Journal of Econometrics 207 (2018) 285–306

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Subvector inference when the true parameter vector may be near or at the boundary[☆]

Philipp Ketz

Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France

ARTICLE INFO

Article history:
Received 12 May 2016
Received in revised form 8 January 2018
Accepted 23 August 2018
Available online 5 September 2018

JEL classification:
C12

Keywords:
Boundary
Asymptotic normality
Admissibility
Random coefficients

ABSTRACT

Extremum estimators are not asymptotically normally distributed when the estimator satisfies the restrictions on the parameter space – such as the non-negativity of a variance parameter – and the true parameter vector is near or at the boundary. This *possible* lack of asymptotic normality makes it difficult to construct tests for testing subvector hypotheses that control asymptotic size in a uniform sense and have good local asymptotic power irrespective of whether the true parameter vector is at, near, or far from the boundary. We propose a novel estimator that is asymptotically normally distributed even when the true parameter vector is near or at the boundary and the objective function is not defined outside the parameter space. The proposed estimator allows the implementation of a new test based on the Conditional Likelihood Ratio statistic that is easy-to-implement, controls asymptotic size, and has good local asymptotic power properties. Furthermore, we show that the test enjoys certain asymptotic optimality properties when the parameter of interest is scalar. In an application of the random coefficients logit model (Berry, Levinsohn and Pakes, 1995) to the European car market, we find that, for most parameters, the new test leads to tighter confidence intervals than the two-sided t-test commonly used in practice.

1. Introduction

It is well-known that extremum estimators are not asymptotically normally distributed when the true parameter vector is at the boundary of the parameter space (see e.g., Geyer, 1994; Andrews, 1999, and references therein). While there exist tests that account for this lack of asymptotic normality (Andrews, 2001), they require knowledge about whether nuisance parameters, which are not specified under the null hypothesis, are at the boundary or not. This knowledge is "required" because the asymptotic distributions of the underlying test statistics display a discontinuity at the boundary. In practice, however, this knowledge is not available and tests suffer from (asymptotic) under- or overrejection, depending on the choice of the test statistic and the exact specification of the testing problem, when the true value of the nuisance parameters is incorrectly specified in the construction of critical values.[1] The problem is amplified by the possibility that the nuisance parameters may be near the boundary *relative* to the sample size. Andrews and Guggenberger (2010b) show, for example, that resampling methods, such as the *m*-out-of-*n* bootstrap, do not control asymptotic size in a uniform sense

☆ A previous version of this paper was circulated under the title "Testing near or at the Boundary of the Parameter Space." I thank an anonymous associate editor and two referees for helpful comments and suggestions that have considerably improved the paper. I also thank Frank Kleibergen, Adam McCloskey, Blaise Melly, and Eric Renault for their advice and helpful discussions. I have also received valuable feedback from Andrew Elzinga, Joachim Freyberger, Bruno Gasperini, Hyojin Han, Pepe Montiel-Olea, and Daniela Scida.

E-mail address: philipp.ketz@psemail.eu.

1 See Section 4 and Appendix D for illustrative examples.

https://doi.org/10.1016/j.jeconom.2018.08.003
0304-4076/

when a scalar nuisance parameter may be near or at the boundary.[2] While it is possible to modify existing tests to ensure asymptotic size control and/or to increase local asymptotic power by adapting critical values, possibly in a data-driven way (see e.g., McCloskey, 2017), such modifications typically are computationally expensive. Perhaps due to these difficulties, researchers sometimes ignore restrictions on nuisance parameters *and* parameters of interest altogether, using critical values that assume that the true parameter vector is in the interior of the parameter space. A prominent example is the random coefficients logit model (Berry, Levinsohn and Pakes, 1995), where researchers typically rely on the standard two-sided t-test to make inference about mean *and* variance parameters, which are known to be non-negative (see e.g., Nevo, 2001; Goeree, 2008).

One main contribution of this paper is to note that, under quite general assumptions, it is possible to obtain an asymptotically normal estimator even when the true parameter vector is near or at the boundary. The proposed estimator does not require the objective function to be defined outside the parameter space and is given by the unconstrained minimizer of the sample analogue of a quadratic approximation to the objective function, which also underlies the asymptotic distribution results in Andrews (1999).[3] It is easy to implement and obtained by a single Newton–Raphson-like iteration starting at the *constrained* extremum estimator.[4] While such an estimator is less attractive than a constrained estimator in terms of point estimation, as it may take on values outside the parameter space, it is useful in the construction of tests; in particular, when interest lies in testing subvector hypotheses, as its asymptotic distribution does not display any discontinuities. In that sense, it makes it unnecessary to consider potentially difficult-to-implement modifications to existing tests.

In order to exploit the information contained in the asymptotically normal estimator as well as in the restrictions on the parameter space for testing subvector hypotheses, we suggest using the Conditional "Likelihood Ratio" (CLR) statistic (named for its asymptotic behavior). The test we propose relies on the conditionality principle of Moreira (2003). The idea is to consider a transformation of the asymptotically normal estimator that yields two orthogonal subvectors, one of which pertains to the parameters of interest while the other constitutes an asymptotically sufficient statistic for the nuisance parameters that are not specified under the null hypothesis. Given the joint asymptotic normality, the orthogonality implies that the two subvectors are asymptotically independent such that asymptotically the conditional null distribution of the CLR statistic given the aforementioned sufficient statistic is nuisance parameter free. Therefore, the CLR test rejects the null hypothesis if the CLR statistic exceeds an appropriately defined *conditional* critical value. We show that the confidence set obtained by inverting the CLR test controls asymptotic size in a uniform sense, by verifying the high-level assumptions in Andrews, Cheng and Guggenberger (2011). Furthermore, we show that, under some conditions, an "asymptotic version" of the CLR test defined in the Gaussian shift model is admissible and *essentially* weighted average power (WAP) maximizing subject to a similarity constraint (Montiel-Olea, 2018) when the parameter of interest is scalar. We, then, show that the CLR test inherits these optimality properties asymptotically, in the sense of Müller (2011).

Recently, Elliott, Müller and Watson (2015) (EMW) and Montiel-Olea (2018) (MO) have suggested alternative tests for testing subvector hypotheses in the Gaussian shift model.[5] EMW propose tests that nearly maximize WAP in the class of level α tests, while MO proposes tests that maximize WAP subject to a similarity constraint. WAP maximizing tests are attractive when a researcher has a particular weight function in mind, where the weight function specifies the alternatives towards which the test directs power. In order to compare our proposed test with the two tests proposed by EMW and MO for certain choices of the weight function, we graphically evaluate their power functions. The analysis is done in the Gaussian shift model and, thus, provides local asymptotic power functions of appropriately defined tests based on the novel, asymptotically normal estimator. The comparison with the test suggested by MO reveals that the weights implicitly underlying the CLR test are attractive. And while the test proposed by EMW has good power, it has the disadvantage of being computationally expensive when the dimension of the nuisance parameter is large. Given its prevalence in applied work, we also analyze an "asymptotic version" of the two-sided t-test that ignores possible boundary effects on the distribution of the underlying *constrained* estimator. We find that it, in many cases, underrejects (see also Andrews and Guggenberger, 2010a, b), but that it can also suffer from overrejection, notably if the dimension of the nuisance parameter is large. Based on these findings, we recommend the CLR test for testing subvector hypotheses in practice: It is easy to implement, controls asymptotic size, has good local asymptotic power properties, and is computationally cheap.

In order to illustrate the usefulness of the CLR test for empirical work, we turn to the random coefficients logit model, which is widely used in the industrial organization and marketing literatures. While there is good reason to believe that not all variance parameters are equal to zero, because in that case the model reduces to the simple multinomial logit model, which is known to suffer from the Independence of Irrelevant Alternatives (IIA), it is seldom known *a priori* which product characteristics interact with a random coefficient. And in a baseline specification, where all product characteristics are interacted with a random coefficient, estimates of variance parameters are often found to be small. The application of the random coefficients logit model to the European car market in Reynaert and Verboven (2014) is no exception, and the

2 Relatedly, Andrews (2000) shows that the asymptotic distribution of a constrained estimator cannot be consistently estimated when the true parameter vector may be near or at the boundary.

3 To be precise, Andrews (1999) considers a quadratic approximation around a *fixed* true parameter, while we follow Andrews and Cheng (2012a) and consider a quadratic approximation around a drifting *sequence* of true parameters; see Section 2 for more details.

4 The proposed estimator has recently been applied by Frazier and Renault (2016) in the context of Indirect-Inference.

5 The tests proposed in Moreira and Moreira (2013) could also be applied. However, given the existence of an asymptotically sufficient statistic for the nuisance parameter, we do not consider them here.

two-sided t-test, which represents common practice, only suggest the presence of consumer heterogeneity with respect to one out of six product characteristics, namely horse power, at the 10% significance level. Using the CLR test, we find evidence of additional consumer heterogeneity with respect to price and height of the car. Furthermore, we obtain shorter confidence intervals for most of the mean parameters. This illustrates that, in practice, the CLR test can offer valuable power gains over the two-sided t-test while ensuring asymptotic size control.

The plan of this paper is as follows. In Section 2, we introduce the new estimator and show that it is asymptotically normal. Section 3 introduces the testing problem and our proposed test. It also presents the optimality results for the CLR test defined in the Gaussian shift model. Section 4 contains the power comparison of the CLR test and several alternative tests in the Gaussian shift model, with some results relegated to Appendix D. Section 5 presents an application to the random coefficients logit model. In this section, we perform a small Monte Carlo study to illustrate the finite sample behavior of the CLR test and present our empirical findings. Proofs are collected in Appendices A and B. Results concerning asymptotic properties of the CLR test are relegated to Appendix C.

Throughout this paper, we use the following notational conventions. For any set A and any $k \in \mathbb{N}$, $A^k = A \times \cdots \times A$ denotes the Cartesian product of k copies of A. Furthermore, I_k denotes the $(k \times k)$ identity matrix. For any $a \in \mathbb{R} \cup \{\pm\infty\}$, $a_k = (a, \ldots, a)'$ denotes the k-dimensional vector whose entries are all equal to a. Sometimes, we write $0 = 0_k$ if the dimension of 0_k is clear from the context. For any two column vectors a and b, we sometimes write (a, b) instead of $(a', b')'$ and let $a \geq b$ denote the element-by-element inequality. For any matrix A, A_{ij} denotes the entry with row index i and column index j. Similarly, a_i denotes the ith entry of vector a. We let $\text{int}(A)$ and $\text{bd}(A)$ denote the interior and the boundary of the set A, respectively. Furthermore, "$\equiv$" denotes "equals by definition". Lastly, $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively, while all limits are taken as "$n \to \infty$", unless otherwise noted.

2. An asymptotically normal estimator

We consider a general class of extremum estimators, including, for example, (Quasi-) Maximum Likelihood (Q-ML) and Generalized Method of Moments (GMM) estimators. The objective function, which is parameterized by the finite-dimensional $(J \times 1)$ parameter θ, is denoted $Q_n(\theta)$ and depends on the data matrix W_n whose column dimension is fixed and whose n rows may be i.i.d., independent and nonidentically distributed, or temporally dependent. Andrews (1999) derives the asymptotic distribution of the *constrained* estimator $\hat{\theta}_n$, which is defined as the (approximate) minimizer of $Q_n(\theta)$ over Θ, i.e., $\hat{\theta}_n \in \Theta$ and

$$Q_n(\hat{\theta}_n) = \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1/n),$$

when the true parameter vector, say $\bar{\theta}$, is at the boundary of Θ, i.e., $\bar{\theta} \in \text{bd}(\Theta)$. Here, $\Theta \subset \mathbb{R}^J$ denotes the true parameter space, i.e., the set of all possible values for $\bar{\theta}$ as specified by the researcher. While, under quite general assumptions, $\hat{\theta}_n$ is $\sqrt{n}$-consistent for $\bar{\theta}$, $\sqrt{n}(\hat{\theta}_n - \bar{\theta})$ is generally not asymptotically normal when $\bar{\theta} \in \text{bd}(\Theta)$.[6] This contrasts with the situation where $\bar{\theta} \in \text{int}(\Theta)$, in which case $\sqrt{n}(\hat{\theta}_n - \bar{\theta})$ is asymptotically normal. This discontinuity in the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \bar{\theta})$ makes it difficult to construct tests for testing subvector hypotheses that control asymptotic size in a uniform sense and have good local asymptotic power irrespective of whether $\bar{\theta}$ or, rather, its subvector that is *not* specified under the null hypothesis is at the boundary, near the boundary, or far from the boundary *relative* to the sample size.

In order to derive asymptotic theory that provides good approximations to the finite sample distributions of estimators and test statistics when the true parameter may be near or at the boundary, we rely on drifting *sequences* of true parameters $\theta_n \in \Theta$ with $\theta_n \to \theta^* \in \Theta$ that are allowed to change with the sample size n. Here and in what follows, the superscript * indicates a *fixed* limit point that does not change with n. Of particular interest are sequences of true parameters that drift towards the boundary, $\theta_n \to \theta^* \in \text{bd}(\Theta)$. For example, if $\Theta = [0, \infty)$, the asymptotic distribution theory obtained for sequences of true parameters of the form $\theta_n = \frac{\mu}{\sqrt{n}}$ with *fixed* $\mu \in [0, \infty)$, such that $\theta_n \to \theta^* = 0$, provides good approximations to the finite sample behavior of, say, an estimator when the true parameter is near (or at) the boundary *relative* to the sample size. The notion of "good approximations" is formalized in results concerning uniformity, where drifting sequences of true parameters have been shown to play a crucial role (see e.g., Andrews and Guggenberger, 2010b).

In most applications, the distribution of the data W_n is not fully specified by θ, but depends on an additional, commonly infinite-dimensional parameter, say ω. For example, in the context of conditional ML estimation ω indexes the distribution of the conditioning variables. The parameter $\gamma \equiv (\theta, \omega)$ fully specifies the distribution of the data W_n and the corresponding true parameter space is compact and of the following form

$$\Gamma = \{\gamma = (\theta, \omega) : \theta \in \Theta, \omega \in \Omega(\theta)\},$$

where Θ is compact and $\Omega(\theta) \subset \Omega \; \forall \theta \in \Theta$ for some compact metric space Ω with a metric that induces weak convergence of $(W_{n,i}, W_{n,i+m})$ for all $i, m \geq 1$, i.e., the metric is such that if $\gamma^1 \to \gamma^2$, then $(W_{n,i}, W_{n,i+m})$ under γ^1 converges in distribution

6 Unless Θ imposes only linear equality constraints.

to $(W_{n,i}, W_{n,i+m})$ under γ^2 for all $i, m \geq 1$.[7] Here, $W_{n,i}$ denotes the ith row of W_n. The assumptions listed below pertain to sequences of true parameters $\gamma_n = (\theta_n, \omega_n)$ in the set

$$\Gamma(\gamma^*) = \{\{\gamma_n \in \Gamma : n \geq 1\} : \gamma_n \to \gamma^* \in \Gamma\},$$

where $\gamma^* = (\theta^*, \omega^*)$. Throughout this paper, we use the shorthand "under $\{\gamma_n\} \in \Gamma(\gamma^*)$" to mean "when the true parameters are $\{\gamma_n\} \in \Gamma(\gamma^*)$ for *any* $\gamma^* \in \Gamma$". This general framework, which makes the dependence on ω explicit and which we borrow from Andrews and Cheng (2012a), facilitates the derivation of asymptotic results that hold uniformly over Γ.

The asymptotic distribution theory in Andrews (1999) relies on a quadratic approximation of the objective function around the true parameter. For the reasons outlined above, we follow Andrews and Cheng (2012a) and consider a quadratic approximation around a drifting sequence of true parameters, where, instead of allowing for lack of identification at some point in the parameter space, we allow the sequence of true parameters to drift towards the boundary, i.e., $\theta_n \to \theta^* \in \mathrm{bd}(\Theta)$, in the spirit of Andrews (1999). In particular, we rely on the following quadratic expansion

$$Q_n(\theta) = Q_n(\theta_n) + DQ_n(\theta_n)'(\theta - \theta_n) + \frac{1}{2}(\theta - \theta_n)'D^2Q_n(\theta_n)(\theta - \theta_n) + R_n(\theta), \tag{1}$$

where the remainder, $R_n(\theta)$, is assumed to satisfy

$$\sup_{\theta \in \Theta : \|\sqrt{n}(\theta - \theta_n)\| \leq \epsilon} |R_n(\theta)| = o_p(1/n)$$

for all constants $0 < \epsilon < \infty$, under $\{\gamma_n\} \in \Gamma(\gamma^*)$. Here, $DQ_n(\theta)$ and $D^2Q_n(\theta)$ denote the generalized first- and second-order partial derivatives of the objective function, respectively. $DQ_n(\theta)$ and $D^2Q_n(\theta)$ are *generalized* in the sense that Eq. (1) does not require them to be partial derivatives. In particular, $Q_n(\theta)$ may not be defined outside the parameter space, as generally the case in random coefficients models, such that, at best, $DQ_n(\theta)$ and $D^2Q_n(\theta)$ may denote first- and second-order left/right partial derivatives (Andrews, 1999).[8] Furthermore, $Q_n(\theta)$ may not be smooth but only "stochastically differentiable" (Pollard, 1985), as for example the case in quantile estimation, and/or given by a GMM (or minimum distance) objective function (see e.g., Pakes and Pollard, 1989).[9,10]

The quadratic expansion in Eq. (1) can be used to derive asymptotic distribution theory for the constrained estimator if the latter satisfies the following high-level assumption.

Assumption 1. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$, $\sqrt{n}(\hat{\theta}_n - \theta_n) = O_p(1)$.

Sufficient conditions for Eq. (1) and Assumption 1 can, for example, be found in Andrews (1999, 2001, 2002), where only minor modifications are required to accommodate drifting sequences of true parameters (cf. Andrews and Cheng, 2012a, b, 2014).

The reason for relying on a quadratic approximation rather than a (approximate) first order condition of the minimization problem is that the latter does not hold (with probability approaching 1) under certain drifting sequences of true parameters that are such that $\theta_n \to \theta^* \in \mathrm{bd}(\Theta)$. This is a direct consequence of the definition of $\hat{\theta}_n$, which is constrained to lie in Θ and whose (asymptotic) distribution is, therefore, subject to boundary effects. While an unconstrained estimator is often unavailable, as $Q_n(\theta)$ may not be defined outside Θ, inspection of Eq. (1) reveals that the quadratic approximation of $Q_n(\theta)$, i.e., $Q_n(\theta) - R_n(\theta)$, does permit an unconstrained minimizer, namely $\theta_n - \left(D^2Q_n(\theta_n)\right)^{-1}DQ_n(\theta_n)$. Furthermore, it is easy to see that this minimizer (appropriately centered and scaled) is asymptotically normally distributed, under $\{\gamma_n\} \in \Gamma(\gamma^*)$, if the following two assumptions hold.

Assumption 2. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$, $\sqrt{n}DQ_n(\theta_n) \xrightarrow{d} N(0, V(\gamma^*))$, where $V(\gamma^*)$ is symmetric and positive-definite.

Assumption 3. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$, $D^2Q_n(\theta_n) \xrightarrow{p} \mathcal{J}(\gamma^*)$, where $\mathcal{J}(\gamma^*)$ is symmetric and nonsingular.

However, this "estimator" is infeasible because it depends on the unknown true parameter, θ_n. In addition, $DQ_n(\theta)$ and $D^2Q_n(\theta)$ may not be given by the first- and second-order (left/right) partial derivatives of $Q_n(\theta)$ and may, thus, also be

[7] Note that Γ is a metric space with metric $d_\Gamma(\gamma_1, \gamma_2) = \|\theta_1 - \theta_2\| + d_\Omega(\omega_1, \omega_2)$, where $\gamma_j = (\theta_j, \omega_j) \in \Gamma$ for $j = 1, 2$ and where d_Ω denotes the metric on Ω.

[8] A function $f(x)$ has left/right partial derivatives (of order one) on $\mathcal{X} \subset \mathbb{R}^k$ for some $k \in \mathbb{N}$, if it (i) has partial derivatives at each $x \in \mathrm{int}(\mathcal{X})$, (ii) partial derivatives at each $x \in \mathrm{bd}(\mathcal{X})$ with respect to coordinates that can be perturbed to the left and the right, and (iii) left (right) partial derivatives at each $x \in \mathrm{bd}(\mathcal{X})$ with respect to coordinates that can only be perturbed to the left (right).

[9] Andrews (1999, 2001) and Andrews (2002) extend the asymptotic distribution theory in Pollard (1985) and Pakes and Pollard (1989), respectively, to allow for the true parameter to be at the boundary and for the objective function to not be defined outside the parameter space. While the respective "stochastic differentiability" and "stochastic equicontinuity" conditions are not given in terms of drifting sequences of true parameters, the required modifications are minor and obtained along the lines of Andrews and Cheng (2012a, b) and Andrews and Cheng (2014), respectively.

[10] Note that Eq. (1) also allows for initial conditions adjustment to the objective function (necessary in certain time series models) as in Andrews (2001), see also (Andrews and Cheng, 2012b).

"unknown". For example, if $Q_n(\theta)$ is a nonsmooth sample average, then $DQ_n(\theta)$ is the "stochastic derivative" of $Q_n(\theta)$, while $D^2Q_n(\theta)$ equals the second-order (left/right) partial derivatives of the expected value of $Q_n(\theta)$ (cf. Andrews and Cheng, 2012a, b).[11] Replacing (possibly) unknowns with estimators, a *feasible* version is given by

$$\tilde{\theta}_n = \hat{\theta}_n - \left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}\widehat{DQ}_n(\hat{\theta}_n), \tag{2}$$

where $\widehat{DQ}_n(\hat{\theta}_n)$ and $\widehat{D^2Q}_n(\hat{\theta}_n)$ are assumed to satisfy the following high-level assumption.

Assumption 4. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$, (i) $\left\|\widehat{DQ}_n(\hat{\theta}_n) - DQ_n(\theta_n) - D^2Q_n(\theta_n)(\hat{\theta}_n - \theta_n)\right\| = o_p(1/\sqrt{n})$ and (ii) $\left\|\widehat{D^2Q}_n(\hat{\theta}_n) - D^2Q_n(\theta_n)\right\| = o_p(1)$.

In many cases, the sufficient conditions for Eq. (1) and Assumption 1 can also be used to verify Assumption 4.[12] If $Q_n(\theta)$ is nonsmooth, $\widehat{DQ}_n(\hat{\theta}_n)$ and $\widehat{D^2Q}_n(\hat{\theta}_n)$ typically involve numerical derivatives (see e.g., Pakes and Pollard, 1989).

The following Theorem states the first main result of this paper.[13]

Theorem 1. *Under $\{\gamma_n\} \in \Gamma(\gamma^*)$ and Assumptions 1–4,*

$$\sqrt{n}(\tilde{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{Z}(\gamma^*) \equiv N(0, \Sigma(\gamma^*)),$$

where $\Sigma(\gamma^*) = \mathcal{J}(\gamma^*)^{-1}V(\gamma^*)\mathcal{J}(\gamma^*)^{-1}$.

The proof of Theorem 1 is given in Appendix A. In many cases, the results in Andrews (1999) suitably adapted imply that, under certain drifting sequences of true parameters (and Assumptions 1–3), the asymptotic distribution of the constrained estimator, $\sqrt{n}(\hat{\theta}_n - \theta_n)$, is given by the unique projection of $\mathcal{Z}(\gamma^*)$ onto a cone, with respect to the norm $\|\lambda\| = (\lambda'\mathcal{J}(\theta^*)\lambda)^{1/2}$. Furthermore, if an unconstrained estimator is available, it is generally asymptotically equivalent to $\tilde{\theta}_n$.[14] Consequently, $\tilde{\theta}_n$ can be considered a *quasi* unconstrained estimator. The advantage of a *quasi* unconstrained estimator over a constrained estimator for testing is twofold. First, it simplifies the testing problem, due to the absence of discontinuities in its asymptotic distribution. Second, it, in some sense, contains more "information". To see this, note that any test statistic based on $\hat{\theta}_n$ can also be constructed using $\tilde{\theta}_n$, simply by considering its appropriately defined projection onto Θ, but not vice versa.

Before turning to our proposed test that exploits this additional "information", we introduce an estimator for $\Sigma(\gamma^*)$ that will be used in its construction. Define

$$\hat{\Sigma} \equiv \left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}\hat{V}_n(\hat{\theta}_n)\left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1},$$

where $\hat{V}_n(\hat{\theta}_n)$ is assumed to satisfy the following high-level assumption.

Assumption 5. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$, $\hat{V}_n(\hat{\theta}_n) \xrightarrow{p} V(\gamma^*)$.

Assumptions 1–5 imply that $\hat{\Sigma} \xrightarrow{p} \Sigma(\gamma^*)$ under $\{\gamma_n\} \in \Gamma(\gamma^*)$ such that $\hat{\Sigma}$ is positive-definite with probability approaching 1. For ease of exposition, we assume $\hat{\Sigma}$ to be positive-definite and, thus, invertible in what follows.

3. A conditional likelihood ratio test

For the remainder of this paper, we assume that Θ is given by a $(J \times 1)$ Cartesian product of intervals equal to $[0, c]$ or $[-c, c]$ for some $c < \infty$, where the use of a common end point is merely for notational ease. The boundary of the parameter space is normalized to be "on the left" and at 0, without loss of generality.[15] If $\theta_j \in [0, c]$ for $j \in \{1, \ldots, J\}$, we say that θ_j is

---

11 In the context of GMM, it is convenient to take $DQ_n(\theta)$ and $D^2Q_n(\theta)$ equal to $G_\theta'WG_n(\theta)$ and $G_\theta'WG_\theta$, respectively, where G_θ denotes the (left/right) partial derivatives of the limit of $G_n(\theta)$ under $\{\gamma_n\} \in \Gamma(\gamma^*)$ evaluated at θ^*, W denotes the limit of the weighting matrix under $\{\gamma_n\} \in \Gamma(\gamma^*)$ (possibly evaluated at θ^*), and $G_n(\theta)$ denotes the sample moment (cf. Andrews and Cheng, 2014).

12 For example, if $Q_n(\theta)$ has *continuous* (left/right) partial derivatives of order two on (a suitably defined superset of) Θ, then Eq. (1) and Assumption 4 can be shown to hold with $\widehat{DQ}_n(\theta) = DQ_n(\theta) = \frac{\partial}{\partial\theta}Q_n(\theta)$ and $\widehat{D^2Q}_n(\theta) = D^2Q_n(\theta) = \frac{\partial^2}{\partial\theta'\partial\theta}Q_n(\theta)$ given the following Assumption: Under $\{\gamma_n\} \in \Gamma(\gamma^*)$,

$$\sup_{\theta\in\Theta:\|\theta-\theta_n\|\le\epsilon_n}\left\|\frac{\partial^2}{\partial\theta'\partial\theta}Q_n(\theta) - \frac{\partial^2}{\partial\theta'\partial\theta}Q_n(\theta_n)\right\| = o_p(1)$$

for all sequences of positive scalar constants $\{\epsilon_n : n \ge 1\}$ for which $\epsilon_n \to 0$ (cf. Lemma 1 in Andrews, 1999). Note that the verification of Assumption 4, in addition, requires that $\hat{\theta}_n$ is consistent, i.e., $\hat{\theta}_n - \theta_n = o_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma^*)$. Furthermore, the above Assumption together with Assumptions 2 and 3 and consistency of $\hat{\theta}_n$ can be shown to imply Assumption 1 (cf. Theorem 1 in Andrews, 1999).

13 So-called one-step estimators, such as the one given in Eq. (2), are also discussed in Newey and McFadden (1994), who consider the standard setting where the true parameter vector is assumed to be in the interior of the parameter space and motivate them by noting that they are "[…] particularly helpful when simple initial estimators can be constructed, but an efficient estimator is more complicated […]". In the nonstandard setting considered here, they are helpful because they are asymptotically normally distributed, whereas the form of the asymptotic variance matrix is only of secondary interest.

14 If the objective function is quadratic in θ, $R_n(\theta) = 0$, then $\tilde{\theta}_n$ coincides with the unconstrained estimator.

15 If the jth entry of the original parameter vector, say ϑ, satisfies $\vartheta_j \in [c_1, c_2]$, where $-\infty < c_1 < c_2 < \infty$, and $\vartheta_j = c_1$ ($\vartheta_j = c_2$) is empirically relevant, take $\theta_j = \vartheta_j - c_1$ ($\theta_j = c_2 - \vartheta_j$).

restricted below by 0. If $\theta_j \in [-c, c]$ we say that θ_j is unrestricted, i.e., c is taken to be large. The reason for considering this form of the parameter space is that it allows us to derive certain optimality results for our proposed test. Furthermore, it is obtained for many models of interest, most notably in the context of random coefficients models.[16]

In what follows, we consider the following partition $\theta = (\beta, \delta)$, where β denotes the $(K \times 1)$ parameter of interest and δ denotes a $(L \times 1)$ nuisance parameter, with $1 \leq K \leq J$ and $L = J - K$. Let $\bar{\theta} = (\bar{\beta}, \bar{\delta})$ and $\Theta = B \times D$ denote the corresponding partitions of the true parameter vector and space, respectively.[17] Formally, we are interested in testing

$$H_0 : \bar{\beta} = \beta_0 \in B, \bar{\delta} \in D \text{ vs. } H_1 : \bar{\beta} \neq \beta_0, \bar{\beta} \in B, \bar{\delta} \in D. \tag{3}$$

We consider the following transformation of the *quasi* unconstrained estimator $X_n \equiv \tilde{\delta}_n - \hat{\Sigma}_{\delta\beta}\hat{\Sigma}_{\beta\beta}^{-1}\tilde{\beta}_n$, where $(\tilde{\beta}_n, \tilde{\delta}_n)$ and

$$\begin{bmatrix} \hat{\Sigma}_{\beta\beta} & \hat{\Sigma}_{\beta\delta} \\ \hat{\Sigma}_{\delta\beta} & \hat{\Sigma}_{\delta\delta} \end{bmatrix}$$

denote conformable partitions of $\tilde{\theta}_n$ and $\hat{\Sigma}$, respectively. For the purpose of testing (3), we suggest using the following Conditional Likelihood Ratio (CLR) statistic

$$\begin{aligned} &\mathrm{CLR}(\beta_0, \tilde{\beta}_n, X_n, \hat{\Sigma}/n, B^{\mathrm{e}}, D^{\mathrm{e}}) \\ &\equiv \inf_{\mathrm{d} \in D^{\mathrm{e}}} \begin{pmatrix} \tilde{\beta}_n - \beta_0 \\ X_n + \hat{\Sigma}_{\delta\beta}\hat{\Sigma}_{\beta\beta}^{-1}\tilde{\beta}_n - \mathrm{d} \end{pmatrix}' \left(\frac{\hat{\Sigma}}{n}\right)^{-1} \begin{pmatrix} \tilde{\beta}_n - \beta_0 \\ X_n + \hat{\Sigma}_{\delta\beta}\hat{\Sigma}_{\beta\beta}^{-1}\tilde{\beta}_n - \mathrm{d} \end{pmatrix} \\ &- \inf_{\mathrm{b} \in B^{\mathrm{e}}, \mathrm{d} \in D^{\mathrm{e}}} \begin{pmatrix} \tilde{\beta}_n - \mathrm{b} \\ X_n + \hat{\Sigma}_{\delta\beta}\hat{\Sigma}_{\beta\beta}^{-1}\tilde{\beta}_n - \mathrm{d} \end{pmatrix}' \left(\frac{\hat{\Sigma}}{n}\right)^{-1} \begin{pmatrix} \tilde{\beta}_n - \mathrm{b} \\ X_n + \hat{\Sigma}_{\delta\beta}\hat{\Sigma}_{\beta\beta}^{-1}\tilde{\beta}_n - \mathrm{d} \end{pmatrix}, \end{aligned} \tag{4}$$

where B^{e} and D^{e} denote "extensions" of B and D such that they equal Cartesian products of intervals equal to $[0, \infty)$ or $(-\infty, \infty)$. This definition of the CLR statistic ensures that only boundaries at 0 are taken into account. The motivation for using the CLR statistic is that it takes the restrictions on the entire parameter vector into account while allowing the construction of a test that controls asymptotic size in a straight-forward manner. The idea of the proposed test relies on the conditioning principle suggested by Moreira (2003). In particular, X_n asymptotically serves as a sufficient statistic for the unknown nuisance parameter $\bar{\delta}$ such that the conditional distribution of the CLR statistic given X_n is approximately nuisance parameter free. An easy-to-implement test is then obtained by using conditional critical values.

Before formally introducing our proposed test, we consider a slight modification of the above CLR statistic. Let s denote a possibly empty subset of $\{1, \ldots, L\}$ with cardinality L^{s}. In what follows, the superscript s also indicates suitably defined subvectors, matrices, and parameter spaces. For example, $D^{\mathrm{s}} = D_1 \times \cdots \times D_i \times \cdots \times D_L$ where $i \in \mathrm{s}$, with D_i denoting the ith coordinate of D. Similarly, $\delta^{\mathrm{s}} = S\delta$, where S denotes a "selection matrix" given by $(e_1, \ldots, e_i, \ldots, e_L)'$ where $i \in \mathrm{s}$ and where e_i denotes the $(L \times 1)$ unit vector with the ith entry equal to 1 and 0s elsewhere. We also let

$$\hat{\Sigma}^{\mathrm{s}} = \begin{bmatrix} \hat{\Sigma}_{\beta\beta} & \hat{\Sigma}_{\beta\delta}^{\mathrm{s}} \\ \hat{\Sigma}_{\delta\beta}^{\mathrm{s}} & \hat{\Sigma}_{\delta\delta}^{\mathrm{s}} \end{bmatrix},$$

where $\hat{\Sigma}_{\beta\delta}^{\mathrm{s}} = \hat{\Sigma}_{\beta\delta}S'$, $\hat{\Sigma}_{\delta\beta}^{\mathrm{s}} = S\hat{\Sigma}_{\delta\beta}$, and $\hat{\Sigma}_{\delta\delta}^{\mathrm{s}} = S\hat{\Sigma}_{\delta\delta}S'$. Then, the "modified" CLR statistic is given by $\mathrm{CLR}(\beta_0, \tilde{\beta}_n, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}})$. The reason for considering this modification is that our asymptotic optimality results below will, in some cases, only hold for certain subsets. Furthermore, our power analysis in Section 4 reveals that it may sometimes be advantageous to consider tests based on subsets. Lastly, note that $\mathrm{CLR}(\beta_0, \tilde{\beta}_n, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}})$ is equal to $\mathrm{CLR}(\beta_0, \tilde{\beta}_n, X_n^{\mathrm{s}'}, \hat{\Sigma}^{\mathrm{s}'}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}'})$ where s′ is such that $D^{\mathrm{e,s}'}$ only includes those entries of $D^{\mathrm{e,s}}$ that equal $[0, \infty)$.[18] In words, if a nuisance parameter is unrestricted, our proposed test statistic and the resulting test are invariant to its true value; a property shared by many standard tests when the true parameter is in the interior of the parameter space.

In what follows, we drop the term "modified" and define the CLR test for testing (3) as follows

$$\varphi_{\mathrm{CLR}}(\beta_0, \tilde{\beta}_n, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}}) = \begin{cases} 1 & \text{if } \mathrm{CLR}(\beta_0, \tilde{\beta}_n, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}}) > \mathrm{cv}_{1-\alpha}(\beta_0, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}}) \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{cv}_{1-\alpha}(\beta_0, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}})$ with $\alpha \in (0, 1)$ denotes the critical value function. Here,

$$\begin{aligned} &\mathrm{cv}_{1-\alpha}(\beta_0, x^{\mathrm{s}}, \boldsymbol{\Sigma}^{\mathrm{s}}, B^{\mathrm{e}}, D^{\mathrm{e,s}}) = \inf\{q \in \mathbb{R} : \\ &P(\mathrm{CLR}(\beta_0, (\hat{\Sigma}_{\beta\beta}/n)^{1/2}Z + \beta_0, X_n^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n, B^{\mathrm{e}}, D^{\mathrm{e,s}}) \leq q | X_n^{\mathrm{s}} = x^{\mathrm{s}}, \hat{\Sigma}^{\mathrm{s}}/n = \boldsymbol{\Sigma}^{\mathrm{s}}) \geq 1 - \alpha\} \end{aligned} \tag{5}$$

16 Including censored panel data models with slope heterogeneity (Abrevaya and Shen, 2014) and GARCH models (Francq and Zakoïan, 2009). See Andrews (1999, 2001) for additional examples. A notable exception is given by random coefficients models that allow the random coefficients to be correlated (Andrews, 2001).

17 Here, we consider *fixed* true parameters in order to introduce our proposed test, but we rely on drifting sequences of true parameters when deriving the corresponding asymptotic theory.

18 This follows from Lemma 1 in Appendix B, assuming $\hat{\Sigma}$ is positive-definite and $\|\tilde{\theta}_n\| < \infty$.

where $Z|X_n^s = x^s$, $\hat{\Sigma}^s/n = \boldsymbol{\Sigma}^s \sim N(0, I_K)$ and where x^s and $\boldsymbol{\Sigma}^s$ denote possible realizations of X_n^s and $\hat{\Sigma}^s/n$, respectively. In Appendix C, we show that, under Assumptions 1–5, the confidence set obtained by inverting the CLR test controls asymptotic size in a uniform sense, over Γ. We also show that, under certain conditions, the stronger result of asymptotic similarity holds.

3.1. Asymptotic optimality results for $K = 1$

While asymptotic size control is a desirable property, we may also ask whether the CLR test enjoys some optimality properties, at least asymptotically. Here, we show that an "asymptotic version" of the CLR test, defined in the relevant Gaussian shift model, is admissible and *essentially* WAP-maximizing subject to a similarity constraint when $K = 1$. In Appendix C, we then show that the CLR test based on $\tilde{\theta}_n$ inherits these optimality properties asymptotically, in the sense of Müller (2011).[19]

We say that $(\sqrt{n}(\tilde{\theta}_n - \theta^*), \hat{\Sigma})$ converges to a Gaussian shift model under $\{\theta_n : n \geq 1\}$ *local* to θ^* if

$$\sqrt{n}(\tilde{\theta}_n - \theta^*) \xrightarrow{d} Y \sim N(\mu, \Sigma) \text{ and } \hat{\Sigma} \xrightarrow{p} \Sigma, \tag{6}$$

where $\mu \in M(\theta^*) \subset \mathbb{R}^J$ and where Σ, suppressing the dependence on γ^*, is positive-definite. Here, μ is *fixed* and denotes the localization parameter, $\mu = \sqrt{n}(\theta_n - \theta^*)$.[20] Solving for $\theta_n = \theta^* + \frac{\mu}{\sqrt{n}}$ explains why we say that Eq. (6) is obtained under $\{\theta_n\}$ *local* to θ^*. Note that $M(\theta^*)$ depends on θ^* in a non-trivial way. In particular, $M(\theta^*)$ equals a Cartesian product of intervals equal to $[0, \infty)$ if $\Theta_j = [0, c]$ and $\theta_j^* = 0$ and $(-\infty, \infty)$ otherwise. We note that Eq. (6) holds under Assumptions 1–5 for suitably defined drifting sequences of true parameters.

Let $\mu = (b, d)$, $\theta_n = (\beta_n, \delta_n)$, $\theta^* = (\beta^*, \delta^*)$, and $M(\theta^*) = B(\beta^*) \times D(\delta^*)$ denote conformable partitions, such that $b = \sqrt{n}(\beta_n - \beta^*)$ and $d = \sqrt{n}(\delta_n - \delta^*)$. Then, the testing problem given in (3) *local* to θ^* can be written as

$$H_0 : b = b_0 \in B(\beta^*), d \in D(\delta^*) \text{ vs. } H_1 : b \neq b_0, b \in B(\beta^*), d \in D(\delta^*), \tag{7}$$

where $b_0 \equiv \sqrt{n}(\beta_{n,0} - \beta^*)$ or, equivalently, $\beta_{n,0} \equiv \beta^* + \frac{b_0}{\sqrt{n}}$ with b_0 fixed, such that $\{\beta_{n,0} : n \geq 1\}$ denotes a drifting sequence of null hypotheses. Here, we consider drifting sequences of true parameters *and* null hypotheses, i.e., $H_0 : \beta_n = \beta_{n,0} = \beta^* + \frac{b_0}{\sqrt{n}}, \delta_n = \delta^* + \frac{d}{\sqrt{n}}$ rather than, say, $H_0 : \bar{\beta} = \beta_0 = \beta^*, \bar{\delta} = \delta^*$, because the testing problem is not invariant to θ^* if $\Theta \neq [-c, c]^J$. In particular, our asymptotic optimality results below cover the case where sequences of null hypotheses with $b_0 > 0$ ($b_0 = 0$) correspond to testing parameter values that are near (or at) the boundary *relative* to the sample size, $B(\beta^*) = [0, \infty)$. Similarly, our results cover the case where nuisance parameters, which are not specified under the null hypothesis, are near (or at) the boundary *relative* to the sample size, $D(\delta^*)_j = [0, \infty)$ for some (or all) $j \in \{1, \ldots, L\}$.

In what follows, we consider tests defined in the Gaussian shift model where Y is observed, with μ unknown and Σ known. Let C_Y denote the class of all tests based on Y for testing (7). Formally, a test is given by a measurable function $\varphi : \mathcal{Y} \to [0, 1]$, where $\mathcal{Y} = \mathbb{R}^J$ denotes the support of Y, and $\varphi(y)$ is to be understood as the probability of rejecting the null hypothesis given a realization of Y, denoted y. Let

$$E_\mu[\varphi(Y)] = \int_{\mathcal{Y}} \varphi(y) dF(y; \theta^*, \mu, \Sigma)$$

denote the power function of φ, where $F(y; \theta^*, \mu, \Sigma)$ denotes the distribution function of Y.[21] Furthermore, let $M_0(\theta^*) = \{b_0\} \times D(\delta^*)$ and $M_1(\theta^*) = M(\theta^*) \setminus M_0(\theta^*)$ such that Eq. (7) can equally be written as $H_0 : \mu \in M_0(\theta^*)$ vs. $H_1 : \mu \in M_1(\theta^*)$. In order to compare tests, we consider the following risk function

$$R_\varphi(\mu) = \begin{cases} E_\mu[\varphi(Y)] & \text{if } \mu \in M_0(\theta^*) \\ 1 - E_\mu[\varphi(Y)] & \text{if } \mu \in M_1(\theta^*), \end{cases}$$

which returns the type I error for $\mu \in M_0(\theta^*)$, i.e., the probability of rejecting the null hypothesis when it is true, and the type II error for $\mu \in M_1(\theta^*)$, i.e., the probability of failing to reject the null hypothesis when it is false. A test φ' is said to *dominate* another test φ if $R_{\varphi'}(\mu) \leq R_\varphi(\mu)$ for all $\mu \in M(\theta^*)$ with strict inequality, $R_{\varphi'}(\mu) < R_\varphi(\mu)$, for some $\mu \in M(\theta^*)$. A test φ is called *admissible* in a class of tests $C_Y^* \subset C_Y$ if there exists no $\varphi' \in C_Y^*$ such that φ' dominates φ. Admissibility is a minimal optimality requirement for a test within a certain class of tests.

The following Theorem states the second main result of this paper. Define $X \equiv Y_d - \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1} Y_b$, where $Y = (Y_b, Y_d)$ and where $\Sigma_{\delta\beta}$ and $\Sigma_{\beta\beta}$ denote the corresponding blocks of Σ. Note that (Y_b, X) is a one-to-one function of Y, with Σ known and positive-definite. Then, the CLR test based on Y or, equivalently, based on (Y_b, X) for testing (7) is given by $\varphi_{\mathrm{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$, where, without loss of generality (by Lemma 1 in Appendix B), s is a subset such that $D(\delta^*)^s = [0, \infty)^{L^s}$.

19 The framework of Müller (2011) is well-suited to derive asymptotic optimality results in the context of extremum estimation as it does not necessitate the specification of a (semi-)parametric statistical model (for W_n), which underlies Le Cam's limit of experiments framework (see e.g., Van der Vaart, 1998).

20 Some authors suppress the dependence of the true parameter on the sample size and write $\mu = \sqrt{n}(\bar{\theta} - \theta^*)$.

21 We write $F(y; \theta^*, \mu, \Sigma)$ to highlight the dependence of (the parameter space for) μ on θ^*.

Theorem 2. *$\varphi_{CLR}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is admissible in the class of tests C_Y if* (i) *$B(\beta^*) = (-\infty, \infty)$ or $B(\beta^*) = [0, \infty)$ and $b_0 > 0$ and if* (ii) *$B(\beta^*) = [0, \infty)$ and $b_0 = 0$, as long as s is a subset for which $\Sigma^s_{\delta\beta} \geq 0$ and $\alpha \leq 0.5$.*

Note that conditions (i) and (ii) in Theorem 2 implicitly impose $K = 1$. The proof of Theorem 2 is given in Appendix B and consists of two steps. First, it is shown that any similar test with convex acceptance sections is admissible. Second, it is shown that, under conditions (i) and (ii), $\varphi_{CLR}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is similar with convex acceptance sections. A test, φ, is said to be *similar* if $E_\mu[\varphi(Y)] = \alpha$ for all $\mu \in M_0(\theta^*)$ and *conditionally similar* if $E_\mu[\varphi(Y)|X = x] = E_\mu[\varphi(Y_b, X)|X = x] = \alpha$ for all $\mu \in M_0(\theta^*)$ and $x \in \mathcal{X}$, where $\mathcal{X} = \mathbb{R}^L$ denotes the support of X. Let $\mathcal{Y}_b = \mathbb{R}$ denote the support of Y_b. A test is said to have *convex acceptance sections* if its acceptance region, $A_\varphi = \{(y_b, x) \in \mathcal{Y}_b \times \mathcal{X} : \varphi(y_b, x) = 0\}$, is measurable and if the acceptance region's x-sections, $A_\varphi(x) = \{y_b \in \mathcal{Y}_b : (y_b, x) \in A_\varphi\}$, are closed and convex in $\mathcal{Y}_b$ for all $x \in \mathcal{X}$. The first step of the proof follows from Matthes and Truax (1967). For the problem at hand, their Theorem 3.1 asserts that the class of similar tests with convex acceptance sections is essentially complete. A class of tests, $C^*_Y \subset C_Y$, is *essentially complete*, if for any $\varphi \notin C^*_Y$, there exists a $\varphi^* \in C^*_Y$ such that $R_{\varphi^*}(\mu) \leq R_\varphi(\mu)$ for all $\mu \in M(\theta^*)$. Admissibility of a similar test with convex acceptance sections can then be derived given that Y_b is scalar. The result relies on X being a complete sufficient statistic for d. The second step of the proof exploits the form of the CLR statistic given in (4).

A comment about condition (ii) is in order. If $B(\beta^*) = [0, \infty)$ and $b_0 = 0$, i.e., if the testing problem is one-sided, $\text{CLR}(0, Y_b, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ given $X^s = x^s$ has a probability mass at zero (see Appendix B). As a result, $\varphi_{\text{CLR}}(0, Y_b, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ may cease to be conditionally similar and, thus, cease to be similar. Choosing s such that $\Sigma^s_{\delta\beta} \geq 0$ ensures that $\varphi_{\text{CLR}}(0, Y_b, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ is (conditionally) similar for $\alpha \leq 0.5$. In fact, under condition (ii), $\varphi_{\text{CLR}}(0, Y_b, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ reduces to the standard one-sided test that rejects if $Y_b/\Sigma_{\beta\beta} > z_{1-\alpha}$, where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of $N(0, 1)$.

The testing problem given in (6) and (7) corresponds to the testing problem studied in Montiel-Olea (2018). Therefore, his Theorem 1 applies, which states that any admissible and similar test is an extended WAP-similar test. As $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is admissible and similar under conditions (i) and (ii), it follows that $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is an extended WAP-similar test. The appeal of an extended WAP-similar test is that there *essentially* exist weights with respect to which the test maximizes WAP subject to a similarity constraint. See Appendix B for details.

4. Power function comparison

In this section, we compare the power functions of $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ (CLR^s) and several other tests based on Y for testing (7). We write CLR if $X^s = X$. Given Eq. (6), the analysis carries over to local asymptotic power functions of appropriately defined tests based on $\tilde{\theta}_n$. Since the distribution functions of the underlying test statistics are non-standard and in many cases do not even have closed form expressions, we resort to a graphical evaluation using simulation. The rejection frequencies are obtained for a 5% nominal level, using 10,000 Monte Carlo draws. Without loss of generality, Σ is normalized to a correlation matrix. Given that the testing problem varies with $B(\beta^*)$, $D(\delta^*)$, and Σ, we focus on several leading examples that allow us to illustrate certain key properties of the tests under study. One test of particular interest is the "naive" t-test that compares $|\hat{b}_{\text{ML}} - b_0|$ to the standard critical value, 1.96, where $\hat{b}_{\text{ML}}$ denotes the maximum likelihood estimator for b in (6). The reason for considering this test is that its power function corresponds to the local asymptotic power function of the two-sided t-test based on $\hat{\beta}_n$ that is used in practice, if $V(\gamma^*) = a\mathcal{J}(\gamma^*)$ for some $0 < a < \infty$. Note that all other tests considered in this section require the novel *quasi* unconstrained estimator when implemented in practice.

4.1. Scalar parameter of interest

4.1.1. Scalar nuisance parameter near the boundary

We first consider the case where b and d are scalar, with $B(\beta^*) = (-\infty, \infty)$ and $D(\delta^*) = [0, \infty)$. Given that the testing problem is invariant to b_0, we take $b_0 = 0$ without loss of generality. Here, Σ varies with a single correlation parameter, $\rho \equiv \Sigma_{12}$. We set $\rho = 0.9$.

Fig. 1 shows the rejection frequency as a function of b of the CLR, the regular t-test (t) that rejects if $|Y_b - b_0| > 1.96$, and the aforementioned "naive" t-test (t_N) for different values of d. For ease of reference, all figures include a horizontal line at the 5% nominal level. We find that the t_N is size-distorted, but does not overreject (for any $d \in [0, \infty)$, see Appendix D).[22] The CLR is by construction size-correct, although not unbiased. In terms of power, the CLR displays advantages over the t_N, except for large values of b when d is small. Compared to the t, which is unbiased, the CLR sacrifices power for $b < 0$, but displays power advantages for $b > 0$.

Fig. 2 shows the rejection frequency of the WAP-similar (WS) test suggested by Montiel-Olea (2018), the Nearly Optimal (NO) test suggested by Elliott, Müller and Watson (2015), and the CLR for ease of reference.[23] The CLR and the WS both

[22] For the testing problem at hand with $B(\beta^*) = (-\infty, \infty)$ and $D(\delta^*) = [0, \infty)$, this result is already known (Andrews and Guggenberger, 2010b). Note, however, that the condition $V(\gamma^*) = a\mathcal{J}(\gamma^*)$ for some $0 < a < \infty$ is crucial for this result. If this condition is not satisfied, the resulting test can easily be shown to overreject for certain "choices" of $V(\gamma^*)$ and $\mathcal{J}(\gamma^*)$.

[23] Both tests require the choice of weights. For the NO test, we use the same weights as Elliott, Müller and Watson (2015) and for the WS test we use the weights suggested in the 2013 working paper version of Montiel-Olea (2018), which yield a closed form expression of the test statistic, indexed by λ. We choose $\lambda = 0.1$. For larger values of λ, the test becomes more skewed with higher power for positive alternatives, $b > 0$, and lower power for negative alternatives, $b < 0$.

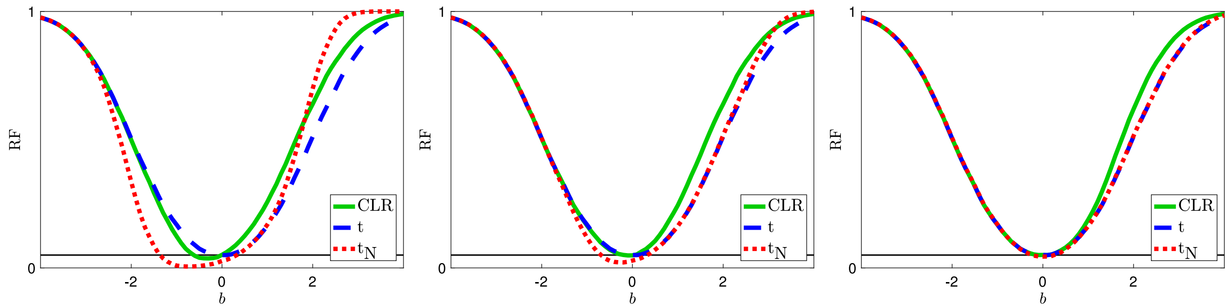

Fig. 1. Rejection frequency as a function of b of CLR (solid), t (dashed), and t_N (dotted) for testing $H_0 : b = 0$ with $d = 0, 1, 2$ from left to right. $\rho = 0.9$.

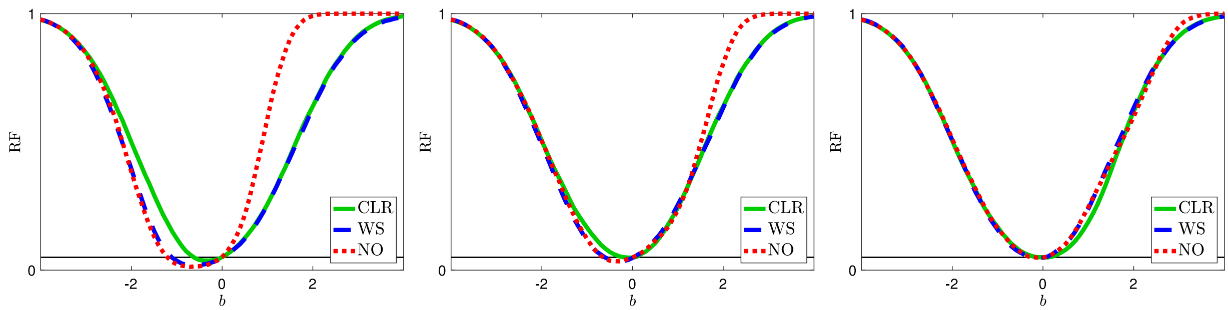

Fig. 2. Rejection frequency as a function of b of CLR (solid), WS (dashed), and NO (dotted) for testing $H_0 : b = 0$ with $d = 0, 1, 2$ from left to right. $\rho = 0.9$.

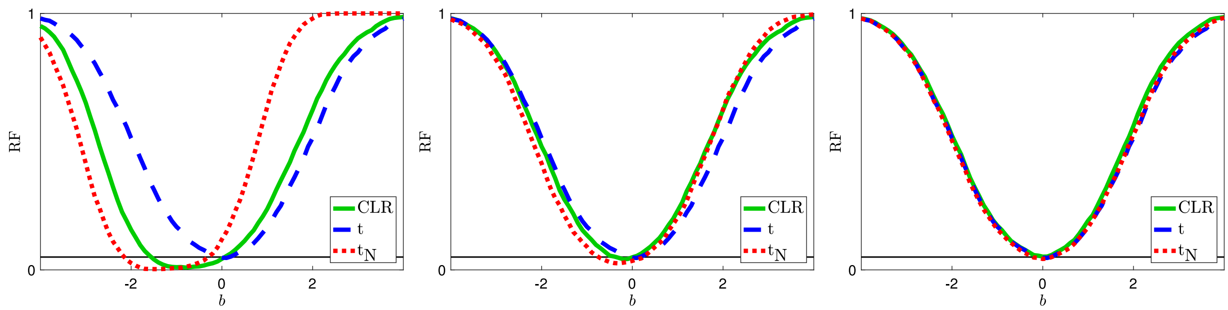

Fig. 3. Rejection frequency as a function of b of CLR (solid), t (dashed), and t_N (dotted) for testing $H_0 : b = 0$ with $d = 0_{10}, 1_{10}, 2_{10}$ from left to right. $\rho = 0.3$.

(essentially) maximize WAP subject to a similarity constraint. Fig. 2 shows that the weights implicitly underlying the CLR are attractive[24] : compared to the WS, the CLR is more powerful for $b < 0$ when d is small and only slightly less powerful for $b > 0$ at intermediate values of d. The NO nearly maximizes WAP among all level α tests. Since the CLR is admissible, the NO does not dominate it. It does, however, offer considerable power gains for $b > 0$ when d is small at the cost of some loss of power for $b < 0$. In Appendix D, we consider several additional tests for the testing problem at hand, notably "asymptotic versions" of the tests considered in Andrews (2001) and two alternative conditional tests.

4.1.2. Multidimensional nuisance parameter near the boundary

Next, we consider the case where d is multidimensional, with $B(\beta^*) = (-\infty, \infty)$ and $D(\delta^*) = [0, \infty)^L$. We impose the following structure on Σ: $\Sigma_{\delta\beta} = \boldsymbol{\rho}_L$ and $\Sigma_{\delta\delta}$, the lower-right submatrix of Σ, equals I_L. We set $L = 10$ and $\rho = 0.3$.

Fig. 3 shows the rejection frequency of the CLR, the t, and the t_N. We no longer implement the NO, because of its high computational cost in the presence of a "high" dimensional nuisance parameter. We also refrain from implementing the WS, although its computational cost is equivalent to that of the CLR, if the weights are chosen such that a closed form expression

[24] While Theorem 1 in Montiel-Olea (2018) implies that there *essentially* exist weights with respect to which admissible and similar tests, such as the CLR, are WAP-maximizing subject to a similarity constraint, it does not provide a characterization result.

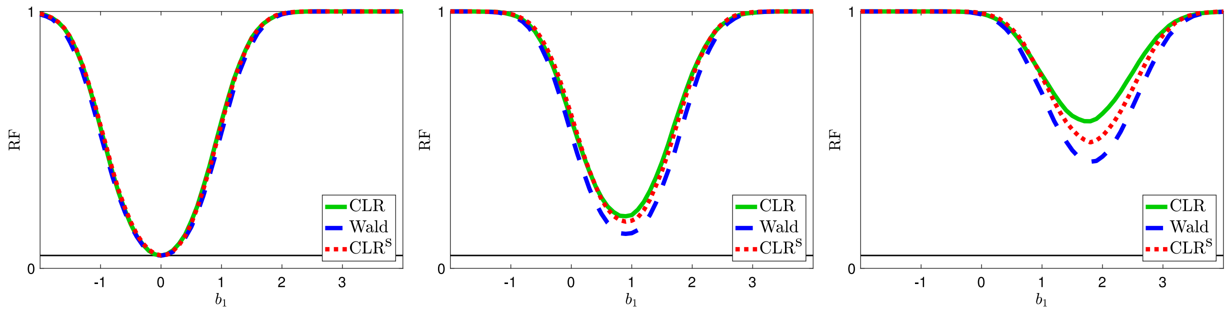

Fig. 4. Rejection frequency as a function of b_1 of CLR (solid), Wald (dashed), and CLR^s (dotted) for testing $H_0 : b_1 = b_2 = 0$ with $b_2 = 0, 1, 2$ from left to right with $d = 0$. $\rho = 0.9$.

of the test statistic is available. Fig. 3 illustrates that the t_N can suffer from overrejection and, thus, does not control size. As in the case of a scalar nuisance parameter near the boundary, the CLR displays greater power than the t for $b > 0$. However, the sacrifice of power for $b < 0$ is much more important and, arguably, in the case at hand the t may well be the preferred choice.

4.2. Two-dimensional parameter of interest

We now turn to the case where b is two-dimensional, $b = (b_1, b_2)$, and $B(\beta^*) = (-\infty, \infty) \times [0, \infty)$. We consider the case of a scalar nuisance parameter near the boundary, $D(\delta^*) = [0, \infty)$, and restrict all off-diagonal entries of Σ to be equal to ρ. We set $\rho = 0.9$ and consider testing $b_0 = (0, 0)$. This null hypothesis corresponds, for example, to testing whether a given regressor can be excluded from a random coefficients regression model. Note, however, that the testing problem is not invariant to b_0.

Fig. 4 shows the rejection frequency as a function of b_1 of the CLR, the regular Wald test based on Y_b (Wald), and the CLR^s with s being the empty subset, i.e., CLR^s only uses Y_b, for different values of b_2. The three tests have equal power when $b_2 = 0$. For $b_2 > 0$, the CLR offers power advantages over the other two tests, with the CLR^s displaying power advantages over the Wald.

4.3. Comments

The above analysis shows that the t_N, representative of Wald tests that are based on a constrained extremum estimator but, nevertheless, use standard χ^2 critical values, is size-distorted: It can suffer from under- as well as overrejection. This provides a strong argument for using tests that employ the *quasi* unconstrained estimator. The NO test displays good power properties, but suffers from high computational costs in the presence of a multidimensional nuisance parameter. The CLR, on the other hand, is computationally cheap, regardless of the dimension of the nuisance parameter, and also displays good power properties. While the t may seem preferable to the CLR in case of a large dimensional nuisance parameter (cf. Section 4.1.2), note that the t or, more generally, the Wald test can be obtained as a CLR^s when $B = [-c, c]^K$ (such that $B(\beta^*) = (-\infty, \infty)^K$ for all $\beta^* \in B$) by choosing s equal to the empty set. We conclude that, given the testing problem, an appropriate choice of s will provide a test that has not only theoretical optimality properties, but also good local asymptotic power properties.

5. Application

The standard formulation of the random coefficients logit model (Berry, Levinsohn and Pakes, 1995) assumes that the random coefficients are independently normally distributed such that the model parameters are given by a vector of means and a vector of variances. Since variances are bounded below by zero, the model readily fits our theoretical framework.[25] Before turning to the empirical application using data on the European car market, we present a small Monte Carlo study to investigate the finite-sample behavior of the CLR test when implemented using the *quasi* unconstrained estimator. Throughout this section, we take s to be the subset that includes all nuisance parameters that are restricted below by zero, unless the parameter of interest is restricted below by zero in which case s is chosen to be the largest subset such that $\hat{\Sigma}^s_{\delta\beta} \geq 0$. For notational convenience, we omit the superscripts.

25 In applications, the model is – to the best of our knowledge – always parameterized with respect to standard deviations rather than variances. In Ketz (2018), we show that Eq. (1) upon which Assumptions 1–5 are based is not or, rather, cannot be satisfied under this alternative parameterization and analyze the consequences this has for inference. Here, we use the parameterization in terms of variances for which Assumptions 1–5 hold given an appropriately defined parameter space, Γ, see Ketz (2018) for details.

Table 1
Monte Carlo — Rejection frequencies.

True values				$H_0: \mu_1 = -2$		$H_0: \mu_2 = 2$		$H_0: \mu_3 = 2$		$H_0: \sigma^2 = 0$	
μ_1	μ_2	μ_3	σ^2	t_N	CLR	t_N	CLR	t_N	CLR	t_N	CLR
−2	2	2	0	0.035	0.037	0.035	0.041	0.032	0.049	0.023	0.049
−2	2	2	0.25	0.037	0.038	0.042	0.049	0.038	0.045	0.264	0.391
−2	2	1.5	0	0.035	0.037	0.030	0.043	0.324	0.383	0.029	0.054
−2	2	2.5	0	0.036	0.038	0.046	0.049	0.393	0.442	0.020	0.045

5.1. Monte Carlo

First, we describe the data generating process and introduce the model in more detail. For ease of reference, we use the same data generating process as Reynaert and Verboven (2014). In what follows, we "recycle" some of the previous notation. Berry, Levinsohn and Pakes (1995) define the demand for product $j \in \{1, \dots, J\}$ in market $t \in \{1, \dots, T\}$ as a function of the product characteristics of all products in that market. Product characteristics can be classified into observables, $x_t = (x_{1t}, \dots, x_{Jt})$, and unobservables, $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})'$, where x_{jt} and ξ_{jt} are vector-valued and scalar, respectively. Here, we consider three product characteristics in x_{jt}, $x_{jt,1}$ through $x_{jt,3}$. However, we only allow for a random coefficient on $x_{jt,3}$. The mean parameters, μ_1 through μ_3, are unrestricted, while the variance parameter, σ^2, is restricted below by 0. Let $\phi(v, \mu, \sigma^2)$ denote the pdf of a normal random variable with mean μ and variance σ^2.[26] Then, the market share for product j in market t is given by

$$s_j(\mu_1, \mu_2, \mu_3, \sigma^2, x_t, \xi_t) = \int_{-\infty}^{\infty} \frac{e^{\mu_1 x_{jt,1} + \mu_2 x_{jt,1} + v x_{jt,3} + \xi_{jt}}}{1 + \sum_{l=1}^{J} e^{\mu_1 x_{lt,1} + \mu_2 x_{lt,1} + v x_{lt,3} + \xi_{lt}}} \phi(v, \mu_3, \sigma^2) dv. \tag{8}$$

Equating model implied market shares, given in (8), with market shares observed in the data, the model parameters are estimated by GMM relying on a zero moment condition that interacts the error term, ξ_{jt}, with a set of instruments. For more details on the model and the estimation procedure, see e.g., Nevo (2000). We follow Reynaert and Verboven (2014) and implement an approximation to the optimal instruments (implying that $V(\gamma^*) = a\mathcal{J}(\gamma^*)$ for some $0 < a < \infty$).

We model $x_{jt,1}$ to be endogenous. In particular, $x_{jt,1}$ is generated as

$$x_{jt,1} = w_{jt}'\pi_1 + z_{jt}'\pi_2 + \zeta_{jt},$$

where $w_{jt} = (x_{jt,2}, x_{jt,3})'$ denotes the exogenous product characteristics and z_{jt} is a three-dimensional vector of instruments. The endogeneity of $x_{jt,1}$ arises, because the error terms, ξ_{jt} and ζ_{jt}, are drawn from a bivariate normal distribution with zero means, unit variances, and correlation equal to 0.7. $x_{jt,2}$ is a constant, $x_{jt,3}$ is uniformly distributed on [1, 2], and z_{jt} is a vector of uniform random variables with support on [0, 1]. The true parameter values are chosen as $\pi_1 = (0.7, 0.7)$, $\pi_2 = (3, 3, 3)$, $\mu_1 = -2$, and $\mu_2 = 2$ while we vary the true parameter values of μ_3 and σ^2 in order to investigate size and power for testing $H_0 : \mu_1 = -2$, $H_0 : \mu_2 = 2$, $H_0 : \mu_3 = 2$, and $H_0 : \sigma^2 = 0$. We choose $T = 25$ and $J = 10$ totaling 250 products over all markets.

Table 1 reports the rejection frequencies of the CLR test and the "naive" t-test that is based on the constrained extremum estimator (t_N) over 1000 Monte Carlo replications. Since market shares, given in (8), and, thus, the GMM objective function cannot be evaluated at negative values of σ^2, the *quasi* unconstrained estimator needs to be employed in order to construct the CLR statistic. Table 1 shows that both tests control size, while the t_N is undersized, and that the CLR test displays power advantages over the t_N. These finite sample results are in line with our asymptotic results (cf. Section 4.1.1) and we conclude that our asymptotic theory provides good approximations.

5.2. Empirical results

We now turn to the empirical application using data from Reynaert and Verboven (2014) (RV). RV estimate the demand for cars in several European countries spanning the years from 1998 to 2010. The product characteristics are price divided by income (Price/Inc.), horse power per weight (Hp/We.), a dummy variable indicating whether the car brand is foreign (Foreign), size (Size) obtained as length times width, height (Height), and fuel efficiency (€/km) given by price in € per kilometer. See RV for more details on the dataset and its construction.

The first two columns of Table 2 show the estimates and the corresponding standard error estimates that RV obtain for the baseline specification that allows for a random coefficient on each product characteristic, see the last two columns of their Table 6.[27] As the estimates of the variance parameters are all in the interior of the parameter space, the CLR statistic can,

[26] $\phi(v, \mu, 0)$ is defined as $\mathbb{1}(v = \mu)$, where $\mathbb{1}(\cdot)$ denotes the indicator function.

[27] As standard in the literature, Table 6 in RV reports estimates and corresponding standard error estimates for standard deviations rather than variances, see also footnote 25 above. The reported standard error estimates in Table 2 are obtained by a simple delta method argument, see Ketz (2018) for details.

Table 2
Confidence intervals based on t_N and CLR.

	Est.	Std. Err.	95% CI				90% CI			
			t_N		CLR		t_N		CLR	
Mean valuations μ										
Price/Inc.	−2.322	0.497	−3.297	−1.347	−3.3100	−1.390	−3.141	−1.504	−3.156	−1.522
Hp/We.	−0.918	1.192	−3.253	1.418	−3.218	1.096	−2.878	1.043	−2.860	0.786
Foreign	−0.853	0.216	−1.275	−0.430	−1.218	−0.425	−1.207	−0.498	−1.134	−0.492
Size	0.667	0.674	−0.654	1.987	−0.650	2.004	−0.441	1.775	−0.452	1.795
Height	0.183	0.052	0.080	0.286	0.087	0.287	0.097	0.269	0.101	0.271
€/km	−3.972	1.344	−6.606	−1.338	−6.391	−1.425	−6.183	−1.761	−5.975	−1.815
Variances σ^2										
Price/Inc.	0.274	0.176	0.000	0.619	0.000	0.623	0.000	0.564	0.045	0.569
Hp/We.	10.252	4.346	1.733	18.770	2.798	18.879	3.103	17.401	3.798	17.531
Foreign	0.515	0.736	0.000	1.958	0.000	1.976	0.000	1.726	0.000	1.748
Size	0.057	0.188	0.000	0.427	0.000	0.431	0.000	0.367	0.000	0.373
Height	0.011	0.006	0.000	0.023	0.000	0.023	0.000	0.021	0.002	0.021
€/km	4.424	19.835	0.000	43.299	0.000	43.795	0.000	37.049	0.000	37.644

Table 3
Wald and CLR for testing $H_0 : \mu_j = \sigma_j^2 = 0$.

	Wald	p	CLR	p
Price/Inc.	30.261	0.000	106.542	0.000
Hp/We.	6.994	0.030	7.695	0.011
Foreign	754.818	0.000	780.757	0.000
Size	0.998	0.607	1.014	0.463
Height	12.496	0.002	14.022	0.000
€/km	11.308	0.004	12.880	0.001

in the application at hand, be constructed using the "constrained" estimates, cf. Eq. (2).[28] The 95% and the 90% confidence intervals based on the CLR test and the t_N are given in the last eight columns of Table 2. The top half of Table 2 shows that the CLR test can yield tighter confidence intervals (CI) for the mean parameters. For example the 95% CI for the mean parameter of HP/We. is around 8% shorter than the CI based on t_N. The bottom half of Table 2 shows that the CLR test and the t_N agree at the 95% significance level in that both indicate that only the variance on Hp/We. is significantly different from zero. At the 90% significance level, however, the CLR test also rejects the null of a zero variance for Price/Inc. and Height, while the t_N continues to reject that null only for Hp/We. Put differently, in the application at hand, the CLR test allows us to detect the presence of additional heterogeneity in consumer preferences, which is not picked up by the standard two-sided t-test.

Table 3 computes the Wald and the CLR statistic for testing $H_0 : \mu_j = \sigma_j^2 = 0$, where j indexes the product characteristics, along with the corresponding p-values. In line with the results in Section 4.1.2, the p-values for the CLR test are smaller. However, here both tests qualitatively yield the same conclusions, at the commonly used significance levels.

Appendix A. Proof of Theorem 1

Proof of Theorem 1. Under $\{\gamma_n\} \in \Gamma(\gamma^*)$ and Assumption 4(i), Eq. (2) satisfies

$$\tilde{\theta}_n - \theta_n = \hat{\theta}_n - \theta_n - \left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}\left(DQ_n(\theta_n) + D^2Q_n(\theta_n)(\hat{\theta}_n - \theta_n) + o_p(1/\sqrt{n})\right).$$

Multiplying both sides by $\sqrt{n}$, we obtain

$$\begin{aligned}\sqrt{n}(\tilde{\theta}_n - \theta_n) = &-\left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}\sqrt{n}DQ_n(\theta_n)\\ &+\left(I_J - \left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}D^2Q_n(\theta_n)\right)\sqrt{n}(\hat{\theta}_n - \theta_n)\\ &-\left(\widehat{D^2Q}_n(\hat{\theta}_n)\right)^{-1}o_p(1).\end{aligned}$$

By Assumptions 1, 3, and 4(ii) together with Slutsky's Theorem, the second and third lines are $o_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma^*)$. The conclusion of the Theorem then follows by Assumptions 2, 3, and 4(ii) together with Slutsky's Theorem. □

[28] The construction of the CLR statistic requires an estimate of the asymptotic variance matrix of the estimator, which is not reported in RV. I thank Mathias Reynaert and Frank Verboven for sharing their estimate of the asymptotic variance matrix with me.

Appendix B. Results for the CLR test based on *Y*

In Appendix B.1, we formally introduce the concept of an extended WAP-similar test (Montiel-Olea, 2018). Appendix B.2 presents several properties of CLR(·), including some Lemmas that are used in the proof of Theorem 2. All proofs, including that of Theorem 2, are collected in Appendix B.3. Throughout this section, we suppress the dependence of Σ on γ^*.

B.1. An extended WAP-similar test

Here and in what follows, we sometimes suppress the dummy variable of integration and write, for example, $\int \varphi dF(\theta^*, \mu, \Sigma)$ for $\int_{\mathcal{Y}} \varphi(y)dF(y;\theta^*,\mu,\Sigma)$. For $\varphi \in C_Y$ define weighted average power over the distribution function $w : M_1(\theta^*) \to [0, 1]$ as

$$\mathrm{WAP}(\varphi, w) = \iint \varphi dF(\theta^*, \mu, \Sigma)dw(\mu). \tag{9}$$

Let $C_Y^{\mathrm{sim}} \subset C_Y$ denote the class of α-similar tests, i.e., $\varphi \in C_Y^{\mathrm{sim}}$ if

$$\int \varphi dF(\theta^*, \mu, \Sigma) = \alpha \text{ for all } \mu \in M_0(\theta^*). \tag{10}$$

Let $\varphi_{\mathrm{WAP}}^{\mathrm{sim},w}$ denote the WAP-similar test with respect to w, i.e., $\varphi_{\mathrm{WAP}}^{\mathrm{sim},w} \in C_Y^{\mathrm{sim}}$ and

$$\mathrm{WAP}(\varphi, w) \leq \mathrm{WAP}(\varphi_{\mathrm{WAP}}^{\mathrm{sim},w}, w)$$

for all $\varphi \in C_Y^{\mathrm{sim}}$. A test $\varphi \in C_Y^{\mathrm{sim}}$ is said to be *extended WAP-similar* (Montiel-Olea, 2018) if for any $\epsilon > 0$ there exists w_ϵ such that

$$\mathrm{WAP}(\varphi_{\mathrm{WAP}}^{\mathrm{sim},w_\epsilon}, w_\epsilon) - \epsilon \leq \mathrm{WAP}(\varphi, w_\epsilon) \leq \mathrm{WAP}(\varphi_{\mathrm{WAP}}^{\mathrm{sim},w_\epsilon}, w_\epsilon). \tag{11}$$

As stated in the main text, $\varphi_{\mathrm{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is extended WAP-similar if condition (i) or (ii) of Theorem 2 is satisfied.

B.2. Properties of CLR(·)

In what follows, we provide several useful results pertaining to $\mathrm{CLR}(b_0, y_b, x^s, \Sigma^s, B, D^s)$ as a function of $(y_b, x^s) \in \mathbb{R}^{K+L^s}$, where Σ and, thus, Σ^s is positive-definite and where, with a slight abuse of notation, B denotes a Cartesian product of intervals equal to $[0, \infty)$ or $(-\infty, \infty)$ and $D^s = [0, \infty)^{L^s}$. This choice of D^s is without loss of generality (w.l.o.g.) in light of the following Lemma.

Lemma 1. *Let $D^{s'}$ be a Cartesian product of intervals equal to $[0, \infty)$ or $(-\infty, \infty)$. Then, $CLR(b_0, y_b, x^{s'}, \Sigma^{s'}, B, D^{s'}) = CLR(b_0, y_b, x^s, \Sigma^s, B, D^s)$, where s is such that D^s only includes those entries of $D^{s'}$ that equal $[0, \infty)$.*

In what follows, we suppress the dependence on (Σ^s, B, D^s) and write, for example, $\mathrm{CLR}(b_0, y_b, x^s)$. We normalize Σ to be a correlation matrix w.l.o.g. Define

$$QF(y_b, x^s, \mathrm{b}, \mathrm{d}) = \begin{pmatrix} y_b - \mathrm{b} \\ x^s + \Sigma^s_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b - \mathrm{d} \end{pmatrix}' \left(\Sigma^s\right)^{-1} \begin{pmatrix} y_b - \mathrm{b} \\ x^s + \Sigma^s_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b - \mathrm{d} \end{pmatrix}. \tag{12}$$

Furthermore, let

$$\mathrm{d}^*(\mathrm{b}, y_b, x^s) = \arg\min_{\mathrm{d} \in D^s} QF(y_b, x^s, \mathrm{b}, \mathrm{d})$$

and

$$(\mathrm{b}^{**}(y_b, x^s), \mathrm{d}^{**}(y_b, x^s)) = \arg\min_{\mathrm{b} \in B, \mathrm{d} \in D^s} QF(y_b, x^s, \mathrm{b}, \mathrm{d}),$$

which are well-defined given convexity of D^s and $B \times D^s$, such that

$$\mathrm{CLR}(b_0, y_b, x^s) = QF(y_b, x^s, b_0, \mathrm{d}^*(b_0, y_b, x^s)) - QF(y_b, x, \mathrm{b}^{**}(y_b, x^s), \mathrm{d}^{**}(y_b, x^s)).$$

The following Lemma is used in the proofs of Lemmas 4 and 5, which, in turn, are used in the proof of Theorem 2.

Lemma 2. *$d^*(b, x^s) \equiv d^*(b, y_b, x^s)$ does not depend on y_b and $QF(y_b, x^s, b, d^*(b, x^s))$ is quadratic in y_b for any given $x^s \in \mathbb{R}^{L^s}$, with partial derivatives given by $2\Sigma^{-1}_{\beta\beta}(y_b - b)$.*

Note that $\mathrm{d}^*(\mathrm{b}, x^s)$ and $(\mathrm{b}^{**}(y_b, x^s), \mathrm{d}^{**}(y_b, x^s))$ are continuous functions of (y_b, x^s). Similarly, $QF(y_b, x^s, b_0, \mathrm{d}^*(b_0, x^s))$, $QF(y_b, x^s, \mathrm{b}^{**}(y_b, x^s), \mathrm{d}^{**}(y_b, x^s))$, and, thus, $\mathrm{CLR}(b_0, y_b, x^s)$ are continuous functions of (y_b, x^s).

Remark 1. Note that $QF(Y_b, X^s, b_0, \mathrm{d}^*(b_0, X^s))$ can be considered a Conditional Score statistic when $X^s = X$. Note, further, that the test that compares $QF(Y_b, X^s, b_0, \mathrm{d}^*(b_0, X^s))$ to its conditional critical value (defined analogously to Eq. (5)) is, by Lemma 2, identical to the regular Wald test based on Y_b that ignores possible restrictions on b.

In what follows, we take w.l.o.g. the first $K - K^c$ elements of B equal to $(-\infty, \infty)$ and the remaining $0 \leq K^c \leq K$ equal to $[0, \infty)$, i.e., $B = (-\infty, \infty)^{K-K^c} \times [0, \infty)^{K^c}$. Furthermore, we let $0_{r\times c}$ denote a matrix of zeros with $r \in \mathbb{N}$ rows and $c \in \mathbb{N}$ columns. The following Lemma, which corresponds to Theorem 5 in Andrews (1999), provides a characterization of $(\mathrm{b}^{**}(y_b, x^s), \mathrm{d}^{**}(y_b, x^s))$.

Lemma 3. *Let $z \equiv (y_b, x^s + \Sigma^s_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b)$. Then, $(b^{**}(y_b, x^s), d^{**}(y_b, x^s)) = P_{A(\hat{j})}z$, where $\hat{j}$ minimizes $z'C_j'(C_j\Sigma^s C_j')^{-1}C_j z$ over $j = 1, \ldots, 2^{K^c+L^s}$ for which $P_{A(j)}z \in B \times D^s$. Here, $A(j) \equiv \{a \in \mathbb{R}^{K+L^s} : C_j a = 0\}$, $P_{A(j)} = I_{K+L^s} - \Sigma^s C_j'(C_j\Sigma^s C_j')^{-1}C_j$, and $\{C_j : j = 1, \ldots, 2^{K^c+L^s}\}$ consists of all different matrices composed of some, possibly zero rows of $[0_{(K^c+L^s)\times(K-K^c)} \;\; -I_{K^c+L^s}]$.*

Note that $P_{A(j)}$ is the projection matrix onto $A(j)$ with respect to the norm $\|a\| = (a'(\Sigma^s)^{-1}a)^{-1/2}$. Since the probability distribution function of (Y_b, X^s) has full support on $\mathbb{R}^{K+L^s}$ for any $(b, d^s) \in B \times D^s$, we have the following Corollary, which provides a sufficient condition for $\varphi_{\mathrm{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ to be similar.

Corollary 1. *For any $(b, d^s) \in B \times D^s$, the (conditional) distribution function of $CLR(b_0, Y_b, X^s)$ (given $X^s = x^s$) is continuous except if $B = [0, \infty)^K$ and $b_0 = 0_K$, in which case it has a discontinuity at zero.*

Lemmas 2 and 3 together yield the following result.

Lemma 4. *If $B = (-\infty, \infty)$, then $b^{**}(y_b'', x^s) > b^{**}(y_b', x^s)$ whenever $y_b'' > y_b'$. If $B = [0, \infty)$, then $b^{**}(y_b, x^s) = 0$ whenever $y_b \leq y_b^*(x^s)$ and $b^{**}(y_b'', x^s) > b^{**}(y_b', x^s)$ whenever $y_b'' > y_b' \geq y_b^*(x^s)$ for some $y_b^*(x^s) \in \mathbb{R}$. Furthermore, if $\Sigma^s_{\delta\beta} \geq 0$, then $y_b^*(x^s) \leq 0 \; \forall x^s \in \mathbb{R}^{L^s}$.*

It follows directly from Lemma 4 that $\mathrm{b}^{**}(y_b, x^s) \to \infty$ as $y_b \to \infty$ and $\mathrm{b}^{**}(y_b, x^s) \to -\infty$ if $B = (-\infty, \infty)$ and $\mathrm{b}^{**}(y_b, x^s) \to 0$ if $B = [0, \infty)$ as $y_b \to -\infty$. Furthermore, unless $B = [0, \infty)$ and $b_0 = 0$, there exists by continuity of $\mathrm{b}^{**}(y_b, x^s)$ a unique $\tilde{y}_b(x^s)$ such that $\mathrm{b}^{**}(\tilde{y}_b(x^s), x^s) = b_0$ and, thus, $\mathrm{CLR}(b_0, \tilde{y}_b(x^s), x^s) = 0$. While, if $B = [0, \infty)$ and $b_0 = 0$, then $\mathrm{b}^{**}(y_b, x^s) = b_0$ and $\mathrm{CLR}(b_0, y_b, x^s) = 0$ for all $y_b \leq y_b^*(x^s)$. Given Lemmas 2 and 4, we then also have the following result, which is used in the proof of Theorem 2 to show that $\varphi_{\mathrm{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ has convex acceptance sections.

Lemma 5. *If $B = (-\infty, \infty)$ or $B = [0, \infty)$ and $b_0 > 0$, then $CLR(b_0, y_b'', x^s) > CLR(b_0, y_b', x^s)$ ($CLR(b_0, y_b'', x^s) < CLR(b_0, y_b', x^s)$) whenever $y_b'' > y_b' > \tilde{y}_b(x^s)$ ($y_b'' < y_b' < \tilde{y}_b(x^s)$). Similarly, if $B = [0, \infty)$ and $b_0 = 0$, then $CLR(b_0, y_b'', x^s) > CLR(b_0, y_b', x^s)$ whenever $y_b'' > y_b' > y_b^*(x^s)$.*

B.3. Proofs

Proof of Lemma 1. Lemma 1 follows from Theorem 4(e) in Andrews (1999). To see this, take $q(\lambda)$ in Andrews (1999) equal to $QF(y_b, x^{s'}, \mathrm{b}, \mathrm{d})$, defined in Eq. (12), with $\lambda = (\mathrm{b}, \mathrm{d})$, $Z = (y_b, x^{s'} + \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b)$, $\mathcal{J} = (\Sigma^{s'})^{-1}$, $p = K + L^s$, $q = L^{s'} - L^s$, and $r = 0$. Then, applying Theorem 4(e) twice, with $\Lambda = \beta_0 \times [0, \infty)^{L^s} \times (-\infty, \infty)^{L^{s'}-L^s}$ (and $\Lambda_\beta = \beta_0 \times [0, \infty)^{L^s}$) and with $\Lambda = B \times D = B \times [0, \infty)^{L^s} \times (-\infty, \infty)^{L^{s'}-L^s}$ (and $\Lambda_\beta = B \times [0, \infty)^{L^s}$), where we w.l.o.g. take the last $L^{s'} - L^s$ elements of $D^{s'}$ to be equal to $(-\infty, \infty)$, yields the desired result, cf. equation (6.4) in Andrews (1999). Note that Assumption 3 in Andrews (1999) (or, rather, the part that is used in the proof of Theorem 4(e)) is satisfied given the positive-definiteness of $(\Sigma^{s'})^{-1}$ and given that $(y_b, x^{s'} + \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b) \in \mathbb{R}^{K+L^{s'}}$, while Assumptions 7 and 8 in Andrews (1999) are satisfied since $\beta_0 \times [0, \infty)^{L^s}$, $B \times [0, \infty)^{L^s}$, and $(-\infty, \infty)^{L^{s'}-L^s}$ equal cones. □

In the proofs of Lemmas 2–5 and Corollary 1, we omit the superscript s for notational convenience.

Proof of Lemma 2. Note that $QF(y_b, x, \mathrm{b}, \mathrm{d})$ can be written as

$$(y_b - \mathrm{b})'(\Sigma_{\beta\beta} - \Sigma_{\beta\delta}\Sigma^{-1}_{\delta\delta}\Sigma_{\delta\beta})^{-1}(y_b - \mathrm{b})$$
$$+(2\Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}\mathrm{b} + x - \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b - \mathrm{d})'(\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta\delta})^{-1}(x + \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}y_b - \mathrm{d}).$$

using the formula for the inverse of a partitioned matrix, i.e.,

$$\Sigma^{-1} = \begin{bmatrix} (\Sigma_{\beta\beta} - \Sigma_{\beta\delta}\Sigma^{-1}_{\delta\delta}\Sigma_{\delta\beta})^{-1} & -\Sigma^{-1}_{\beta\beta}\Sigma_{\beta\delta}(\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta\delta})^{-1} \\ -(\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta\delta})^{-1}\Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta} & (\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta\delta})^{-1} \end{bmatrix}.$$

Evaluating this function at $y_b' = y_b + a$, where $a \in \mathbb{R}^K$, and rearranging terms, we obtain

$$\begin{aligned}&(y_b + a - \mathrm{b})'(\Sigma_{\beta\beta} - \Sigma_{\beta\delta}\Sigma_{\delta\delta}^{-1}\Sigma_{\delta\beta})^{-1}(y_b + a - \mathrm{b})\\ &+(2\Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}\mathrm{b} + x - \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}y_b - \mathrm{d})'(\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\delta})^{-1}(x + \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}y_b - \mathrm{d})\\ &-(2\Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}(y_b - \mathrm{b}) + \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}a)'(\Sigma_{\delta\delta} - \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\delta})^{-1}(\Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}a).\end{aligned}$$

Inspection of the last display shows that the minimization problem at $y_b + a$ equals the minimization problem at y_b plus a constant, such that the arg min, $\mathrm{d}^*(\mathrm{b}, y_b, x)$, is the same in both cases and we write $\mathrm{d}^*(\mathrm{b}, x)$. It follows that $QF(y_b, x, \mathrm{b}, \mathrm{d}^*(\mathrm{b}, x))$ is quadratic in y_b, given $x \in \mathbb{R}^L$. To see this, note that the first-order partial derivatives are given by

$$\frac{\partial QF(y_b, x, \mathrm{b}, \mathrm{d}^*(\mathrm{b}, x))}{\partial y_b} = 2\begin{pmatrix} I_K \\ \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}\end{pmatrix}' \Sigma^{-1}\begin{pmatrix} y_b - \mathrm{b} \\ x + \Sigma_{\delta\beta}\Sigma_{\beta\beta}^{-1}y_b - \mathrm{d}^*(\mathrm{b}, x)\end{pmatrix} = 2\Sigma_{\beta\beta}^{-1}(y_b - \mathrm{b}). \quad \square$$

Proof of Lemma 3. Lemma 3 follows directly from Theorem 5 in Andrews (1999). To see this, take $\lambda_\beta = (\mathrm{b}, \mathrm{d})$, $\Lambda_\beta = B \times D$ such that $\Gamma_b = [0_{(K^c+L)\times(K-K^c)}\ -I_{K^c+L}]$ (Γ_a does not appear), $Z_\beta = z$, $\mathcal{J}_* = \Sigma^{-1}$, $H = I_{K+L}$, and $q = r = 0$ (such that $p = K+L$). Note that Assumption 3 in Andrews (1999) (as in the proof of Lemma 1) is satisfied given the positive-definiteness of Σ^{-1} and given that $z \in \mathbb{R}^{K+L}$, while Assumptions 7–9 in Andrews (1999) are satisfied given the definition of $B \times D$. $\square$

Proof of Corollary 1. Lemma 3 implies that $\mathrm{b}^{**}(y_b, x)$ is a piecewise linear function of y_b with "intercepts" that depend on x and with "slopes" given by $(\Sigma - \Sigma C_j'(C_j\Sigma C_j')^{-1}C_j\Sigma)_{(1:K)(1:K)}\Sigma_{\beta\beta}^{-1}$ where $\{C_j : j = 1, \ldots, 2^{K^c+L}\}$ consists of all different matrices composed of some, possibly zero rows of $[0_{(K^c+L)\times(K-K^c)}\ -I_{K^c+L}]$. Here, the subscript $(1:K)(1:K)$ denotes the Kth order leading principal submatrix. By positive-definiteness of Σ, it follows that, for any given $x \in \mathbb{R}^L$, $\mathrm{b}^{**}(y_b, x)$ is a nontrivial function of y_b almost everywhere (a.e.) unless $K^c = K$, i.e., $B = [0, \infty)^K$, in which case $\mathrm{b}^{**}(y_b, x) = 0$ on a set with positive Lebesgue measure. Therefore, given any $x \in \mathbb{R}^L$, $\mathrm{CLR}(b_0, y_b, x)$ is a nontrivial function of y_b a.e. unless $B = [0, \infty)^K$ and $b_0 = 0$. Since the probability distribution function of (Y_b, X) has full support on $\mathbb{R}^{K+L}$ for any $(b, d) \in B \times D$, the conclusion of the Lemma follows. $\square$

Proof of Lemma 4. Let $B = (-\infty, \infty)$. As in the proof of Corollary 1, Lemma 3, with $K = 1$ and $K^c = 0$ such that $\{C_j : j = 1, \ldots, 2^L\}$ consists of all different matrices composed of some, possibly zero rows of $[0\ -I_L]$, implies that $\mathrm{b}^{**}(y_b, x)$ is a piecewise linear function of y_b with "slopes" given by $(\Sigma - \Sigma C_j'(C_j\Sigma C_j')^{-1}C_j\Sigma)_{11}$ which is strictly positive for all j, by positive-definiteness of Σ, all its principal submatrices that include $\Sigma_{\beta\beta}(= 1)$, and their respective inverses. Therefore, $\mathrm{b}^{**}(y_b'', x) > \mathrm{b}^{**}(y_b', x)$ whenever $y_b'' > y_b'$.

Now, let $B = [0, \infty)$. Then, the set of possible solutions for $\mathrm{b}^{**}(y_b, x)$, given above, is augmented by $\mathrm{b}^{**}(y_b, x) = 0$, as Lemma 3 applies with $K = K' = 1$ such that $\{C_j : j = 1, \ldots, 2^{1+L}\}$ consists of all different matrices composed of some, possibly zero rows of $-I_{1+L}$. For any given $x \in \mathbb{R}^L$, however, there exists y_b large enough such that *all* non-zero candidate solutions for $\mathrm{b}^{**}(y_b, x)$ become feasible, since they are *all* linear in y_b with strictly positive slopes. Therefore, for y_b large enough we have $\mathrm{b}^{**}(y_b, x) > 0$.

Now, let $\Sigma_{\beta\delta} \geq 0$. Assume $\mathrm{b}^{**}(y_b, x) = 0$ for some $y_b > 0$. Then, Lemma 2 implies that $\mathrm{d}^{**}(y_b, x) = \mathrm{d}^*(0, x)$ and, thus, $(\mathrm{b}^{**}(y_b, x), \mathrm{d}^{**}(y_b, x)) = (0, \mathrm{d}^*(0, x))$. But if $\Sigma_{\beta\delta} \geq 0$, all constraints become less binding as y_b increases, cf. Eq. (12). Therefore, $\mathrm{b}^{**}(y_b, x) = 0$ implies that the additional slackness is not picked up, even though $QF(y_b, x, \mathrm{b}^{**}(y_b, x), \mathrm{d}^{**}(y_b, x)) = QF(y_b, x, 0, \mathrm{d}^*(0, x))$ is, by Lemma 2, strictly increasing in y_b for all $y_b > 0$. A contradiction.

The conclusion of the Lemma then follows by continuity of $\mathrm{b}^{**}(y_b, x)$. $\square$

Proof of Lemma 5. Consider $y_b'' > y_b' > \tilde{y}_b(x)$. From Lemmas 2 and 4, it follows that $QF(y_b, x, b_0, \mathrm{d}^*(b_0, x))$ and $QF(y_b, x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x))$ (using $\mathrm{d}^{**}(y_b', x) = \mathrm{d}^*(\mathrm{b}^{**}(y_b', x), x)$) are quadratic functions of y_b given $x \in \mathbb{R}^L$, where the derivative of the latter is strictly smaller than the derivative of the former. Therefore, we have

$$QF(y_b'', x, b_0, \mathrm{d}^*(b_0, x)) - QF(y_b', x, b_0, \mathrm{d}^*(b_0, x)) > QF(y_b'', x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x)) - QF(y_b', x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x))$$

for all $y_b'' > y_b'$ such that

$$QF(y_b'', x, b_0, \mathrm{d}^*(b_0, x)) - QF(y_b'', x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x)) > QF(y_b', x, b_0, \mathrm{d}^*(b_0, x)) - QF(y_b', x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x)).$$

Since by definition $QF(y_b'', x, \mathrm{b}^{**}(y_b'', x), \mathrm{d}^{**}(y_b'', x)) \leq QF(y_b'', x, \mathrm{b}^{**}(y_b', x), \mathrm{d}^{**}(y_b', x))$, it follows that $\mathrm{CLR}(b_0, y_b'', x) > \mathrm{CLR}(b_0, y_b', x)$. The other conclusions of the Lemma follow by a similar argument. $\square$

Proof of Theorem 2. As described in the main text, the proof proceeds in two steps. The first step follows from the argument given in Section 4 of Matthes and Truax (1967). For sake of completeness, we reproduce it here. Let $C_Y^{\mathrm{sim,cas}}$ denote the class of similar tests with convex acceptance sections. Let φ be any test in $C_Y^{\mathrm{sim,cas}}$. Assume that φ is not admissible in the class C_Y. Then there exists a test $\varphi' \in C_Y$ that dominates φ. Using completeness of X, which follows from Theorem 4.3.1 in Lehmann and Romano (2005), it can be shown that a test that dominates φ needs to satisfy the following two equations for all $x \in \mathcal{X}$:

$$\int_{\mathcal{Y}_b} (\varphi'(y_b, x) - \varphi(y_b, x)) f(y_b|b_0) dy_b = 0 \tag{13}$$

and

$$\int_{\mathcal{Y}_b} y_b(\varphi'(y_b, x) - \varphi(y_b, x))f(y_b|b_0)dy_b = 0, \tag{14}$$

where $f(y_b|b) = f(y_b|x, b, d)$ is the pdf of a normal with mean b and variance 1.[29] Eq. (13) implies that φ' needs to have the same size as φ conditional on $X = x$. Eq. (14) implies that φ' needs to have the same "center of gravity" as φ conditional on $X = x$, i.e., the same conditional expected value over the rejection region. Since $C_Y^{\text{sim,cas}}$ is essentially complete, which follows from Theorem 3.1 in Matthes and Truax (1967), we can assume w.l.o.g. that $\varphi' \in C_Y^{\text{sim,cas}}$, i.e., φ' is similar with convex acceptance sections. Since y_b is scalar, the conditional acceptance regions of φ and φ' are given by intervals on the real line, which are uniquely determined by the size and the conditional expected value. Consequently, $\varphi' = \varphi$, a contradiction.

Next, we show that, under condition (i), $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is similar with convex acceptance sections. Similarity follows from Corollary 1 together with the observation that, under the null hypothesis, $\text{CLR}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ conditional on $X^s = x^s$ is, due to the independence of Y_b and X^s, distributed as $\text{CLR}(b_0, Z + b_0, x^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$, where $Z|X^s = x^s \sim N(0, 1)$, whose $1-\alpha$ quantile is given by $\text{cv}_{1-\alpha}(b_0, x^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$. Note that the acceptance region of $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is measurable with respect to the Borel σ-algebra on $\mathcal{Y} \times \mathcal{X}$ and closed, by continuity of $\text{CLR}(b_0, y_b, x^s)$ in (y_b, x), suppressing the dependence on $(\Sigma^s, B(\beta^*), D(\delta^*)^s)$. Furthermore, its x-sections are convex, as $\text{CLR}(b_0, y_b, x^s)$ is strictly quasiconvex on $\mathcal{Y}_b$ given $x \in \mathcal{X}$. To see this, note that, by Lemma 5, we have that for all $\lambda \in (0, 1)$ and all $y_{b1}, y_{b2} \in \mathcal{Y}$ with $y_{b1} < y_{b2}$, $\text{CLR}(b_0, \lambda y_{b1} + (1-\lambda)y_{b2}, x^s) < \max\{\text{CLR}(b_0, y_{b1}, x^s), \text{CLR}(b_0, y_{b2}, x^s)\}$.

Lastly, we show that, under condition (ii), $\varphi_{\text{CLR}}(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)$ is also similar with convex acceptance sections. From Lemmas 4 and 5, it follows that $\text{CLR}(0, y_b, x^s)$ is non-decreasing in y_b for all $y_b \in \mathcal{Y}_b$ and all $x \in \mathcal{X}$ and strictly increasing in y_b for all $y_b \geq 0$ and all $x \in \mathcal{X}$, as long as $\Sigma^s_{\delta\beta} \geq 0$. This together with the definition of $\text{cv}_{1-\alpha}(0, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ implies that for $\alpha \leq 0.5$ $\varphi_{\text{CLR}}(0, Y_b, X^s, \Sigma^s, [0, \infty), D(\delta^*)^s)$ equals the one-sided test that rejects if $Y_b > z_\alpha$, which is similar with convex acceptance sections.[30] □

Appendix C. Asymptotic results for the CLR test

In Appendix C.1, we show that the confidence set (CS) obtained by inverting the CLR test controls asymptotic size and, under some conditions, is asymptotically similar, uniformly over Γ. In Appendix C.2, we show that the CLR test asymptotically inherits the optimality properties of the CLR test based on Y, in the sense of Müller (2011). Proofs are collected in Appendix C.3. Throughout this section, we let $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{\infty\}$.

C.1. Asymptotic size control

In this section, we heavily borrow notation from Andrews, Cheng and Guggenberger (2011). The nominal $1-\alpha$ CS obtained by inverting the CLR test is given by

$$\begin{aligned}\text{CS}_n =& \{\beta : \varphi_{\text{CLR}}(\beta, \tilde{\beta}_n, X_n^s, \hat{\Sigma}^s/n, B^e, D^{e,s}) = 0\}\\ =& \{\beta : \text{CLR}(\beta, \tilde{\beta}_n, X_n^s, \hat{\Sigma}^s/n, B^e, D^{e,s}) \leq \text{cv}_{1-\alpha}(\beta, X_n^s, \hat{\Sigma}^s/n, B^e, D^{e,s})\}.\end{aligned}$$

The coverage probability of a CS for the true parameter vector $\bar{\beta}$, under $\bar{\gamma} = (\bar{\theta}, \bar{\omega}) \in \Gamma$ with $\bar{\theta} = (\bar{\beta}, \bar{\delta})$, is given by

$$\text{CP}_n(\bar{\gamma}) \equiv P_{\bar{\gamma}}(\bar{\beta} \in \text{CS}_n) = P_{\bar{\gamma}}(\text{CLR}(\bar{\beta}, \tilde{\beta}_n, X_n^s, \hat{\Sigma}^s/n, B^e, D^{e,s}) \leq \text{cv}_{1-\alpha}(\bar{\beta}, X_n^s, \hat{\Sigma}^s/n, B^e, D^{e,s})).$$

Here, the subscript on $P(\cdot)$ indicates the parameter value under which the probability is evaluated. The asymptotic size, which approximates finite sample size, is given by

$$\text{AsySz} = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} \text{CP}_n(\gamma).$$

Similarly, the asymptotic maximum coverage probability is given by

$$\text{AsyMaxCP} = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} \text{CP}_n(\gamma).$$

Next, we provide a characterization result for AsySz and AsyMaxCP using results in Andrews, Cheng and Guggenberger (2011). To that end, we first introduce some additional notation. Let β^c denote the subvector of β whose entries are restricted below by 0 and assume $\beta^c = (\beta_1, \ldots, \beta_{K^c})'$ w.l.o.g., i.e., β^c collects the first K^c elements of β where $0 \leq K^c \leq K$. As before, we assume w.l.o.g. that all elements of δ^s are restricted below by 0. Define

$$H = \{h = (b^c, d^s, \gamma^*) : \sqrt{n}\beta_n^c \to b^c \in \mathbb{R}_{+,\infty}^{K^c}, \sqrt{n}\delta_n^s \to d^s \in \mathbb{R}_{+,\infty}^{L^s}, \text{ and } \gamma_n \to \gamma^* \text{ for some } \{\gamma_n \in \Gamma : n \geq 1\}\}$$

[29] $f(y_b|b)$ equals $f(y_b|x, b, d)$ because X is a sufficient statistic for d and Y_b and X are independent.

[30] Alternatively, note that the one-sided test is admissible by Theorem 4.4.1 and Problem 4.1 in Lehmann and Romano (2005). Note, further, that admissibility of the one-sided test is obtained, although Y_b, considered as an estimator of b, is inadmissible when $b = 0$, cf. Tripathi and Kumar (2007).

and let $h_n(\gamma) = (\sqrt{n}\beta^c, \sqrt{n}\delta^s, \gamma)$. With a slight abuse of notation, let $b = (g(b^c), 0_{K-K^c})$, where $g : \mathbb{R}^{K^c}_{+,\infty} \to \mathbb{R}^{K^c}_{+}$ with $g_j(x) = x_j$ if $x_j < \infty$ and $g_j(x) = 0$ otherwise $\forall j \in \{1, \dots, K^c\}$. Correspondingly, let $B_\infty = B_{\infty,1} \times \cdots \times B_{\infty,K^c} \times (-\infty, \infty)^{K-K^c}$ where $B_{\infty,j} = [0, \infty)$ if $b^c_j < \infty$ and $B_{\infty,j} = (-\infty, \infty)$ otherwise $\forall j \in \{1, \dots, K^c\}$. Let s' denote the possibly empty subset of s for which $d^s_j < \infty$, where $j \in \{1, \dots, L^s\}$, such that $d^{s'}_j < \infty$ for all $j \in \{1, \dots, L^{s'}\}$. And let $D^{s'}_\infty = [0, \infty)^{L^{s'}}$. Then, for any sequence $\{\gamma_n \in \Gamma : n \ge 1\}$ for which $h_n(\gamma_n) \to h \in H$, $\mathrm{CP}_n(\gamma_n) \to \mathrm{CP}(h)$, where suppressing the dependence of Σ on γ^*

$$\mathrm{CP}(h) = P(\mathrm{CLR}(b, Y_b, X^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_\infty, D^{s'}_\infty) \le \mathrm{cv}^{\mathrm{alt}}_{1-\alpha}(b, X^{s'}, \Sigma^{s'}, B_\infty, D^{s'}_\infty))$$

with

$$\begin{pmatrix} Y_b \\ X^{s'} \end{pmatrix} \sim N\left(\begin{pmatrix} b \\ d^{s'} \end{pmatrix}, \begin{pmatrix} \Sigma_{\beta\beta} & 0 \\ 0 & \Sigma^{s'}_{\delta\delta} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}\Sigma^{s'}_{\beta\delta} \end{pmatrix}\right).$$

Here and in what follows, $\mathrm{cv}^{\mathrm{alt}}_{1-\alpha}(b, X^{s'}, \Sigma^{s'}, B_\infty, D^{s'}_\infty) \equiv \mathrm{cv}_{1-\alpha}(b, X^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_\infty, D^{s'}_\infty)$, i.e., $\mathrm{cv}^{\mathrm{alt}}_{1-\alpha}(b, X^{s'}, \Sigma^{s'}, B_\infty, D^{s'}_\infty)$ denotes the conditional $1-\alpha$ critical value of $\mathrm{CLR}(\cdot)$ given the first summand of the second entry and the third entry of $\mathrm{cv}_{1-\alpha}(\cdot)$.[31] $\mathrm{CP}(h)$ is derived in the proof of the following Lemma.

Lemma 6. *Under Assumptions 1–5,* $\mathrm{AsySz} = \inf_{h\in H} CP(h)$ *and* $\mathrm{AsyMaxCP} = \sup_{h\in H} CP(h)$.

Lemma 6 follows from Corollary 2.1(b) in Andrews, Cheng and Guggenberger (2011) by verifying their Assumptions B1, B2*, C1, and C2.

The following Corollary shows that the CS obtained by inverting the CLR test controls asymptotic size and, under certain conditions, is asymptotically similar.

Corollary 2. *Suppose Assumptions 1–5 hold. Then, AsySz $\ge 1-\alpha$. Furthermore, if $B \ne [0, c]^K$ or $B = [0, c]$, $\hat{\Sigma}^s_{\delta\beta} \ge 0$, and $\alpha \le 0.5$, then AsySz = AsyMaxCP = $1-\alpha$.*[32]

Note that the condition for the CS obtained by inverting the CLR test to be asymptotically similar if $B = [0, c]$ reflects that the "asymptotic version" of the underlying test, as outlined in the main text, may cease to be similar when the testing problem is one-sided and the corresponding condition is not satisfied. Note that similar but less succinct conditions apply if $B = [0, c]^K$ with $K > 1$.

C.2. Asymptotic optimality

In this section, we heavily rely on the framework used in Müller (2011). Let $G_n(\theta^*, \mu, \Sigma, m)$ denote the distribution function of the data, W_n, under a given model m, parameterized by (θ^*, μ, Σ).[33] The model m takes a similar role as ω above. As before, we suppress the dummy variable of integration and, thus, the argument of the distribution function, for notational convenience. Let φ_W denote a test for testing (7) based on W_n and let C_W denote the class of all such tests. We are interested in the asymptotic properties of φ_W over a large set of models, say $\mathcal{M}$. In particular, we consider the set of models that satisfy Eq. (6), i.e.,

$$\mathcal{M} = \{m : F_n(\theta^*, \mu, \Sigma, m) \to F(\theta^*, \mu, \Sigma)\},$$

where $F_n(\theta^*, \mu, \Sigma, m)$ and $F(\theta^*, \mu, \Sigma)$ denote the distribution functions of $f_n(W_n) \equiv (\sqrt{n}(\tilde{\theta}_n - \theta^*), \hat{\Sigma})$ and (Y, Σ), respectively. As Σ is known in the limit, we also let $F(\theta^*, \mu, \Sigma)$, with a slight abuse of notation, denote the distribution function of Y, cf. Section 3.1. Note that $F_n(\theta^*, \mu, \Sigma, m) = f_n G_n(\theta^*, \mu, \Sigma, m)$, using the notational conventions in Müller (2011).

Following Müller (2011), let $C^{\mathrm{lev}}_W \subset C_W$ denote the class of tests that is asymptotically level α for all models $m \in \mathcal{M}$, i.e., $\varphi_W \in C^{\mathrm{lev}}_W$ if

$$\limsup_{n\to\infty} \int \varphi_W dG_n(\theta^*, \mu, \Sigma, m) \le \alpha \text{ for all } \mu \in M_0(\theta^*), m \in \mathcal{M}. \tag{15}$$

A test φ_W is said to be *asymptotically admissible* in a class of tests $C^*_W \subset C^{\mathrm{lev}}_W$ if $\varphi_W \in C^*_W$ and if for all $\varphi'_W \in C^*_W$

$$\liminf_{n\to\infty} \int \varphi'_W dG_n(\theta^*, \mu, \Sigma, m) \ge \lim_{n\to\infty} \int \varphi_W dG_n(\theta^*, \mu, \Sigma, m) \text{ for all } \mu \in M_1(\theta^*), m \in \mathcal{M}$$

---

31 In the case at hand, $\Sigma^{s'}$ is non-random such that

$$\mathrm{cv}^{\mathrm{alt}}_{1-\alpha}(b, x^{s'}, \Sigma^{s'}, B_\infty, D^{s'}_\infty) = \inf\{q \in \mathbb{R} : P(\mathrm{CLR}(b, \Sigma^{1/2}_{\beta\beta}Z + b, X^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_\infty, D^{s'}_\infty) \le q | X^{s'} = x^{s'}) \ge 1-\alpha\},$$

where $Z|X^{s'} = x^{s'} \sim N(0, I_K)$ and where $x^{s'}$ denotes a possible realization of $X^{s'}$.

32 Note that s may change with n, to ensure that $\hat{\Sigma}^s_{\delta\beta} \ge 0$ for all n.

33 Müller (2011) writes $F_T(m, \theta, \gamma)$ (see p. 414): Here, G_n "replaces" F_T, μ "replaces" θ, and Σ "replaces" γ. Note that Müller (2011) suppresses the dependence (of μ) on θ^* (see e.g., his unit root test example on p. 400, where $\theta^* = 1$). Below, we will also let f_n "replace" h_T.

implies

$$\lim_{n\to\infty}\int \varphi'_W dG_n(\theta^*,\mu,\Sigma,m)=\lim_{n\to\infty}\int \varphi_W dG_n(\theta^*,\mu,\Sigma,m) \text{ for all } \mu\in M_1(\theta^*), m\in\mathcal{M}.$$

Note that this definition of asymptotic admissibility is weaker than the definition of admissibility given in Section 3, as we only compare asymptotic rejection frequencies under the alternative and, thus, restrict asymptotic admissibility to (a subclass of) C_W^{lev} rather than C_W. We employ this weaker definition because it allows us to directly apply the results in Müller (2011).

Let $C_W^{\text{sim}}\subset C_W$ denote the class of asymptotically similar tests, i.e., $\varphi_W\in C_W^{\text{sim}}$ if

$$\lim_{n\to\infty}\int \varphi_W dG_n(\theta^*,\mu,\Sigma,m)=\alpha \text{ for all } \mu\in M_0(\theta^*), m\in\mathcal{M}. \tag{16}$$

Define WAP for $\varphi_W\in C_W$ in a given model m, analogous to Eq. (9), as

$$\text{WAP}_n(\varphi_W,w,m)=\iint \varphi_W dG_n(\theta^*,\mu,\Sigma,m)dw(\mu).$$

A test $\varphi_W\in C_W$ is said to be *asymptotically extended WAP-similar* if $\varphi_W\in C_W^{\text{sim}}$ and if for all $\epsilon>0$ there exist weights w_ϵ such that for all $m\in\mathcal{M}$

$$\limsup_{n\to\infty}\text{WAP}_n(\varphi'_W,w_\epsilon,m)\le\lim_{n\to\infty}\text{WAP}_n(\varphi_W,w_\epsilon,m)+\epsilon$$

for all $\varphi'_W\in C_W^{\text{sim}}$.

Lemma 7. *Assume that $f_n(\cdot)$ is measurable. Let $K=1$ and take s to be a subset for which $\hat{\Sigma}^s_{\delta\beta}\ge 0$ and $\alpha\le 0.5$ if $B=[0,c]$ and $b_0=0$. Then, $\varphi_{CLR}(\beta_{n,0},\tilde{\beta}_n,X_n^s,\hat{\Sigma}^s/n,B^e,D^{e,s})$ is* (i) *asymptotically admissible in the class of tests C_W^{lev} and* (ii) *asymptotically extended WAP-similar.*

The proof of Lemma 7 follows from Theorem 1(ii) and Section 3.1 in Müller (2011) together with the observation that, under $m\in\mathcal{M}$, $\varphi_{\text{CLR}}(\beta_{n,0},\tilde{\beta}_n,X_n^s,\hat{\Sigma}^s/n,B^e,D^{e,s})$ is asymptotically equivalent to $\varphi_{\text{CLR}}(b_0,Y_b,X^{s'},\Sigma^{s'},B(\beta^*),D(\delta^*)^{s'})$ for some s'. Note that Lemma 7 is obtained for all θ^* (with a possible exception at $\beta^*=0$) such that the CLR test enjoys the above asymptotic optimality properties, in some sense, uniformly.

C.3. Proofs

Proof of Lemma 6. As mentioned above, the conclusion of Lemma 6 follows from Corollary 2.1(b) in Andrews, Cheng and Guggenberger (2011) (ACG). First, we derive CP(h) which verifies Assumption B1 in ACG. Let $\beta_n^\dagger\equiv(g'(\beta_n^c,b^c),0_{K-K^c})$, where $g':\mathbb{R}_+^{K^c}\times\mathbb{R}_{+,\infty}^{K^c}\to\mathbb{R}_+^{K^c}$ with $g'_j(\beta_n^c,b^c)=\beta_{n,j}^c$ if $b_j^c<\infty$ and $g'_j(\beta_n^c,b^c)=0$ otherwise for all $j\in\{1,\dots,K^c\}$. Let $\delta^{s'}=(\delta_1^s,\dots,\delta_{L^{s'}}^s)'$ w.l.o.g., i.e., $\delta^{s'}$ collects the first $L^{s'}$ elements of δ^s where $0\le L^{s'}\le L^s$. Let $\delta_n^{s\dagger}$ be a $(L^s\times 1)$ vector where $\delta_{n,j}^{s\dagger}=0$ if $j\in\{1,\dots,L^{s'}\}$ and $\delta_{n,j}^{s\dagger}=\delta_{n,j}^s$ otherwise. Furthermore, let $X_n^s(\beta)\equiv\tilde{\delta}_n-\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}(\tilde{\beta}_n-\beta)=X_n^s+\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}\beta$. Then, for any sequence $\{\gamma_n\in\Gamma:n\ge 1\}$ for which $h_n(\gamma_n)\to h\in H$, we can write $\text{CP}_n(\gamma_n)$ as follows

$$\begin{aligned}
&P_{\gamma_n}(\text{CLR}(\beta_n,\tilde{\beta}_n,X_n^s,\hat{\Sigma}^s/n,B^e,D^{e,s})\le \text{cv}_{1-\alpha}(\beta_n,X_n^s,\hat{\Sigma}^s/n,B^e,D^{e,s}))\\
=&P_{\gamma_n}(\text{CLR}(\sqrt{n}\beta_n,\sqrt{n}\tilde{\beta}_n,\sqrt{n}X_n^s(\beta_n)-\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}\sqrt{n}\beta_n,\hat{\Sigma}^s,\sqrt{n}B^e,\sqrt{n}D^{e,s})\le \text{cv}^{\text{alt}}_{1-\alpha}(\sqrt{n}\beta_n,\sqrt{n}X_n^s(\beta_n),\hat{\Sigma}^s,\sqrt{n}B^e,\sqrt{n}D^{e,s}))\\
=&P_{\gamma_n}(\text{CLR}(\sqrt{n}\beta_n^\dagger,\sqrt{n}(\tilde{\beta}_n-\beta_n+\beta_n^\dagger),\sqrt{n}X_n^s(\beta_n)-\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}\sqrt{n}\beta_n^\dagger,\hat{\Sigma}^s,\sqrt{n}(B^e-\beta_n+\beta_n^\dagger),\sqrt{n}D^{e,s})\\
&\quad\le \text{cv}^{\text{alt}}_{1-\alpha}(\sqrt{n}\beta_n^\dagger,\sqrt{n}X_n^s(\beta_n),\hat{\Sigma}^s,\sqrt{n}(B^e-\beta_n+\beta_n^\dagger),\sqrt{n}D^{e,s})).
\end{aligned}$$

Here, $\sqrt{n}(B^e-\beta_n+\beta_n^\dagger)=\{\text{b}\in\mathbb{R}^K:\text{b}=\sqrt{n}(\text{b}'-\beta_n+\beta_n^\dagger)\text{ for some b}'\in B^e\}$. Other transformations of sets, above and below, are defined similarly. Note that by Theorem 1, the Continuous Mapping Theorem (CMT), and Lemma 1

$$\begin{aligned}
&\text{CLR}(\sqrt{n}\beta_n^\dagger,\sqrt{n}(\tilde{\beta}_n-\beta_n+\beta_n^\dagger),\sqrt{n}X_n^s(\beta_n)-\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}\sqrt{n}\beta_n^\dagger,\hat{\Sigma}^s,\sqrt{n}(B^e-\beta_n+\beta_n^\dagger),\sqrt{n}D^{e,s})\\
=&\text{CLR}(\sqrt{n}\beta_n^\dagger,\sqrt{n}(\tilde{\beta}_n-\beta_n+\beta_n^\dagger),\sqrt{n}(X_n^s(\beta_n)-\delta_n^{s\dagger})-\hat{\Sigma}^s_{\delta\beta}\hat{\Sigma}^{-1}_{\beta\beta}\sqrt{n}\beta_n^\dagger,\hat{\Sigma}^s,\sqrt{n}(B^e-\beta_n+\beta_n^\dagger),\sqrt{n}(D^{e,s}-\delta_n^{s\dagger}))\\
\xrightarrow{d}&\text{CLR}(b,Y_b,X^s-\Sigma^s_{\delta\beta}\Sigma^{-1}_{\beta\beta}b,\Sigma^s,B_\infty,D^s_\infty)\\
\sim&\text{CLR}(b,Y_b,X^{s'}-\Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b,\Sigma^{s'},B_\infty,D^{s'}_\infty),
\end{aligned}$$

where

$$\sqrt{n}\begin{pmatrix}\tilde{\beta}_n-\beta_n+\beta_n^\dagger\\ X_n^s(\beta_n)-\delta_n^{s\dagger}\end{pmatrix}=\sqrt{n}\begin{pmatrix}\tilde{\beta}_n-\beta_n+\beta_n^\dagger\\ X_n^{s'}(\beta_n)\\ X_n^{s-s'}(\beta_n)-\delta_n^{s-s'}\end{pmatrix}\xrightarrow{d}\begin{pmatrix}Y_b\\ X^{s'}\\ X^{s-s'}\end{pmatrix}\equiv\begin{pmatrix}Y_b\\ X^s\end{pmatrix},$$

under any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ for which $h_n(\gamma_n) \to h \in H$. Here, $D^{s}_{\infty} = D^{s'}_{\infty} \times (-\infty, \infty)^{L^s - L^{s'}}$ and $X^{s-s'}$ has mean zero, where the superscript s−s′ indicates the subvector that contains all entries of the corresponding vector with superscript s whose indexes are in $s\backslash s'$, e.g., $X^s = (X^{s'}, X^{s-s'})$.

Next, we derive the asymptotic distribution of $cv^{alt}_{1-\alpha}(\sqrt{n}\beta^{\dagger}_n, \sqrt{n}X^s_n(\beta_n), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta_n + \beta^{\dagger}_n), \sqrt{n}D^{e,s})$. By Lemma 5 in Andrews and Guggenberger (2010b),

$$cv^{alt}_{1-\alpha}(\sqrt{n}\beta^{\dagger}_n, x^s, \boldsymbol{\Sigma}^s, \sqrt{n}(B^e - \beta_n + \beta^{\dagger}_n), \sqrt{n}D^{e,s}) \to cv^{alt}_{1-\alpha}(b, x^s, \boldsymbol{\Sigma}^s, B_{\infty}, D^{s'}_{\infty} \times [0, \infty)^{L^s - L^{s'}}),$$

where x^s and $\boldsymbol{\Sigma}^s$ denote possible realizations of $\sqrt{n}X^s(\beta_n)$ and $\hat{\Sigma}^s$, respectively, and where $cv^{alt}_{1-\alpha}(b, x^s, \boldsymbol{\Sigma}^s, B_{\infty}, D^{s'}_{\infty} \times [0, \infty)^{L^s - L^{s'}})$ is continuous in x^s and $\boldsymbol{\Sigma}^s$, by arguments similar to those given at the bottom of page 38 in ACG. By continuity in x^s and Lemma 1

$$\begin{aligned}
&cv^{alt}_{1-\alpha}(b, x^s, \boldsymbol{\Sigma}^s, B_{\infty}, D^{s'}_{\infty} \times [0, \infty)^{L^s - L^{s'}})\\
=&cv^{alt}_{1-\alpha}(b, x^s - (0, x^{s-s'}), \boldsymbol{\Sigma}^s, B_{\infty}, D^{s'}_{\infty} \times ([0, \infty)^{L^s - L^{s'}} - x^{s-s'}))\\
\to &cv^{alt}_{1-\alpha}(b, x^{s'}, \boldsymbol{\Sigma}^{s'}, B_{\infty}, D^{s'}_{\infty})
\end{aligned}$$

as $x^{s-s'} \to \infty_{L^s - L^{s'}}$. We conclude that

$$cv^{alt}_{1-\alpha}(\sqrt{n}\beta^{\dagger}_n, \sqrt{n}X^s_n(\beta_n), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta_n + \beta^{\dagger}_n), \sqrt{n}D^{e,s}) \xrightarrow{d} cv^{alt}_{1-\alpha}(b, X^{s'}, \Sigma^{s'}, B_{\infty}, D^{s'}_{\infty}).$$

As the above convergence results hold jointly, we obtain CP(h) which verifies Assumption B1 in ACG. Assumption B2* in ACG is satisfied given the definition of Γ and $h_n(\gamma)$. To conclude the proof, note that Assumptions C1 and C2 in ACG are satisfied by CP(h).[34] □

Proof of Corollary 2. The first part of Corollary 2 follows immediately from the observation that, due to the independence of Y_b and $X^{s'}$, $CLR(b, Y_b, X^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_{\infty}, D^{s'}_{\infty})$ conditional on $X^{s'} = x^{s'}$ is distributed as $CLR(b, \Sigma^{1/2}_{\beta\beta}Z + b, x^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_{\infty}, D^{s'}_{\infty})$, where $Z|X^{s'} = x^{s'} \sim N(0, I_K)$, whose $1 - \alpha$ quantile is given by $cv^{alt}_{1-\alpha}(b, x^{s'}, \Sigma^{s'}, B_{\infty}, D^{s'}_{\infty})$. The second part follows from the fact that the $1-\alpha$ quantile of $CLR(b, Y_b, X^{s'} - \Sigma^{s'}_{\delta\beta}\Sigma^{-1}_{\beta\beta}b, \Sigma^{s'}, B_{\infty}, D^{s'}_{\infty})$ conditional on $X^{s'} = x^{s'}$ is unique for all $x^{s'} \in \mathbb{R}^{s'}$. If $B \neq [0, c]^K$, this follows from Corollary 1. If $B = [0, c]$, $\Sigma^s_{\delta\beta} \geq 0$, and $\alpha \leq 0.5$, this follows from Lemmas 4 and 5. □

Proof of Lemma 7. First, note that, by continuity, $\varphi^n_{CLR} \equiv \varphi_{CLR}(\beta_{n,0}, \tilde{\beta}_n, X^s_n, \hat{\Sigma}^s/n, B^e, D^{e,s})$ is a measurable function of W_n given that $f_n(W_n)$ is measurable by assumption. The following steps closely follow the proof of Lemma 6. The rejection frequency of φ^n_{CLR} for testing (7), under $\{\theta_n\}$ *local* to θ^*, is given by

$$\begin{aligned}
&P(CLR(\beta_{n,0}, \tilde{\beta}_n, X^s_n, \hat{\Sigma}^s/n, B^e, D^{e,s}) > cv_{1-\alpha}(\beta_{n,0}, \tilde{\beta}_n, X^s_n, \hat{\Sigma}^s/n, B^e, D^{e,s}))\\
=&P(CLR(b_0, \sqrt{n}(\tilde{\beta}_n - \beta^*), \sqrt{n}X^s_n(\beta^*), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta^*), \sqrt{n}D^{e,s})\\
&\quad > cv_{1-\alpha}(b_0, \sqrt{n}(\tilde{\beta}_n - \beta^*), \sqrt{n}X^s_n(\beta^*), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta^*), \sqrt{n}D^{e,s})),
\end{aligned}$$

where $X^s_n(\beta)$ is defined as in the proof of Lemma 6 and where we have suppressed the dependence of $P(\cdot)$ on the true data generating process. Under $m \in \mathcal{M}$, we have

$$\begin{aligned}
&CLR(b_0, \sqrt{n}(\tilde{\beta}_n - \beta^*), \sqrt{n}X^s_n(\beta^*), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta^*), \sqrt{n}D^{e,s})\\
=&CLR(b_0, \sqrt{n}(\tilde{\beta}_n - \beta^*), \sqrt{n}(X^s_n(\beta^*) - \delta^*), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta^*), \sqrt{n}(D^{e,s} - \delta^*))\\
\xrightarrow{d}&CLR(b_0, Y_b, X^s, \Sigma^s, B(\beta^*), D(\delta^*)^s)\\
\sim&CLR(b_0, Y_b, X^{s'}, \Sigma^{s'}, B(\beta^*), D(\delta^*)^{s'}),
\end{aligned}$$

where s′ is such that $D(\delta^*)^{s'}$ only includes those entries of $D(\delta^*)^s$ that equal $[0, \infty)$, by Lemma 1. Similarly, we have

$$cv_{1-\alpha}(b_0, \sqrt{n}(\tilde{\beta}_n - \beta^*), \sqrt{n}X^s_n(\beta^*), \hat{\Sigma}^s, \sqrt{n}(B^e - \beta^*), \sqrt{n}D^{e,s}) \xrightarrow{d} cv_{1-\alpha}(b_0, X^{s'}, \Sigma^{s'}, B(\beta^*), D(\delta^*)^{s'}).$$

As the above convergence results hold jointly, we have that φ^n_{CLR} is asymptotically equivalent to $\varphi_{CLR} \equiv \varphi_{CLR}(b_0, Y_b, X^{s'}, \Sigma^{s'}, B(\beta^*), D(\delta^*)^{s'})$, under $m \in \mathcal{M}$. As in the proof of Theorem 1(i) in Müller (2011), it follows, by the CMT, that

$$\lim_{n\to\infty} \int \varphi^n_{CLR} dF_n(\theta^*, \mu, \Sigma, m) = \int \varphi_{CLR} dF(\theta^*, \mu, \Sigma) \text{ for all } \mu \in M(\theta^*) \tag{17}$$

34 Note that $CP(h) = CP^-(h) = CP^+(h)$ using the notation in ACG.

and, consequently, by dominated convergence that

$$\lim_{n\to\infty} \mathrm{WAP}_n(\varphi^n_{\mathrm{CLR}}, w, m) = \mathrm{WAP}(\varphi_{\mathrm{CLR}}, w). \tag{18}$$

The proof of part (i) follows the reasoning of Comment 3 in Müller (2011). Consider the following constraint on $\varphi_W \in C_W$

$$\liminf_{n\to\infty} \int \varphi_W dG_n(\theta^*, \mu, \Sigma, m) \geq \pi(\mu) \text{ for all } \mu \in M_1(\theta^*), m \in \mathcal{M}, \tag{19}$$

where

$$\pi(\mu) = \int \varphi_{\mathrm{CLR}} dF(\theta^*, \mu, \Sigma).$$

The analogous constraint in the Gaussian shift model is given by

$$\int \varphi dF(\theta^*, \mu, \Sigma) \geq \pi(\mu) \text{ for all } \mu \in M_1(\theta^*). \tag{20}$$

Let $C_Y^{\mathrm{lev}} \subset C_Y$ denote the class of level α tests based on Y, i.e., $\varphi \in C_Y^{\mathrm{lev}}$ if

$$\int \varphi dF(\theta^*, \mu, \Sigma) \leq \alpha \text{ for all } \mu \in M_0(\theta^*). \tag{21}$$

Since φ_{CLR} is admissible in C_Y and, thus, in C_Y^{lev} it maximizes WAP subject to (20) and (21) for any weight function w. Therefore, repeatedly applying Theorem 1(ii) and Section 3.1 in Müller (2011) with w putting all mass on a single point $\mu \in M_1(\theta^*)$ implies that for any $\varphi_W \in C_W^{\mathrm{lev}}$ that satisfies (19) we have[35]

$$\limsup_{n\to\infty} \int \varphi_W dG_n(\theta^*, \mu, \Sigma, m) \leq \pi(\mu) \text{ for all } \mu \in M_1(\theta^*), m \in \mathcal{M}.$$

and, thus,

$$\lim_{n\to\infty} \int \varphi_W dG_n(\theta^*, \mu, \Sigma, m) = \pi(\mu) \text{ for all } \mu \in M_1(\theta^*), m \in \mathcal{M}.$$

This together with (17) implies that φ^n_{CLR} is asymptotically admissible in C_W^{lev}.

To prove part (ii), fix $\epsilon > 0$ and let w_ϵ be such that Eq. (11) holds for φ_{CLR}. By Theorem 1(ii) and Section 3.1 in Müller (2011), we have that for all $\varphi_W \in C_W^{\mathrm{sim}}$

$$\limsup_{n\to\infty} \mathrm{WAP}_n(\varphi_W, w_\epsilon, m) \leq \mathrm{WAP}(\varphi_{\mathrm{WAP}}^{\mathrm{sim}, w_\epsilon}, w_\epsilon) \text{ for all } m \in \mathcal{M}.$$

To see this, note that Eqs. (10) and (16) correspond to equations (6) and (11) in Müller (2011), respectively, with $\bar{\Theta}_0$ equal to $M_0(\theta^*)$ and $\mathcal{F}_s = \{1\}$. Using Eq. (11), we have that for any $\epsilon > 0$ there exists w_ϵ such that for all $\varphi_W \in C_W^{\mathrm{sim}}$

$$\limsup_{n\to\infty} \mathrm{WAP}_n(\varphi_W, w_\epsilon, m) \leq \mathrm{WAP}(\varphi_{\mathrm{CLR}}, w_\epsilon) + \epsilon \text{ for all } m \in \mathcal{M}.$$

The desired result then follows given Eq. (18). □

Appendix D. Additional tests for Section 4.1.1

Here, we consider several versions of the trinity of tests for the testing problem considered in Section 4.1.1. Fig. 5 shows the rejection frequency of the Likelihood Ratio (LR), Score (S), and Wald (t) tests, using "naive" (N) critical values and using "boundary imposed" (BI) critical values, e.g., $\mathrm{LR_N}$ denotes the test that compares the LR statistic to N critical values. For the testing problem at hand, the Score statistic, suppressing the dependence on $(B(\beta^*), D(\delta^*), \Sigma)$, is given by

$$\mathrm{S}(b_0, Y) = \min_{\mathrm{d}\in[0,\infty)} \begin{pmatrix} Y_b - b_0 \\ Y_d - \mathrm{d} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} Y_b - b_0 \\ Y_d - \mathrm{d} \end{pmatrix}$$

and the Likelihood Ratio statistic is given by

$$\mathrm{LR}(b_0, Y) = \mathrm{S}(b_0, Y) - \min_{\mathrm{b}\in(-\infty,\infty), \mathrm{d}\in[0,\infty)} \begin{pmatrix} Y_b - \mathrm{b} \\ Y_d - \mathrm{d} \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} Y_b - \mathrm{b} \\ Y_d - \mathrm{d} \end{pmatrix}.$$

The N critical values are given by the $1 - \alpha$ quantile of a $\chi^2(1)$ (or its square root in case of the t) and the BI critical values are given by the $1 - \alpha$ quantiles of the test statistics under $\mu = (b_0, 0)$.

[35] To see this, note that Eqs. (15), (19), and (20) correspond to equations (8), (10) and (5) in Müller (2011), respectively.

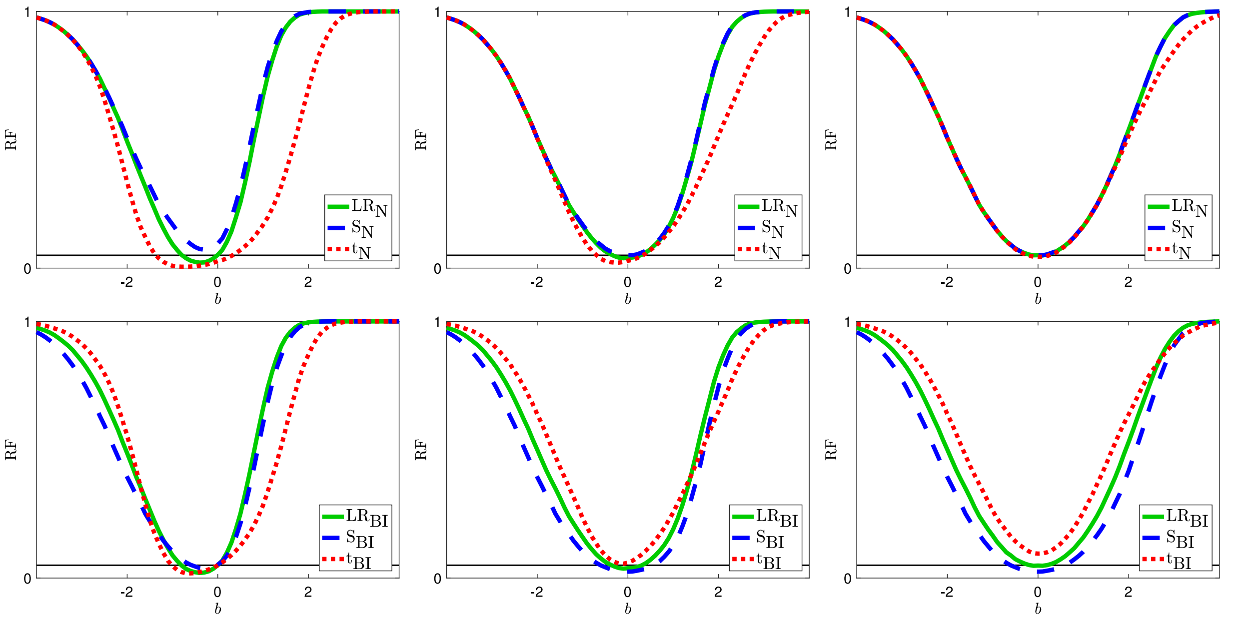

Fig. 5. Rejection frequency as a function of b for testing $H_0 : b = 0$ with $d = 0, 1, 2$ from left to right. $\rho = 0.9$. 1st row: $\mathrm{LR_N}$ (solid), $\mathrm{S_N}$ (dashed), and $\mathrm{t_N}$ (dotted). 2nd row: $\mathrm{LR_{BI}}$ (solid), $\mathrm{S_{BI}}$ (dashed), and $\mathrm{t_{BI}}$ (dotted).

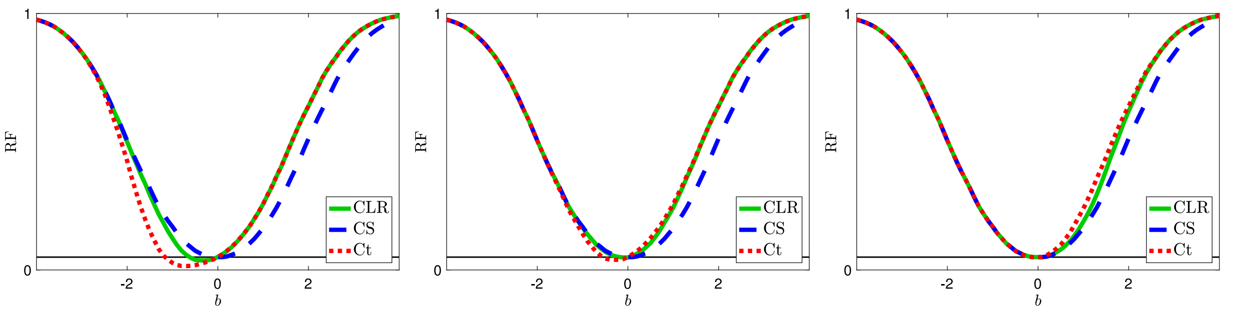

Fig. 6. Rejection frequency as a function of b of CLR (solid), CS (dashed), and Ct (dotted) for testing $H_0 : b = 0$ with $d = 0, 1, 2$ from left to right. $\rho = 0.9$.

Note that the LR and t statistics constitute "asymptotic versions" of the (rescaled) Quasi-Likelihood Ratio and Wald statistics considered in Andrews (2001) when $V(\gamma^*) = a\mathcal{J}(\gamma^*)$ for some $0 < a < \infty$. Since the Wald and Score statistics considered in Andrews (2001) are asymptotically equivalent when $V(\gamma^*) = a\mathcal{J}(\gamma^*)$ for some $0 < a < \infty$, the t statistic also constitutes an "asymptotic version" of his Score statistic. The tests in Andrews (2001) use BI critical values when the nuisance parameter is at the boundary, $d = 0$, and N critical values when they are in the interior, $d = \infty$. In practice, it is seldom known whether the nuisance parameter is at the boundary or not. Furthermore, the nuisance parameter may be near the boundary, $0 < d < \infty$. We may, therefore, mistakenly use BI critical values when the true parameter is not at the boundary, $d > 0$, or mistakenly use N critical values when the true parameter is near or at the boundary, $d < \infty$. Fig. 5 analyzes these mistakes.

Fig. 5 shows that $\mathrm{LR_N}$ and $\mathrm{S_N}$ overreject when $d = 0$, although it is, in case of the former, hard to see with naked eyes.[36] This overrejection is a direct consequence of the fact that for both statistics, LR and S, the boundary, $d = 0$, turns out to be the *unique* least favorable configuration. While $\mathrm{S_N}$ overrejects for all $d < \infty$, as the 95th quantile of the null distribution of the S statistic is strictly decreasing in d, the $\mathrm{LR_N}$ starts to be undersized for $d \gtrapprox 0.018$; the 95th quantile of the null distribution of the LR statistic is U-shaped in d. Equivalently, the corresponding tests that rely on BI critical values underreject for $d > 0$. Although not shown here, the $\mathrm{LR_{BI}}$ and the $\mathrm{S_{BI}}$ can have very poor power when d is far from the boundary and L, the dimension of d, is large. In the case at hand, the *unique* least favorable configuration for the t is given by $d = \infty$. As a result, we find that the $\mathrm{t_N}$ underrejects for all $d < \infty$. Similarly, the $\mathrm{t_{BI}}$ overrejects for all $d > 0$, as the 95th quantile of the null distribution of the t statistic is strictly increasing in d.

36 Kopylev and Sinha (2011) also document overrejection of $\mathrm{LR_N}$ for several other constellations of the testing problem.

Fig. 6 shows the rejection frequency of the Conditional Score test (CS—with a slight abuse of notation), the Conditional t-test (Ct), and the CLR for ease of reference. The CS compares the Conditional Score statistic, introduced in Remark 1, and the Ct compares $|b^{**}(Y_b, X) - b_0|$ using the notation of Appendix B to their respective conditional $1-\alpha$ critical values (defined analogously to Eq. (5)). As noted in Remark 1, the CS equals the regular t-test, cf. Fig. 1. The Ct has slightly higher power than the CLR for $b > 0$ when $d > 0$, at the cost of considerably lower power for $b < 0$.

References

Abrevaya, J., Shen, S., 2014. Estimation of censored panel-data models with slope heterogeneity. J. Appl. Econ. 29 (4), 523–548.
Andrews, D.W.K., 1999. Estimation when a parameter Is on a boundary. Econometrica 67 (6), 1341–1383.
Andrews, D.W.K., 2000. Inconsistency of the bootstrap when a parameter Is on the boundary of the parameter space. Econometrica 68 (2), 399–405.
Andrews, D.W.K., 2001. Testing when a parameter Is on the boundary of the maintained hypothesis. Econometrica 69 (3), 683–734.
Andrews, D.W.K., 2002. Generalized method of moments estimation when a parameter Is on a boundary. J. Bus. Econ. Stat. 20 (4), 530–544.
Andrews, D.W.K., Cheng, X., 2012a. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80 (5), 2153–2211.
Andrews, D.W.K., Cheng, X., 2012b. Supplement to 'Estimation and inference with weak, semi-strong, and strong identification'. Econ. Suppl. Mater. 80 (5).
Andrews, D.W.K., Cheng, X., 2014. GMM estimation and uniform subvector inference with possible identification failure. Econometric Theory 30 (2), 287–333.
Andrews, D.W.K., Cheng, X., Guggenberger, P., 2011. Generic results for establishing the asymptotic size of confidence sets and tests, Cowles Foundation Discussion Paper No. 1813.
Andrews, D.W.K., Guggenberger, P., 2010a. Applications of subsampling, hybrid, and size-correction methods. J. Econometrics 158 (2), 285–305.
Andrews, D.W.K., Guggenberger, P., 2010b. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26 (2), 426–468.
Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.
Elliott, G., Müller, U.K., Watson, M.W., 2015. Nearly optimal tests when a nuisance parameter Is present under the null hypothesis. Econometrica 83 (2), 771–811.
Francq, C., Zakoïan, J.-M., 2009. Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons. J. Am. Stat. Assoc. 104 (485), 313–324.
Frazier, D.T., Renault, E., 2016. Indirect Inference With(Out) Constraints, arXiv preprint arXiv:1607.06163. Working paper.
Geyer, C.J., 1994. On the asymptotics of constrained M-estimation. Ann. Statist. 22 (4), 1993–2010.
Goeree, M.S., 2008. Limited information and advertising in the US personal computer industry. Econometrica 76 (5), 1017–1074.
Ketz, P., 2018. A Simple Solution to Invalid Inference in the Random Coefficients Logit Model. Working Paper.
Kopylev, L., Sinha, B., 2011. On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary. Sankhya B 73 (1), 20–41.
Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses. Springer.
Matthes, T.K., Truax, D.R., 1967. Tests of composite hypotheses for the multivariate exponential family. Ann. Math. Stat. 38 (3), 681–697.
McCloskey, A., 2017. Bonferroni-based size-correction for nonstandard testing problems. J. Econometrics 200 (1), 17–35.
Montiel-Olea, J., 2018. Admissible, Similar Tests: A Characterization. Working Paper.
Moreira, H., Moreira, M.J., 2013. Contributions to the Theory of Optimal Tests. Working paper.
Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.
Müller, U.K., 2011. Efficient tests under a weak convergence assumption. Econometrica 79 (2), 395–435.
Nevo, A., 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. J. Econ. Manag. Strategy 9 (4), 513–548.
Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69 (2), 307–342.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, pp. 2111–2245 (Chapter 36).
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.
Pollard, D., 1985. New ways to prove central limit theorems. Econometric Theory 1 (3), 295–313.
Reynaert, M., Verboven, F., 2014. Improving the performance of random coefficients demand models: The role of optimal instruments. J. Econometrics 179 (1), 83–98.
Tripathi, Y.M., Kumar, S., 2007. Estimating a positive normal mean. Stat. Pap. 48 (4), 609–629.
Van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press.