in turn implies $0 \in N$ by closedness of N). Third, given Assumption 5 holds, the sets $\left\{y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) = 0\right\}$ and $\left\{y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) \neq 0\right\}$ are invariant under the transformations in $G(\mathfrak{M})$, and the set

$$N^* = N \cup \left\{y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) = 0\right\} \tag{24}$$

is a closed $\lambda_{\mathbb{R}^n}$-null set that is also invariant under the transformations in $G(\mathfrak{M})$; cf. Lemma F.1 in Appendix F. Hence, the set $\left\{ y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) = 0 \right\}$ could in principle have been absorbed into N in the above assumption; however, we shall not do so since keeping the exceptional set N as small as possible will lead to stronger results. Furthermore, $\mathfrak{M} \subseteq N^*$ always holds. To see this note that $\mathfrak{M} \subseteq N \subseteq N^*$ holds if N is not empty as noted above; in case N is empty, $\check{\Omega}(y)$ is well-defined for every y and $\check{\Omega}(0) = \check{\Omega}(\alpha 0) = \alpha^2 \check{\Omega}(0)$ must hold, implying $\check{\Omega}(0) = 0$ and thus also $\check{\Omega}(X\gamma) = \check{\Omega}(\alpha 0 + X\gamma) = \alpha^2 \check{\Omega}(0) = 0$. In particular, this shows that either $\check{\Omega}$ is not defined on $\mathfrak{M}$ or is zero on $\mathfrak{M}$.

Given estimators $\check{\beta}$ and $\check{\Omega}$ satisfying Assumption 5 we define the test statistic

$$
T(y) = \begin{cases} (R\check{\beta}(y) - r)'\check{\Omega}^{-1}(y)(R\check{\beta}(y) - r), & y \in \mathbb{R}^n \backslash N^*, \\ 0, & y \in N^*. \end{cases} \tag{25}
$$

We note that assigning the test statistic the value zero at points $y \in \mathbb{R}^n$ for which either $y \in N$ or $\det(\check{\Omega})(y) = 0$ holds is arbitrary, but has no effect on the rejection probabilities of the test, since N^* is a $\lambda_{\mathbb{R}^n}$-null set as noted above and since all relevant probability measures $P_{\mu, \sigma^2 \Sigma}$ are absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^n$.

In line with the interpretation of $\check{\Omega}$ as an estimator for a variance covariance matrix, the leading case is when $\check{\Omega}$ is positive definite almost everywhere. However, sometimes we encounter situations where this is not guaranteed for a given fixed sample size (cf. Subsection 3.2), although typically the probability of being positive definite will go to one for each fixed value of the parameters as sample size increases. In order to be able to accommodate also such cases, Assumption 5 does not contain a requirement that $\check{\Omega}$ is positive definite almost everywhere. Nevertheless, in light of what has just been said, we shall consider the rejection region to be of the form $\{y \in \mathbb{R}^n : T(y) \geq C\}$ for C a real number satisfying $0 < C < \infty$.

For some of the results that follow we shall need further conditions on $\check{\Omega}$ which, however, are much weaker than the almost everywhere positive definiteness requirement just mentioned.

Assumption 6. *There exists $v \in \mathbb{R}^q$, $v \neq 0$, and a $y \in \mathbb{R}^n \backslash N^*$ such that $v'\check{\Omega}^{-1}(y)v > 0$ holds.*

Since under Assumption 5 the matrix $\check{\Omega}^{-1}(y)$ is continuous on $\mathbb{R}^n \backslash N^*$, it follows that Assumption 6 in fact implies that $v'\check{\Omega}^{-1}(y)v > 0$ holds on an open set of y's. The condition expressed in the next assumption is also certainly satisfied if $\check{\Omega}$ is positive definite almost everywhere. At first glance it may seem that this condition rules out the case where $\check{\Omega}(y)$ is allowed to be indefinite on a set of positive Lebesgue measure, but this is not so as v is not allowed to depend on y in this condition.

Assumption 7. *For every $v \in \mathbb{R}^q$ with $v \neq 0$ we have $\lambda_{\mathbb{R}^n} \left(\left\{ y \in \mathbb{R}^n \backslash N^* : v'\check{\Omega}^{-1}(y)v = 0 \right\} \right) = 0$.*

The following lemma collects some properties of the test statistic that will be useful in the sequel.

Lemma 5.15. *Suppose Assumption 5 is satisfied and let T be the test statistic defined in (25). Then the following holds:*

1. *The set $\mathbb{R}^n \backslash N^*$ is invariant under the elements of $G(\mathfrak{M})$.*

2. *The test statistic T is continuous on $\mathbb{R}^n \backslash N^*$; in particular, T is $\lambda_{\mathbb{R}^n}$-almost surely continuous on $\mathbb{R}^n$.*

3. The test statistic T is invariant under the group $G(\mathfrak{M}_0)$. Consequently, the rejection region $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ and its complement are invariant under $G(\mathfrak{M}_0)$.

4. The set $\{y \in \mathbb{R}^n : T(y) = C\}$ is a $\lambda_{\mathbb{R}^n}$-null set for every $0 < C < \infty$.

5. Suppose $0 < C < \infty$ holds. Then $\{y \in \mathbb{R}^n \backslash N^* : T(y) > C\}\, (= \{y \in \mathbb{R}^n : T(y) > C\})$ is an open set in $\mathbb{R}^n$, which is guaranteed to be non-empty under Assumption 6. Consequently, under Assumption 6 the rejection region $W(C)$ contains a non-empty open set and thus satisfies $\lambda_{\mathbb{R}^n}(W(C)) > 0$.

6. Suppose $0 < C < \infty$ holds. Then $\{y \in \mathbb{R}^n \backslash N^* : T(y) < C\}$ is a non-empty open set in $\mathbb{R}^n$. Consequently, the complement of the rejection region $W(C)$ contains a non-empty open set and thus satisfies $\lambda_{\mathbb{R}^n}(\mathbb{R}^n \backslash W(C)) > 0$.

7. Suppose Assumption 7 and $0 < C < \infty$ hold. Then, for every $\mu_0 \in \mathfrak{M}_0$, every sequence $\nu_m \in \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mathfrak{M}_1 - \mu_0)$ with $\|\nu_m\| \to \infty$, and for every sequence Φ_m of positive definite symmetric $n \times n$ matrices with $\Phi_m \to \Phi$, Φ a positive definite matrix, we have that

$$
\begin{aligned}
\liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)) &= \inf_{v \in A((\nu_m)_{m \geq 1})} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right) \\
&= \inf_{v \in A((\nu_m)_{m \geq 1})} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v > 0\right) \quad (26)
\end{aligned}
$$

where $A((\nu_m)_{m \geq 1})$ is the set of all accumulation points of the sequence

$$
R(X'X)^{-1} X'\nu_m / \left\| R(X'X)^{-1} X'\nu_m \right\|.
$$

A lower bound that does not depend on the sequence ν_m is as follows:

$$
\begin{aligned}
\liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)) &\geq \inf_{v \in \mathbb{R}^q, \|v\|=1} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right) \\
&= \inf_{v \in \mathbb{R}^q, \|v\|=1} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v > 0\right) \\
&\geq \Pr\left(\check{\Omega}(\Phi^{1/2}\mathbf{G}) \text{ is nonnegative definite}\right). \quad (27)
\end{aligned}
$$

In particular, if $\check{\Omega}$ is nonnegative definite $\lambda_{\mathbb{R}^n}$-almost everywhere (implying that Assumption 7 is satisfied), this lower bound is 1.

Remark 5.16. (i) Because $A((\nu_m)_{m \geq 1})$ is a closed subset of the unit ball in $\mathbb{R}^q$ and because the map $v \mapsto \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right)$ is continuous on the unit ball under Assumption 7, we see that the expressions in (26) are positive if and only if

$$
\lambda_{\mathbb{R}^n}\left(\{y \in \mathbb{R}^n \backslash N^* : v'\check{\Omega}^{-1}(y)v \geq 0\}\right) > 0 \quad (28)
$$

holds for every $v \in A((\nu_m)_{m \geq 1})$. Under Assumption 7 we have $\lambda_{\mathbb{R}^n}\left(\{y \in \mathbb{R}^n \backslash N^* : v'\check{\Omega}^{-1}(y)v \geq 0\}\right) = \lambda_{\mathbb{R}^n}\left(\{y \in \mathbb{R}^n \backslash N^* : v'\check{\Omega}^{-1}(y)v > 0\}\right)$ for every $v \neq 0$ and hence, by continuity of $\check{\Omega}^{-1}(y)$ on $\mathbb{R}^n \backslash N^*$, condition (28) for some $v \neq 0$ is in turn equivalent to $v'\check{\Omega}^{-1}(y)v > 0$ for some $y = y(v) \in \mathbb{R}^n \backslash N^*$.

(ii) Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumption 5, let T be the test statistic defined in (25), and suppose that we now use a "random" critical value $\check{C} = \check{C}(y) > 0$ for $y \in \mathbb{R}^n$. Suppose that $\check{C}$ satisfies the invariance condition $\check{C}(\alpha y + X\gamma) = \check{C}(y)$ for every $y \in \mathbb{R}^n \backslash N$, every $\alpha \neq 0$, and for every $\gamma \in \mathbb{R}^k$. Rewriting the rejection region $\{y \in \mathbb{R}^n : T(y) \geq \check{C}\}$ as $\{y \in \mathbb{R}^n : T(y)/\check{C} \geq 1\}$ and observing that $\bar{\Omega}(y) = \check{C}(y)\check{\Omega}(y)$ satisfies Assumption 5 shows that the results of this subsection also apply to the test with rejection region $\{y \in \mathbb{R}^n : T(y) \geq \check{C}\}$.

As a corollary to Theorem 5.7, we now obtain negative size and power results for tests of the form (25). The semicontinuity conditions in Theorem 5.7 are implied by continuity properties of the estimators $\check{\Omega}$ and $\check{\beta}$ used in the construction of the test. The sufficient conditions so obtained are easy to verify in practice and become particularly simple in the practically relevant case where $\dim(\mathcal{Z}) = 1$, cf. the remark following the corollary.

Corollary 5.17. *Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumption 5 and let T be the test statistic defined in (25). Furthermore, let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ with $0 < C < \infty$ be the rejection region. Suppose that $\mathcal{Z}$ is a concentration space of the covariance model $\mathfrak{C}$. Recall that N is the exceptional set in Assumption 5 and that N^* is given by (24). Then the following holds:*

1. *Suppose we have for some $\mu_0^* \in \mathfrak{M}_0$ that $z \in \mathbb{R}^n \backslash N^*$ and $T(\mu_0^* + z) > C$ hold simultaneously $\lambda_{\mathcal{Z}}$-almost surely. Then*
$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C)) = 1$$
holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

2. *Suppose we have for some $\mu_0^* \in \mathfrak{M}_0$ that $z \in \mathbb{R}^n \backslash N^*$ and $T(\mu_0^* + z) < C$ hold simultaneously $\lambda_{\mathcal{Z}}$-almost surely. Then*
$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C)) = 0$$
holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$, and hence
$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W(C)) = 0,$$
holds for every $0 < \sigma^2 < \infty$. In particular, the test is biased (except in the trivial case where its size is zero). Furthermore, the nuisance-infimal rejection probability at every point $\mu_1 \in \mathfrak{M}_1$ is zero, i.e.,
$$\inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W(C)) = 0.$$
In particular, the infimal power of the test is equal to zero.

3. *Suppose $\check{\Omega}$ is nonnegative definite on $\mathbb{R}^n \backslash N$. If $z \in \mathbb{R}^n \backslash N$, $\check{\Omega}(z) = 0$, and $R\check{\beta}(z) \neq 0$ hold simultaneously $\lambda_{\mathcal{Z}}$-almost surely, then*
$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C)) = 1$$
holds for every $\mu_0 \in \mathfrak{M}_0$ and every $0 < \sigma^2 < \infty$. In particular, the size of the test is equal to one.

42

Remark 5.18. (i) Since T in the above corollary is invariant under $G(\mathfrak{M}_0)$, the condition in the corollary does not depend on the particular choice of $\mu_0^* \in \mathfrak{M}_0$. Furthermore, if $\mathcal{Z}$ is one-dimensional, the invariance of T shows that $T(\mu_0^* + z) > C$ already holds for *all* $z \in \mathcal{Z}$ with $z \neq 0$ provided it holds for one $z \in \mathcal{Z}$ with $z \neq 0$. In a similar vein, Part 1 of Lemma 5.15 implies for one-dimensional $\mathcal{Z}$ that $z \in \mathbb{R}^n \backslash N^*$ holds for *all* $z \in \mathcal{Z}$ with $z \neq 0$ if and only if $z \in \mathbb{R}^n \backslash N^*$ holds for at least one $z \in \mathcal{Z}$ with $z \neq 0$. In view of Assumption 5 a similar statement also applies to the relations $z \in \mathbb{R}^n \backslash N$, $\check{\Omega}(z) = 0$, and $R\check{\beta}(z) \neq 0$.

(ii) In case the covariance model $\mathfrak{C}$ contains AR(1) correlation matrices, a remark analogous to Remark 5.9 also applies here. Furthermore, note that the concentration spaces derived from the AR(1) correlation matrices are one-dimensional, and hence the discussion in (i) above applies.

The negative result in the preceding corollary does not apply if substantial portions of $\mathcal{Z}$ belong to the exceptional set N (which in particular occurs if $\mathcal{Z} \subseteq \mathfrak{M}$ holds and N is not empty as then $\mathcal{Z} \subseteq \mathfrak{M} \subseteq N$). For this case we provide a further negative result which is applicable provided (29) given below holds. For example, if $\mathcal{Z} = \mathrm{span}\,(e_+)$ and the design matrix contains an intercept, we immediately obtain $\mathcal{Z} \subseteq \mathfrak{M}$, and (29) holds if and only if the column in R corresponding to the intercept is nonzero. The significance of the subsequent theorem is that it provides an *upper* bound K_1 for the power in certain directions which is less than or equal to a *lower* bound for the size. This will typically imply biasedness of the test (except if equality holds in (30)). Furthermore, note that the result implies that the test has size 1 in case $\check{\Omega}$ is positive definite $\lambda_{\mathbb{R}^n}$-almost everywhere. The condition on the covariance model $\mathfrak{C}$ is often satisfied, see Remark 5.20 following the theorem.

Theorem 5.19. *Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumptions 5 and 7, let T be the test statistic defined in (25), and let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ with $0 < C < \infty$ be the rejection region. Assume that there is a sequence $\Sigma_m \in \mathfrak{C}$ such that $\Sigma_m \to \bar{\Sigma}$ for $m \to \infty$ where $\bar{\Sigma}$ is singular with $l := \dim \mathrm{span}(\bar{\Sigma}) > 0$. Suppose that for some sequence of positive real numbers s_m the matrix $D_m = \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp}/s_m$ converges to a matrix D, which is regular on $\mathrm{span}(\bar{\Sigma})^\perp$, and that $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})}/s_m^{1/2} \to 0$. Suppose further that $\mathrm{span}(\bar{\Sigma}) \subseteq \mathfrak{M}$, and let Z be a matrix, the columns of which form a basis for $\mathrm{span}(\bar{\Sigma})$. Assume also that*

$$R\hat{\beta}(z) \neq 0 \qquad \lambda_{\mathrm{span}(\bar{\Sigma})}\text{-}a.e. \tag{29}$$

is satisfied. Then for every $\mu_0 \in \mathfrak{M}_0$, every σ with $0 < \sigma < \infty$, and every $M \geq 0$ we have

$$\inf_{\gamma \in \mathbb{R}^l, \|\gamma\| \geq M} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_0 + Z\gamma, \sigma^2 \Sigma}\left(W(C)\right) \leq K_1 \leq K_2 \leq \sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}\left(W(C)\right). \tag{30}$$

The constants K_1 and K_2 are given by

$$K_1 = \inf_{\gamma \in \mathbb{R}^l} \Pr\left(\bar{\xi}\,(\gamma) \geq 0\right) = \inf_{\|\gamma\|=1} \Pr\left(\bar{\xi}\,(\gamma) \geq 0\right)$$

and

$$K_2 = \int \Pr\left(\bar{\xi}\,(\gamma) \geq 0\right) dP_{0,A}(\gamma)$$

with the random variable $\bar{\xi}\,(\gamma)$ given by

$$\bar{\xi}\,(\gamma) = \left(R\hat{\beta}\,(Z\gamma)\right)' \check{\Omega}^{-1}\left(\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G}\right) R\hat{\beta}\,(Z\gamma)$$

43

on the event $\left\{\left(\bar{\Sigma}^{1/2} + D^{1/2}\right) \boldsymbol{G} \in \mathbb{R}^n \backslash N^*\right\}$ and by $\bar{\xi}(\gamma) = 0$ otherwise, where $\boldsymbol{G}$ is a standard normal n-vector. The matrix A denotes $(Z'Z)^{-1} Z'\bar{\Sigma}Z (Z'Z)^{-1}$, which is nonsingular, and $P_{0,A}$ denotes the Gaussian distribution on $\mathbb{R}^l$ with mean zero and variance covariance matrix A.

Remark 5.20. Suppose the covariance model $\mathfrak{C}$ contains $\mathfrak{C}_{AR(1)}$, or, more generally, $\mathfrak{C}$ contains AR(1) correlation matrices $\Lambda(\rho_m)$ for some sequence $\rho_m \in (-1, 1)$ with $\rho_m \to 1$ ($\rho_m \to -1$, respectively). Then all the conditions on the covariance model in the preceding theorem are satisfied with $\bar{\Sigma} = e_+e_+'$, $\mathrm{span}(\bar{\Sigma}) = \mathrm{span}(e_+)$, and $Z = e_+$ ($\bar{\Sigma} = e_-e_-'$, $\mathrm{span}(\bar{\Sigma}) = \mathrm{span}(e_-)$, and $Z = e_-$, respectively); cf. Lemma G.1 in Appendix G. Furthermore, condition (29) simplifies to $R\hat{\beta}(e_+) \neq 0$ ($R\hat{\beta}(e_-) \neq 0$, respectively).

The subsequent theorem specializes the positive result given in Theorems 5.10 and 5.12 to the class of test considered in the present subsection.

Theorem 5.21. Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumptions 5, 6, and 7. Let T be the test statistic defined in (25). Furthermore, let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ with $0 < C < \infty$ be the rejection region. Suppose further that

$$T(y + z) = T(y) \qquad \text{for every } y \in \mathbb{R}^n \text{ and every } z \in J(\mathfrak{C}). \tag{31}$$

Assume that $\mathfrak{C}$ is bounded (as a subset of $\mathbb{R}^{n \times n}$). Assume also that for every sequence $\Sigma_m \in \mathfrak{C}$ converging to a singular $\bar{\Sigma}$ there exists a subsequence $(m_i)_{i \in \mathbb{N}}$ and a sequence of positive real numbers s_{m_i} such that the sequence of matrices $D_{m_i} = \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_{m_i} \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} / s_{m_i}$ converges to a matrix D which is regular on the orthogonal complement of $\mathrm{span}(\bar{\Sigma})$. Then the following holds:

1. The size of the rejection region $W(C)$ is strictly less than 1, i.e.,

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma} (W(C)) < 1.$$

   Furthermore,

$$\inf_{\mu_0 \in \mathfrak{M}_0} \inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma} (W(C)) > 0.$$

2. Suppose that $\lambda_{\mathbb{R}^n} \left(\left\{y \in \mathbb{R}^n \backslash N^* : v'\check{\Omega}^{-1}(y)v \geq 0\right\}\right) > 0$ for every $v \in \mathbb{R}^q$ with $\|v\| = 1$. Then the infimal power is bounded away from zero, i.e.,

$$\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{0 < \sigma^2 < \infty} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_1, \sigma^2 \Sigma}(W(C)) > 0.$$

3. Suppose that $\check{\Omega}$ is nonnegative definite $\lambda_{\mathbb{R}^n}$-almost everywhere. Then for every $0 < c < \infty$

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, 0 < \sigma^2 < \infty \\ d(\mu_1, \mathfrak{M}_0)/\sigma \geq c}} P_{\mu_1, \sigma^2 \Sigma_m}(W(C)) \to 1$$

   holds for $m \to \infty$ and for any sequence $\Sigma_m \in \mathfrak{C}$ satisfying $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a singular matrix. Furthermore, for every sequence $0 < c_m < \infty$

$$\inf_{\substack{\mu_1 \in \mathfrak{M}_1, \\ d(\mu_1, \mathfrak{M}_0) \geq c_m}} P_{\mu_1, \sigma_m^2 \Sigma_m}(W(C)) \to 1$$

44

holds for $m \to \infty$ whenever $0 < \sigma_m^2 < \infty$, $c_m/\sigma_m \to \infty$, and the sequence $\Sigma_m \in \mathfrak{C}$ satisfies $\Sigma_m \to \bar{\Sigma}$ with $\bar{\Sigma}$ a positive definite matrix. [The very last statement even holds without recourse to condition (31) and the condition on $\mathfrak{C}$ following (31).]

4. For every δ, $0 < \delta < 1$, there exists a $C(\delta)$, $0 < C(\delta) < \infty$, such that

$$\sup_{\mu_0 \in \mathfrak{M}_0} \sup_{0 < \sigma^2 < \infty} \sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W(C(\delta))) \leq \delta.$$

Remark 5.22. (i) In case the covariance model $\mathfrak{C}$ equals $\mathfrak{C}_{AR(1)}$, a remark analogous to Remark 5.14 also applies here.

(ii) Under the assumptions of the preceding theorem, the additional condition in Part 2 of the theorem is equivalent to $v'\check{\Omega}^{-1}(y)v > 0$ for every $v \in \mathbb{R}^q$ with $\|v\| = 1$ and a suitable $y = y(v) \in \mathbb{R}^n \backslash N^*$. Cf. Remark 5.16(i).

We now discuss when the preceding theorem can be expected to apply and how the crucial condition (31) can be enforced. As already noted prior to Theorem 5.10, a sufficient condition for (31) to be satisfied for *any* test statistic T of the form (25), based on estimators $\check{\beta}$ and $\check{\Omega}$ satisfying Assumption 5, is that $J(\mathfrak{C}) \subseteq \mathfrak{M}_0 - \mu_0$ for some (and hence all) $\mu_0 \in \mathfrak{M}_0$ holds. This sufficient condition is equivalent to $J(\mathfrak{C}) \subseteq \mathfrak{M}$ and $R\hat{\beta}(z) = 0$ for every $z \in J(\mathfrak{C})$, because $\mathfrak{M}_0 - \mu_0$ coincides with the set $\left\{ \mu \in \mathfrak{M} : R\hat{\beta}(\mu) = 0 \right\}$. [Note that replacing $J(\mathfrak{C})$ by $\mathrm{span}\,(J(\mathfrak{C}))$ in the preceding two sentences leads to equivalent statements because $\mathfrak{M}_0 - \mu_0$ as well as $\mathfrak{M}$ are linear spaces.] Now consider the general case where $J(\mathfrak{C})$, or equivalently $\mathrm{span}\,(J(\mathfrak{C}))$, may not be a subset of $\mathfrak{M}_0 - \mu_0$: If there exists a $z \in \mathrm{span}\,(J(\mathfrak{C})) \cap \mathfrak{M}$ with $z \notin \mathfrak{M}_0 - \mu_0$ (i.e., with $R\hat{\beta}(z) \neq 0$), then *any* test statistic T of the form (25), based on estimators $\check{\beta}$ and $\check{\Omega}$ satisfying Assumptions 5 and 7, does *not* satisfy the invariance condition (31), see Lemma F.3 in Appendix F. Hence, $\mathrm{span}\,(J(\mathfrak{C})) \cap \mathfrak{M} \subseteq \mathfrak{M}_0 - \mu_0$, or in other words $R\hat{\beta}(z) = 0$ for every $z \in \mathrm{span}\,(J(\mathfrak{C})) \cap \mathfrak{M}$, is a necessary condition for (31) to be satisfied for *some* T as above. We next show how a test statistic of the form (25) satisfying the crucial invariance condition (31) can in fact be constructed if we impose this necessary condition.

Proposition 5.23. *Let $\mathfrak{C}$ be a covariance model and suppose that $\mathrm{span}\,(J(\mathfrak{C})) \cap \mathfrak{M} \subseteq \mathfrak{M}_0 - \mu_0$ holds.*

1. *Let $\bar{\mathfrak{M}}$ be the linear space spanned by $J(\mathfrak{C}) \cup \mathfrak{M}$. Define $\bar{X} = (X, \bar{x}_1, \ldots, \bar{x}_p)$ where $\bar{x}_i \in \mathrm{span}\,(J(\mathfrak{C}) \cup (\mathfrak{M}_0 - \mu_0))$ are chosen in such a way that the columns of $\bar{X}$ form a basis of $\bar{\mathfrak{M}}$. Assume that $k < k + p < n$ holds. Suppose $\bar{\theta}$ and $\bar{\Omega}$ are estimators satisfying the analogue of Assumption 5 but with k replaced by $k + p$, X replaced by $\bar{X}$, and $\mathfrak{M}$ replaced by $\bar{\mathfrak{M}}$. Let $\bar{N}$ denote the null set appearing in that analogue of Assumption 5 and $\bar{N}^* = \bar{N} \cup \left\{ y \in \mathbb{R}^n \backslash \bar{N} : \det \bar{\Omega}(y) = 0 \right\}$. Define $\bar{\beta} = (I_k, 0) \bar{\theta}$. Then $\bar{\beta}$ and $\bar{\Omega}$ satisfy the original Assumption 5 (with N given by $\bar{N}$), and the test statistic $\bar{T}$ given by*

$$\bar{T}(y) = \begin{cases} (R\bar{\beta}(y) - r)'\bar{\Omega}^{-1}(y)(R\bar{\beta}(y) - r), & y \in \mathbb{R}^n \backslash \bar{N}^*, \\ 0, & y \in \bar{N}^*. \end{cases}$$

satisfies the invariance condition (31).

2. *Let $\bar{\mathfrak{M}}$ and $\bar{X}$ be as above and $k < k + p < n$. Suppose $\bar{\theta}(y) = (\bar{X}'\bar{X})^{-1}\bar{X}'y$ is the least squares estimator based on $\bar{X}$. Then the requirements on $\bar{\theta}$ postulated in the above mentioned analogue of Assumption 5 are satisfied, and $R\bar{\beta}(z) = 0$ holds for every $z \in \mathrm{span}\,(J(\mathfrak{C}))$. Furthermore, if $X^* = (X, x_1^*, \ldots, x_p^*)$ is obtained in the same way as is $\bar{X}$ but for another choice of elements $x_i^* \in \mathrm{span}\,(J(\mathfrak{C}) \cup (\mathfrak{M}_0 - \mu_0))$ and if θ^* denotes the least squares estimator w.r.t. the design matrix X^*, then $R\bar{\beta}(y) = R\beta^*(y)$ holds for every $y \in \mathbb{R}^n$ with β^* denoting $(I_k, 0)\,\theta^*$.*

We next discuss ways of choosing $\bar{x}_1, \ldots, \bar{x}_p$ such that they satisfy the requirements in the preceding proposition: One natural way is to first find $z_1 \ldots, z_r$ in $J(\mathfrak{C})$ that form a basis of $\mathrm{span}\, J(\mathfrak{C})$. From these vectors then select $\bar{x}_1 = z_{i_1} \ldots, \bar{x}_p = z_{i_p}$ to complement the columns of X to a basis of $\bar{\mathfrak{M}}$. An alternative way is based on the observation that adding elements of $\mathfrak{M}_0 - \mu_0$ to each of the previously found z_{i_j} obviously gives rise to another feasible choice of $\bar{x}_i$. It hence follows that an alternative feasible choice for the $\bar{x}_i$ is to use the projections of the z_{i_j} onto the orthogonal complement of $\mathfrak{M}_0 - \mu_0$. Of course, if the estimator $\bar{\theta}$ is chosen to be the least squares estimator, then Part 2 of the preceding proposition informs us that the particular choice of the $\bar{x}_i$ has no effect on $R\bar{\beta}(y)$ since it is invariant under the choice of the $\bar{x}_i$. [In particular, it follows that $\bar{\beta}(y) = \hat{\beta}(y)$ holds for the choice $\bar{x}_i = z_{i_j} - \Pi_{\mathfrak{M}_0 - \mu_0} z_{i_j}$ just discussed.]

Remark 5.24. (i) Suppose that the assumptions of Proposition 5.23 hold, except that now $p = 0$ holds. Then $J(\mathfrak{C}) \subseteq \mathfrak{M}$ and hence $R\hat{\beta}(z) = 0$ holds for every $z \in J(\mathfrak{C})$, implying that actually the sufficient condition mentioned prior to the proposition is satisfied. Consequently, as discussed above, the invariance condition (31) is already satisfied for *every* T of the form (25) based on estimators $\check{\beta}$ and $\check{\Omega}$ satisfying Assumption 5.

(ii) Suppose that the assumptions of Proposition 5.23 hold, except that now $k + p = n$ holds (note that $k + p \leq n$ always holds). Suppose further that T is a test statistic of the form (25) based on estimators $\check{\beta}$ and $\check{\Omega}$ satisfying Assumptions 5 and 6. Then T can never satisfy (31) and hence Theorem 5.21 does not apply in this situation. This can be seen as follows: Because of $k + p = n$ it follows that every $y \in \mathbb{R}^n$ can be written as a linear combination of finitely many $z_i \in J(\mathfrak{C})$ plus an element μ in $\mathfrak{M}$. Because invariance w.r.t. addition of elements $z \in J(\mathfrak{C})$ is equivalent to invariance w.r.t. addition of elements $z \in \mathrm{span}\,(J(\mathfrak{C}))$ (cf. Remark 5.11(i)) we see that $T(y) = T(\mu)$ would have to hold under (31). As noted after the introduction of Assumption 5, either $\mathfrak{M} \subseteq N \subseteq N^*$ holds or N is empty. In the second case we have that $\check{\Omega}(\mu) = 0$ as a consequence of equivariance. Hence in both cases we arrive at $\mu \in N^*$ and thus at $T(\mu) = 0$. But this shows that T is constant equal to zero, contradicting Part 5 of Lemma 5.15.

(iii) Proposition 5.23 uses the auxiliary matrix $\bar{X}$ and the associated estimators $\bar{\theta}$ to construct an estimator $\bar{\beta}$ for the parameter β in the originally given regression model (1) and this estimator $\bar{\beta}$ is then used to construct a test statistic $\bar{T}$ for the testing problem (4) to which Theorem 5.21 can be applied. In an alternative view we can consider the auxiliary model $\mathbf{Y} = \bar{X}\theta + \mathbf{U}$ with $\theta = (\beta', \zeta')'$ as a model in its own right. [Of course, if we maintain model (1) then $\zeta = 0$ must hold in the auxiliary model.] Define the $q \times (k + p)$ matrix $\bar{R} = R(I_k, 0)$, define $\bar{\mathfrak{M}}_0 = \{\mu \in \bar{\mathfrak{M}} : \mu = \bar{X}\theta, \bar{R}\theta = r\}$ and set $\bar{\mathfrak{M}}_1 = \bar{\mathfrak{M}} \backslash \bar{\mathfrak{M}}_0$, and define a null hypothesis $\bar{H}_0$ and an alternative hypothesis $\bar{H}_1$ analogously as in (4). Proposition 5.23 can now be viewed as stating that condition (31) is satisfied for the test statistic which is obtained by using (25) based on the restriction matrix $\bar{R}$ and on the estimators $\bar{\theta}$ and $\bar{\Omega}$ figuring in Proposition 5.23. Consequently, Theorem 5.21 can be directly applied to this test statistic (provided $\bar{\Omega}$ satisfies Assumptions 6 and 7). It should be noted that the so-obtained result now applies to the problem of testing $\bar{H}_0$ versus $\bar{H}_1$. However, since $\mathfrak{M}_0 \subseteq \bar{\mathfrak{M}}_0$ and $\mathfrak{M}_1 \subseteq \bar{\mathfrak{M}}_1$

hold and since T is invariant under translation by elements in span $(J(\mathfrak{C}))$, we essentially recover the same result as before.

5.5 Non-Gaussian distributions

As already noted in Section 2, the negative results given in this paper immediately extend in a trivial way without imposing the Gaussianity assumption on the error vector $\mathbf{U}$ in (1) as long as the assumptions on the feasible error distributions is weak enough to ensure that the implied set of distributions for $\mathbf{Y}$ contains the set $\left\{ P_{\mu,\sigma^2\Sigma} : \mu \in \mathfrak{M}, 0 < \sigma^2 < \infty, \Sigma \in \mathfrak{C} \right\}$, but possibly contains also other distributions.

Another, less trivial, extension is as follows: Suppose that $\mathbf{U}$ is elliptically distributed in the sense that it has the same distribution as $\varrho\sigma\Sigma^{1/2}\mathbf{E}$ where $0 < \sigma < \infty$, $\Sigma \in \mathfrak{C}$, $\mathbf{E}$ is a random vector uniformly distributed on the unit sphere S^{n-1}, and ϱ is a random variable distributed independently of $\mathbf{E}$ satisfying $\Pr(\varrho > 0) = 1$. [If ϱ is distributed as the square root of a chi-square with n degrees of freedom we recover the Gaussian situation described in Section 2.] If φ is a test that is invariant under the group $G(\mathfrak{M}_0)$ then it is easy to see that for $\mu_0 \in \mathfrak{M}_0$

$$\mathbb{E}(\varphi(\mu_0 + \varrho\sigma\Sigma^{1/2}\mathbf{E})) = \mathbb{E}(\varphi(\mu_0 + \Sigma^{1/2}\mathbf{E}))$$

holds.[16] Since this does not depend on the distribution of ϱ at all, we learn that the rejection probability under the null hypothesis is therefore the same as in the Gaussian case. As a consequence, all results concerning only the null behavior of φ obtained under Gaussianity in the paper extend immediately to regression models in which the disturbance vector $\mathbf{U}$ is elliptically distributed in the above sense. Furthermore, all results concerning rejection probabilities under the alternative which are obtained from the behavior of the null rejection probabilities by an approximation argument (e.g., Parts 2 and 3 of Theorem 5.7 as well as of Corollary 5.17, and the corresponding applications of these results in Sections 3 and 4) also go through in view of Scheffé's lemma provided the density of $\varrho\mathbf{E}$ exists and is almost surely continuous.

A   Appendix: Proofs for Subsection 3.1

Proof of Lemma 3.1: Observe that $\hat{\Omega}_w(y) = B(y)\mathcal{W}_n B'(y)$. Given that $\mathcal{W}_n$ is positive definite due to Assumption 2, this immediately establishes Parts 1-3 of the Lemma. It remains to prove Part 4. Let s be as in Assumption 3 and consider first the case where this assumption is satisfied, i.e., where rank $\left(R(X'X)^{-1}X'\left(\neg(i_1,\ldots i_s)\right) \right) = q$ holds. If now y is such that $\hat{\Omega}_w(y)$ is singular it follows, in view of the equivalent condition rank $(B(y)) < q$, that $\hat{u}_l(y) = 0$ 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 $e'_l(n)\left(I_n - X(X'X)^{-1}X' \right)y = 0$. Since $e'_l(n)\left(I_n - X(X'X)^{-1}X' \right) \neq 0$ by construction of l, it follows that the set of y for which $\hat{\Omega}_w(y)$ is singular is contained in a finite union of proper linear subspaces, and hence is a $\lambda_{\mathbb{R}^n}$-null set. Next consider the case where Assumption 3 is not satisfied. Observe that then $s > 0$ must hold. Note that $\hat{u}_i(y) = 0$ holds for all $y \in \mathbb{R}^n$ and all $i \in \{i_1,\ldots i_s\}$ by construction of $\{i_1,\ldots i_s\}$.

---

[16]Under an additional absolute continuity assumption this is also true for almost invariant tests φ.

But then for every $y \in \mathbb{R}^n$

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

is satisfied where $A(y)$ is obtained from $\operatorname{diag}\left(\hat{u}_1(y),\ldots,\hat{u}_n(y)\right)$ by deleting rows and columns i with $i \in \{i_1,\ldots i_s\}$. This completes the proof. ∎

Lemma A.1. *Suppose Assumptions 2 and 3 are satisfied. Then $\hat{\beta}$ and $\hat{\Omega}_w$ satisfy Assumption 5, 6, and 7 with $N = \emptyset$. In fact, $\hat{\Omega}_w(y)$ is nonnegative definite for every $y \in \mathbb{R}^n$, and is positive definite $\lambda_{\mathbb{R}^n}$-almost everywhere. The test statistic T defined in (7), with $\hat{\Psi}_w$ as in (6), is invariant under the group $G\left(\mathfrak{M}_0\right)$ and the rejection probabilities $P_{\mu,\sigma^2\Sigma}(T \geq C)$ depend on $\left(\mu,\sigma^2,\Sigma\right) \in \mathfrak{M} \times (0,\infty) \times \mathfrak{C}$ only through $\left((R\beta - r)/\sigma,\Sigma\right)$ (in fact, only through $\left(\langle(R\beta - r)/\sigma\rangle,\Sigma\right)$), where β corresponds to μ via $\mu = X\beta$.*

Proof. Clearly, $\hat{\beta}$ and $\hat{\Omega}_w$ are well-defined and continuous on $\mathbb{R}^n$, hence we may set $N = \emptyset$ in Assumption 5. Symmetry of $\hat{\Omega}_w$ as well as the required equivariance properties of $\hat{\beta}$ and $\hat{\Omega}_w$ are obviously satisfied. By Assumption 2 $\hat{\Omega}_w(y)$ is nonnegative definite for every $y \in \mathbb{R}^n$. By Assumptions 2 and 3 and Lemma 3.1 the matrix $\hat{\Omega}_w$ is nonsingular (and hence positive definite) $\lambda_{\mathbb{R}^n}$-almost everywhere. Hence Assumptions 5, 6, and 7 are satisfied which proves the first claim. The remaining claims follow immediately from Lemma 5.15 and Proposition 5.4. ∎

Proof of Theorem 3.3: By Lemma A.1 we know that $\hat{\beta}$ and $\hat{\Omega}_w$ satisfy Assumption 5 and that $\hat{\Omega}_w(y)$ is nonnegative definite for every $y \in \mathbb{R}^n$. Furthermore, in view of this lemma and because $N = \emptyset$, the set N^* in Corollary 5.17 is precisely the set of y for which $\operatorname{rank}(B(y)) < q$, cf. Lemma 3.1. By Assumption 1 the spaces $\mathcal{Z}_+ = \operatorname{span}(e_+)$ and $\mathcal{Z}_- = \operatorname{span}(e_-)$ are concentration spaces of $\mathfrak{C}$. The theorem now follows by applying Corollary 5.17 and Remark 5.18(i) to $\mathcal{Z}_+$ as well as to $\mathcal{Z}_-$ and by noting that $e_+ \in \mathbb{R}^n \backslash N^*$ translates into $\operatorname{rank}(B(e_+)) = q$ with a similar translation if e_+ is replaced by e_-. Also note that the size of the test can not be zero in view of Part 5 of Lemma 5.15 and Lemma A.1. ∎

Proof of Proposition 3.5: (1) Define the matrix $B_X^*(y) = \left(\det(X'X)\right)^2 B_X(y)$ and observe that (for given y) every element of this matrix is a multivariate polynomial in the elements x_{ti} of X because $(X'X)^{-1}$ can be written as $\left(\det(X'X)\right)^{-1}\operatorname{adj}(X'X)$ (with the convention that $\operatorname{adj}(X'X) = 1$ if $k = 1$). Because $\det(X'X) \neq 0$ for $X \in \mathfrak{X}_0$ holds, we have

$$
\mathfrak{X}_1(e_+) = \mathfrak{X}_0 \cap \left\{X \in \mathbb{R}^{n \times k} : \det\left(B_X^*(e_+)B_X^{*\prime}(e_+)\right) = 0\right\}.
$$

The set to the right of the intersection operation in the above display is obviously the zero-set of a multivariate polynomial in the variables x_{ti}. Thus it is an algebraic set, and hence is either a $\lambda_{\mathbb{R}^{n \times k}}$-null set or is all of $\mathbb{R}^{n \times k}$. However, the latter case can not arise because we can choose an $n \times k$ matrix $X^\# \in \mathfrak{X}_0$, say, such that all its columns are orthogonal to e_+ (this being possible since $k < n$ by assumption) and this matrix then satisfies $\operatorname{rank}\left(B_{X^\#}^*(e_+)\right) = q$. This shows that $\mathfrak{X}_1(e_+)$ is a $\lambda_{\mathbb{R}^{n \times k}}$-null set. Next consider $\mathfrak{X}_2(e_+)$: Observe that for $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_1(e_+)$ we have $\det(\hat{\Omega}_{w,X}(e_+)) \neq 0$ and hence for $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_1(e_+)$ the relation $T_X(e_+ + \mu_0^*) = C$ can equivalently be written as

$$
(R\operatorname{adj}(X'X)X'e_+)'\operatorname{adj}\hat{\Omega}_{w,X}(e_+)(R\operatorname{adj}(X'X)X'e_+) - \left(\det(X'X)\right)^2 \det(\hat{\Omega}_{w,X}(e_+))C = 0.
$$

Furthermore, for $X \in \mathfrak{X}_0$ we can write $\hat{\Omega}_{w,X}(e_+)$ as $\left(\det(X'X)\right)^{-4}B_X^*(e_+)\mathcal{W}_n B_X^{*\prime}(e_+)$. Note that $B_X^*(e_+)\mathcal{W}_n B_X^{*\prime}(e_+)$ is a multivariate polynomial in the variables x_{ti}. Consequently, for $X \in$

$\mathfrak{X}_0\backslash\mathfrak{X}_1\left(e_+\right)$ the relation $T_X(e_+ + \mu_0^*) = C$ can, after multiplication by $(\det(X'X))^{4q-2}$, which is nonzero for $X \in \mathfrak{X}_0$, equivalently be written as

$$(\det(X'X))^2 \left(R\operatorname{adj}(X'X)X'e_+\right)' \operatorname{adj}\left(B_X^*\left(e_+\right)\mathcal{W}_n B_X^{*\prime}\left(e_+\right)\right)\left(R\operatorname{adj}(X'X)X'e_+\right)$$
$$-\det(B_X^*\left(e_+\right)\mathcal{W}_n B_X^{*\prime}\left(e_+\right))C \quad = \quad 0.$$

The left-hand side of the above display is now a multivariate polynomial in the elements x_{ti}. The polynomial does not vanish on all of $\mathbb{R}^{n\times k}$ since the matrix $X^{\#}$ constructed before provides an element in $\mathfrak{X}_0\backslash\mathfrak{X}_1\left(e_+\right)$ for which $T_{X^{\#}}(e_+ + \mu_0^*) = 0 < C$ holds. The proofs for $\mathfrak{X}_1\left(e_-\right)$ and $\mathfrak{X}_2\left(e_-\right)$ are completely analogous, as is the proof for the fact that $\mathbb{R}^{n\times k}\backslash\mathfrak{X}_0$ is a $\lambda_{\mathbb{R}^{n\times k}}$-null set. Finally, that the set of all design matrices $X \in \mathfrak{X}_0$ for which Theorem 3.3 does not apply is a subset of $\left(\mathfrak{X}_1\left(e_+\right)\cup\mathfrak{X}_2\left(e_+\right)\right)\cap\left(\mathfrak{X}_1\left(e_-\right)\cup\mathfrak{X}_2\left(e_-\right)\right)$ is obvious upon observing that the set of all $X \in \mathfrak{X}_0$ which do not satisfy Assumption 3 is contained in $\mathfrak{X}_1\left(e_+\right)$ as well as in $\mathfrak{X}_1\left(e_-\right)$.

(2) Similar arguments as in the proof of Part 1 show that $\tilde{\mathfrak{X}}_1\left(e_-\right)$ and $\tilde{\mathfrak{X}}_2\left(e_-\right)$ are each contained in an algebraic set. Define the matrix $X^{\sharp} = \left(e_+, \tilde{X}^{\sharp}\right)$ where the columns of $\tilde{X}^{\sharp}$ are $k-1$ linearly independent unit vectors that are orthogonal on e_+ as well as e_-. It is then easy to see that $\tilde{X}^{\sharp} \in \tilde{\mathfrak{X}}_0\backslash\tilde{\mathfrak{X}}_1\left(e_-\right)$, implying that $\tilde{\mathfrak{X}}_1\left(e_-\right)$ does not coincide with all of $\tilde{\mathfrak{X}}_0$. Furthermore, simple computation shows that $T_{X^{\sharp}}(e_- + \mu_0^*) = 0 < C$ by the assumption on R, which implies that $\tilde{\mathfrak{X}}_2\left(e_-\right)$ is a proper subset of $\tilde{\mathfrak{X}}_0\backslash\tilde{\mathfrak{X}}_1\left(e_-\right)$. It follows now as above that $\tilde{\mathfrak{X}}_1\left(e_-\right)$ and $\tilde{\mathfrak{X}}_2\left(e_-\right)$ are $\lambda_{\mathbb{R}^{n\times(k-1)}}$-null sets. The rest of the proof now proceeds as before.

(3) See Example 3.1. ∎

Proof of Theorem 3.6: We verify the assumptions of Theorem 5.21. By Lemma A.1 Assumptions 5, 6, and 7 are satisfied. Because of $\mathfrak{C} = \mathfrak{C}_{AR(1)}$ we have that $J\left(\mathfrak{C}\right) = \operatorname{span}(e_+)\cup\operatorname{span}(e_-)$, see Lemma G.1, and because $e_+, e_- \in \mathfrak{M}$ is assumed we conclude that $J\left(\mathfrak{C}\right) \subseteq \mathfrak{M}$. The assumption $R\hat{\beta}(e_+) = R\hat{\beta}(e_-) = 0$ then implies that even $J\left(\mathfrak{C}\right) \subseteq \mathfrak{M}_0 - \mu_0$ holds. The invariance condition (31) in Theorem 5.21 is thus satisfied, because T is $G\left(\mathfrak{M}_0\right)$-invariant by Lemma 5.15. The assumptions on $\mathfrak{C}$ in Theorem 5.21 are satisfied in view of Lemma G.1. Finally the assumptions on $\hat{\Omega}_w$ in Parts 2 and 3 of Theorem 5.21 are satisfied because $\hat{\Omega}_w$ is positive definite $\lambda_{\mathbb{R}^n}$-almost everywhere as shown in Lemma A.1. The theorem now follows from Theorem 5.21 using a standard subsequence argument for Part 3. The claim in parenthesis in Part 3 follows from the corresponding claim in parenthesis in Theorem 5.21 and the observation that the conditions on e_+ and e_- in the theorem were only used to verify condition (31). ∎

Proof of Theorem 3.7: Similar as in the preceding proof verify the assumptions of Theorem 5.21 but now for $\bar{\beta}$ and $\bar{\Omega}_w$ by making use of Proposition 5.23. Note that the condition $\operatorname{span}\left(J(\mathfrak{C})\right)\cap \mathfrak{M} \subseteq \mathfrak{M}_0 - \mu_0$ is satisfied in all five parts of the theorem. This is obvious for Parts 1-3. For Part 4 this follows from the following argument: Observe that $e_- = \delta e_+ + X\gamma$ must hold by the assumptions of Part 4. Now suppose $m \in \operatorname{span}\left(J(\mathfrak{C})\right)\cap\mathfrak{M}$. Then $\alpha_+ e_+ + \alpha_- e_- = m = X\gamma^*$ must hold. These relations together imply $(\alpha_+ + \alpha_-\delta)e_+ = X(\gamma^* - \alpha_-\gamma)$. Because $e_+ \notin \mathfrak{M}$, it follows that $\gamma^* - \alpha_-\gamma = 0$. Thus

$$R\left(X'X\right)^{-1}X'm = R\gamma^* = \alpha_- R\gamma = \alpha_-\bar{R}(\gamma':\delta)' = \alpha_-\bar{R}\left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'e_- = 0,$$

which establishes that $m \in \mathfrak{M}_0 - \mu_0$. The verification for Part 5 is completely analogous. ∎

Proof of Lemma 3.10: Since $\hat{\Omega}_w\left(y\right) = nB\left(y\right)\mathcal{W}_n^* B'\left(y\right)$, Parts 1-3 of the Lemma follow immediately from nonnegative definiteness of $\mathcal{W}_n^*$. To prove Part 4 observe that $\hat{\Omega}_w\left(y\right)$ is singular if and only if $\det\left(B\left(y\right)\mathcal{W}_n^* B'\left(y\right)\right) = 0$. Now observe that the l.h.s. of this equation is a multivariate

49

polynomial in y, hence the solution set is an algebraic set and thus is either a $\lambda_{\mathbb{R}^n}$-null set or all of $\mathbb{R}^n$. ■

Proof of Theorem 3.11: The proof is completely analogous to the proof of Theorem 3.3 using Lemma G.2 in case $\nu \in (0, \pi)$. ■

B Appendix: Proofs for Subsection 3.2

Proof of Lemma 3.13: The inclusion $\mathfrak{M} \subseteq N_0(a_1, a_2)$ is trivial since $\hat{u}(y) = 0$ for $y \in \mathfrak{M}$. Because $a_1 \in \{1, 2\}$, $a_2 \in \{n-1, n\}$ with $a_1 \leq a_2$ holds, $N_0(a_1, a_2)$ is contained in $N_1(a_1, a_2)$, establishing the first claim. Closedness of $N_1(a_1, a_2)$ is obvious. Given the just established inclusion $N_0(a_1, a_2) \subseteq N_1(a_1, a_2)$ the alternative description of $N_1(a_1, a_2)$ given in the second claim is also immediately seen to be true. Continuity of $\hat{\rho}$ on $\mathbb{R}^n \backslash N_0(a_1, a_2)$ is obvious. Assume now that $k \leq a_2 - a_1$ holds. If $a_1 = 1$, $a_2 = n$, i.e., $\hat{\rho} = \hat{\rho}_{YW}$, we have $N_1(a_1, a_2) = N_0(a_1, a_2) = \mathfrak{M}$ because $\hat{\rho}_{YW}$ is well-defined and bounded away from one in modulus on $\mathbb{R}^n \backslash \mathfrak{M}$ as shown in Remark 3.12(i). Hence, $N_1(a_1, a_2)$ is a $\lambda_{\mathbb{R}^n}$- null set in this case as $k < n$ holds by assumption. To establish this result also for the other choices of a_1 and a_2 note that $N_1(a_1, a_2)$ is the zero set of a multivariate polynomial in y. It hence is a $\lambda_{\mathbb{R}^n}$- null set, provided we can show that the polynomial is not identically zero. Observe that we now have $n - k \geq n - a_2 + a_1 \geq 2$ (as we have already disposed off the case $a_1 = 1$, $a_2 = n$). Let $y^{(1)}, \ldots, y^{(n-k)}$ be a basis for $\mathfrak{M}^\perp$. The submatrix obtained from $(y^{(1)}, \ldots, y^{(n-k)})$ by selecting the rows with index j satisfying $j < a_1$ as well as the rows with $j > a_2$ has dimension $(n - a_2 + a_1 - 1) \times (n - k)$ and thus has rank at most $n - a_2 + a_1 - 1 < n - k$. Consequently, we can find constants $c_1, \ldots, c_{n-k}$, not all equal to zero, such that the j-th component of $y_0 = \sum_{i=1}^{n-k} c_i y^{(i)}$ is zero whenever $j < a_1$ or $j > a_2$. Because $y_0 \in \mathfrak{M}^\perp$ and $y_0 \neq 0$ by construction, we have $y_0 \in \mathbb{R}^n \backslash \mathfrak{M} = \mathbb{R}^n \backslash N_0(1, n)$. Because $y_0 \neq 0$ and because the j-th component of $y_0 = \hat{u}(y_0)$ is zero whenever $j < a_1$ or $j > a_2$, we also have $y_0 \in \mathbb{R}^n \backslash N_0(a_1, a_2)$. Hence, $\hat{\rho}(y_0)$ as well as $\hat{\rho}_{YW}(y_0)$ are well-defined. Furthermore, they coincide in view of the construction of $y_0 = \hat{u}(y_0)$. By what was said above for the Yule-Walker estimator it follows that $|\hat{\rho}(y_0)| = |\hat{\rho}_{YW}(y_0)| < 1$. Hence $y_0 \in \mathbb{R}^n \backslash N_1(a_1, a_2)$, and the polynomial is not identically equal to zero. ■

Lemma B.1. *Suppose $\hat{\rho}$ satisfies Assumption 4.*

1. *The sets $\mathbb{R}^n \backslash N_0(a_1, a_2)$, $\mathbb{R}^n \backslash N_1(a_1, a_2)$, and $\mathbb{R}^n \backslash N_2(a_1, a_2)$ are invariant under the group of transformations $y \mapsto \alpha y + X\gamma$ where $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$.*

2. *The estimators $\tilde{\beta}$, $\tilde{\sigma}^2$, and $\tilde{\Omega}$ are well-defined and continuous on $\mathbb{R}^n \backslash N_2(a_1, a_2)$. They satisfy the equivariance conditions $\tilde{\beta}(\alpha y + X\gamma) = \alpha \tilde{\beta}(y) + \gamma$, $\tilde{\sigma}^2(\alpha y + X\gamma) = \alpha^2 \tilde{\sigma}^2(y)$, and $\tilde{\Omega}(\alpha y + X\gamma) = \alpha^2 \tilde{\Omega}(y)$ for $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$, and $y \in \mathbb{R}^n \backslash N_2(a_1, a_2)$. The estimator $\tilde{\Omega}(y)$ is (well-defined and) nonsingular if and only if $y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$. The sets $N_2(a_1, a_2)$ and $N_2^*(a_1, a_2)$ are closed. If $k \leq a_2 - a_1$ holds, $N_2(a_1, a_2)$ and $N_2^*(a_1, a_2)$ are $\lambda_{\mathbb{R}^n}$-null sets.*

3. *The estimator $\hat{\Omega}$ is well-defined and continuous on $\mathbb{R}^n \backslash N_0(a_1, a_2)$, whereas $\hat{\beta}$ and $\hat{\sigma}^2$ are well-defined and continuous on all of $\mathbb{R}^n$. They satisfy the equivariance conditions $\hat{\beta}(\alpha y + X\gamma) = \alpha \hat{\beta}(y) + \gamma$, $\hat{\sigma}^2(\alpha y + X\gamma) = \alpha^2 \hat{\sigma}^2(y)$ for $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$, and $y \in \mathbb{R}^n$, as well as $\hat{\Omega}(\alpha y + X\gamma) = \alpha^2 \hat{\Omega}(y)$ for $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$, and $y \in \mathbb{R}^n \backslash N_0(a_1, a_2)$. Furthermore, $\hat{\sigma}^2(y) > 0$ holds for $y \in \mathbb{R}^n \backslash \mathfrak{M} \supseteq \mathbb{R}^n \backslash N_0^*(a_1, a_2)$, and hence $\hat{\Omega}(y)$ is (well-defined and) nonsingular if and only if $y \in \mathbb{R}^n \backslash N_0^*(a_1, a_2)$. The set $N_0^*(a_1, a_2)$ is closed. If $k \leq a_2 - a_1$ holds, $N_0^*(a_1, a_2)$*

is a $\lambda_{\mathbb{R}^n}$-null set. [Recall from Remark 3.12(iv) that $N_0(a_1, a_2)$ is always a closed set, and is a $\lambda_{\mathbb{R}^n}$-null set in case $k \leq a_2 - a_1$.]

Proof. (1) The invariance of the first two sets follows since $\hat{u}(\alpha y + X\gamma) = \alpha \hat{u}(y)$ holds for every $y \in \mathbb{R}^n$, $\alpha \neq 0$, and $\gamma \in \mathbb{R}^k$. This property of the residual vector implies $\hat{\rho}(\alpha y + X\gamma) = \hat{\rho}(y)$ for every $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$ and $y \in \mathbb{R}^n \backslash N_0(a_1, a_2) \supseteq \mathbb{R}^n \backslash N_1(a_1, a_2)$. Together with the already established invariance of $\mathbb{R}^n \backslash N_1(a_1, a_2)$ this implies invariance of $\mathbb{R}^n \backslash N_2(a_1, a_2)$ upon observing that $\Lambda^{-1}(\hat{\rho}(y))$ is well-defined for $y \in \mathbb{R}^n \backslash N_1(a_1, a_2)$. The latter holds because for $b \in \mathbb{R}$, $|b| \neq 1$ the matrix $\Lambda(b)$ is nonsingular. [This can, e.g., be seen from the fact that its inverse is given by the symmetric tridiagonal matrix with diagonal equal to $\left(1, 1 + b^2, \ldots, 1 + b^2, 1\right) / \left(1 - b^2\right)$ and with the elements next to the diagonal given by $-b / \left(1 - b^2\right)$.]

(2) Using Lemma 3.13 and the just established fact that $\Lambda^{-1}(\hat{\rho}(y))$ is well-defined for $y \in \mathbb{R}^n \backslash N_1(a_1, a_2)$, we see that $\tilde{\beta}$, $\tilde{\sigma}^2$, and $\tilde{\Omega}$ are well-defined and continuous on $\mathbb{R}^n \backslash N_2(a_1, a_2) \subseteq \mathbb{R}^n \backslash N_1(a_1, a_2)$. Observing that $\hat{\rho}(\alpha y + X\gamma) = \hat{\rho}(y)$ holds for $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$, and $y \in \mathbb{R}^n \backslash N_0(a_1, a_2) \supseteq \mathbb{R}^n \backslash N_1(a_1, a_2)$, the claimed equivariance of $\tilde{\beta}$, $\tilde{\sigma}^2$, and $\tilde{\Omega}$ follows. The third claim is obvious, and the fourth claim follows easily from Lemma 3.13. We next prove the last claim for the Yule-Walker estimator, i.e., for $a_1 = 1$ and $a_2 = n$: For this it suffices to show that $N_2^*(1, n) \subseteq \mathfrak{M}$ since $\mathfrak{M}$ is a proper subspace of $\mathbb{R}^n$ in view of the assumption $k < n$. Now for arbitrary $y \notin \mathfrak{M} = N_0(1, n)$ we have that $\hat{\rho}_{YW}(y)$ is well-defined and satisfies $|\hat{\rho}_{YW}(y)| < 1$ (cf. Remark 3.12(i)) implying $y \in \mathbb{R}^n \backslash N_1(1, n)$ as well as positive definiteness of $\Lambda(\hat{\rho}_{YW}(y))$. But this gives positive definiteness, and hence nonsingularity, of $X'\Lambda^{-1}(\hat{\rho}_{YW}(y))X$, implying that $y \in \mathbb{R}^n \backslash N_2(1, 2)$. It also delivers positive definiteness of $R(X'\Lambda^{-1}(\hat{\rho}(y))X)^{-1}R'$. Furthermore, $y \notin \mathfrak{M}$ implies $y - X\tilde{\beta}(y) \neq 0$ and thus $\tilde{\sigma}^2(y) > 0$ in view of the just established positive definiteness of $\Lambda(\hat{\rho}_{YW}(y))$. But this gives $y \in \mathbb{R}^n \backslash N_2^*(1, 2)$, completing the proof for the case $a_1 = 1$ and $a_2 = n$. To prove the claim for the remaining values of a_1 and a_2 we first show that $N_2(a_1, a_2)$ is a $\lambda_{\mathbb{R}^n}$-null set: observe that $N_2(a_1, a_2)$ is the union of $N_1(a_1, a_2)$ and $\left\{y \in \mathbb{R}^n \backslash N_1(a_1, a_2) : \det\left(X'\Lambda^{-1}(\hat{\rho}(y))X\right) = 0\right\}$. In view of Lemma 3.13 it hence suffices to show that the latter set is a $\lambda_{\mathbb{R}^n}$-null set. Using the relation $D^{-1} = \operatorname{adj}(D) / \det(D)$ (with the convention that $\operatorname{adj}(D) = 1$ if D is 1×1) and noting that $\det(\Lambda(\hat{\rho}(y))) \neq 0$ for $y \in \mathbb{R}^n \backslash N_1(a_1, a_2)$ the set in question can be rewritten as

$$A = \left\{y \in \mathbb{R}^n \backslash N_1(a_1, a_2) : \det\left(X' \operatorname{adj}\left(\Lambda(\hat{\rho}(y))\right)X\right) = 0\right\}.$$

Note that the equation in the set in the above display is polynomial in $\hat{\rho}(y)$. Upon multiplying the equation defining A by $\left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(y)\right)^d$, which is non-zero on $\mathbb{R}^n \backslash N_1(a_1, a_2)$, where $d = (n-1)^2 k$, the set A is seen to be the intersection of $\mathbb{R}^n \backslash N_1(a_1, a_2)$ with the zero-set of a multivariate polynomial in y. Hence, A is a $\lambda_{\mathbb{R}^n}$-null set provided we can establish that the polynomial is not identically zero. For this it suffices to find an $y \in \mathbb{R}^n \backslash N_1(a_1, a_2)$ such that $\det\left(X'\Lambda^{-1}(\hat{\rho}(y))X\right) \neq 0$: Set $y = y_0$ where y_0 has been constructed in the proof of Lemma 3.13. Observe that $\hat{\rho}(y_0) = \hat{\rho}_{YW}(y_0)$ for the estimator $\hat{\rho}$ specified by a_1 and a_2 and hence $y_0 \in \mathbb{R}^n \backslash N_1(a_1, a_2) \subseteq \mathbb{R}^n \backslash \mathfrak{M}$ since $|\hat{\rho}_{YW}(y_0)| < 1$ holds. But then $\det\left(X'\Lambda^{-1}(\hat{\rho}(y_0))X\right) \neq 0$ holds because $\Lambda(\hat{\rho}_{YW}(y))$ is always positive definite (whenever it is defined) as has been established before. This shows that $N_2(a_1, a_2)$ is a $\lambda_{\mathbb{R}^n}$-null set. It remains to show that $N_2^*(a_1, a_2)$ is a $\lambda_{\mathbb{R}^n}$-null set. For this it suffices to show that

$$B = \left\{y \in \mathbb{R}^n \backslash N_2(a_1, a_2) : \tilde{\sigma}^2(y) = 0\right\}$$

as well as

$$C = \left\{y \in \mathbb{R}^n \backslash N_2(a_1, a_2) : \det\left(R(X'\Lambda^{-1}(\hat{\rho}(y))X)^{-1}R'\right) = 0\right\}$$

are $\lambda_{\mathbb{R}^n}$-null sets. Noting that $\det\left(\Lambda(\hat\rho(y))\right) \neq 0$ as well as $\det\left(X' \operatorname{adj}\left(\Lambda(\hat\rho(y))\right) X\right) \neq 0$ hold for $y \in \mathbb{R}^n \backslash N_2(a_1, a_2)$, the set B can be rewritten as

$$B = \left\{ y \in \mathbb{R}^n \backslash N_2(a_1, a_2) : \det\left(\Lambda(\hat\rho(y))\right) \left[\det\left(X' \operatorname{adj}\left(\Lambda(\hat\rho(y))\right) X\right)\right]^2 \tilde\sigma^2(y) = 0 \right\}.$$

Again the equation in the set in the above display is polynomial in y and $\hat\rho(y)$. Upon multiplying this by $\left(\sum_{t=a_1}^{a_2} \hat u_t^2(y)\right)^d$, which is non-zero on $\mathbb{R}^n \backslash N_2(a_1, a_2)$, where $d = (n-1)^2(2k+1)$, one sees that B is the intersection of $\mathbb{R}^n \backslash N_2(a_1, a_2)$ with the zero-set of a multivariate polynomial in y. To establish that B is a $\lambda_{\mathbb{R}^n}$-null set it thus suffices to find an $y \in \mathbb{R}^n \backslash N_2(a_1, a_2)$ with $\tilde\sigma^2(y) > 0$. Choose y_0 as above. Then we know that $y_0 \in \mathbb{R}^n \backslash N_1(a_1, a_2)$ and $\det\left(X'\Lambda^{-1}(\hat\rho(y_0))X\right) \neq 0$ hold, i.e., $y_0 \in \mathbb{R}^n \backslash N_2(a_1, a_2)$. Furthermore, as shown before $\Lambda(\hat\rho(y_0))$ is positive definite (since $\Lambda(\hat\rho(y_0)) = \Lambda(\hat\rho_{YW}(y_0))$) and $y_0 - X\tilde\beta(y_0) \neq 0$ holds (since $y_0 \notin \mathfrak{M}$). Consequently, $\tilde\sigma^2(y_0) > 0$ holds. The proof for C is very similar.

(3) Well-definedness is trivial and continuity follows from continuity of $\hat\rho$ on the open set $\mathbb{R}^n \backslash N_0(a_1, a_2)$ (cf. Lemma 3.13). Equivariance of $\hat\beta$ and $\hat\sigma^2$ is obvious, while the equivariance property of $\hat\Omega$ follows from invariance of $\mathbb{R}^n \backslash N_0(a_1, a_2)$ and the equivariance of $\hat\rho$ established in (1). The third claim is obvious. Closedness of $N_0^*(a_1, a_2)$ follows from the continuity property of $\hat\rho$ established in Lemma 3.13. To prove the final claim observe that $N_0^*(a_1, a_2)$ is the union of the $\lambda_{\mathbb{R}^n}$-null set $N_0(a_1, a_2)$ with

$$\left\{ y \in \mathbb{R}^n \backslash N_0(a_1, a_2) : \det\left(R(X'X)^{-1} X'\Lambda(\hat\rho(y))X(X'X)^{-1}R'\right) = 0 \right\}.$$

Multiplying the equation defining this set by $\left(\sum_{t=a_1}^{a_2} \hat u_t^2(y)\right)^{q(n-1)}$, which is non-zero on $\mathbb{R}^n \backslash N_0(a_1, a_2)$, one sees that the above set is the intersection of $\mathbb{R}^n \backslash N_0(a_1, a_2)$ with the zero-set of a multivariate polynomial in y. Again perusing y_0 constructed before shows that the polynomial is not identically zero, which then delivers the desired result. $\blacksquare$

The following lemma is an immediate consequence of Lemma B.1.

Lemma B.2. *Suppose $\hat\rho$ satisfies Assumption 4 and $k \leq a_2 - a_1$ holds. Then $\tilde\beta$ and $\tilde\Omega$ satisfy Assumption 5 with $N = N_2(a_1, a_2)$, and the set N^* (cf. equation (24)) is given by $N_2^*(a_1, a_2)$. Similarly, $\hat\beta$ and $\hat\Omega$ satisfy Assumption 5 with $N = N_0(a_1, a_2)$ and the set N^* is given by $N_0^*(a_1, a_2)$. The sets $N_0^*(a_1, a_2)$ and $N_2^*(a_1, a_2)$ are invariant under the group of transformations $y \mapsto \alpha y + X\gamma$ where $\alpha \neq 0$, $\gamma \in \mathbb{R}^k$.*

Proof. The lemma except for the last claim follows from B.1. The last claim then follows form Lemma F.1 in Appendix F, cf. also the discussion following Assumption 5. $\blacksquare$

Lemma B.3. *Suppose $\hat\rho$ satisfies Assumption 4 and $k \leq a_2 - a_1$ holds. Then $\tilde\Omega$ and $\hat\Omega$ satisfy Assumptions 6 and 7 with $N^* = N_2^*(a_1, a_2)$ in case of $\tilde\Omega$ and with $N^* = N_0^*(a_1, a_2)$ in case of $\hat\Omega$.*

Proof. Consider first the case of the Yule-Walker estimator, i.e., $a_1 = 1$ and $a_2 = n$. Then $\Lambda(\hat\rho_{YW}(y))$ is positive definite for every $y \notin N_0(a_1, a_2)$. Hence $\tilde\Omega(y)$ is positive definite for $y \notin N_2^*(a_1, a_2)$ and $\hat\Omega(y)$ is positive definite for $y \notin N_0^*(a_1, a_2)$. Consequently, Assumptions 6 and 7 are clearly satisfied. Next consider the case where $a_1 \neq 1$ or $a_2 \neq n$. Then y_0 constructed in the proof of Lemma 3.13 satisfies $y_0 \in \mathbb{R}^n \backslash N_0^*(a_1, a_2)$ as well as $y_0 \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$ as shown in the proof of Lemma B.1. Because of $\hat\rho(y_0) = \hat\rho_{YW}(y_0)$, we also see that $\tilde\Omega(y_0)$ as well as $\hat\Omega(y_0)$ are positive definite (as the variance covariance estimators based on $\hat\rho$ coincide with the ones based on the Yule-Walker estimator). This shows that Assumption 6 is satisfied for

52

$\tilde{\Omega}$ and $\hat{\Omega}$. It remains to establish Assumption 7: Let $v \neq 0$, $v \in \mathbb{R}^q$ be arbitrary. The preceding argument has shown $y_0 \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$ and $y_0 \in \mathbb{R}^n \backslash N_0^*(a_1, a_2)$ and also shows that $v'\tilde{\Omega}^{-1}(y_0)v > 0$ and $v'\hat{\Omega}^{-1}(y_0)v > 0$ hold. To complete the proof it suffices to show that the set $\left\{y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2) : v'\tilde{\Omega}^{-1}(y)v = 0\right\}$ is the intersection of $y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$ with the zero-set of a multivariate polynomial, and similarly for $\left\{y \in \mathbb{R}^n \backslash N_0^*(a_1, a_2) : v'\hat{\Omega}^{-1}(y)v = 0\right\}$. This is proved in a similar manner as in the proof of Lemma B.1 by rewriting all inverse matrices appearing in $v'\tilde{\Omega}^{-1}(y)v$ ($v'\hat{\Omega}^{-1}(y)v$, respectively) in terms of the adjoints and determinants and observing that the determinants are all non-zero for $y \in \mathbb{R}^n \backslash N_2^*(a_1, a_2)$ ($y \in \mathbb{R}^n \backslash N_0^*(a_1, a_2)$, respectively). This shows that $v'\tilde{\Omega}^{-1}(y)v = 0$ ($v'\hat{\Omega}^{-1}(y)v = 0$, respectively) can be rewritten as a polynomial equation in $\hat{\rho}(y)$. Multiplying this polynomial equation by a suitable power of $\sum_{t=a_1}^{a_2} \hat{u}_t^2(y)$, which is non-zero on $\mathbb{R}^n \backslash N_2^*(a_1, a_2)$ ($\mathbb{R}^n \backslash N_0^*(a_1, a_2)$, respectively) shows that these equations can be rewritten as polynomial equations in y. $\blacksquare$

Proof of Theorem 3.14: We first verify the assumptions of Corollary 5.17. Assumption 5 is satisfied for $\tilde{\beta}$ and $\tilde{\Omega}$ (with $N = N_2(a_1, a_2)$ and $N^* = N_2^*(a_1, a_2)$) as well as for $\hat{\beta}$ and $\hat{\Omega}$ (with $N = N_0(a_1, a_2)$ and $N^* = N_0^*(a_1, a_2)$) in view of Lemma B.2. In view of Assumption 1 we conclude from Lemma G.1 that $\mathcal{Z}_+ = \text{span}(e_+)$ as well as $\mathcal{Z}_- = \text{span}(e_-)$ are concentration spaces of $\mathfrak{C}$. Applying Parts 1 and 2 of Corollary 5.17 and Remark 5.18(i) to $\mathcal{Z}_+$ as well as to $\mathcal{Z}_-$ establishes (1) and (2) of the theorem as well as the corresponding parts of (4), if we also note that the size of the test can not be zero in view of Part 5 of Lemma 5.15 and Lemma B.3. In order to prove (3) of the theorem, we apply Theorem 5.19. First note that $\hat{\Omega}$ satisfies Assumption 7 because of Lemma B.3. Furthermore, choose as the sequence Σ_m in that theorem $\Sigma_m = \Lambda(\rho_m)$ for some sequence $\rho_m \to 1$, $\rho_m \in (-1, 1)$. Then $\bar{\Sigma} = e_+ e_+$ by Lemma G.1, which also provides the matrix D and its required properties. Hence $l = 1$ and $\text{span}(\bar{\Sigma}) = \text{span}(e_+)$ which is contained in $\mathfrak{M}$ since $e_+ \in \mathfrak{M}$ has been assumed and $\mathfrak{M}$ is a linear space. Condition (29) in Theorem 5.19 is satisfied in view of the assumption $R\hat{\beta}(e_+) \neq 0$ since $\text{span}(\bar{\Sigma}) = \text{span}(e_+)$. Inspection of the constants K_1 and K_2 in Theorem 5.19 reveal that $K_1 = K_2 =: K_{FGLS}(e_+)$ since in the present case γ is one-dimensional. That $K_{FGLS}(e_+)$ depends only on the quantities given in the theorem is obvious from the formulas for K_1 and K_2. Furthermore, if $\hat{\rho} \equiv \hat{\rho}_{YW}$, then $\tilde{\Omega}$ is always positive definite on $\mathbb{R}^n \backslash N_2(1, n) = \mathbb{R}^n \backslash \mathfrak{M}$, because $|\hat{\rho}_{YW}| < 1$ holds implying that $\Lambda(\hat{\rho}_{YW})$ is positive definite on $\mathbb{R}^n \backslash N_2(1, n)$. Inspection of the constants K_1 and K_2 then reveals $K_1 = K_2 = 1$ in that case. The claims in (3) with e_+ replaced by e_- are proved analogously, and so are the remaining claims in (4). $\blacksquare$

Proof of Proposition 3.15: (1) First consider $\mathfrak{X}_{1,FGLS}(e_+)$. The condition $e_+ \in N_{2,X}^*(a_1, a_2)$ is equivalent to $e_+ \in N_{1,X}(a_1, a_2)$, or $e_+ \in \mathbb{R}^n \backslash N_{1,X}(a_1, a_2)$ but $\det(X'\Lambda^{-1}(\hat{\rho}_X(e_+))X) = 0$, or to $e_+ \in \mathbb{R}^n \backslash N_{1,X}(a_1, a_2)$ and $\det(X'\Lambda^{-1}(\hat{\rho}_X(e_+))X) \neq 0$ but $\tilde{\sigma}_X^2(e_+)\det(R(X'\Lambda^{-1}(\hat{\rho}_X(e_+))X)^{-1}R') = 0$. The first one of these three conditions can be written as

$$\left(\sum_{t=2}^{n} \hat{u}_t(e_+)\hat{u}_{t-1}(e_+)\right)^2 = \left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+)\right)^2. \tag{32}$$

Since $\det(X'X) \neq 0$ holds for $X \in \mathfrak{X}_0$, the set of $X \in \mathfrak{X}_0$ satisfying (32) is – after multiplication of both sides of (32) by the fourth power of $\det(X'X)$ – seen to be included in the zero-set of a multivariate polynomial in the variables x_{ti}. Observing that $\det(\Lambda(\hat{\rho}_X(e_+))) \neq 0$ and $\sum_{t=a_1}^{a_2} \hat{u}_{t,X}^2(e_+) \neq 0$

for $e_+ \in \mathbb{R}^n \backslash N_{1,X}(a_1, a_2)$, the second one of the above conditions takes the equivalent form

$$
\left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+) \right)^{k(n-1)^2} \det \left(X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) X \right) = 0, \left(\sum_{t=2}^{n} \hat{u}_t(e_+) \hat{u}_{t-1}(e_+) \right)^2 \neq \left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+) \right)^2.
\tag{33}
$$

For $X \in \mathfrak{X}_0$ satisfying the inequality in (33), the left-hand side of the equation in the preceding display is easily seen to be a polynomial in the variables x_{ti} and $\hat{u}_t(e_+)$. Since $\det(X'X)\hat{u}_t(e_+)$ is polynomial in the variables x_{ti} and $\det(X'X) \neq 0$ for $X \in \mathfrak{X}_0$, we may rewrite the equation in the preceding display by multiplying it by the $4k(n-1)^2$-th power of $\det(X'X)$. The resulting equivalent equation is obviously a polynomial in the variables x_{ti}. This shows that the set of $X \in \mathfrak{X}_0$ satisfying (33) is (a subset of) the zero-set of a multivariate polynomial. Recalling that $\det(\Lambda(\hat{\rho}_X(e_+))) \neq 0$ and $\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+) \neq 0$ for $e_+ \in \mathbb{R}^n \backslash N_{1,X}(a_1, a_2)$, and that $\det\left(X'\Lambda^{-1}(\hat{\rho}_X(e_+))X \right) \neq 0$ implies $\det\left(X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) X \right) \neq 0$, the third one of the above conditions takes the equivalent form

$$
\left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+) \right)^{(n-1)^2(2k+1+q(k-1))} f(X)' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) f(X)g(X) = 0
\tag{34}
$$

subject to

$$
\left(\sum_{t=2}^{n} \hat{u}_t(e_+)\hat{u}_{t-1}(e_+) \right)^2 \neq \left(\sum_{t=a_1}^{a_2} \hat{u}_t^2(e_+) \right)^2, \det\left(X'\Lambda^{-1}(\hat{\rho}_X(e_+))X \right) \neq 0,
\tag{35}
$$

where

$$
\begin{aligned}
f(X) &= \left[\det\left(X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) X \right) I_n - X \operatorname{adj} \left(X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) X \right) X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) \right] e_+ \\
g(X) &= \det\left(R \operatorname{adj}(X' \operatorname{adj} \left(\Lambda(\hat{\rho}_X(e_+)) \right) X)R' \right).
\end{aligned}
$$

The left-hand side of the equation in (34) is a polynomial in the variables x_{ti} as well as $\hat{u}_{t,X}(e_+)$ for all $X \in \mathfrak{X}_0$ satisfying the inequality in (35). After multiplying the left-hand side of the equation in (34) by a suitable power of $\det(X'X)$, which is non-zero for $X \in \mathfrak{X}_0$, (34) can be equivalently recast as an equation that is polynomial in x_{ti}, showing that the set of $X \in \mathfrak{X}_0$ satisfying (34) and (35) is a subset of the zero-set of a multivariate polynomial. It follows that $\mathfrak{X}_{1,FGLS}(e_+)$ is a $\lambda_{\mathbb{R}^{n \times k}}$-null set provided we can show that each of the three polynomials in the variables x_{ti} mentioned before is not trivial. For this it certainly suffices to construct a matrix $X \in \mathfrak{X}_0$ such that $e_+ \notin N_{2,X}^*(a_1, a_2)$ holds: Consider first the case $n \geq 3$. Let the first column $x_{\cdot 1}^*$ of X^* be equal to $(1, 0, \ldots, 0, 1)'$, and choose the remaining columns linearly independent in the orthogonal complement of the space spanned by $x_{\cdot 1}^*$ and e_+. Then $X^* \in \mathfrak{X}_0$ holds and $\hat{u}_{X^*}(e_+) = (0, 1, 1, \ldots, 1, 1, 0)'$ and hence $\hat{\rho}_{X^*}(e_+)$ is well-defined and equals $\hat{\rho}_{YW,X^*}(e_+)$, which is always less than 1 in absolute value. Consequently, $e_+ \in \mathbb{R}^n \backslash N_{1,X^*}(a_1, a_2)$ holds. Furthermore, $\Lambda(\hat{\rho}_{X^*}(e_+))$ is then positive definite and hence $\det\left(X^{*\prime}\Lambda^{-1}(\hat{\rho}_{X^*}(e_+))X^* \right) \neq 0$ and $\det\left(R(X^{*\prime}\Lambda^{-1}(\hat{\rho}_{X^*}(e_+))X^*)^{-1}R' \right) \neq 0$ hold; also $\tilde{\sigma}_{X^*}^2(e_+) > 0$ follows from positive definiteness of $\Lambda(\hat{\rho}_{X^*}(e_+))$ and the fact that $e_+ \notin \operatorname{span}(X^*)$. But this establishes $e_+ \in \mathbb{R}^n \backslash N_{2,X^*}^*(a_1, a_2)$ in case $n \geq 3$. Next consider the case $n = 2$. Then $k = 1$ must hold. The assumption $k \leq a_2 - a_1$ entails $a_2 = n = 2$ and $a_1 = 1$, i.e., $\hat{\rho}$ must be the Yule-Walker estimator implying that $N_{2,X^*}^*(a_1, a_2) = \operatorname{span}(X^*)$. Choose X^* as an arbitrary vector linearly independent of e_+ (which is possible since $n = 2 > 1 = k$). Then $X^* \in \mathfrak{X}_0$ and

$e_+ \in \mathbb{R}^n \backslash N^*_{2,X^*}(a_1, a_2)$ are satisfied. The proof for $\mathfrak{X}_{1,FGLS}(e_-)$ is completely analogous where in case $n \geq 3$ the matrix X^* is now chosen in such a way that $x^*_{\cdot 1}$ is equal to $(-1, 0, \ldots, 0, (-1)^n)'$ and e_- takes the rôle of e_+ in the construction of the remaining columns. Next consider the set $\mathfrak{X}_{2,FGLS}(e_+)$. Observe that for $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ the relation $T_{FGLS,X}(e_+ + \mu^*_0) = C$ can equivalently be written as

$$(R\tilde{\beta}_X(e_+))'\tilde{\Omega}^{-1}_X(e_+)(R\tilde{\beta}_X(e_+)) - C = 0. \tag{36}$$

Similar arguments as above show that for $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ this equation can equivalently be stated as $p(X) = 0$ where $p(X)$ is a polynomial in the variables x_{ti}. But this shows that the set of $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ satisfying (36) is (a subset of) an algebraic set. It follows that $\mathfrak{X}_{2,FGLS}(e_+)$ is a $\lambda_{\mathbb{R}^{n \times k}}$-null set provided the polynomial p is not trivial, or in other words that there exists a matrix $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,FGLS}(e_+)$ that violates (36). But this is guaranteed by the provision in the theorem. The result for $\mathfrak{X}_{2,FGLS}(e_-)$ is proved in exactly the same manner. The remaining claims of Part 1 are now obvious.

(2) Similar arguments as in the proof of Part 1 show that $\tilde{\mathfrak{X}}_{1,FGLS}(e_-)$ and $\tilde{\mathfrak{X}}_{2,FGLS}(e_-)$ are each contained in an algebraic set. By the assumed provision it follows immediately that $\tilde{\mathfrak{X}}_{2,FGLS}(e_-)$ is a $\lambda_{\mathbb{R}^{n \times (k-1)}}$-null set. The same conclusion holds for $\tilde{\mathfrak{X}}_{1,FGLS}(e_-)$ if we can find a matrix $X^* = \left(e_+, \tilde{X}^*\right)$ such that $e_- \notin N^*_{2,X^*}(a_1, a_2)$. To this end let the $n \times 1$ vector $a = (-1, 0, \ldots, 0, (-1)^n)'$ be the first column of $\tilde{X}^*$ and choose the remaining $k - 2$ columns linearly independent in the orthogonal complement of the space spanned by e_+, e_-, and a (which is possible since $k < n$). Simple computation now shows that $\hat{u}_{X^*}(e_-) \neq 0$ (note that $n \geq 4$ has been assumed) and that the first and last entry of $\hat{u}_{X^*}(e_-)$ is zero. Consequently, $\hat{\rho}_{X^*}(e_-)$ is well-defined and equals $\hat{\rho}_{YW,X^*}(e_-)$, which is always less than 1 in absolute value, and the same argument as in the proof of Part 1 shows that $e_- \notin N^*_{2,X^*}(a_1, a_2)$ is indeed satisfied. The remaining claims of Part 2 are now obvious.

(3) First consider $\mathfrak{X}_{1,OLS}(e_+)$. The condition $e_+ \in N^*_{0,X}(a_1, a_2)$ is equivalent to $\sum^{a_2}_{t=a_1} \hat{u}^2_t(e_+) = 0$, or $\sum^{a_2}_{t=a_1} \hat{u}^2_t(e_+) \neq 0$ but $\det\left(R(X'X)^{-1}X'\Lambda(\hat{\rho}(y))X(X'X)^{-1}R'\right) = 0$. Similar arguments as in (1) then show that $\mathfrak{X}_{1,OLS}(e_+)$ is a subset of an algebraic set. The matrix X^* constructed in (1) is easily seen to satisfy $e_+ \in \mathbb{R}^n \backslash N^*_{0,X^*}(a_1, a_2)$. Thus $\mathfrak{X}_{1,OLS}(e_+)$ is a $\lambda_{\mathbb{R}^{n \times k}}$-null set. The proof for $\mathfrak{X}_{1,OLS}(e_-)$ is exactly the same. Next consider $\mathfrak{X}_{2,OLS}(e_+)$. Observe that for $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,OLS}(e_+)$ the relation $T_{OLS,X}(e_+ + \mu^*_0) = C$ can equivalently be written as

$$(R\hat{\beta}_X(e_+))'\hat{\Omega}^{-1}_X(e_+)(R\hat{\beta}_X(e_+)) - C = 0. \tag{37}$$

The same argument as in the proof of Part (1) shows that the set of $X \in \mathfrak{X}_0 \backslash \mathfrak{X}_{1,OLS}(e_+)$ satisfying (37) is (a subset of) the zero-set of a multivariate polynomial in the variables x_{ti}. It follows that $\mathfrak{X}_{2,OLS}(e_+)$ is a $\lambda_{\mathbb{R}^{n \times k}}$-null set under the maintained provision that it is a proper subset of $\mathfrak{X}_0 \backslash \mathfrak{X}_{1,OLS}(e_+)$. The proof for $\mathfrak{X}_{2,OLS}(e_-)$ is the same. The proof for $\tilde{\mathfrak{X}}_{1,OLS}(e_-)$ and $\tilde{\mathfrak{X}}_{2,OLS}(e_-)$ is similar to the proof for $\tilde{\mathfrak{X}}_{1,FGLS}(e_-)$ and $\tilde{\mathfrak{X}}_{2,FGLS}(e_-)$.

(4) Note that the assumptions obviously imply $e_+ \in \mathfrak{M}$ and $R\hat{\beta}(e_+) \neq 0$. $\blacksquare$

Remark B.4. In case $a_1 = 1$ and $a_2 = n$ the argument in the above proof simplifies due to the fact that $N^*_{2,X}(1, n) = N^*_{0,X}(1, n) = \text{span}(X)$.

Proof of Theorem 3.16: We apply Theorem 5.21. That $(\tilde{\beta}, \tilde{\Omega})$ as well as $(\hat{\beta}, \hat{\Omega})$ satisfy Assumptions 5, 6, and 7 has been shown in Lemmata B.2 and B.3. The covariance model $\mathfrak{C}_{AR(1)}$

satisfies the properties required in Theorem 5.21 as shown in Lemma G.1. Furthermore, we have $J\left(\mathfrak{C}_{AR(1)}\right) = \mathrm{span}(e_{+}) \cup \mathrm{span}(e_{-})$, see Lemma G.1, and because $e_{+}, e_{-} \in \mathfrak{M}$ is assumed we conclude that $J\left(\mathfrak{C}_{AR(1)}\right) \subseteq \mathfrak{M}$. The assumption $R\hat{\beta}(e_{+}) = R\hat{\beta}(e_{-}) = 0$ then implies that even $J\left(\mathfrak{C}_{AR(1)}\right) \subseteq \mathfrak{M}_{0} - \mu_{0}$ holds. The invariance condition (31) in Theorem 5.21 is thus satisfied, because T is $G\left(\mathfrak{M}_{0}\right)$-invariant by Lemma 5.15. We next show that the additional condition in Part 2 of Theorem 5.21 is satisfied. This is trivial in case the Yule-Walker estimator is used (i.e., if $a_{1} = 1$ and $a_{n} = n$) since then $\tilde{\Omega}(y)$ is positive definite for $y \notin N_{2}^{*}(a_{1}, a_{2})$ and $\hat{\Omega}(y)$ is positive definite for $y \notin N_{0}^{*}(a_{1}, a_{2})$ (see the proof of Lemma B.1) and since $N_{2}^{*}(a_{1}, a_{2})$ and $N_{0}^{*}(a_{1}, a_{2})$ are $\lambda_{\mathbb{R}^{n}}$-null sets by Lemma B.1. If $a_{1} \neq 1$ or $a_{n} \neq n$, then y_{0} constructed in the proof of Lemma 3.13 satisfies $y_{0} \in \mathbb{R}^{n} \backslash N_{2}^{*}(a_{1}, a_{2})$ and $y_{0} \in \mathbb{R}^{n} \backslash N_{0}^{*}(a_{1}, a_{2})$ (cf. proof of Lemma B.1) as well as $\hat{\rho}(y_{0}) = \hat{\rho}_{YW}(y_{0})$, implying that $\tilde{\Omega}(y_{0})$ as well as $\hat{\Omega}(y_{0})$ are positive definite. As shown in Lemma B.1, the matrix $\tilde{\Omega}$ is, in particular, continuous on the open set $\mathbb{R}^{n} \backslash N_{2}^{*}(a_{1}, a_{2})$ and the matrix $\hat{\Omega}$ is continuous on the open set $\mathbb{R}^{n} \backslash N_{0}^{*}(a_{1}, a_{2})$. Consequently, $\tilde{\Omega}$ and $\hat{\Omega}$ are positive definite in a neighborhood of y_{0} and thus the additional condition in Part 2 of Theorem 5.21 is satisfied. Finally, the condition $a_{1} = 1$ and $a_{2} = n$ implies that $\tilde{\Omega}$ and $\hat{\Omega}$ are $\lambda_{\mathbb{R}^{n}}$-almost everywhere positive definite (since then $\hat{\rho} = \hat{\rho}_{YW}$), verifying the extra condition in Part 3 of Theorem 5.21. ∎

C   Appendix: Proofs for Section 4

Proof of Theorem 4.2: First observe that $\hat{\beta}$ and $\hat{\Omega}_{Het}$ satisfy Assumptions 5 and 6 with $N = \emptyset$. In fact, $\hat{\Omega}_{Het}(y)$ is nonnegative definite for every $y \in \mathbb{R}^{n}$, and is positive definite $\lambda_{\mathbb{R}^{n}}$-almost everywhere under Assumption 3 by Lemma 4.1. Furthermore, in view of this lemma and because $N = \emptyset$, the set N^{*} in Corollary 5.17 is precisely the set of y for which rank $(B(y)) < q$. It is trivial that $\mathcal{Z}_{i} = \mathrm{span}(e_{i}(n))$ is a concentration space of $\mathfrak{C}$ for every $i = 1, \ldots, n$. The theorem now follows by applying Corollary 5.17 and Remark 5.18(i) to $\mathcal{Z}_{i}$ and by noting that $e_{i}(n) \in \mathbb{R}^{n} \backslash N^{*}$ translates into rank $(B(e_{i}(n))) = q$. Also note that the size of the test can not be zero in view of Part 5 of Lemma 5.15. ∎

D   Appendix: Proofs for Subsection 5.1

Proof of Proposition 5.2: Since

$$\begin{aligned}
\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}\left(g_{\alpha,\nu,\nu'}(y) - \nu_{*}\right) &= \alpha\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y-\nu) + \Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(\nu'-\nu_{*}) \\
&= \alpha\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y-\nu) = \alpha\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y-\nu_{*}),
\end{aligned}$$

invariance of h follows, and hence h is constant on the orbits of $G(\mathfrak{M})$. Now suppose that $h(y) = h(y')$. If $h(y) = h(y') = 0$ holds, it follows that

$$\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y-\nu_{*}) = \Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y'-\nu_{*}) = 0.$$

Consequently, $y' - y$ is of the form $\nu^{*} - \nu_{*}$ for some $\nu^{*} \in \mathfrak{M}$. But this gives $y' = (y - \nu_{*}) + \nu^{*} = g_{1,\nu_{*},\nu^{*}}(y)$, showing that y' is in the same orbit as y. Next consider the case where $h(y) = h(y') \neq 0$. Then

$$\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}\left(\left\|\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y'-\nu_{*})\right\|(y-\nu_{*}) - c\left\|\Pi_{(\mathfrak{M}-\nu_{*})^{\perp}}(y-\nu_{*})\right\|(y'-\nu_{*})\right) = 0$$

where $c = \pm 1$. It follows that the argument inside the projection operator is of the form $\nu^* - \nu_*$ for some $\nu^* \in \mathfrak{N}$. Elementary calculations give

$$y' = \frac{\left\| \Pi_{(\mathfrak{N} - \nu_*)^\perp} (y' - \nu_*) \right\|}{c \left\| \Pi_{(\mathfrak{N} - \nu_*)^\perp} (y - \nu_*) \right\|} (y - \nu_*) + \left(\nu_* + \frac{1}{c \left\| \Pi_{(\mathfrak{N} - \nu_*)^\perp} (y - \nu_*) \right\|} (\nu_* - \nu^*) \right).$$

Since the last term in parenthesis on the right-hand side above is obviously an element of $\mathfrak{N}$, we have obtained $y' = g(y)$ for some $g \in G(\mathfrak{N})$, i.e., y' is in the same orbit as y. This shows that h is a maximal invariant. ∎

Proof of Proposition 5.4: (1) From (11) and its extension discussed subsequently to (11), as well as from the transformation theorem for integrals we obtain

$$E_{\mu, \sigma^2 \Phi} \left(\varphi(y) \right) = E_{\alpha(\mu - \mu_0) + \mu_0', \alpha^2 \sigma^2 \Phi} \left(\varphi \left(g_{\alpha, \mu_0, \mu_0'}^{-1} (y) \right) \right).$$

By almost invariance of φ we have that $\varphi(y) = \varphi \left(g_{\alpha, \mu_0, \mu_0'}^{-1} (y) \right)$ for all $y \in \mathbb{R}^n \backslash N$ with $\lambda_{\mathbb{R}^n}(N) = 0$ (where N may depend on $g_{\alpha, \mu_0, \mu_0'}^{-1}$). Since Φ is positive definite, also $P_{\alpha(\mu - \mu_0) + \mu_0', \alpha^2 \sigma^2 \Phi}(N) = 0$ holds, and thus the right-hand side of the above display equals $E_{\alpha(\mu - \mu_0) + \mu_0', \alpha^2 \sigma^2 \Phi} \left(\varphi(y) \right)$ which proves the first claim. The claim in parenthesis follows similarly observing that for invariant φ the exceptional set is empty.

(2) Setting $\alpha = 1$ in (12) shows that the rejection probability is invariant under addition of elements that belong to $\mathfrak{M}_0 - \mu_0$. Since $\mu = \Pi_{(\mathfrak{M}_0 - \mu_0)}(\mu - \mu_0) + \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mu - \mu_0) + \mu_0$ we thus conclude that $E_{\mu, \sigma^2 \Phi}(\varphi) = E_{\nu + \mu_0, \sigma^2 \Phi}(\varphi)$ where $\nu = \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mu - \mu_0) \in \mathfrak{M}$. Now applying (12) with $\alpha = \sigma^{-1}$ and $\mu_0' = \mu_0$ to $E_{\nu + \mu_0, \sigma^2 \Phi}(\varphi)$ establishes the first equality in (13). The second equality follows by the same argument by setting $\alpha = \pm \sigma^{-1}$, the sign equaling the sign of the first non-zero component of ν if $\nu \neq 0$, and the choice of sign being irrelevant if $\nu = 0$.

(3) The first claim is an immediate consequence of (13). For the second claim it suffices to show that $\mu - X \hat{\beta}_{rest}(\mu)$ (for $\mu \in \mathfrak{M}$) is an injective linear function of $R\beta - r$, bijectivity of this mapping following from dimension considerations. To this end note that

$$
\begin{aligned}
\mu - X \hat{\beta}_{rest}(\mu) &= X\beta - X \left[\hat{\beta}(\mu) - (X'X)^{-1} R' \left(R (X'X)^{-1} R' \right)^{-1} \left(R \hat{\beta}(\mu) - r \right) \right] \\
&= X\beta - X \left[\beta - (X'X)^{-1} R' \left(R (X'X)^{-1} R' \right)^{-1} (R\beta - r) \right] \\
&= (X'X)^{-1} R' \left(R (X'X)^{-1} R' \right)^{-1} (R\beta - r)
\end{aligned}
$$

and that the matrix premultiplying $R\beta - r$ is of full column rank q. ∎

Proof of Proposition 5.6: Set $\bar{h}(\mu, \sigma^2) = \left\langle \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp} (\mu - \mu_0) / \sigma \right\rangle$. The invariance of $\left(\bar{h}(\mu, \sigma^2), \Sigma \right)$ follows from a simple computation similar to the one in the proof of Proposition 5.2. Now assume that $\left(\bar{h}(\mu, \sigma^2), \Sigma \right) = \left(\bar{h}(\mu', \sigma'^2), \Sigma' \right)$. We immediately get $\bar{h}(\mu, \sigma^2) = \bar{h}(\mu', \sigma'^2)$ and $\Sigma = \Sigma'$. The former implies

$$\Pi_{(\mathfrak{M}_0 - \mu_0)^\perp} \left((\mu - \mu_0) - c(\sigma/\sigma')(\mu' - \mu_0) \right) = 0$$

where $c = \pm 1$. Similar calculations as in the proof of Proposition 5.2 give

$$\mu' = c(\sigma'/\sigma)(\mu - \mu^*) + \mu_0$$

57

for some $\mu^* \in \mathfrak{M}_0$. Together with $\Sigma = \Sigma'$ this shows that $(\mu', \sigma'^2, \Sigma')$ is in the same orbit under the associated group as is (μ, σ^2, Σ). ∎

E   Appendix: Proofs and Auxiliary Results for Subsections 5.2 and 5.3

The next lemma is a simple consequence of a continuity property of the characteristic function of a multivariate Gaussian probability measure and of the Portmanteau theorem.

Lemma E.1. *Let Φ_m be a sequence of nonnegative definite symmetric $n \times n$ matrices converging to an $n \times n$ matrix Φ as $m \to \infty$, where Φ may be singular, and let $\mu_m \in \mathbb{R}^n$ be a sequence converging to $\mu \in \mathbb{R}^n$ as $m \to \infty$. Then P_{μ_m, Φ_m} converges weakly to $P_{\mu, \Phi}$. If, in addition, $A \in \mathcal{B}(\mathbb{R}^n)$ satisfies $\lambda_{\mu + \mathrm{span}(\Phi)}(\mathrm{bd}(A)) = 0$, then $P_{\mu_m, \Phi_m}(A) \to P_{\mu, \Phi}(A)$.*

Proof of Theorem 5.7: (1) Since $\mathcal{Z}$ is a concentration space of $\mathfrak{C}$, there exists a sequence $(\Sigma_m)_{m \in \mathbb{N}}$ in $\mathfrak{C}$ converging to $\bar{\Sigma}$ such that $\mathrm{span}(\bar{\Sigma}) = \mathcal{Z}$. Note that $\mu_0 + \mathcal{Z}$ is a $\lambda_{\mathbb{R}^n}$-null set because $\dim(\mathcal{Z}) < n$ in view of Definition 2.1. Because Σ_m is positive definite, we thus have

$$P_{\mu_0, \sigma^2 \Sigma_m}(W) = P_{\mu_0, \sigma^2 \Sigma_m}(W \cup (\mu_0 + \mathcal{Z})).$$

By Lemma E.1 we then have that $P_{\mu_0, \sigma^2 \Sigma_m}(W \cup (\mu_0 + \mathcal{Z}))$ converges to $P_{\mu_0, \sigma^2 \bar{\Sigma}}(W \cup (\mu_0 + \mathcal{Z}))$. But the later probability is not less than $P_{\mu_0, \sigma^2 \bar{\Sigma}}(\mu_0 + \mathcal{Z})$ which equals 1 since $P_{\mu_0, \sigma^2 \bar{\Sigma}}$ is supported by $\mu_0 + \mathcal{Z}$. To prove the claim in parentheses observe that $T(\mu_0 + z) > C$ and lower semicontinuity of T at $\mu_0 + z$ implies that $T(w) > C$ holds for all w in a neighborhood of $\mu_0 + z$; hence such points $\mu_0 + z$ belong to $\mathrm{int}(W) \subseteq \mathrm{int}(W \cup (\mu_0 + \mathcal{Z}))$, and consequently do not belong to $\mathrm{bd}(W \cup (\mu_0 + \mathcal{Z}))$. But this establishes (15).

(2) Apply the same argument as above to $\mathbb{R}^n \backslash W$. Also note that $P_{\mu_0, \sigma^2 \Sigma}(W)$ can be approximated arbitrarily closely by $P_{\mu_1, \sigma^2 \Sigma}(W)$ for suitable $\mu_1 \in \mathfrak{M}_1$, since $\|P_{\mu_0, \sigma^2 \Sigma} - P_{\mu_1, \sigma^2 \Sigma}\|_{TV} \to 0$ for $\mu_1 \to \mu_0$ holds by Scheffé's Lemma as $\sigma^2 \Sigma$ is positive definite.

(3) Choose an arbitrary $\mu_1 \in \mathfrak{M}_1$. By assumption we have $\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2 \Sigma}(W) = 0$ for a suitable $\sigma^2 > 0$. It hence suffices to show that for every $\Sigma \in \mathfrak{C}$

$$P_{\mu_0, \sigma^2 \Sigma}(W) - P_{\mu_1, \sigma^2 \tau^2 \Sigma}(W) \to 0$$

holds for $\tau \to \infty$. By almost invariance of W under $G(\{\mu_0\})$ we have that $W \triangle (\tau W + (1 - \tau)\mu_0)$ is a $\lambda_{\mathbb{R}^n}$-null set. Hence, by the reproductive property of the normal distribution

$$P_{\mu_1, \sigma^2 \tau^2 \Sigma}(W) = P_{\mu_1, \sigma^2 \tau^2 \Sigma}(\tau W + (1 - \tau)\mu_0) = P_{\mu_0 + \tau^{-1}(\mu_1 - \mu_0), \sigma^2 \Sigma}(W).$$

But, since $\sigma^2 \Sigma$ is positive definite, we have by an application of Scheffé's Lemma

$$\|P_{\mu_0, \sigma^2 \Sigma} - P_{\mu_0 + \tau^{-1}(\mu_1 - \mu_0), \sigma^2 \Sigma}\|_{TV} \to 0$$

as $\tau \to \infty$, and hence $P_{\mu_0, \sigma^2 \Sigma}(W) - P_{\mu_0 + \tau^{-1}(\mu_1 - \mu_0), \sigma^2 \Sigma}(W) \to 0$. The claim in parenthesis is obvious. ∎

58

Lemma E.2. *Let $\varphi : \mathbb{R}^n \to [0,1]$ be a Borel-measurable function that is almost invariant under $G(\mathfrak{M}_0)$. Suppose Φ_m is a sequence of positive definite symmetric $n \times n$ matrices converging to a positive definite matrix Φ, suppose $\mu_m \in \mathfrak{M}$, and suppose the sequence σ_m^2 satisfies $0 < \sigma_m^2 < \infty$. Then*

$$\lim_{m \to \infty} E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi) = E_{\nu + \mu_0, \Phi}(\varphi)$$

provided $\nu_m^* = \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mu_m - \mu_0)/\sigma_m$ for some $\mu_0 \in \mathfrak{M}_0$ converges to an element $\nu \in \mathbb{R}^n$ (which then necessarily belongs to $\mathfrak{M}$). [Note that ν_m^*, and thus the result, does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$.]

Proof. By Proposition 5.4 we have that $E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi) = E_{\nu_m^* + \mu_0, \Phi_m}(\varphi)$. Since $\nu_m^* \to \nu$ and since $\Phi_m \to \Phi$, with Φ positive definite, the result follows from total variation distance convergence of $P_{\nu_m^* + \mu_0, \Phi_m}$ to $P_{\nu + \mu_0, \Phi}$. ∎

Remark E.3. (i) Consider the case where $\nu_m^* = \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mu_m - \mu_0)/\sigma_m$ does not converge. Then, as long as the sequence ν_m^* is bounded, the above lemma can be applied by passing to subsequences along which ν_m^* converges. In the case where the sequence ν_m^* is unbounded, then, along subsequences such that the norm of ν_m^* diverges, one would expect $E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi) = E_{\nu_m^* + \mu_0, \Phi_m}(\varphi)$ to converge to 1 for any reasonable test since $\nu_m^* + \mu_0$ moves farther and farther away from $\mathfrak{M}_0$ (and Φ_m stabilizes at a positive definite matrix). Indeed, such a result can be shown for a large class of tests, see Lemma 5.15.

(ii) In the special case where $\mu_m \equiv \mu$ it is easy to see, using Proposition 5.4, that the limit in the above lemma is $E_{\mu, \sigma^2 \Phi}(\varphi)$ if $\sigma_m^2 \to \sigma^2 \in (0, \infty)$ and $\mu \in \mathfrak{M}_1$, is $E_{\mu_0, \Phi}(\varphi)$ if $\sigma_m^2 \to \infty$ and $\mu \in \mathfrak{M}_1$, and is $E_{\mu, \Phi}(\varphi)$ if $\mu \in \mathfrak{M}_0$.

Lemma E.4. *Let $\varphi : \mathbb{R}^n \to [0,1]$ be a Borel-measurable function that is almost invariant under $G(\mathfrak{M}_0)$. Suppose Φ_m is a sequence of positive definite symmetric $n \times n$ matrices converging to a singular matrix Φ, suppose $\mu_m \in \mathfrak{M}$, and σ_m^2 is a sequence satisfying $0 < \sigma_m^2 < \infty$. Assume further that $\varphi(x + z) = \varphi(x)$ holds for every $x \in \mathbb{R}^n$ and every $z \in \operatorname{span}(\Phi)$. Suppose that for some sequence of positive real numbers s_m the matrix $D_m = \Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}/s_m$ converges to a matrix D, which is regular on the orthogonal complement of $\operatorname{span}(\Phi)$. Then*

$$\lim_{m \to \infty} E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi) = E_{\nu + \mu_0, D + \Phi}(\varphi) = E_{\nu + \mu_0, D}(\varphi)$$

provided $\nu_m^{**} = \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mu_m - \mu_0)/\left(\sigma_m s_m^{1/2}\right)$ for some $\mu_0 \in \mathfrak{M}_0$ converges to an element $\nu \in \mathbb{R}^n$ (which then necessarily belongs to $\mathfrak{M}$). [Note that ν_m^{**}, and thus the result, does not depend on the choice of $\mu_0 \in \mathfrak{M}_0$.] Furthermore, the matrix $D + \Phi$ is positive definite.

Proof. Because $\Pi_{\operatorname{span}(\Phi)}(x - \mu_m) \in \operatorname{span}(\Phi)$, we obtain by the assumed invariance w.r.t. addition of $z \in \operatorname{span}(\Phi)$

$$\varphi(x) = \varphi(\mu_m + \Pi_{\operatorname{span}(\Phi)^\perp}(x - \mu_m) + \Pi_{\operatorname{span}(\Phi)}(x - \mu_m)) = \varphi(\mu_m + \Pi_{\operatorname{span}(\Phi)^\perp}(x - \mu_m))$$

for every x. By the transformation theorem we then have on the one hand

$$
\begin{aligned}
E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi(\cdot)) &= E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi(\mu_m + \Pi_{\operatorname{span}(\Phi)^\perp}(\cdot - \mu_m))) = E_{\mu_m, \sigma_m^2 \Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}}(\varphi(\cdot)) \\
&= E_{\mu_m, \sigma_m^2 s_m D_m}(\varphi(\cdot)).
\end{aligned}
\tag{38}
$$

On the other hand, by the same invariance property of φ

$$E_{\mu_m, \sigma_m^2 s_m D_m}(\varphi(\cdot)) = E_{\mu_m, \sigma_m^2 s_m D_m}(\varphi(\cdot + z))$$

holds for every $z \in \operatorname{span}(\Phi)$. Integrating this w.r.t. a normal distribution $P_{0, \sigma_m^2 s_m \Phi}$ (in the variable z) and using the reproductive property of the normal distribution gives

$$E_{\mu_m, \sigma_m^2 s_m D_m}(\varphi(x)) = E_{Q_m}(\varphi(x + z)) = E_{\mu_m, \sigma_m^2 s_m (D_m + \Phi)}(\varphi(x)) \tag{39}$$

where Q_m denotes the product of the normal distributions $P_{\mu_m, \sigma_m^2 s_m D_m}$ and $P_{0, \sigma_m^2 s_m \Phi}$. Observe that $D + \Phi$ as well as $D_m + \Phi$ are positive definite. An application of Lemma E.2 now gives

$$\lim_{m \to \infty} E_{\mu_m, \sigma_m^2 s_m (D_m + \Phi)}(\varphi) = E_{\nu + \mu_0, D + \Phi}(\varphi).$$

The same argument that has led to (39) now shows that $E_{\nu + \mu_0, D + \Phi}(\varphi) = E_{\nu + \mu_0, D}(\varphi)$. Combining this with (38) completes the proof of the display in the theorem. The positive definiteness of $D + \Phi$ is obvious as noted earlier in the proof. ∎

Remark E.5. (i) A remark similar to Remark E.3(i) also applies here. In particular, we typically can expect $E_{\mu_m, \sigma_m^2 \Phi_m}(\varphi)$ to converge to 1 in case the norm of ν_m^{**} diverges.

(ii) In the special case where $\mu_m \equiv \mu$ it is easy to see, using Proposition 5.4, that the limit in the above lemma is $E_{\mu, \kappa(D + \Phi)}(\varphi) = E_{\mu, \kappa D}(\varphi)$ if $\sigma_m^2 s_m \to \kappa \in (0, \infty)$ and $\mu \in \mathfrak{M}_1$, is $E_{\mu_0, (D + \Phi)}(\varphi) = E_{\mu_0, D}(\varphi)$ if $\sigma_m^2 s_m \to \infty$ and $\mu \in \mathfrak{M}_1$, and is $E_{\mu, (D + \Phi)}(\varphi) = E_{\mu, D}(\varphi)$ if $\mu \in \mathfrak{M}_0$.

Remark E.6. (i) If s_m and s_m^* are two positive scaling factors such that $\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}/s_m \to D$ and $\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}/s_m^* \to D^*$ with both D and D^* being regular on the orthogonal complement of $\operatorname{span}(\Phi)$, then s_m/s_m^* must converge to a positive finite number, i.e., the scaling sequence is essentially uniquely determined.

(ii) Typical choices for s_m are $s_m^{(1)} = \left\|\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}\right\|$ (for some choice of norm) or $s_m^{(2)} = \operatorname{tr}(\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp})$; note that $s_m^{(1)}$ as well as $s_m^{(2)}$ are positive, since Φ_m is positive definite and Φ is singular. With both choices convergence of $\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}/s_m$ (at least along suitable subsequences) is automatic. Furthermore, since for any choice of norm we have $c_1 \left\|\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}\right\| \leq \operatorname{tr}(\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}) \leq c_2 \left\|\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}\right\|$ for suitable $0 < c_1 \leq c_2 < \infty$, we have convergence of $s_m^{(1)}/s_m^{(2)}$ to a positive finite number (at least along suitable subsequences). Hence, which of the normalization factors $s_m^{(i)}$ is used in an application of the above lemma, typically does not make a difference.

Proof of Theorem 5.10: (1) By the invariance properties of the rejection probability expressed in Proposition 5.4 it suffices to show for an arbitrary fixed $\mu_0 \in \mathfrak{M}_0$ that

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \Sigma}(\varphi) < 1$$

in order to establish the first claim in Part 1. To this end let $\Sigma_m \in \mathfrak{C}$ be a sequence such that $E_{\mu_0, \Sigma_m}(\varphi)$ converges to $\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \Sigma}(\varphi)$. Since $\mathfrak{C}$ is assumed to be bounded, we may assume without loss of generality that Σ_m converges to a matrix $\bar{\Sigma}$ (not necessarily in $\mathfrak{C}$). If $\bar{\Sigma}$ is positive definite, it follows from Lemma E.2 applied to $E_{\mu_0, \Sigma_m}(\varphi)$ that $\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \Sigma}(\varphi) = E_{\mu_0, \bar{\Sigma}}(\varphi)$ (since $\nu = 0$). But $E_{\mu_0, \bar{\Sigma}}(\varphi)$ is less than 1 since $\varphi \leq 1$ is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 1. If $\bar{\Sigma}$ is

singular, then in view of the assumptions of the theorem we can pass to the subsequence Σ_{m_i} and then apply Lemma E.4 to $E_{\mu_0, \Sigma_{m_i}}(\varphi)$ to obtain that $\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0, \Sigma}(\varphi) = E_{\mu_0, D+\bar{\Sigma}}(\varphi)$ (since again $\nu = 0$) for a matrix D with the properties as given in the theorem. But $E_{\mu_0, D+\bar{\Sigma}}(\varphi)$ is less than 1, since $D + \bar{\Sigma}$ is positive definite (as noted in Lemma E.4) and since $\varphi \leq 1$ is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 1. This proves the first claim of Part 1 of the theorem. To prove the second claim in Part 1, observe that for the same invariance reasons it suffices to show that for an arbitrary fixed $\mu_0 \in \mathfrak{M}_0$

$$\inf_{\Sigma \in \mathfrak{C}} E_{\mu_0, \Sigma}(\varphi) > 0$$

holds. Now the same argument as before shows that this infimum either equals $E_{\mu_0, \bar{\Sigma}}(\varphi)$ for some positive definite $\bar{\Sigma}$, or equals $E_{\mu_0, D+\bar{\Sigma}}(\varphi)$ for some positive definite $D + \bar{\Sigma}$. Since $\varphi \geq 0$, but φ is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 0 by assumption, the result follows.

(2) Let $\mu_m \in \mathfrak{M}_1$, $0 < \sigma_m^2 < \infty$, and $\Sigma_m \in \mathfrak{C}$ be sequences such that $E_{\mu_m, \sigma_m^2 \Sigma_m}(\varphi)$ converges to $\inf_{\mu_1 \in \mathfrak{M}_1} \inf_{\sigma^2 > 0} \inf_{\Sigma \in \mathfrak{C}} E_{\mu_1, \sigma^2 \Sigma}(\varphi)$. Since $\mathfrak{C}$ is assumed to be bounded, we may assume without loss of generality that Σ_m converges to a matrix $\bar{\Sigma}$.

Consider first the case where $\bar{\Sigma}$ is positive definite: Set $\nu_m^* = \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu_m - \mu_0)/\sigma_m$. If this sequence is bounded, we may pass to a subsequence m' such that $\nu_{m'}^*$ converges to some ν. Applying Lemma E.2 then shows that $E_{\mu_{m'}, \sigma_{m'}^2 \Sigma_{m'}}(\varphi)$ converges to $E_{\nu+\mu_0, \bar{\Sigma}}(\varphi)$, which is positive since $\varphi \geq 0$ is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 0 and since $\bar{\Sigma}$ is positive definite. If the sequence ν_m^* is unbounded, we may pass to a subsequence m' such that $\|\nu_{m'}^*\| \to \infty$. Since $E_{\mu_{m'}, \sigma_{m'}^2 \Sigma_{m'}}(\varphi) = E_{\nu_{m'}^*+\mu_0, \Sigma_{m'}}(\varphi)$ by Proposition 5.4, it follows from assumption (19) that $\lim_{m'} E_{\mu_{m'}, \sigma_{m'}^2 \Sigma_{m'}}(\varphi)$ is positive.

Next consider the case where $\bar{\Sigma}$ is singular: Pass to the subsequence m_i mentioned in the theorem and set now $\nu_{m_i}^{**} = \Pi_{(\mathfrak{M}_0-\mu_0)^\perp}(\mu_{m_i} - \mu_0)/\left(\sigma_{m_i} s_{m_i}^{1/2}\right)$. If this sequence is bounded, we may pass to a subsequence m_i' of m_i such that $\nu_{m_i'}^{**}$ converges to some ν. Applying Lemma E.4 then shows that $E_{\mu_{m_i'}, \sigma_{m_i'}^2 \Sigma_{m_i'}}(\varphi)$ converges to $E_{\nu+\mu_0, D+\bar{\Sigma}}(\varphi)$, which is positive since $\varphi \geq 0$ is not $\lambda_{\mathbb{R}^n}$-almost everywhere equal to 0 and since $D + \bar{\Sigma}$ is positive definite. If the sequence $\nu_{m_i}^{**}$ is unbounded, we may pass to a subsequence m_i' of m_i such that $\left\|\nu_{m_i'}^{**}\right\| \to \infty$. Since

$$E_{\mu_{m_i'}, \sigma_{m_i'}^2 \Sigma_{m_i'}}(\varphi) = E_{\mu_{m_i'}, \sigma_{m_i'}^2 s_{m_i'}\left(D_{m_i'}+\bar{\Sigma}\right)}(\varphi) = E_{\nu_{m_i'}^{**}+\mu_0, D_{m_i'}+\bar{\Sigma}}(\varphi)$$

by (38), (39), and Proposition 5.4, it follows from assumption (19) and positive definiteness of $D+\bar{\Sigma}$ that $\lim_{i \to \infty} E_{\mu_{m_i'}, \sigma_{m_i'}^2 \Sigma_{m_i'}}(\varphi)$ is positive. Taken together the preceding arguments establish Part 2 of the theorem.

(3) To prove the first claim of Part 3 of the theorem observe that we can find $\mu_m \in \mathfrak{M}_1$ and σ_m^2 with $0 < \sigma_m^2 < \infty$ with $d(\mu_m, \mathfrak{M}_0)/\sigma_m \geq c$ such that the expression left of the arrow in (20) differs from $E_{\mu_m, \sigma_m^2 \Sigma_m}(\varphi)$ only by a sequence converging to zero. Let m' denote an arbitrary subsequence. We can then find a further subsequence m_i' such that the corresponding matrix $D_{m_i'}$ satisfies the assumptions of the theorem. Note that the sequence $s_{m_i'}$ corresponding to $D_{m_i'}$ necessarily converges to zero. But then the norm of $\nu_{m_i'}^{**}$ defined above must diverge since $d\left(\mu_{m_i'}, \mathfrak{M}_0\right)/\sigma_m \geq c$ and since $\Pi_{(\mathfrak{M}_0-\mu_0)^\perp}$ is the projection onto the orthogonal complement of $\mathfrak{M}_0 - \mu_0$. Because

$$E_{\mu_{m_i'}, \sigma_{m_i'}^2 \Sigma_{m_i'}}(\varphi) = E_{\nu_{m_i'}^{**}+\mu_0, D_{m_i'}+\bar{\Sigma}}(\varphi)$$

61

in view of (38), (39), and Proposition 5.4, the result then follows from the assumption that the limit inferior in (19) is equal to 1, noting that $D_{m'_i} + \bar{\Sigma}$ is positive definite and converges to the positive definite matrix $D + \bar{\Sigma}$.

We next prove the second claim in Part 3. Choose $\mu_m \in \mathfrak{M}_1$ with $d(\mu_m, \mathfrak{M}_0) \geq c_m$ such that the expression to the left of the arrow in (21) differs from $E_{\mu_m, \sigma_m^2 \Sigma_m}(\varphi)$ only by a sequence converging to zero. Since

$$E_{\mu_m, \sigma_m^2 \Sigma_m}(\varphi) = E_{\nu_m^* + \mu_0, \Sigma_m}(\varphi)$$

by Proposition 5.4 where ν_m^* was defined above and since $\|\nu_m^*\| \geq c_m/\sigma_m \to \infty$ clearly holds, the result follows from the assumption that the limit inferior in (19) is equal to 1. [Note that we have not made use of condition (18) and the condition on $\mathfrak{C}$ following (18).] ∎

Proof of Theorem 5.12: By invariance properties of the rejection probability (cf. Proposition 5.4) it suffices to show for the particular $\mu_0^* \in \mathfrak{M}_0$ appearing in (23) that for every δ, $0 < \delta < 1$, there exists $k_0 = k_0(\delta)$ such that

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_{k_0}) \leq \delta. \tag{40}$$

For this it suffices to show that $\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_k)$ converges to zero for $k \to \infty$. Let $\Sigma_k \in \mathfrak{C}$ be a sequence such that for all $k \geq 1$

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_k) \leq E_{\mu_0^*, \Sigma_k}(\varphi_k) + k^{-1}. \tag{41}$$

Since $\mathfrak{C}$ is assumed to be bounded, we can find for every subsequence k^* a further subsubsequence k' such that $\Sigma_{k'}$ converges to a matrix $\bar{\Sigma}$ (not necessarily in $\mathfrak{C}$). Let $\varepsilon > 0$ be given. We then distinguish two cases:

Case 1: $\bar{\Sigma}$ is positive definite. By (23) we can then find a k'_0 in the subsequence such that

$$E_{\mu_0^*, \bar{\Sigma}}(\varphi_{k'_0}) < \varepsilon/2$$

holds. But then by (41) and by the monotonicity expressed in (23)

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_{k'}) \leq E_{\mu_0^*, \Sigma_{k'}}(\varphi_{k'}) + k'^{-1} \leq E_{\mu_0^*, \Sigma_{k'}}(\varphi_{k'_0}) + k'^{-1} \tag{42}$$

holds for all $k' \geq k'_0$. Now Lemma E.2 together with Remark E.3(ii) may clearly be applied to the subsequence k', showing that $E_{\mu_0^*, \Sigma_{k'}}(\varphi_{k'_0})$ converges to $E_{\mu_0^*, \bar{\Sigma}}(\varphi_{k'_0}) < \varepsilon/2$. But this shows that

$$\limsup_{k' \to \infty} \sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_{k'}) < \varepsilon. \tag{43}$$

Case 2: $\bar{\Sigma}$ is singular. Then we can find a subsequence k'_i of k' and normalization constants $s_{k'_i}$ such that the resulting matrices $D_{k'_i}$ converge to a matrix D with the properties specified in Theorem 5.10. Because $D + \bar{\Sigma}$ is positive definite, we can in view of (23) find a $k'_{i(0)}$ in the subsequence k'_i such that

$$E_{\mu_0^*, D + \bar{\Sigma}}(\varphi_{k'_{i(0)}}) < \varepsilon/2.$$

Now applying Lemma E.4 together with Remark E.5(ii) to the subsequence k'_i shows that $E_{\mu_0^*, \Sigma_{k'_i}}(\varphi_{k'_{i(0)}})$ converges to $E_{\mu_0^*, D + \bar{\Sigma}}(\varphi_{k'_{i(0)}}) < \varepsilon/2$. But by (41) and (23)

$$\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_{k'_i}) \leq E_{\mu_0^*, \Sigma_{k'_i}}(\varphi_{k'_i}) + k'^{-1}_i \leq E_{\mu_0^*, \Sigma_{k'_i}}(\varphi_{k'_{i(0)}}) + k'^{-1}_i$$

holds for all $i \geq i(0)$. This shows that

$$\limsup_{i \to \infty} \sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_{k_i'}) < \varepsilon.$$

Taken together we have shown that $\sup_{\Sigma \in \mathfrak{C}} E_{\mu_0^*, \Sigma}(\varphi_k)$ must converge to zero along the original sequence k which proves (40). $\blacksquare$

F Appendix: Proofs and Auxiliary Results for Subsection 5.4

Lemma F.1. *Suppose Assumption 5 holds. Then the sets*

$$A_1 = \big\{ y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) = 0 \big\} \quad and \quad A_2 = \big\{ y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) \neq 0 \big\}$$

are invariant under $G(\mathfrak{M})$, the former set being closed in the relative topology on $\mathbb{R}^n \backslash N$. The set

$$N^* = N \cup \big\{ y \in \mathbb{R}^n \backslash N : \det \check{\Omega}(y) = 0 \big\}$$

is a closed $\lambda_{\mathbb{R}^n}$-null set in $\mathbb{R}^n$ that is invariant under $G(\mathfrak{M})$.

Proof. The invariance of A_1 and A_2 follows immediately from the invariance of $\mathbb{R}^n \backslash N$ and the equivariance of $\check{\Omega}(y)$. The relative closedness of A_1 is an immediate consequence of the continuity of $\check{\Omega}(y)$ on $\mathbb{R}^n \backslash N$. The invariance of N^* follows from invariance of N discussed after Assumption 5 and the just established invariance of A_1. Because N is a $\lambda_{\mathbb{R}^n}$-null set and because $\check{\Omega}(y)$ is $\lambda_{\mathbb{R}^n}$-almost surely nonsingular on $\mathbb{R}^n \backslash N$, it follows that N^* is a $\lambda_{\mathbb{R}^n}$-null set. Finally, we establish closedness of N^*: let $y_i \in N^*$ be a sequence with limit y_0. If $y_0 \in N$, we are done. If $y_0 \in \mathbb{R}^n \backslash N$, by openness of this set also $y_i \in \mathbb{R}^n \backslash N$ for all but finitely many i must hold and thus $\det \check{\Omega}(y_i) = 0$. But then continuity of $\check{\Omega}$ on $\mathbb{R}^n \backslash N$ implies $\det \check{\Omega}(y_0) = 0$, and hence $y_0 \in N^*$. $\blacksquare$

Proof of Lemma 5.15: (1) Follows from the discussion preceding the lemma and Lemma F.1.

(2) Follows immediately from the observation that T coincides on the open set $\mathbb{R}^n \backslash N^*$ with $(R\check{\beta}(y) - r)' \check{\Omega}^{-1}(y)(R\check{\beta}(y) - r)$ which is continuous on this set by Assumption 5.

(3) Since N^* is invariant under the elements of $G(\mathfrak{M})$, it is in particular invariant under $G(\mathfrak{M}_0)$. The result $T(g(y)) = T(y) = 0$ for $g \in G(\mathfrak{M}_0)$ then follows trivially for $y \in N^*$. Now suppose $y \in \mathbb{R}^n \backslash N^*$. Then also $g_{\alpha, \mu_0^{(1)}, \mu_0^{(2)}}(y) \in \mathbb{R}^n \backslash N^*$ for $\alpha \neq 0$, $\mu_0^{(i)} \in \mathfrak{M}_0$ $(i = 1, 2)$ by invariance of $\mathbb{R}^n \backslash N^*$. The invariance of T then follows immediately from the equivariance properties of $\check{\beta}$ and $\check{\Omega}$ expressed in Assumption 5, using that $\mu_0^{(i)} \in \mathfrak{M}_0$ implies $R\gamma^{(i)} = r$ for uniquely defined vectors $\gamma^{(i)}$ satisfying $\mu_0^{(i)} = X\gamma^{(i)}$.

(4) Set $O = \{y \in \mathbb{R}^n : T(y) = C\}$ and note that $O \subseteq \mathbb{R}^n \backslash N^*$ since $C > 0$ by assumption. We can then write

$$O = \bigcup_{y_2 \in \mathfrak{M}^\perp} \left(\{ y_1 \in \mathfrak{M} : y_1 + y_2 \in \mathbb{R}^n \backslash N^*, T(y_1 + y_2) = C \} + y_2 \right) = \bigcup_{y_2 \in \mathfrak{M}^\perp} \left(O(y_2) + y_2 \right).$$

Note that O as well as $O(y_2)$ are clearly measurable sets. By the already established invariance of $\mathbb{R}^n \backslash N^*$, the fact that $\mathbb{R}^n \backslash N^* \subseteq \mathbb{R}^n \backslash N$, and by the equivariance properties of $\check{\beta}$ and $\check{\Omega}$ maintained

in Assumption 5, the set $O(y_2)$ equals

$$\left\{ y_1 \in \mathfrak{M} : \left(R\left(\check{\beta}(y_2) + (X'X)^{-1} X'y_1 \right) - r \right)' \check{\Omega}^{-1}(y_2) \left(R\left(\check{\beta}(y_2) + (X'X)^{-1} X'y_1 \right) - r \right) = C \right\}$$

if $y_2 \in (\mathbb{R}^n \backslash N^*) \cap \mathfrak{M}^\perp$, and it is empty if $y_2 \in N^* \cap \mathfrak{M}^\perp$ (since $C > 0$). If $y_2 \in (\mathbb{R}^n \backslash N^*) \cap \mathfrak{M}^\perp$, the set $O(y_2) \subseteq \mathfrak{M}$ is the image of

$$\bar{O}(y_2) = \left\{ \gamma \in \mathbb{R}^k : \left(R\left(\check{\beta}(y_2) + \gamma \right) - r \right)' \check{\Omega}^{-1}(y_2) \left(R\left(\check{\beta}(y_2) + \gamma \right) - r \right) = C \right\}$$

under the invertible linear map $\gamma \mapsto X\gamma$ from $\mathbb{R}^k$ onto $\mathfrak{M}$. Now $\bar{O}(y_2)$ is the zero-set of a multivariate real polynomial (in the components of γ). The polynomial does not vanish everywhere on $\mathbb{R}^k$ because the quadratic form making up the polynomial is unbounded on $\mathbb{R}^k$ (because $\check{\Omega}^{-1}(y_2)$ is symmetric and well-defined if $y_2 \in \mathbb{R}^n \backslash N^*$ and because $\text{rank}(R) = q$ holds). Consequently, $\bar{O}(y_2)$ has k-dimensional Lebesgue measure zero and hence $\lambda_\mathfrak{M}(O(y_2)) = 0$ for every $y_2 \in (\mathbb{R}^n \backslash N^*) \cap \mathfrak{M}^\perp$. We conclude that $\lambda_\mathfrak{M}(O(y_2)) = 0$ for every $y_2 \in \mathfrak{M}^\perp$.

We now identify $\mathbb{R}^n$ with $\mathfrak{M} \times \mathfrak{M}^\perp$ and view Lebesgue measure $\lambda_{\mathbb{R}^n}$ on $\mathbb{R}^n$ as $\lambda_\mathfrak{M} \otimes \lambda_{\mathfrak{M}^\perp}$. Hence, y is identified with $(y_1, y_2) \in \mathfrak{M} \times \mathfrak{M}^\perp$ satisfying $y = y_1 + y_2$. Fubini's Theorem then shows

$$
\begin{aligned}
\lambda_{\mathbb{R}^n}(O) &= \lambda_{\mathfrak{M} \otimes \mathfrak{M}^\perp}(O) = \int_{\mathfrak{M} \times \mathfrak{M}^\perp} \mathbf{1}_O((y_1, y_2)) d\lambda_{\mathfrak{M} \times \mathfrak{M}^\perp}(y_1, y_2) \\
&= \int_{\mathfrak{M}^\perp} \int_\mathfrak{M} \mathbf{1}_{O(y_2)}(y_1) d\lambda_\mathfrak{M}(y_1) d\lambda_{\mathfrak{M}^\perp}(y_2) = \int_{\mathfrak{M}^\perp} \lambda_\mathfrak{M}(O(y_2)) d\lambda_{\mathfrak{M}^\perp}(y_2) = 0.
\end{aligned}
$$

(5&6) First observe that $\{y \in \mathbb{R}^n \backslash N^* : T(y) > C\} = \{y \in \mathbb{R}^n : T(y) > C\}$ holds in view of $C > 0$ and the definition of T. By continuity of T on $\mathbb{R}^n \backslash N^*$ established in Part 2 and by openness of $\mathbb{R}^n \backslash N^*$, the openness of $\{y \in \mathbb{R}^n \backslash N^* : T(y) > C\}$ and $\{y \in \mathbb{R}^n \backslash N^* : T(y) < C\}$ follows. It hence suffices to show that these two sets are non-empty: Choose an arbitrary $y \in \mathbb{R}^n \backslash N^*$ and set $y(\gamma) = y + X\gamma$ for $\gamma \in \mathbb{R}^k$. Then $y(\gamma) \in \mathbb{R}^n \backslash N^*$ by invariance of $\mathbb{R}^n \backslash N^*$ under $G(\mathfrak{M})$. Now by the equivariance properties of $\check{\beta}$ and $\check{\Omega}$ expressed in Assumption 5

$$T(y(\gamma)) = \left(R\gamma + R\check{\beta}(y) - r \right)' \check{\Omega}^{-1}(y) \left(R\gamma + R\check{\beta}(y) - r \right).$$

Define $\bar{\gamma} = \bar{\beta} - \check{\beta}(y)$ for some $\bar{\beta}$ satisfying $R\bar{\beta} = r$. Then $T(y(\bar{\gamma})) = 0 < C$ holds showing that $\{y \in \mathbb{R}^n \backslash N^* : T(y) < C\}$ is non-empty. Finally choose $y \in \mathbb{R}^n \backslash N^*$ and v as in Assumption 6. Choose δ such that $v = R\delta$. Then set $\gamma = c\delta + \bar{\beta} - \check{\beta}(y)$ where $\bar{\beta}$ is as before and c is a real number. Observe that then $T(y(\gamma)) = c^2 v' \check{\Omega}^{-1}(y)v$. Choosing c sufficiently large shows that $T(y(\gamma)) > C$ can be achieved, establishing that $\{y \in \mathbb{R}^n \backslash N^* : T(y) > C\}$ is non-empty.

(7) Let $\mathbf{G}$ be a standard normal $n \times 1$ random vector. Then

$$P_{\nu_m + \mu_0, \Phi_m}(W(C)) = \Pr \left(T(\nu_m + \mu_0 + \Phi_m^{1/2} \mathbf{G}) - C \geq 0 \right). \tag{44}$$

Set $\gamma_m = (X'X)^{-1} X'\nu_m$ and $\gamma_0 = (X'X)^{-1} X'\mu_0$. Observe that $R\gamma_0 = r$ while $\|R\gamma_m\| \to \infty$ as $m \to \infty$ in view of $\nu_m \in \Pi_{(\mathfrak{M}_0 - \mu_0)^\perp}(\mathfrak{M}_1 - \mu_0)$ and $\|\nu_m\| \to \infty$. For $\Phi_m^{1/2} \mathbf{G} \in \mathbb{R}^n \backslash N^*$ (an

event which has probability 1 because N^* is a $\lambda_{\mathbb{R}^n}$-null set and Φ_m is positive-definite) we may use equivariance of $\check{\beta}$ and $\check{\Omega}$ and obtain that $T(\nu_m + \mu_0 + \Phi_m^{1/2}\mathbf{G}) - C$ coincides on this event with

$$(R\gamma_m + R\check{\beta}(\Phi_m^{1/2}\mathbf{G}))'\check{\Omega}^{-1}(\Phi_m^{1/2}\mathbf{G})(R\gamma_m + R\check{\beta}(\Phi_m^{1/2}\mathbf{G})) - C. \tag{45}$$

Observe that $\Phi_m^{1/2}\mathbf{G} \to \Phi^{1/2}\mathbf{G}$ as $m \to \infty$ with probability 1. Furthermore, $\check{\beta}$ and $\check{\Omega}^{-1}$ are continuous on $\mathbb{R}^n \backslash N^*$, a set which has probability 1 under the law of $\Phi^{1/2}\mathbf{G}$ (since N^* is a $\lambda_{\mathbb{R}^n}$-null set and Φ is positive-definite). From the continuous mapping theorem we conclude that $R\check{\beta}(\Phi_m^{1/2}\mathbf{G})$ and $\check{\Omega}^{-1}(\Phi_m^{1/2}\mathbf{G})$ converge almost surely to $R\check{\beta}(\Phi^{1/2}\mathbf{G})$ and $\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})$, respectively. Now let $v \in A((\nu_m)_{m \geq 1})$ and let m_i be a subsequence such that $\left\|R\gamma_{m_i}\right\|^{-1} R\gamma_{m_i} \to v$. It follows that

$$\left[(R\gamma_{m_i} + R\check{\beta}(\Phi_{m_i}^{1/2}\mathbf{G}))'\check{\Omega}^{-1}(\Phi_{m_i}^{1/2}\mathbf{G})(R\gamma_{m_i} + R\check{\beta}(\Phi_{m_i}^{1/2}\mathbf{G})) - C\right] / \left\|R\gamma_{m_i}\right\|^2$$

converges to

$$v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v$$

with probability 1. Since $\Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v = 0\right)$ by Assumption 7, it follows that

$$\Pr\left(T(\nu_{m_i} + \mu_0 + \Phi_{m_i}^{1/2}\mathbf{G}) - C \geq 0\right) \to \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right).$$

This shows that

$$\begin{aligned}
\liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)) &\leq \liminf_{i \to \infty} P_{\nu_{m_i} + \mu_0, \Phi_{m_i}}(W(C)) \\
&= \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right),
\end{aligned}$$

implying that

$$\liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)) \leq \inf_{v \in A((\nu_m)_{m \geq 1})} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right).$$

Conversely, let m_i be a subsequence such that

$$P_{\nu_{m_i} + \mu_0, \Phi_{m_i}}(W(C)) \to \liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)).$$

Since the unit ball in $\mathbb{R}^q$ is compact, we may assume that $\left\|R\gamma_{m_{i(j)}}\right\|^{-1} R\gamma_{m_{i(j)}}$ converges to some $v \in A((\nu_m)_{m \geq 1})$ along a suitable subsequence $m_{i(j)}$. The same arguments as above then shows that

$$\begin{aligned}
\liminf_{m \to \infty} P_{\nu_m + \mu_0, \Phi_m}(W(C)) &= \liminf_{j \to \infty} P_{\nu_{m_{i(j)}} + \mu_0, \Phi_{m_{i(j)}}}(W(C)) \\
&= \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right) \\
&\geq \inf_{v \in A((\nu_m)_{m \geq 1})} \Pr\left(v'\check{\Omega}^{-1}(\Phi^{1/2}\mathbf{G})v \geq 0\right).
\end{aligned}$$

Given Assumption 7, the remaining equalities and inequalities in (26) and (27) are now obvious. $\blacksquare$

65

Proof of Corollary 5.17: (1) If $z \in \mathbb{R}^n \backslash N^*$ then $\mu_0 + z \in \mathbb{R}^n \backslash N^*$ and T is continuous at $\mu_0 + z$ for every $\mu_0 \in \mathfrak{M}_0$ by Parts 1 and 2 of Lemma 5.15. If $T(\mu_0^* + z) > C$ holds, then by the invariance of T established in Part 3 of Lemma 5.15, we have $T(\mu_0 + z) = T(\mu_0^* + z) > C$ for every $\mu_0 \in \mathfrak{M}_0$. Hence the sufficient conditions in Part 1 of Theorem 5.7 are satisfied and an application of this theorem delivers the result.

(2) Completely analogous to the proof of (1) noting that the invariance of T required in Part 3 of Theorem 5.7 is clearly satisfied.

(3) Since N^* is a $\lambda_{\mathbb{R}^n}$-null set the test statistic T is $\lambda_{\mathbb{R}^n}$-almost surely equal to the test statistic

$$T^*(y) = \begin{cases} T(y) & y \in \mathbb{R}^n \backslash N^*, \\ \infty, & y \in N^*. \end{cases}$$

We verify that the sufficient conditions in Part 1 of Theorem 5.7 are satisfied for T^*. To that end fix $\mu_0 \in \mathfrak{M}_0$ and let $\mathcal{Z}' \subseteq \mathcal{Z}$ denote the set of all z such that $z \in \mathbb{R}^n \backslash N$, $\check{\Omega}(z) = 0$, and $R\check{\beta}(z) \neq 0$ hold. By invariance of N (cf. discussion after Assumption 5) and equivariance of $\check{\Omega}$ we see that $z \in \mathcal{Z}'$ implies $\mu_0 + z \in \mathbb{R}^n \backslash N$ and $\check{\Omega}(\mu_0 + z) = 0$, and thus $T^*(\mu_0 + z) = \infty > C$ holds for every $z \in \mathcal{Z}'$ by definition of T^*. We next show that T^* is lower semicontinuous at $\mu_0 + z$ for every $z \in \mathcal{Z}'$. Let y_m be a sequence converging to $\mu_0 + z$. Since $\mathbb{R}^n \backslash N$ is open, we may assume that this sequence entirely belongs to $\mathbb{R}^n \backslash N$. If $\det \check{\Omega}(y_m) = 0$ eventually holds, we are done since then $T^*(y_m) = \infty$ eventually by construction. By a standard subsequence argument we may thus assume that $\det \check{\Omega}(y_m) > 0$ eventually holds since $\check{\Omega}$ is nonnegative definite on $\mathbb{R}^n \backslash N$ by assumption. Now note that

$$T^*(y_m) = T(y_m) = (R\check{\beta}(y_m) - r)'\check{\Omega}(y_m)^{-1}(R\check{\beta}(y_m) - r) \geq \lambda_{max}^{-1}(\check{\Omega}(y_m))\|R\check{\beta}(y_m) - r\|.$$

Since $\check{\beta}$ is continuous on $\mathbb{R}^n \backslash N$ by assumption, we have $R\check{\beta}(y_m) \to R\check{\beta}(\mu_0 + z) = R\check{\beta}(z) + r \neq r$ where we have made use of equivariance of $\check{\beta}(z)$ and of $\mu_0 \in \mathfrak{M}_0$. Hence $\|R\check{\beta}(y_m) - r\| \to \|R\check{\beta}(z)\| > 0$. Furthermore, $\check{\Omega}$ is continuous on $\mathbb{R}^n \backslash N$ by assumption, hence $\check{\Omega}(y_m) \to \check{\Omega}(\mu_0 + z) = 0$. Consequently, $T^*(y_m) \to \infty$, establishing lower semicontinuity of T^*. We may now apply Part 1 of Theorem 5.7 together with Remark 5.8(i) to conclude the proof. ∎

Lemma F.2. *Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumption 5, let T be the test statistic defined in (25), and let $W(C) = \{y \in \mathbb{R}^n : T(y) \geq C\}$ with $0 < C < \infty$ be the rejection region. Let Φ_m be symmetric positive definite $n \times n$ matrices such that $\Phi_m \to \Phi$ for $m \to \infty$ where Φ is singular with $l := \dim \operatorname{span}(\Phi) > 0$. Suppose that for some sequence of positive real numbers s_m the matrix $D_m = \Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)^\perp}/s_m$ converges to a matrix D, which is regular on $\operatorname{span}(\Phi)^\perp$, and that $\Pi_{\operatorname{span}(\Phi)^\perp} \Phi_m \Pi_{\operatorname{span}(\Phi)}/s_m^{1/2} \to 0$. Suppose further that $\operatorname{span}(\Phi) \subseteq \mathfrak{M}$. Let Z be a matrix, the columns of which form a basis for $\operatorname{span}(\Phi)$ and let $\boldsymbol{G}$ be a standard normal n-vector. Then:*

1. *For every $\mu_0 \in \mathfrak{M}_0$, $\gamma \in \mathbb{R}^l$, $0 < \sigma < \infty$ we have*

$$s_m \left[T\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G} \right) - C \right] \xrightarrow{d} \xi(\gamma, \sigma)$$

for $m \to \infty$ where the random variable $\xi(\gamma, \sigma)$ is given by

$$\left(R\hat{\beta}\left(\sigma^{-1}Z\gamma + \Phi^{1/2}\boldsymbol{G} \right) \right)' \check{\Omega}^{-1} \left(\left(\Phi^{1/2} + D^{1/2} \right)\boldsymbol{G} \right) \left(R\hat{\beta}\left(\sigma^{-1}Z\gamma + \Phi^{1/2}\boldsymbol{G} \right) \right)$$

for $\left(\Phi^{1/2} + D^{1/2} \right)\boldsymbol{G} \notin N^*$, which is an event that has probability 1 under the law of $\boldsymbol{G}$, and where $\xi(\gamma, \sigma) = 0$ else.

66

2. *If additionally Assumption 7 holds and*

$$R\hat{\beta}(z) \neq 0 \qquad \lambda_{\text{span}(\Phi)}\text{-}a.e.$$

is satisfied, then

$$P_{\mu_0+Z\gamma,\sigma^2\Phi_m}(W(C)) = \Pr\left(T\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G}\right) \geq C\right) \to \Pr\left(\xi\left(\gamma,\sigma\right) \geq 0\right)$$

as $m \to \infty$.

Proof. (1) Observe that $\mu_0 + Z\gamma \in \mathfrak{M}$, that the columns of $\Phi^{1/2}$ as well of $\Pi_{\text{span}(\Phi)}\Phi_m^{1/2}$ belong to $\mathfrak{M}$, and that $\mathbb{R}^n\backslash N^*$ is invariant under the group $G(\mathfrak{M})$. Hence, using the equivariance properties of $\check{\beta}$ and $\check{\Omega}$ expressed in Assumption 5 repeatedly, we obtain that on the event $\left\{\Phi_m^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right\}$

$$
\begin{aligned}
R\check{\beta}\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G}\right) - r &= R\check{\beta}\left(\mu_0 + Z\gamma + \sigma\Pi_{\text{span}(\Phi)}\Phi_m^{1/2}\boldsymbol{G} + \sigma\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\boldsymbol{G}\right) - r \\
&= R\left(BZ\gamma + \sigma B\Pi_{\text{span}(\Phi)}\Phi_m^{1/2}\boldsymbol{G} + \sigma s_m^{1/2}\check{\beta}\left(s_m^{-1/2}\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\boldsymbol{G}\right)\right) \\
&= \sigma R\left(\sigma^{-1}BZ\gamma + \boldsymbol{K}_m + s_m^{1/2}\check{\beta}\left(\boldsymbol{L}_m\right)\right)
\end{aligned}
$$

holds, where B is shorthand for $(X'X)^{-1}X'$, $\boldsymbol{K}_m = B\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2} - s_m^{1/2}\Phi^{1/2}\right)\boldsymbol{G}$, and $\boldsymbol{L}_m = \Phi^{1/2}\boldsymbol{G} + s_m^{-1/2}\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\boldsymbol{G})$. Similarly, we obtain

$$
\begin{aligned}
\check{\Omega}\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G}\right) &= \sigma^2\check{\Omega}\left(\Phi_m^{1/2}\boldsymbol{G}\right) = \sigma^2\check{\Omega}\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2}\boldsymbol{G} + \Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\boldsymbol{G}\right) \\
&= \sigma^2\check{\Omega}\left(\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\boldsymbol{G}\right) = \sigma^2 s_m\check{\Omega}(\boldsymbol{L}_m)
\end{aligned}
$$

on the event $\left\{\Phi_m^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right\}$. Hence, on this event we have

$$
\begin{aligned}
s_m\left[T\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G}\right) - C\right] &= \left(R\left(\sigma^{-1}BZ\gamma + \boldsymbol{K}_m + s_m^{1/2}\check{\beta}\left(\boldsymbol{L}_m\right)\right)\right)'\check{\Omega}^{-1}\left(\boldsymbol{L}_m\right) \\
&\quad \times R\left(\sigma^{-1}BZ\gamma + \boldsymbol{K}_m + s_m^{1/2}\check{\beta}\left(\boldsymbol{L}_m\right)\right) - s_m C.
\end{aligned}
$$

Clearly, $\boldsymbol{K}_m$ and $\boldsymbol{L}_m$ are jointly normal with mean zero and second moments given by

$$E\left(\boldsymbol{K}_m\boldsymbol{K}_m'\right) = B\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2} - s_m^{1/2}\Phi^{1/2}\right)\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2} - s_m^{1/2}\Phi^{1/2}\right)'B',$$

$$E\left(\boldsymbol{L}_m\boldsymbol{L}_m'\right) = \Phi + D_m,$$

and

$$E\left(\boldsymbol{K}_m\boldsymbol{L}_m'\right) = B\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2} - s_m^{1/2}\Phi^{1/2}\right)\left(\Phi^{1/2} + s_m^{-1/2}\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\right)'.$$

It is easy to see that $E\left(\boldsymbol{K}_m\boldsymbol{K}_m'\right)$ converges to $B\Phi B'$ because $s_m \to 0$, while $E\left(\boldsymbol{L}_m\boldsymbol{L}_m'\right)$ converges to $\Phi + D$. Furthermore, $E\left(\boldsymbol{K}_m\boldsymbol{L}_m'\right)$ converges to $B\Phi$ because

$$
\begin{aligned}
& B\left(\Pi_{\text{span}(\Phi)}\Phi_m^{1/2} - s_m^{1/2}\Phi^{1/2}\right)\left(s_m^{-1/2}\Pi_{\text{span}(\Phi)^\perp}\Phi_m^{1/2}\right)' \\
&= B\left(\Pi_{\text{span}(\Phi)^\perp}\Phi_m\Pi_{\text{span}(\Phi)}/s_m^{1/2}\right)' - B\Phi^{1/2}\Phi_m^{1/2}\Pi_{\text{span}(\Phi)^\perp} \\
&\to -B\Phi\Pi_{\text{span}(\Phi)^\perp} = -B\left(\Pi_{\text{span}(\Phi)^\perp}\Phi\right)' = 0
\end{aligned}
$$

where we have made use of the assumption $\Pi_{\mathrm{span}(\Phi)^\perp}\Phi_m\Pi_{\mathrm{span}(\Phi)}/s_m^{1/2} \to 0$ and of symmetry of Φ. Hence we have (cf. Lemma E.1) that

$$\left(\begin{array}{c} \boldsymbol{K}_m \\ \boldsymbol{L}_m \end{array}\right) \xrightarrow{d} N\left(0, \left[\begin{array}{cc} B\Phi B' & B\Phi \\ \Phi B' & \Phi + D \end{array}\right]\right).$$

Note that this limiting normal distribution is also the joint distribution of $\boldsymbol{K} = B\Phi^{1/2}\boldsymbol{G}$ and $\boldsymbol{L} = \left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G}$. [Observe that $\Phi^{1/2} + D^{1/2} = (\Phi + D)^{1/2}$ since $\Phi D = D\Phi = 0$ as D vanishes on $\mathrm{span}(\Phi)$ by construction.] Now consider the map f on $\mathbb{R}^{n+k}$ given by $f(x, y) = (f_1(x), f_2(y), f_3(y))$ where $f_1(x) = x$ for $x \in \mathbb{R}^k$, and where $f_2(y) = \check{\beta}(y)$, $f_3(y) = \check{\Omega}^{-1}(y)$ for $y \in \mathbb{R}^n\backslash N^*$ and are zero else. Observe that the set of discontinuity points, F say, of f is contained in $\mathbb{R}^k \times N^*$. But

$$\Pr\left((\boldsymbol{K}, \boldsymbol{L}) \in F\right) \leq \Pr\left((\boldsymbol{K}, \boldsymbol{L}) \in \mathbb{R}^k \times N^*\right) = \Pr\left(\boldsymbol{L} \in N^*\right) = 0 \tag{46}$$

because N^* is a $\lambda_{\mathbb{R}^n}$-null set and the distribution of $\boldsymbol{L}$ is equivalent to Lebesgue measure on $\mathbb{R}^n$ as $\Phi + D$ is positive definite. This shows that $f(\boldsymbol{K}_m, \boldsymbol{L}_m)$ converges in distribution to $f(\boldsymbol{K}, \boldsymbol{L})$ as $m \to \infty$. Now

$$s_m\left[T\left(\mu_0 + Z\gamma + \sigma\Phi_m^{1/2}\boldsymbol{G}\right) - C\right] = \left(R\left(\sigma^{-1}BZ\gamma + f_1(\boldsymbol{K}_m) + s_m^{1/2}f_2(\boldsymbol{L}_m)\right)\right)' f_3(\boldsymbol{L}_m)$$
$$\times R\left(\sigma^{-1}BZ\gamma + f_1(\boldsymbol{K}_m) + s_m^{1/2}f_2(\boldsymbol{L}_m)\right) - s_m C$$

holds everywhere (note that $\boldsymbol{L}_m \in \mathbb{R}^n\backslash N^*$ if and only if $\Phi_m^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*$ by $G(\mathfrak{M})$-invariance of $\mathbb{R}^n\backslash N^*$). Because $s_m^{1/2}f_2(\boldsymbol{L}_m)$ converges to zero in probability and $s_m C \to 0$ we immediately see that the random variable in the preceding display converges in distribution to

$$\left(R\left(\sigma^{-1}BZ\gamma + f_1\left(B\Phi^{1/2}\boldsymbol{G}\right)\right)\right)' f_3\left(\left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G}\right) R\left(\sigma^{-1}BZ\gamma + f_1\left(B\Phi^{1/2}\boldsymbol{G}\right)\right)$$

which coincides with $\xi(\gamma, \sigma)$. Finally, the claim that $\left\{\left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right\}$ is a probability 1 event has already been established in (46).

(2) This follows from Part 1 if we can establish that $\Pr(\xi(\gamma, \sigma) = 0) = 0$. Now observe that $\check{\Omega}^{-1}\left(\left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G}\right) = \check{\Omega}^{-1}(D^{1/2}\boldsymbol{G})$ by equivariance and that $\left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n\backslash N^*$ if and only if $D^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*$. Hence

$$\Pr\left(\xi(\gamma, \sigma) = 0\right) = \Pr\left(\xi(\gamma, \sigma) = 0, \left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right)$$
$$= \Pr\left(\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + \Phi^{1/2}\boldsymbol{G}\right)\right)' R'\check{\Omega}^{-1}(D^{1/2}\boldsymbol{G})\right.$$
$$\left.\times R\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + \Phi^{1/2}\boldsymbol{G}\right)\right) = 0, D^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right)$$
$$= \int \Pr\left(\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + x\right)\right)' R'\check{\Omega}^{-1}(D^{1/2}\boldsymbol{G})\right.$$
$$\left.\times R\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + x\right)\right) = 0, D^{1/2}\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right) dP_{0,\Phi}(x)$$
$$= \int \Pr\left(\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + x\right)\right)' R'\check{\Omega}^{-1}\left(\left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G}\right)\right.$$
$$\left.\times R\left(\hat{\beta}\left(\sigma^{-1}Z\gamma + x\right)\right) = 0, \left(\Phi^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n\backslash N^*\right) dP_{0,\Phi}(x)$$
$$= \int P_{0,\Phi+D}\left(\left\{y \in \mathbb{R}^n\backslash N^* : v(x)'\check{\Omega}^{-1}(y)v(x) = 0\right\}\right) dP_{0,\Phi}(x) \tag{47}$$

with $v(x) = R\hat{\beta}\left(\sigma^{-1}Z\gamma + x\right)$, the third equality in the preceding display being true since $\Phi^{1/2}\boldsymbol{G}$ and $D^{1/2}\boldsymbol{G}$ are independent as

$$E\left(\Phi^{1/2}\boldsymbol{G}\left(D^{1/2}\boldsymbol{G}\right)'\right) = \Phi^{1/2}D^{1/2} = 0.$$

Now the integrand in the last line of (47) is zero by Assumption (7) for every x except when $v(x) = 0$. Hence, we are done if we can establish that $P_{0,\Phi}\left(v(x) = 0\right) = 0$. Because span($\Phi$) equals the span of the columns of Z, we can make the change of variables $x = Zc$ and obtain

$$P_{0,\Phi}\left(v(x) = 0\right) = P_{0,A}\left(v(Zc) = 0\right) = P_{0,A}\left(R\left(\hat{\beta}\left(Z\left(\sigma^{-1}\gamma + c\right)\right)\right) = 0\right)$$

where $A = \left(Z'Z\right)^{-1}Z'\Phi Z\left(Z'Z\right)^{-1}$. Because A is non-singular, this probability is zero if the event has $\lambda_{\mathbb{R}^l}$-measure zero. But

$$\lambda_{\mathbb{R}^l}\left(\left\{c : R\hat{\beta}\left(Z\left(\sigma^{-1}\gamma + c\right)\right) = 0\right\}\right) = \lambda_{\mathrm{span}(\Phi)}\left(\left\{z : R\hat{\beta}\left(z\right) = 0\right\}\right) = 0$$

by our assumptions. ∎

Proof of Theorem 5.19: Fix $\mu_0 \in \mathfrak{M}_0$ and σ, $0 < \sigma < \infty$. Then for every $\gamma \in \mathbb{R}^l$ we have

$$P_{\mu_0 + Z\gamma, \sigma^2\Sigma_m}\left(W(C)\right) = \Pr\left(s_m\left[T\left(\mu_0 + Z\gamma + \sigma\Sigma_m^{1/2}\boldsymbol{G}\right) - C\right] \geq 0\right)$$

which converges to $\Pr\left(\xi\left(\gamma, \sigma\right) \geq 0\right)$ as shown in the preceding lemma (with Σ_m and $\bar{\Sigma}$ playing the rôles of Φ_m and Φ, respectively). Consequently, for every $\gamma \in \mathbb{R}^l$

$$\inf_{\Sigma \in \mathfrak{C}} P_{\mu_0 + Z\gamma, \sigma^2\Sigma}\left(W(C)\right) \leq \Pr\left(\xi\left(\gamma, \sigma\right) \geq 0\right).$$

But now

$$\liminf_{M \to \infty} \inf_{\|\gamma\| \geq M} \Pr\left(\xi\left(\gamma, \sigma\right) \geq 0\right) \leq \liminf_{M \to \infty} \inf_{R\hat{\beta}(Z\gamma) \neq 0, \|\gamma\| \geq M} \Pr\left(\xi\left(\gamma, \sigma\right) \geq 0\right)$$

$$= \liminf_{M \to \infty} \inf_{R\hat{\beta}(Z\gamma) \neq 0, \|\gamma\| \geq M} \Pr\left(\xi\left(\gamma, \sigma\right) / \|\gamma\|^2 \geq 0\right)$$

$$\leq \liminf_{M \to \infty} \inf_{R\hat{\beta}(Z\gamma) \neq 0, \|\gamma\| = M} \Pr\left(\xi\left(\gamma, \sigma\right) / \|\gamma\|^2 \geq 0\right)$$

$$\leq \inf_{\|c\| = 1, R\hat{\beta}(Zc) \neq 0} \liminf_{M \to \infty} \Pr\left(\bar{\xi}\left(c, M, \sigma\right) \geq 0\right)$$

where

$$\bar{\xi}\left(c, M, \sigma\right) = \left(R\left(\hat{\beta}\left(Zc\right) + \sigma\hat{\beta}\left(\bar{\Sigma}^{1/2}\boldsymbol{G}\right)/M\right)\right)'\check{\Omega}^{-1}\left(\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G}\right)$$
$$\times R\left(\hat{\beta}\left(Zc\right) + \sigma\hat{\beta}\left(\bar{\Sigma}^{1/2}\boldsymbol{G}\right)/M\right)$$

on the event where $\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n\backslash N^*$ and is zero else. The random variable $\bar{\xi}\left(c, M, \sigma\right)$ converges in probability to the random variable $\bar{\xi}\left(c\right)$ as $M \to \infty$. Hence

$$\liminf_{M \to \infty} \Pr\left(\bar{\xi}\left(c, M, \sigma\right) \geq 0\right) = \Pr\left(\bar{\xi}\left(c\right) \geq 0\right)$$

holds for every $c \in \mathbb{R}^l$ satisfying $\|c\| = 1$ and $R\hat{\beta}\left(Zc\right) \neq 0$, because $\Pr\left(\bar{\xi}\left(c\right) = 0\right) = 0$ for such c in view of Assumption 7 observing that $P_{0,\bar{\Sigma}+D}$ is equivalent to $\lambda_{\mathbb{R}^n}$ as $\bar{\Sigma} + D$ is nonsingular. This proves that

$$\liminf_{M \to \infty} \inf_{\|\gamma\| \geq M} \inf_{\Sigma \in \mathfrak{C}} P_{\mu_0 + Z\gamma, \sigma^2\Sigma}\left(W(C)\right) \leq \inf_{\|c\|=1, R\hat{\beta}(Zc)\neq 0} \Pr\left(\bar{\xi}\left(c\right) \geq 0\right)$$

$$= \inf_{\|c\|=1} \Pr\left(\bar{\xi}\left(c\right) \geq 0\right) = \inf_{c \in \mathbb{R}^l} \Pr\left(\bar{\xi}\left(c\right) \geq 0\right) = K_1,$$

the first two equalities holding because $\bar{\xi}\left(c\right) \equiv 0$ if $R\hat{\beta}\left(Zc\right) = 0$ (and in particular if $c = 0$) and because $\Pr\left(\bar{\xi}\left(c\right) \geq 0\right)$ is homogenous in c. This establishes the first inequality in (30) because the left-most expression in (30) is monotonically increasing in M. Furthermore,

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2\Sigma}\left(W(C)\right) \geq P_{\mu_0, \sigma^2\Sigma_m}\left(W(C)\right),$$

and hence we obtain from Lemma F.2 that

$$\sup_{\Sigma \in \mathfrak{C}} P_{\mu_0, \sigma^2\Sigma}\left(W(C)\right) \geq \Pr\left(\xi\left(0, \sigma\right) \geq 0\right)$$

$$= \Pr\left(\left(R\hat{\beta}\left(\bar{\Sigma}^{1/2}\boldsymbol{G}\right)\right)' \check{\Omega}^{-1}\left(\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G}\right)R\hat{\beta}\left(\bar{\Sigma}^{1/2}\boldsymbol{G}\right) \geq 0,\right.$$

$$\left.\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n \backslash N^*\right).$$

Now observe that $\check{\Omega}^{-1}\left(\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G}\right) = \check{\Omega}^{-1}(D^{1/2}\boldsymbol{G})$ by equivariance and that $\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n \backslash N^*$ if and only if $D^{1/2}\boldsymbol{G} \in \mathbb{R}^n \backslash N^*$. Then by the same arguments as in (47) we obtain

$$\Pr\left(\xi\left(0, \sigma\right) \geq 0\right) = \int \Pr\left(\left(R\hat{\beta}\left(x\right)\right)' \check{\Omega}^{-1}\left(\left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G}\right)\right.$$

$$\times R\left(\hat{\beta}\left(x\right)\right) \geq 0, \left(\bar{\Sigma}^{1/2} + D^{1/2}\right)\boldsymbol{G} \in \mathbb{R}^n \backslash N^*\right) dP_{0,\bar{\Sigma}}(x)$$

$$= \int \Pr\left(\bar{\xi}\left(\gamma\right) \geq 0\right) dP_{0,A}(\gamma) = K_2,$$

the last equality resulting from the variable change $x = Z\gamma$ which is possible since $\text{span}(\bar{\Sigma})$ equals the space spanned by Z. Finally, the inequality $K_1 \leq K_2$ is obvious from the definition of these constants. ∎

Proof of Theorem 5.21: Define $\varphi = \mathbf{1}\left(W(C)\right)$ and note that invariance of φ under $G(\mathfrak{M}_0)$ as well as the fact that φ is $\lambda_{\mathbb{R}^n}$-almost surely neither equal to 0 or 1 follows from Lemma 5.15. Part 1 of Theorem 5.10 then implies Part 1 of the theorem. Similarly, Parts 2 and 3 of the theorem follow from Parts 2 and 3 of Theorem 5.10, respectively, because condition (19) follows from Part 7 of Lemma 5.15 combined with Remark 5.16 and because the lower bound in (27) equals 1 under the assumptions of Part 3. To prove Part 4 we use Theorem 5.12. Choose a sequence C_k, $0 < C_k < \infty$, that diverges monotonically to infinity and set $\varphi_k = \mathbf{1}\left(W(C_k)\right)$. Then (23) is satisfied by the monotone convergence theorem and the result follows from Theorem 5.12 upon setting $C(\delta) = C_{k_0(\delta)}$. ∎

Lemma F.3. *Let $\check{\beta}$ and $\check{\Omega}$ satisfy Assumptions 5 and 7. Let T be the test statistic defined in (25) and let $\mathfrak{C}$ be a covariance model. If there is a $z \in \text{span}\left(J(\mathfrak{C})\right) \cap \mathfrak{M}$ with $z \notin \mathfrak{M}_0 - \mu_0$ (i.e., with $R\hat{\beta}(z) \neq 0$), then T does not satisfy the invariance condition (31).*

Proof. Choose $z \in \operatorname{span}(J(\mathfrak{C})) \cap \mathfrak{M}$ with $z \notin \mathfrak{M}_0 - \mu_0$. Because $\mathfrak{M}$ is a linear space, we also have $cz \in \operatorname{span}(J(\mathfrak{C})) \cap \mathfrak{M}$ for every $c \in \mathbb{R}$. Now $cz \in \mathfrak{M}$ entails that $y \in \mathbb{R}^n \backslash N^*$ implies $y + cz \in \mathbb{R}^n \backslash N^*$. Using the definition of T and Assumption 5 we obtain

$$T(y + cz) = T(y) + 2c\left(R\check{\beta}(y) - r\right)' \check{\Omega}^{-1}(y) R\hat{\beta}(z) + c^2 \left(R\hat{\beta}(z)\right)' \check{\Omega}^{-1}(y) \left(R\hat{\beta}(z)\right)$$

for every $y \in \mathbb{R}^n \backslash N^*$. Because $R\hat{\beta}(z) \neq 0$, we can in view of Assumption 7 find an $y \in \mathbb{R}^n \backslash N^*$ such that

$$\left(R\hat{\beta}(z)\right)' \check{\Omega}^{-1}(y) \left(R\hat{\beta}(z)\right) \neq 0$$

holds. Hence $T(y + cz) = T(y)$ cannot hold for the so-chosen y and all $c \neq 0$. Because $cz \in \operatorname{span}(J(\mathfrak{C}))$, Remark 5.11(i) implies that condition (31) is not satisfied. $\blacksquare$

Proof of Proposition 5.23: (1) By the assumed equivariance (invariance, respectively) of $\bar{\theta}$, $\bar{\Omega}$, and $\bar{N}$ (and hence of $\bar{N}^*$) w.r.t. the transformations $y \mapsto \alpha y + \bar{X}\eta$, the equivariance (invariance, respectively) of $\bar{\beta}$, $\bar{\Omega}$, and $\bar{N}$ required in the original Assumption 5 is clearly satisfied. Now choose $z \in J(\mathfrak{C})$ and $y \in \mathbb{R}^n$. If $y \in \bar{N}^*$ then so is $y + z$ because of invariance of $\bar{N}^*$ and because $z \in J(\mathfrak{C}) \subseteq \bar{\mathfrak{M}}$ holds by construction. Hence, $T(y) = 0 = T(y + z)$ is satisfied in this case. Now let $y \in \mathbb{R}^n \backslash \bar{N}^*$ (and hence also $y + z \in \mathbb{R}^n \backslash \bar{N}^*$). Note that $\bar{\Omega}(y) = \bar{\Omega}(y + z)$ holds by equivariance. It remains to show that $R\bar{\beta}(y) = R\bar{\beta}(y + z)$. Because $z \in J(\mathfrak{C}) \subseteq \bar{\mathfrak{M}}$ we have $z = X\gamma + (\bar{x}_1, \ldots, \bar{x}_p)\delta$ and thus obtain

$$R\bar{\beta}(y + z) = (R, 0)\bar{\theta}(y + z) = (R, 0)\left(\bar{\theta}(y) + (\gamma', \delta')'\right) = R\bar{\beta}(y) + R\gamma, \tag{48}$$

where we have made use of equivariance of $\bar{\theta}$. Now observe that $(\bar{x}_1, \ldots, \bar{x}_p)\delta \in \operatorname{span}(J(\mathfrak{C}) \cup (\mathfrak{M}_0 - \mu_0))$ by construction of the $\bar{x}_i$. Hence, we can find an element $\mu_0^{\#} \in \mathfrak{M}_0$ such that $(\bar{x}_1, \ldots, \bar{x}_p)\delta - \left(\mu_0^{\#} - \mu_0\right) \in \operatorname{span}(J(\mathfrak{C}))$. Consequently, we obtain

$$z - \left((\bar{x}_1, \ldots, \bar{x}_p)\delta - \left(\mu_0^{\#} - \mu_0\right)\right) = X\gamma + \left(\mu_0^{\#} - \mu_0\right).$$

The left-hand side is obviously an element of $\operatorname{span}(J(\mathfrak{C}))$, while the right-hand side belongs to $\mathfrak{M}$, implying that the right-hand side is in $\operatorname{span} J(\mathfrak{C}) \cap \mathfrak{M}$ which is a subset of $\mathfrak{M}_0 - \mu_0$ by assumption. Because $\mu_0^{\#} - \mu_0 \in \mathfrak{M}_0 - \mu_0$, we have established that $X\gamma \in \mathfrak{M}_0 - \mu_0$, or in other words, that $R\gamma = 0$.

(2) The very first claim is obvious. If $z \in \operatorname{span}(J(\mathfrak{C}))$ then again we have $z = X\gamma + (\bar{x}_1, \ldots, \bar{x}_p)\delta$ and $\bar{\theta}(z) = (\gamma', \delta')'$. Now $R\bar{\beta}(z) = (R, 0)\bar{\theta}(z) = R\gamma$ and exactly the same argument as above shows that $R\gamma = 0$. For the last claim note that $\bar{X}\bar{\theta}(y) = X^*\theta^*(y)$ holds because $\bar{X}$ and X^* span the same space. This equality can be written as

$$X\bar{\beta}(y) - X\beta^*(y) = \sum_{i=1}^{p} x_i^* \theta_{k+i}^*(y) - \sum_{i=1}^{p} \bar{x}_i \bar{\theta}_{k+i}(y).$$

Because the right-hand side of the above equation belongs to $\operatorname{span}(J(\mathfrak{C}) \cup (\mathfrak{M}_0 - \mu_0))$ we can find $\mu_0^{\#} \in \mathfrak{M}_0$ such that the right-hand side of

$$X\left(\bar{\beta}(y) - \beta^*(y)\right) - \left(\mu_0^{\#} - \mu_0\right) = \sum_{i=1}^{p} x_i^* \theta_{k+i}^*(y) - \sum_{i=1}^{p} \bar{x}_i \bar{\theta}_{k+i}(y) - \left(\mu_0^{\#} - \mu_0\right)$$

belongs to $\operatorname{span}(J(\mathfrak{C}))$ while the left-hand side belongs to $\mathfrak{M}$. Arguing now similarly as in the proof of Part 1, we conclude that $R\bar{\beta}(y) = R\beta^*(y)$. $\blacksquare$

G   Appendix: Properties of AR-Correlation Matrices

Lemma G.1.   *1. Suppose the covariance model $\mathfrak{C}$ contains $\Lambda(\rho_m)$ for some sequence $\rho_m \in (-1, 1)$ with $\rho_m \to 1$ ($\rho_m \to -1$, respectively). Then $\mathrm{span}\,(e_+)$ ($\mathrm{span}\,(e_-)$, respectively) is a concentration space of $\mathfrak{C}$.*

2. $\mathfrak{C}_{AR(1)}$ has $\mathrm{span}\,(e_+)$ and $\mathrm{span}\,(e_-)$ as its only concentration spaces. Consequently, $J(\mathfrak{C}_{AR(1)}) = \mathrm{span}(e_+) \cup \mathrm{span}(e_-)$.

3. If $\rho_m \in (-1, 1)$ is a sequence converging to 1 then $\Sigma_m = \Lambda(\rho_m)$ satisfies $\Sigma_m \to \bar{\Sigma} = e_+ e'_+$ and $D_m = \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} / s_m \to D$ as well as $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} / s_m^{1/2} \to 0$ where $s_m = \mathrm{tr}\left(\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \right)$ converges to zero and D is the matrix with (i, j)-th element $-n\,|i - j| / \sum_{i,j} |i - j|$ pre- and postmultiplied by $\left(I_n - n^{-1} e_+ e'_+ \right)$. Furthermore, D is regular on $\mathrm{span}\left(\bar{\Sigma} \right)^\perp$.

4. If $\rho_m \in (-1, 1)$ is a sequence converging to -1 then $\Sigma_m = \Lambda(\rho_m)$ satisfies $\Sigma_m \to \bar{\Sigma} = e_- e'_-$ and $D_m = \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} / s_m \to D$ as well as $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} / s_m^{1/2} \to 0$ where $s_m = \mathrm{tr}\left(\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \right)$ converges to zero and D is the matrix with (i, j)-th element $n(-1)^{|i-j|+1} |i - j| / \sum_{i,j} |i - j|$ pre- and postmultiplied by $\left(I_n - n^{-1} e_- e'_- \right)$. Furthermore, D is regular on $\mathrm{span}\left(\bar{\Sigma} \right)^\perp$.

Proof. (1) and (2) are obvious.

(3) Because $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp}$ is nonnegative definite, but obviously different from the zero matrix (recall that $n > 1$ is assumed), we see that s_m is always positive. Clearly, $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})^\perp}$ converges to $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \bar{\Sigma} \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} = 0$ and hence $s_m \to 0$. By l'Hopital's rule the limit of D_m can be obtained as the limit of $\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} (d\Lambda/d\rho) (\rho_m) \Pi_{\mathrm{span}(\bar{\Sigma})^\perp}$ divided by the limit of

$$\mathrm{tr}\left(\Pi_{\mathrm{span}(\bar{\Sigma})^\perp} (d\Lambda/d\rho) (\rho_m) \Pi_{\mathrm{span}(\bar{\Sigma})^\perp} \right)$$

provided the latter is nonzero. The second limit now equals

$$
\begin{aligned}
\mathrm{tr}\left(\left(I_n - n^{-1} e_+ e'_+ \right) (d\Lambda/d\rho) (1) \left(I_n - n^{-1} e_+ e'_+ \right) \right) &= \mathrm{tr}\left((d\Lambda/d\rho) (1) \left(I_n - n^{-1} e_+ e'_+ \right) \right) \\
&= \mathrm{tr}\left((d\Lambda/d\rho) (1) \right) - n^{-1} \mathrm{tr}\left(e'_+ (d\Lambda/d\rho) (1) e_+ \right).
\end{aligned}
$$

Observe that the (i, j)-th element of the matrix $(d\Lambda/d\rho) (1)$ is given by $|i - j|$. Hence, the above expression equals

$$-n^{-1} \mathrm{tr}\left(e'_+ (d\Lambda/d\rho) (1) e_+ \right) = -n^{-1} \sum_{i,j} |i - j|,$$

which is clearly nonzero. The first limit exists and equals

$$\left(I_n - n^{-1} e_+ e'_+ \right) (d\Lambda/d\rho) (1) \left(I_n - n^{-1} e_+ e'_+ \right)$$

72

which shows that D is of the form as claimed in the lemma. We next show that D is regular on $\mathrm{span}\left(\bar{\Sigma}\right)^{\perp} = \mathrm{span}(e_+)^{\perp}$. This is equivalent to showing that the equation system

$$(d\Lambda/d\rho)\,(1)x + \lambda e_+ = 0$$
$$e'_+ x = 0$$

has $x = 0$, $\lambda = 0$ as its only solution. We hence need to show that the $(n+1) \times (n+1)$ matrix

$$A = \left[\begin{array}{cc} (d\Lambda/d\rho)\,(1) & e_+ \\ e'_+ & 0 \end{array}\right]$$

has rank $n+1$. Let B be the $(n+1) \times (n+1)$ matrix given by

$$B = \left[\begin{array}{cc} B_{11} & 0 \\ 0 & 1 \end{array}\right]$$

where the $n \times n$ matrix B_{11} has 1 everywhere on the main diagonal, -1 everywhere on the first off-diagonal above the main diagonal, and zeroes elsewhere. Let the $(n+1) \times (n+1)$ matrices B^* and B^{**} be given by

$$B^* = \left[\begin{array}{cc} 0 & 1 \\ I_n & 0 \end{array}\right], \qquad B^{**} = \left[\begin{array}{cc} I_n & 0 \\ f & 1 \end{array}\right],$$

where $f = -(n-1, n-2, n-3, \ldots, 1, 0)$. Observe that B, B^*, as well as B^{**} are non-singular and that

$$B^* B A B^{**} = C = \left[\begin{array}{cc} C_{11} & 0 \\ 0 & 1 \end{array}\right]$$

where C_{11} is an $n \times n$ matrix that has 1 everywhere on and above the diagonal and -1 everywhere below the diagonal. Obviously, C is nonsingular and hence A is so. Finally, we show that the limit of $\Pi_{\mathrm{span}(\bar{\Sigma})^{\perp}} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})} / s_m^{1/2}$ equals zero. Because $s_m \to 0$, it suffices to show that the limit of $\Pi_{\mathrm{span}(\bar{\Sigma})^{\perp}} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})} / s_m$ exists and is finite. Now the same arguments as above show that the latter limit is equal to $\left(I_n - n^{-1}e_+ e'_+\right)(d\Lambda/d\rho)\,(1) n^{-1} e_+ e'_+$ divided by $-n^{-1}\sum_{i,j}|i - j|$.

(4) For the same reasons as in (3) s_m is positive and converges to zero. By the same argument as in (3) the limit of D_m is

$$\left[\left(I_n - n^{-1}e_- e'_-\right)(d\Lambda/d\rho)\,(-1)\left(I_n - n^{-1}e_- e'_-\right)\right] / \operatorname{tr}\left(\left(I_n - n^{-1}e_- e'_-\right)(d\Lambda/d\rho)\,(-1)\left(I_n - n^{-1}e_- e'_-\right)\right).$$

Note that the denominator is equal to

$$\operatorname{tr}\left((d\Lambda/d\rho)\,(-1)\right) - n^{-1}\operatorname{tr}\left(e'_-\,(d\Lambda/d\rho)\,(-1)e_-\right) = n^{-1}\sum_{i,j}|i - j| \neq 0,$$

observing that the (i,j)-th element of $(d\Lambda/d\rho)\,(-1)$ is given by $(-1)^{|i-j|+1}|i - j|$. We next show that D is regular on $\mathrm{span}\left(\bar{\Sigma}\right)^{\perp} = \mathrm{span}(e_-)^{\perp}$. This is equivalent to showing that the equation system

$$(d\Lambda/d\rho)\,(-1)x + \lambda e_- = 0$$
$$e'_- x = 0$$

has $x = 0$, $\lambda = 0$ as its only solution. We hence need to show that the $(n + 1) \times (n + 1)$ matrix

$$A^{\#} = \left[\begin{array}{cc} (d\Lambda/d\rho)(-1) & e_- \\ e'_- & 0 \end{array} \right]$$

has rank $n + 1$. Note that this is equivalent to establishing that the matrix

$$A^{\dagger} = \left[\begin{array}{cc} (d\Lambda/d\rho)(-1) & (-1)^{n+1}e_- \\ (-1)^{n+1}e'_- & 0 \end{array} \right]$$

is nonsingular. Now note that

$$A^{\dagger} = -EAE$$

where A is as in (3) and E is an $(n+1) \times (n+1)$ diagonal matrix with the i-th diagonal element given by $(-1)^i$. This proves regularity of D on span $(\bar{\Sigma})^{\perp}$. The claim for $\Pi_{\mathrm{span}(\bar{\Sigma})^{\perp}} \Sigma_m \Pi_{\mathrm{span}(\bar{\Sigma})} / s_m^{1/2}$ is proved as in (3). ∎

Lemma G.2. *For every $\nu \in [0, \pi]$ there exists a sequence $\Sigma_m \in \mathfrak{C}_{AR(2)}$ converging to $E(\nu)E(\nu)'$.*

Proof. For $\nu = 0$ ($\nu = \pi$, respectively) the matrix $E(\nu)E(\nu)'$ equals $e_+e'_+$ ($e_-e'_-$, respectively), and the result thus follows from Lemma G.1. Hence assume that $\nu \in (0, \pi)$. Consider for $0 < r < 1$ the AR(2)-spectral density

$$f_r(\omega) = (2\pi)^{-1} c(r) \left| 1 - 2r\cos(\nu)\exp(-\iota\omega) + r^2 \exp(-2\iota\omega) \right|^{-2}$$

where

$$c(r) = \left(1 - r^2\right) \left(\left(1 + r^2\right)^2 - 4r^2\cos^2(\nu) \right) \left(1 + r^2\right)^{-1}.$$

Observe that $\int f_r(\omega)\, d\omega = 1$ where the integral extends over $[-\pi, \pi]$. Hence the $n \times n$ variance covariance matrix $\Sigma(r)$ corresponding to f_r belongs to $\mathfrak{C}_{AR(2)}$. Let $\varepsilon > 0$ be given and set $A(\varepsilon) = \{\omega \in [-\pi, \pi] : |\omega - \nu| \geq \varepsilon\} \cup \{\omega \in [-\pi, \pi] : |\omega + \nu| \geq \varepsilon\}$. Then it is easy to see that

$$\sup_{\omega \in A(\varepsilon)} |f_r(\omega)| \to 0 \quad \text{for} \quad r \to 1.$$

Consequently, for every $\delta > 0$ and every $\varepsilon > 0$ there exists an $0 < r(\varepsilon, \delta) < 1$ such that

$$\int_{[-\pi,\pi] \backslash A(\varepsilon)} f_r(\omega)\, d\omega > 1 - \delta$$

holds for all r satisfying $r(\varepsilon, \delta) < r < 1$. In view of symmetry of f_r around $\omega = 0$, this shows that for r sufficiently close to 1 the spectral density f_r is arbitrarily small outside of the union of the neighborhoods $|\omega - \nu| < \varepsilon$ and $|\omega + \nu| < \varepsilon$ and puts mass arbitrarily close to 1/2 on each one of the two neighborhoods. A standard argument then shows for every continuous function g on $[-\pi, \pi]$ that

$$\int_{[-\pi,\pi]} g(\omega) f_r(\omega)\, d\omega \to 0.5g(\nu) + 0.5g(-\nu) = \int_{[-\pi,\pi]} g(\omega)\, d\left(0.5\delta_\nu + 0.5\delta_{-\nu}\right)$$

where δ_x denotes unit pointmass at x. Specializing to $g(\omega) = \exp(-\iota l\omega)$ shows that $\Sigma(r)$ converges to $E(\nu)E(\nu)'$. ∎

74

Using the arguments in the above proof it is actually not difficult to show that the closure of the set of AR(2)-spectral densities in the weak topology is the class of AR(2)-spectral densities plus all spectral measures of the form $0.5\delta_\nu + 0.5\delta_{-\nu}$ for $\nu \in [0, \pi]$. This result extends in an obvious way to higher-order autoregressive models and has an appropriate generalization to (multivariate) autoregressive moving average models, see Theorem 4.1 in Deistler and Pötscher (1984).

References

ANDERSON, T. W. (1971). *The statistical analysis of time series.* Wiley Series in Probability and Mathematical Statistics, Wiley New York.

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica,* **59** 817–858.

ANDREWS, D. W. K. and MONAHAN, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica,* **60** 953–966.

BAKIROV, N. and SZÉKELY, G. (2005). Student's t-test for gaussian scale mixtures. *Zapiski Nauchnyh Seminarov POMI,* **328** 5–19.

BANERJEE, A. N. and MAGNUS, J. R. (2000). On the sensitivity of the usual t- and F-tests to covariance misspecification. *Journal of Econometrics,* **95** 157 – 176.

BARTLETT, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika,* **37** 1–16.

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics,* **2** 489–502.

CRIBARI-NETO, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis,* **45** 215 – 233.

DEISTLER, M. and PÖTSCHER, B. M. (1984). The behaviour of the likelihood function for ARMA models. *Advances in Applied Probability,* **16** 843–866.

DEN HAAN, W. J. and LEVIN, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust Inference* (G. Maddala and C. Rao, eds.), vol. 15 of *Handbook of Statistics.* Elsevier, 299 – 342.

DUFOUR, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica,* **65** 1365–1387.

DUFOUR, J.-M. (2003). Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique,* **36** 767–808.

EICKER, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Ann. Math. Statist.,* **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).* Univ. California Press, Berkeley, Calif., Vol. I: Statistics, pp. 59–82.

FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.*, **38** 1034–1070.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical analysis of stationary time series*. John Wiley & Sons, New York.

HANNAN, E. (1970). *Multiple time series*. Wiley Series in Probability and Mathematical Statistics, Wiley New York.

HANNAN, E. J. (1957). The variance of the mean of a stationary process. *Journal of the Royal Statistical Society. Series B*, **19** 282–285.

HANSEN, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica*, **60** 967–972.

HEIDELBERGER, P. and WELCH, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM*, **24** 233–245.

IBRAGIMOV, R. and MÜLLER, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics*, **28** 453–468.

JANSSON, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory*, **18** 1449–1459.

JANSSON, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, **72** 937–946.

JOWETT, G. H. (1955). The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society. Series B*, **17** 208–227.

KEENER, R. W., KMENTA, J. and WEBER, N. C. (1991). Estimation of the covariance matrix of the least-squares regression coefficients when the disturbance covariance matrix is of unknown form. *Econometric Theory*, **7** 22–45.

KELEJIAN, H. H. and PRUCHA, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, **140** 131 – 154.

KELEJIAN, H. H. and PRUCHA, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, **157** 53 – 67.

KIEFER, N. M. and VOGELSANG, T. J. (2002a). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, **70** 2093–2095.

KIEFER, N. M. and VOGELSANG, T. J. (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, **18** 1350–1366.

KIEFER, N. M. and VOGELSANG, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, **21** 1130–1164.

KIEFER, N. M., VOGELSANG, T. J. and BUNZEL, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, **68** 695–714.

KRÄMER, W. (1989). On the robustness of the F-test to autocorrelation among disturbances. *Economics Letters*, **30** 37 – 40.

KRÄMER, W. (2003). The robustness of the F-test to spatial autocorrelation among regression disturbances. *Statistica (Bologna)*, **63** 435–440 (2004).

KRÄMER, W., KIVIET, J. and BREITUNG, J. (1990). The null distribution of the F-test in the linear regression model with autocorrelated disturbances. *Statistica (Bologna)*, **50** 503–509.

KRÄMER, W. and HANCK, C. (2009). More on the F-test under nonspherical disturbances. In *Statistical Inference, Econometric Analysis and Matrix Algebra* (B. Schipp and W. Krämer, eds.). Physica-Verlag HD, 179–184.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing statistical hypotheses.* 3rd ed. Springer Texts in Statistics, Springer, New York.

LONG, J. S. and ERVIN, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, **54** 217–224.

MAGEE, L. (1989). An Edgeworth test size correction for the linear model with AR(1) errors. *Econometrica*, **57** 661–674.

MARTELLOSIO, F. (2010). Power properties of invariant tests for spatial autocorrelation in linear regression. *Econometric Theory*, **26** 152–186.

NEAVE, H. R. (1970). An improved formula for the asymptotic variance of spectrum estimates. *The Annals of Mathematical Statistics*, **41** 70–77.

NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55** 703–708.

NEWEY, W. K. and WEST, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, **61** 631–653.

PARK, R. E. and MITCHELL, B. M. (1980). Estimating the autocorrelated error model with trended data. *Journal of Econometrics*, **13** 185–201.

PERRON, P. and REN, L. (2011). On the irrelevance of impossibility theorems: the case of the long-run variance. *J. Time Ser. Econom.*, **3** Art. 1, 34.

PHILLIPS, P. C. (2005). HAC estimation by automated regression. *Econometric Theory*, **21** 116–142.

PHILLIPS, P. C., SUN, Y. and JIN, S. (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review*, **47** 837–894.

PHILLIPS, P. C., SUN, Y. and JIN, S. (2007). Long run variance estimation and robust regression testing using sharp origin kernels with no truncation. *Journal of Statistical Planning and Inference*, **137** 985–1023.

Politis, D. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, **27** 703–744.

Preinerstorfer, D. (2013). On size and power of heteroscedasticity and autocorrelation robust tests: Further results. *Working Paper, Department of Statistics, University of Vienna. In preparation.*

Preinerstorfer, D. and Pötscher, B. M. (2013). On the power of invariant tests for hypotheses on a covariance matrix. *Working Paper, Department of Statistics, University of Vienna. In preparation.*

Pötscher, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica*, **70** 1035–1065.

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

Sun, Y. (2012). Heteroskedasticity and autocorrelation robust F test using orthonormal series variance estimators. *Working Paper, Department of Economics, UC San Diego.*

Sun, Y. and Kaplan, D. M. (2012). Fixed-smoothing asymptotics and accurate F approximation using vector autoregressive covariance matrix estimators. *Working Paper, Department of Economics, UC San Diego.*

Sun, Y., Phillips, P. C. and Jin, S. (2011). Power maximization and size control in heteroskedasticity and autocorrelation robust tests with exponentiated kernels. *Econometric Theory*, **27** 1320–1368.

Sun, Y., Phillips, P. C. B. and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, **76** 175–194.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, **70** 1055–1096.

Velasco, C. and Robinson, P. M. (2001). Edgeworth expansions for spectral density estimates and studentized sample mean. *Econometric Theory*, **17** 497–539.

Vogelsang, T. J. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, **166** 303 – 319.

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

Zhang, X. and Shao, X. (2013). Fixed-smoothing asymptotics for time series. *Ann. Statist., forthcoming.*