UNIFORM CRITICAL VALUES FOR LIKELIHOOD RATIO TESTS IN BOUNDARY PROBLEMS

Giuseppe Cavaliere[a], Adam McCloskey[b], Rasmus S. Pedersen[c] and Anders Rahbek[c]

September 3, 2026

Abstract

Limit distributions of likelihood ratio statistics are well-known to be discontinuous in the presence of nuisance parameters at the boundary of the parameter space, which leads to size distortions when standard critical values are used for testing. In this paper, we propose a new and simple way of constructing critical values that yields uniformly correct asymptotic size, regardless of whether nuisance parameters are at, near or far from the boundary of the parameter space. Importantly, the proposed critical values are trivial to compute and at the same time provide powerful tests in most settings. In comparison to existing size-correction methods, the new approach exploits the monotonicity of the two components of the limiting distribution of the likelihood ratio statistic, in conjunction with rectangular confidence sets for the nuisance parameters, to gain computational tractability.

Uniform validity is established for likelihood ratio tests based on the new critical values, and we provide illustrations of their construction in two key examples: (i) testing a coefficient of interest in the classical linear regression model with non-negativity constraints on control coefficients, and, (ii) testing for the presence of exogenous variables in autoregressive conditional heteroskedastic models (ARCH) with exogenous regressors.

Keywords: Bonferroni correction; Boundary problems; Likelihood ratio test; Nuisance parameters; Uniform inference.

---

[a]Department of Economics, University of Bologna, Italy and Department of Economics, University of Exeter, UK.
[b]Department of Economics, University of Colorado, USA.
[c]Department of Economics, University of Copenhagen, Denmark.
Correspondence to: Adam McCloskey, email adam.mccloskey@colorado.edu. Address: 256 UCB, Boulder, Colorado 80309-0256, USA

Introduction

As is now well-known, the asymptotic distribution of a likelihood ratio (LR) statistic is discontinuous under the null hypothesis being tested at points for which a vector nuisance parameter lies on the boundary of its parameter space (Self and Liang, 1987; Andrews, 2001). This discontinuity of the LR statistic's asymptotic distribution implies that its finite-sample distribution under the null hypothesis depends upon how close the vector nuisance parameter is to its boundary, making it unclear which null distribution critical values (CVs) ought to be constructed from. In addition, the upper quantiles of these null distributions often exceed those of the chi-squared distributions from which standard CVs are derived, even in large samples, so that standard chi-squared-based CVs do not control asymptotic size. These features complicate the proper formation of CVs that control the asymptotic size of the LR test uniformly over the true parameter space. In this setting, we propose a simple and computationally-tractable method of CV formation (Algorithm 1) that controls the asymptotic size of a test using the standard LR statistic. The correct uniform asymptotic size of the resulting test (Theorem 3.3) ensures that the test also has good size properties in finite samples whether the nuisance parameter is at, near or far away from the boundary of its parameter space.

A few recent advances in the literature have now produced hypothesis tests that are uniformly valid over the nuisance parameter space (e.g., Andrews and Guggenberger, 2009; McCloskey, 2017). While these tests are computationally involved, our focus in this paper is to introduce a hypothesis test that is simple to compute even when the dimension of the vector of nuisance parameters is not small, without sacrificing much power. To attain this goal, we consider the simple-to-compute standard LR statistic as it is one of the most widely-used tests in statistics and is efficient in finite samples under correct specification when nuisance parameters are far from the boundary. This latter property does not generally hold for other widely-used tests such as those based upon Wald and Lagrange Multiplier statistics. However, the standard χ^2-based CVs for LR statistics are invalid when nuisance parameters are on or near the boundary. In contrast, the CVs we propose are based upon the asymptotic distributions of this test statistic under drifting parameter sequences. These asymptotic distributions possess a unique component-wise stochastic monotonicity property that we exploit to obtain computationally-tractable CVs yielding correct uniform asymptotic size, providing further motivation for focusing on the LR statistic.

In order to establish uniform asymptotic size control, we derive the null asymptotic distribution of LR statistics under a comprehensive class of parameter sequences that drift the nuisance parameter vector toward the boundary at any sample size-dependent rate. When evaluated at the rate corresponding to the square root of the sample size, the asymptotic distribution of the LR statistic depends upon a local nuisance parameter vector and, for the relevant value of this nuisance parameter, this local asymptotic distribution approximates the finite-sample distribution of the LR statistic. The value of this local nuisance parameter corresponds to how far the nuisance parameter is from its boundary in the finite-sample problem. However, because the relevant value of this local nuisance parameter is equal to the limit of the square root of the sample size multiplied by the true value of the (drifting) nuisance parameter, it is not consistently estimable. Nevertheless, it can still be "estimated" by an asymptotically Gaussian random vector centered at its true value, allowing us to construct a rectangular confidence set for its true value.

Our key insight is that because the two components of the local asymptotic distribution of the LR statistic are stochastically monotone in the underlying local nuisance parameter, if we evaluate the two components separately, the first at a value that is "too small" and the second at a value that is "too large", the distribution that results from adding the two components together will stochastically dominate the relevant local asymptotic distribution. Therefore, we can use the two opposing corners of the rectangular confidence set as the "too small" and "too large" values to obtain stochastic dominance with high probability and use the upper quantiles of the dominating distribution as a CV that controls size with high probability. Finally, we adjust the quantile level of the dominating distribution according to a standard Bonferroni correction to account for the uncertainty corresponding to the level of the rectangular confidence set to obtain uniform asymptotic size control.

The main theoretical result of our paper (Theorem 3.3) proves that our algorithm for CV formation (Algorithm 1) controls asymptotic size uniformly in a general framework encompassing numerous and varied applications. Examples include tests of a one-dimensional sub-vector of the mean in the multivariate Gaussian location model with a restricted mean vector, tests of a regression coefficient in the linear regression model when some coefficients have a known sign and tests of parameters in random coefficients models such as the workhorse empirical industrial organization model of Berry, Levinsohn and Pakes (1995). We verify that the conditions of our main theoretical result hold in applications of tests on regression coefficients and specification tests in ARCH-type models, the latter being a pervasive example of tests suffering from boundary problems in the recent literature on conditional volatility models, see, e.g., Francq and Zakoïan (2009), Cavaliere, Nielsen and Rahbek (2017), Cavaliere, Nielsen, Pedersen and Rahbek (2022) and Cavaliere, Perera and Rahbek (2024). We contribute to this literature by providing the first hypothesis test we are aware of with proven uniform asymptotic validity for specification testing for the presence of exogenous variables in ARCH-type models.

Several papers in the literature derive asymptotic distributions of estimators and test statistics at the boundary of the parameter space; see, e.g., Self and Liang (1987), Shapiro (1989), Geyer (1994), Silvapulle and Silvapulle (1995), and Andrews (2001). More recently, Cavaliere et al. (2022) propose a bootstrap-based CV construction that implicitly uses an estimator to switch between the quantiles of the asymptotic distribution at and far away from the boundary in large samples. However, using CVs that correspond to distributions at the boundary and/or far away from the boundary does not ensure uniform asymptotic size control since null distributions can have larger quantiles in intermediate ranges near the boundary of the parameter space.

This latter fact has been recognized in the literature, leading to inference methods that control asymptotic null rejection probabilities uniformly across all parameter sequences that may drift toward the boundary of the parameter space. In particular, Andrews and Guggenberger (2009) propose CVs in a general testing framework that are asymptotically equivalent to least-favorable CVs which find the maximal quantile of a test statistic's null asymptotic distribution over the nuisance parameter that cannot be consistently estimated. Recent alternative approaches allowing for quite general constraints, including Hong and Li (2020) and Li (2025), advocate similar least favorable-type CVs that are derived from sophisticated bootstrap implementations. Recognizing the conservative nature of tests using these CVs, and its negative consequence on power, McCloskey (2017) proposes CVs that do not maximize these quantiles over the entire nuisance parameter space but rather a first stage confidence set for the nuisance parameter; see also Berger and Boos (1994) and Silvapulle (1996) for parametric finite-sample versions of this approach. In the context of boundary problems, Mitchell, Allman and Rhodes (2019) apply McCloskey's (2017) approach to CV formation for LR statistics although, unlike the current paper, they focus on problems that may also feature singularities (see Drton, 2009). Fan and Shi (2023) also take this approach to CV formation for Wald and Quasi-Likelihood Ratio statistics in more complicated boundary problems that require an initial first step of identifying an implicit nuisance parameter.

Although McCloskey's (2017) method tends to produce better power properties than the least favorable approach of Andrews and Guggenberger (2009) and related papers, both methods are in general computationally intractable when the nuisance parameter is not low-dimensional because they generally require the optimization of a function whose values must be simulated at each parameter value. The approach we propose in this paper is similar in spirit to the approach of McCloskey (2017), but by exploiting the monotonicity properties of the two components of the LR statistic, remains computationally tractable in the presence of nuisance parameters that are not low-dimensional because it does not require optimization of a function calculated via simulation. Our approach is also conceptually related to the inference approaches of Romano, Shaikh and Wolf (2014) and McCloskey (2020), who construct critical values by simulating the quantiles of asymptotic distributions of statistics at the corners of rectangular confidence sets in the settings of moment inequality models and inference after model selection, respectively. However, in these two settings, the authors do not work separately on components of the underlying test statistics as we do here. Rather, the asymptotic distributions of the test statistics are themselves monotone in the local nuisance parameter in those settings.

Finally, Ketz (2018) and Ketz and McCloskey (2025) propose computationally tractable and uniformly valid approaches to inference when parameters may be on or near the boundary of the parameter space. Although these approaches are different from that taken in this paper, one commonality these two works share with our approach here is that they use one-step estimators, discussed e.g., in Newey and McFadden (1994), to attain asymptotically Gaussian estimation as input to form valid inference regardless of where the true parameter lies in relation to the boundary of their parameter space.

Structure of the paper. The remainder of the paper is organized as follows. Section 1 provides basic intuition for our proposed method. Section 2 introduces the general setting and assumptions as well as the ongoing examples on hypothesis testing in a linear regression model with positivity constrained coefficients and specification testing in ARCH models. Section 3 contains our proposed Algorithm 1 for CV construction and proves the validity of the CVs. Section 4 considers the performance of our proposed method by means of Monte Carlo simulations, and Section 5 contains a small empirical illustration in relation to specification testing in ARCH models for various stock market indices. All proofs are collected in Supplemental Appendix A. Supplemental Appendix B provides a description of the conditional likelihood ratio (CLR) test of Ketz (2018) while tables appear at the very end of the document.

Notation. The following notation is used throughout the paper. The set of positive definite matrices with eigenvalues bounded below by some $\kappa > 0$, and above by κ^{-1} is denoted by Υ. For a square $m \times m$ matrix A, let $\text{diag}(A)$ denote the diagonal matrix with the same diagonal of A. Moreover, let $\text{diagv}(A)$ denote the m-dimensional column vector with the diagonal elements of A as entries. $\mathbb{I}(\cdot)$ is the indicator function. With A a positive definite matrix and x a vector, $\|x\|_A = \sqrt{x'Ax}$.

1 Main ideas in a simple Gaussian model

To obtain the basic intuition underlying our proposed CV construction, consider the case of a scalar parameter of interest γ and a scalar nuisance parameter b, where $\gamma \geq 0$ and $b \geq 0$. We consider testing $\mathsf{H}_0 : \gamma = 0$, against the alternative, $\mathsf{H}_1 : \gamma > 0$. For simplicity, we frame this in terms of a bivariate normal model for $Y \in \mathbb{R}^2$, where

$$Y = (Y_\gamma, Y_b)' \sim \mathrm{N}(\lambda, \Sigma),$$

with $\lambda = (\gamma, b)'$ and Σ a known positive definite covariance matrix. In standard settings, this model corresponds to a limiting experiment. For testing H_0, consider the standard LR statistic as defined by

$$\mathsf{LR} = \inf_{\lambda \in \{0\} \times [0,\infty]} (\lambda - Y)'\Sigma^{-1}(\lambda - Y) - \inf_{\lambda \in [0,\infty] \times [0,\infty]} (\lambda - Y)'\Sigma^{-1}(\lambda - Y).$$

Under H_0, by standard arguments

$$\mathsf{LR} \sim \inf_{\lambda \in \{0\} \times [-b,\infty]} Q(\lambda) - \inf_{\lambda \in [0,\infty] \times [-b,\infty]} Q(\lambda) \equiv \mathcal{L}(b, b), \tag{1.1}$$

for $Q(\lambda) = (\lambda - Z)'\Sigma^{-1}(\lambda - Z)$ with $Z \sim \mathrm{N}(0, \Sigma)$.

The difficulty with controlling the size of a test using the LR statistic in this setting comes down to the fact that b is unknown and therefore the null distribution $\mathcal{L}(b, b)$ in (1.1) is unknown. Since the null distribution is unknown, it is not obvious how to define a CV such that the test has size no greater than a nominal level $\alpha \in (0, 1)$. A naive approach which sets the CV to the $1 - \alpha$ quantile of (1.1) at some given value of b can lead to size distortions. For example, the standard CV is equal to the $1 - \alpha$ quantile of (1.1) for $b = \infty$, corresponding to a $\max\{\mathrm{N}(0, 1), 0\}^2$-distribution. However, a test using this CV will over-reject if the $1 - \alpha$ quantile of the distribution in (1.1) is larger for some value of $b < \infty$. To illustrate this, the figure below contains the 95-percentile of $\mathcal{L}(b, b)$ for the case where Σ is a correlation matrix with correlation parameter ρ.

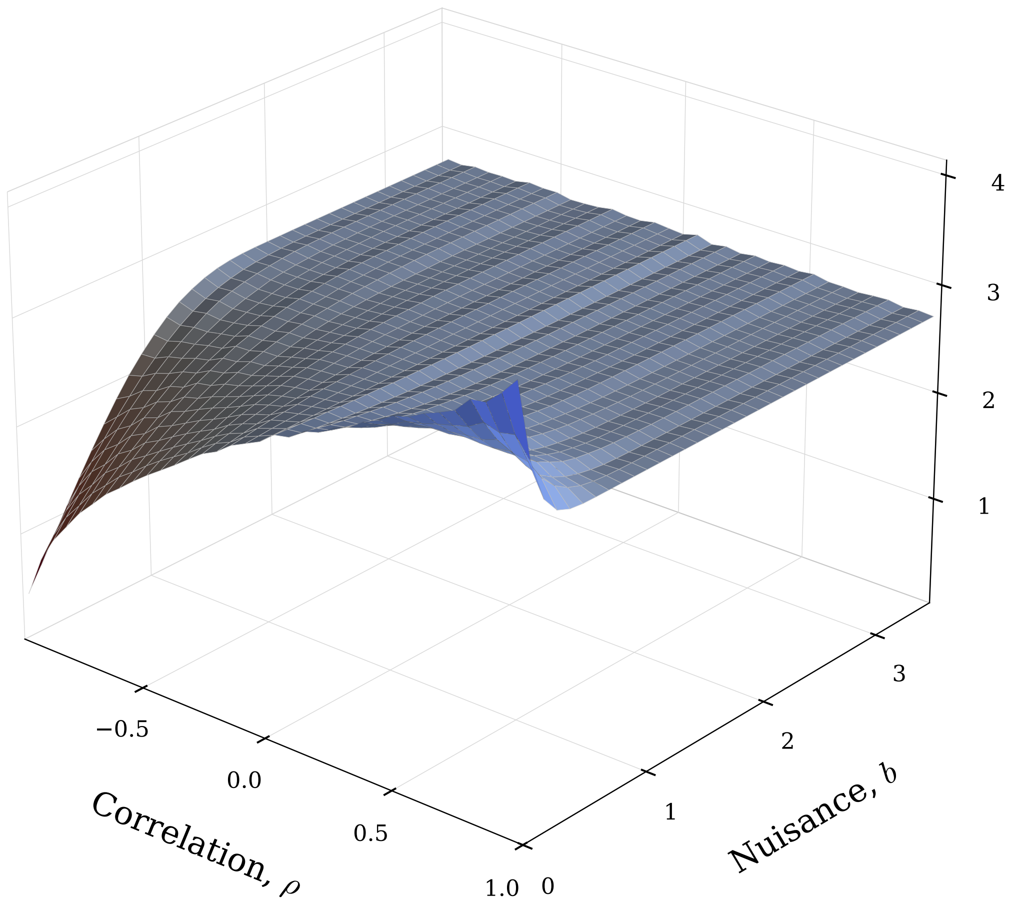

Figure: Simulated 95%-quantiles of $\mathcal{L}(b, b)$ as a function of $\rho \in (-1, 1)$ and $b \geq 0$.

A solution, referred to as the "least favorable approach" of Andrews and Guggenberger (2009), is based on using the largest $1 - \alpha$ quantile of (1.1) across all possible values of $b \in [0, \infty]$ and therefore has correct size. Although not a major concern in the present example for which b is one-dimensional, the least favorable approach becomes computationally intractable as the dimension of the nuisance parameter grows beyond two or so. Furthermore, this approach can be very conservative, leading to poor power when the true value of b is not close to the value used to compute the CV.

Our alternative CV is constructed as follows. Define

$$\mathcal{L}(x, y) = \inf_{\lambda \in \{0\} \times [-x, \infty]} Q(\lambda) - \inf_{\lambda \in [0, \infty] \times [-y, \infty]} Q(\lambda),$$

where $\mathcal{L}(x, y)$ is stochastically decreasing in x, while increasing in y. Furthermore, let

$\mathsf{CV}_q(x,y)$ denote the qth quantile of $\mathcal{L}(x,y)$ and note: (i) under H_0, $\mathsf{LR} \sim \mathcal{L}(b,b)$, and, (ii) for any values $b_L < b_U$ with $b \in (b_L, b_U)$, $\mathsf{CV}_q(b,b) \leq \mathsf{CV}_q(b_L, b_U)$. We use these observations to choose b_L and b_U in a judicious data-dependent manner to feasibly control the size of the LR test. Specifically, choose $\eta \in (0,\alpha)$ and let $z_{1-\eta/2}$ denote the $(1-\eta/2)^{th}$ quantile of $\mathrm{N}(0,1)$. With Σ_{22} denoting the second diagonal entry of Σ and

$$b_L = Y_b - \sqrt{\Sigma_{22}} z_{1-\eta/2}, \quad b_U = Y_b + \sqrt{\Sigma_{22}} z_{1-\eta/2}, \tag{1.2}$$

it holds that $\mathbb{P}(b \in (b_L, b_U)) = 1-\eta$. Thus, our proposal is to use $\mathsf{CV}_\alpha = \mathsf{CV}_{1-\alpha+\eta}(b_L, b_U)$ as a CV. To see how it controls size, note that for any $b \geq 0$ the probability of rejecting under H_0 can be bounded above using Bonferroni's inequality:

$$\mathbb{P}\left(\mathcal{L}(b,b) > \mathsf{CV}_\alpha\right) \leq \mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) > \mathsf{CV}_\alpha\right) + \mathbb{P}\left(\mathcal{L}(b,b) > \mathsf{CV}_{1-\alpha+\eta}(b,b)\right) \leq \alpha.$$

The last inequality follows by noting that

$$\mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) > \mathsf{CV}_\alpha\right) = 1 - \mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) \leq \mathsf{CV}_\alpha\right) \leq 1 - \mathbb{P}(b \in (b_L, b_U)) = \eta,$$

and, moreover, $\mathbb{P}(\mathcal{L}(b,b) > \mathsf{CV}_{1-\alpha+\eta}(b,b)) \leq \alpha - \eta$.

Note that since the above argument for size control when using CV_α relies on Bonferroni's inequality (which is not necessarily sharp) the null rejection probabilities of our proposed test can lie below the nominal level α, whereas this is not true for similar tests with exact size such as the CLR test of Ketz (2018). Indeed, this property is manifest in the simulation results of Section 4. Nevertheless, by helping to determine where the nuisance parameter is likely to lie within its parameter space via b_L and b_U when forming a CV, our method can lead to efficiency gains relative to, e.g., the least favorable approach.

As we demonstrate in Theorem 3.3, CV_α can be generalized to control size as a CV for LR statistics when (a) the parameter space for γ is not constrained, (b) the dimension of b exceeds one, and, (c) (the estimators) Y_γ and Y_b are not normally distributed in finite samples. For (a), the minimization space $[0,\infty]$ in the expressions for LR and $\mathcal{L}(b,b)$ are modified to properly reflect the parameter space for γ under H_1. For (b), in order to use the multivariate version of the monotonicity property (ii), b_L and b_U are replaced by multivariate counterparts derived from rectangular confidence sets for the mean of a multivariate normal distribution. For (c), the CVs described in Section 3 below do not require normally distributed estimators but rather estimators that are asymptotically Gaussian. We describe how to obtain such estimators in the presence of boundary constraints via a one-step Newton-Raphson iteration below.

2 General setup and framework

2.1 Setup

For a given set of observations $\{W_t\}_{t=1}^n$, consider the criterion function,

$$L_n(\theta) = f_n(\{W_t\}_{t=1}^n; \theta), \tag{2.1}$$

in terms of the parameter vector $\theta = (\gamma', \beta', \delta')' \in \Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$, where $\gamma \in \mathbb{R}^{d_\gamma}$ is the parameter (vector) of interest, while $\beta \in \mathbb{R}^{d_\beta}$ and $\delta \in \mathbb{R}^{d_\delta}$ are nuisance parameters. The set Θ is the *optimization* parameter space used to define the restricted and unrestricted estimators below. It is distinct from the *true* parameter space that we assume the true value of the parameter vector θ lies within. Formally (without loss of generality), $\Theta_\gamma = [\gamma_L, \gamma_U]^s \times [0, \gamma_U]^{d_\gamma - s}$ and $\Theta_\beta = [0, \beta_U]^{d_\beta}$, with $-\infty \leq \gamma_L < 0 < \gamma_U \leq \infty$ and $0 < \beta_U \leq \infty$, where $0 \leq s \leq d_\gamma$ and the first s components of γ are unrestricted in sign while the remaining $d_\gamma - s$ components are constrained to be nonnegative. The set $\Theta_\delta \subset (-\infty, \infty)^{d_\delta}$ is a compact subset with nonempty interior. Finally, as in Andrews and Cheng (2012), we introduce an additional parameter ϕ used here to capture any features of the distribution of the data $\{W_t\}$ not explicit from the formulation in (2.1). This way, $\psi = (\theta, \phi)$ completely determines the distribution of the data. The full model parameter space is $\Psi = \{\psi = (\theta, \phi) : \theta \in \Theta, \phi \in \Phi(\theta)\}$, where $\Phi(\theta) \subset \Phi$ and Φ is a compact metric space such that $\phi_n \to \phi \in \Phi$ implies weak convergence of the bivariate distribution of (W_t, W_{t+s}), for all $t, s \geq 1$.

We separately specify the maintained set of possible true parameters. Let $\Theta_\gamma^\star = [\underline{\gamma}, \bar{\gamma}]^s \times [0, \bar{\gamma}]^{d_\gamma - s}$, $\Theta_\beta^\star = [0, \bar{\beta}]^{d_\beta}$ and $\Theta_\delta^\star$ be a nonempty compact subset of $\mathrm{int}(\Theta_\delta)$, where $\underline{\gamma}$, $\bar{\gamma}$, and $\bar{\beta}$ satisfy $\gamma_L < \underline{\gamma} < 0 < \bar{\gamma} < \gamma_U \quad \text{and} \quad 0 < \bar{\beta} < \beta_U$. The true parameter space for θ is $\Theta^\star = \Theta_\gamma^\star \times \Theta_\beta^\star \times \Theta_\delta^\star$ and the true full model parameter space is $\Psi^\star = \{\psi = (\theta, \phi) \in \Psi :$

$\theta \in \Theta^\star\}$, which we assume is nonempty and compact. Specifically, we assume throughout that the true value of δ is an interior point of Θ_δ, while the true value of β is allowed to be in the interior, or at (near) the lower boundary of Θ_β. Furthermore, for γ we assume s of the components are in the "interior", while the remaining $d_\gamma - s$ are allowed to be on the (lower) "boundary", $0 \leq s \leq d_\gamma$. We are interested in testing the null hypothesis

$$\mathsf{H}_0 : \gamma = \gamma_0,$$

where $\gamma_{0,i} \in [\underline{\gamma}, \bar{\gamma}]$ for $i = 1, \ldots, s$ and $\gamma_{0,i} = 0$ for $i = s+1, \ldots, d_\gamma$ so that the true parameter space under the null is given by $\Psi_0 = \{\psi = (\theta, \phi) \in \Psi^\star : \gamma = \gamma_0\}$, which we assume is nonempty. It is a compact subset of $\Psi^\star$.

Remark 2.1 *The boundary of the nuisance parameter space Θ_β in our setup is given by vectors β that have at least one component equal to zero. However, our setup can often accommodate more complicated boundary constraints via suitable reparameterization. For example, if the boundary of the nuisance parameter space in the original problem is characterized by the set of points β for which at least one component of $f(\beta)$ is equal to zero for some smooth function f, if one can find a reparameterization such that $f(\beta)$ (i) becomes a parameter in the reparameterized model, (ii) admits a full-rank Jacobian for active constraints and (iii) entails a constrained set that maps locally to an orthant, then the reparameterized model would fit the setup outlined above. Moreover, if the function f is sufficiently smooth, standard chain rule and delta method arguments ensure that the assumptions we introduce below continue to hold for the reparameterized model when they hold for the original model.*

We are interested in testing the hypothesis H_0 using the (quasi-) LR statistic. That is, with the unrestricted $\hat{\theta}_n$ and restricted $\tilde{\theta}_n$ estimators as defined by

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta) \quad \text{and} \quad \tilde{\theta}_n = \arg\max_{\theta \in \Theta_{\mathsf{H}_0}} L_n(\theta), \tag{2.2}$$

where $\Theta_{\mathsf{H}_0} = \{\theta = (\gamma', \beta', \delta')' \in \Theta : \gamma = \gamma_0\}$, the statistic is given by,

$$\mathsf{LR}_n = 2(L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)). \tag{2.3}$$

To provide context and intuition, we include as running examples linear regression models with sign-restricted coefficients as well as a time series ARCH model with explanatory covariates.

Running example: Regression. With $W_t = (y_t, x_t')'$ consider the regression equation,

$$y_t = \theta' x_t + \varepsilon_t, \text{ for } t = 1, 2, \ldots, n, \tag{2.4}$$

where $x_t = (x_{1,t}, x_{\beta,t}')'$ and $\theta = (\gamma, \beta')' \in \Theta = \Theta_\gamma \times \Theta_\beta = [0, \infty]^{1+d_\beta}$. We are interested in testing $\mathsf{H}_0 : \gamma = 0$, which we emphasize is a non-standard testing problem as some entries of β may be zero-valued such that β is a boundary point of Θ_β.

Given a sample $\{W_t\}_{t=1}^n$, the least-squares objective function is given by

$$L_n(\theta) = -\tfrac{1}{2} \sum_{t=1}^{n} (y_t - x_t'\theta)^2 = -\tfrac{n}{2} \left(S_{yy} + \theta' S_{xx} \theta - 2\theta' S_{xy} \right),$$

with $S_{yy} = n^{-1} \sum_{t=1}^n y_t^2$, $S_{xx} = n^{-1} \sum_{t=1}^n x_t x_t'$, and $S_{xy} = n^{-1} \sum_{t=1}^n x_t y_t$. With the ordinary least-squares estimator given by $\hat{\theta}_{LS} = S_{xx}^{-1} S_{xy}$, it follows that the unrestricted estimator is given by

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} (\theta - \hat{\theta}_{LS})' S_{xx} (\theta - \hat{\theta}_{LS}),$$

see Lemma A.5 in Supplemental Appendix A, while the restricted estimator is given by

$$\tilde{\theta}_n = \arg\min_{\theta \in \Theta_{\mathsf{H}_0}} (\theta - \hat{\theta}_{LS})' S_{xx} (\theta - \hat{\theta}_{LS}),$$

where $\Theta_{\mathsf{H}_0} = \{\theta = (\gamma, \beta')' \in \Theta : \gamma = 0\}$, and the LR statistic is given by (2.3).

Finally, we note that, in this example, the additional parameter ϕ determines the joint distribution of x_t and ε_t. □

Running example: ARCH. For the second example, we consider hypothesis testing in ARCH models augmented with non-negative explanatory (or exogenous, X) covariates. Let $y_t \in \mathbb{R}$ be given by

$$y_t = \sigma_t \varepsilon_t, \quad t \in \mathbb{Z}, \tag{2.5}$$

$$\sigma_t^2 = \theta' F_{t-1} \tag{2.6}$$

where $F_t = (X_t', Y_t', 1)'$, with $X_t = (x_{1,t}, \ldots, x_{p,t})'$ and $Y_t = (y_t^2, \ldots y_{t-q+1}^2)'$. We are interested in testing whether the covariates are needed for the conditional variance σ_t^2. With θ partitioned as $\theta = (\gamma', \beta', \delta)'$, write (2.6) as a function of θ as

$$\sigma_t^2(\theta) = \theta' F_{t-1} = \delta + \beta' Y_{t-1} + \gamma' X_{t-1},$$

with $\theta \in \Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$, $\Theta_\gamma = [0, \gamma_U]^p$, $\Theta_\beta = [0, \beta_U]^q$, and $\Theta_\delta = [\delta_L, \delta_U]$, where $0 < \gamma_U, \beta_U < \infty$ and $0 < \delta_L < \delta_U < \infty$. As before, the objective is to test the hypothesis $\mathsf{H}_0 : \gamma = 0$, which is a non-standard testing problem as some of the β's can be on the boundary.

Given a sample $\{W_t\}_{t=1}^n = \{(y_t^2, F_{t-1}')\}_{t=1}^n$, the Gaussian quasi-log-likelihood function is (up to a constant) given by

$$L_n(\theta) = \sum_{t=1}^n l_t(\theta), \qquad l_t(\theta) = -\frac{1}{2}\left(\log \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right). \tag{2.7}$$

The estimator $\hat{\theta}_n$ ($\tilde{\theta}_n$) is any maximizer of $L_n(\theta)$ over Θ ($\Theta_{\mathsf{H}_0} = \{\theta \in \Theta : \gamma = 0\}$), and

$$\mathsf{LR}_n(\mathsf{H}_0) = 2\left[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)\right] \tag{2.8}$$

is the LR statistic.

Finally, in this example the additional parameter $\phi = \phi(\theta)$ determines the joint distribution of F_{t-1} and ε_t. □

2.2 Drifting sequences and assumptions

In order to establish uniform validity of our proposed CVs in the possible presence of nuisance parameters near or on the lower boundary, we study the limiting behavior of the LR statistic in (2.3) under drifting sequences $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n = (\theta_n, \phi_n) \to \psi_0 = (\theta_0, \phi_0)$. Specifically, the θ component of such a sequence is $\theta_n = (\gamma_0', \beta_n', \delta_n')'$, that is, a sequence along which H_0 holds and which satisfies $\theta_n \to \theta_0 = (\gamma_0', \beta_0', \delta_0')'$ as $n \to \infty$. In order to allow β to be either an interior point or near/at the boundary, in addition to $\beta_n = \left(\beta_{n,1}, \ldots, \beta_{n,d_\beta}\right)' \to \beta_0$ we require that $\sqrt{n}\beta_n \to b = (b_1, b_2, \ldots, b_{d_\beta})' \in [0, \infty]^{d_\beta}$, as $n \to \infty$. This includes sequences of true parameters converging to an interior point – e.g., $b_i = \infty$ – or to a boundary point at the standard $\sqrt{n}$ rate – e.g., $b_i \in [0, \infty)$. Moreover, parameters converging to zero at a rate slower (faster) than $\sqrt{n}$ corresponds to, e.g., $b_i = \infty$ ($b_i = 0$). Henceforth, the distribution of the (stationary) random vectors $\{W_t : t \geq 1\}$ is determined by the true parameter ψ; expectations, variances and probabilities computed under ψ are denoted as $\mathbb{E}_\psi$, $\mathbb{V}_\psi$ and $\mathbb{P}_\psi$, respectively.

In order to state the limiting distribution of LR_n, we make the following assumptions which are similar to Andrews (2001); see also Ketz (2018) and Fan and Shi (2023). These assumptions are stated for any drifting sequence $\{\psi_n\} \subset \Psi_0$ satisfying

$$\psi_n \to \psi_0 \quad \text{and} \quad \sqrt{n}\beta_n \to b = (b_1, b_2, \ldots, b_{d_\beta})' \in [0, \infty]^{d_\beta}. \tag{2.9}$$

Assumption 1 *$\hat{\theta}_n - \theta_n = o_p(1)$ and $\tilde{\theta}_n - \theta_n = o_p(1)$ as $n \to \infty$.*[1]

Assumption 2 *(i) $L_n(\theta)$ has continuous left/right partial derivatives of order two on Θ for all $n \geq 1$ (almost surely); (ii) for all deterministic $\epsilon_n \to 0$,*

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

Recall that Υ denotes the set of positive definite matrices with eigenvalues bounded below by κ and above by κ^{-1} for some $\kappa > 0$.

Assumption 3 *The Hessian satisfies*

$$n^{-1} \frac{\partial^2 L_n(\theta_n)}{\partial \theta \partial \theta'} \xrightarrow{p} -\Omega_0 \in \Upsilon.$$

Assumption 4 *The score satisfies*

$$n^{-1/2} \frac{\partial L_n(\theta_n)}{\partial \theta} \xrightarrow{d} N(0, \Sigma_0), \quad \text{with } \Sigma_0 \in \Upsilon$$

Remark 2.2 *Note that Assumptions 1–4 imply that $(\hat{\theta}_n - \theta_n)$ and $(\tilde{\theta}_n - \theta_n)$ are $O_p(n^{-1/2})$ under H_0 (by an extension of Andrews, 1999, Lemma 1 and Theorem 1).*

Running example: Regression. We consider here a setting where the data $\{W_t\}$ are assumed to be i.i.d.; see Remark 3.1 below for the extension to the more general case of time series data and triangular arrays. Specifically, we make throughout the following standard assumptions.

[1] As is common in the boundary-robust literature, our assumptions rule out settings allowing for weak identification, an interesting topic for potential future research.

Assumption LinIID 1 *For all* $\psi \in \Psi_0$*,*

1. $\{(x_t', \varepsilon_t)'\}_{t=1,2,\ldots}$ *are i.i.d., with* $\mathbb{E}_\psi[\varepsilon_t|x_t] = 0$ *almost surely for all* t*;*

2. $\mathbb{E}_\psi[x_t x_t'] \in \Upsilon$ *and* $\mathbb{E}_\psi[x_t x_t' \varepsilon_t^2] \in \Upsilon$*,*

3. $\mathbb{E}_\psi|x_{j,t}|^{2+\nu} \leq c$ *and* $\mathbb{E}_\psi|x_{j,t}\varepsilon_t|^{2+\nu} \leq c$ *for* $j = 1, \ldots, d_\beta + 1$ *and constants* $c < \infty$ *and* $\nu > 0$ *(not depending on* ψ*).*

Note that under Assumption 1 and any drifting sequence $\{\psi_n\} \subset \Psi_0$ satisfying (2.9), it holds that $\mathbb{E}_{\psi_n}[x_t x_t'] \to \mathbb{E}_{\psi_0}[x_t x_t'] = \Omega_0$ and $\mathbb{E}_{\psi_n}[x_t x_t' \varepsilon_t^2] \to \mathbb{E}_{\psi_0}[x_t x_t' \varepsilon_t^2] = \Sigma_0$. We have the following result that ensures that the high-level Assumptions 1–4 hold for linear regressions.

Proposition 2.1 *Assumption LinIID 1 implies that Assumptions 1–4 hold with* $\Omega_0 = \mathbb{E}_{\psi_0}[x_t x_t']$ *and* $\Sigma_0 = \mathbb{E}_{\psi_0}[x_t x_t' \varepsilon_t^2]$*.* □

Running example: ARCH. Recall that $\psi = (\theta, \phi)$ where ϕ denotes the joint distribution of F_{t-1} and ε_t. In particular, under the assumptions below, Ω_0 and Σ_0 in Assumptions 3–4 are determined entirely by ψ.

Assumption ARCH 1 *For any* $\psi \in \Psi_0$*, the process* $\{(y_t^2, F_{t-1}')'\}_{t\in\mathbb{Z}}$ *is stationary and* α*-mixing.*

The assumption is similar to that used in most of the literature on GARCH models, such as Francq and Thieu (2019) where the data-generating process (DGP) is assumed to be strictly stationary and ergodic. Here we impose the stronger assumption of strong mixing, to apply a weak law of large numbers (LLN) for triangular arrays.

To ensure identification, we make the following assumption.

Assumption ARCH 2 *For any non-zero constant vector* $v \in \mathbb{R}^{d_\beta + d_\gamma}$ *and constant* $c \in \mathbb{R}$*, it holds that*

$$\mathbb{P}_\psi((Y_t', X_t')v \neq c) > 0 \quad \textit{for all } \psi \in \Psi_0.$$

Assumption ARCH 3 *For all $\psi \in \Psi_0$ the innovation ε_t is independent of the σ-field*

$$\mathcal{F}_{t-1} = \sigma\{F_s : s \leq t-1\}. \tag{2.10}$$

Moreover, $\mathbb{E}_\psi[\varepsilon_t^2 - 1] = 0$, and there exist constants $\nu_1, c_1 \in (0,\infty)$ such that $\mathbb{E}_\psi[|\varepsilon_t|^{4(1+\nu_1)}] \leq c_1$ for all $\psi \in \Psi_0$. There exists a constant $\kappa \in (0,\infty)$ such that $\mathbb{E}_\psi[\varepsilon_t^4 - 1] = \kappa < \infty$ for all $\psi \in \Psi_0$.

Assumption ARCH 4 *There exist constants $\nu_2, c_2 \in (0,\infty)$ such that*

$$\mathbb{E}_\psi[\left(y_t^2 \|F_{t-1}\|^3\right)^{1+\nu_2}] \leq c_2 \quad \textit{for all } \psi \in \Psi_0.$$

We note that the space Θ is constrained such that $\{(y_t, F_{t-1}')'\}_{t\in\mathbb{Z}}$ is stationary for any $\theta_0 \in \Theta$. We have the following result.

Proposition 2.2 *Assumptions ARCH 1-4 imply that Assumptions 1-4 hold.* □

2.3 The asymptotic distribution of the LR statistic

The following lemma states the limiting distribution of the LR statistic when some of the nuisance parameters can be on or near the boundary. It extends existing results in the literature (e.g., Theorem 4 of Andrews, 2001) as we consider drifting sequences and hence nuisance parameters local to the boundary of the parameter space. To state the results, define the generic quadratic form

$$Q(\lambda) = \|\lambda - HZ\|^2_{(H\Omega_0^{-1}H')^{-1}}, \tag{2.11}$$

where $\lambda \in \mathbb{R}^{d_\gamma + d_\beta}$ and H is a $(d_\gamma + d_\beta) \times d_\theta$ selection matrix such that $H\theta = (\gamma', \beta')'$. Moreover, Z is $N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1})$ distributed, with the matrices Ω_0 and Σ_0 provided, respectively, by Assumptions 3 and 4.

Lemma 2.1 *Under Assumptions 1–4, and any sequence $\{\psi_n\} \subset \Psi_0$ satisfying (2.9),*

$$\mathsf{LR}_n \xrightarrow{d} \mathcal{L}_\infty(b,b)$$

with

$$\mathcal{L}_\infty(x,y) = \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(x)} Q(\lambda) - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(y)} Q(\lambda), \tag{2.12}$$

for $x, y \in [0, \infty]^{d_\beta}$, *where*

$$\Lambda_\gamma = [-\infty, \infty]^s \times [0, \infty]^{d_\gamma - s} \quad \textit{and} \quad \Lambda_\beta(b) = [-b_1, \infty] \times \cdots \times [-b_{d_\beta}, \infty].$$

The distribution of the limiting random variable $\mathcal{L}_\infty(b, b)$ approximates the finite-sample distribution of LR_n for $\beta_n = b/\sqrt{n}$. Notice that this distribution is unknown in practice, as it is unknown whether the nuisance parameters in β are interior points or on/near the boundary. Put differently, it is not feasible to consistently estimate the quantiles of $\mathcal{L}_\infty(b, b)$ for use as CVs in this context because b is not consistently estimable along $\{\psi_n\}$ sequences due to the $\sqrt{n}$ scaling of β_n.

Indeed, controlling the asymptotic size of a LR test using LR_n as the test statistic requires the use of a CV that asymptotically bounds the rejection probability under all parameter sequences $\psi_n \to \psi_0$ satisfying (2.9) as $n \to \infty$ by the nominal level of the test $\alpha \in (0, 1)$ (see, e.g., Andrews and Guggenberger, 2009, or McCloskey, 2017). This motivates us to examine an alternative method of CV construction that can feasibly control the asymptotic size of the test. This we do next.

3 Feasible uniform CV construction

In this section we detail how to construct uniformly valid CVs for the LR statistic. As we argue below, these CVs are simple to construct, allow asymptotic size control regardless of whether nuisance parameters are on the boundary or not, and provide power gains with respect to existing methods.

As outlined in Section 1 and detailed below, because $\mathcal{L}_\infty(b, b)$ is given by (2.12) with $x = y = b$, we exploit properties of $\mathcal{L}_\infty(x, y)$ in order to select x and y in a data-driven (and simple) way such that the asymptotic size is bounded above by the user-chosen nominal level α.

We proceed as follows. In Section 3.1 we present a key monotonicity property of $\mathcal{L}_\infty(x, y)$ and discuss initially a naive (inefficient) method to construct valid CVs. In Section 3.2 we introduce some preliminaries needed for constructing our proposed CV. Our main Algorithm 1 to construct the CVs is given in Section 3.3, where we also prove the uniform validity of our procedure.

3.1 Monotonicity of $\mathcal{L}_\infty(x, y)$ and some naive CVs

A key property of $\mathcal{L}_\infty(x, y)$, which we exploit throughout, is given in the following lemma.

Lemma 3.1 *Let $\underline{b}, b, \bar{b} \in [0, \infty]^{d_\beta}$ satisfy (element-wise) $\underline{b} \leq b \leq \bar{b}$, then*

$$\mathcal{L}_\infty(b, b) \leq \mathcal{L}_\infty(\underline{b}, \bar{b}) \ a.s.$$

Given the monotonicity property of Lemma 3.1, a naive and straightforward way of constructing a uniformly valid test is to use a CV based on the distribution of $\mathcal{L}_\infty(0^{d_\beta}, \infty^{d_\beta})$, which satisfies $\mathcal{L}_\infty(b, b) \leq \mathcal{L}_\infty(0^{d_\beta}, \infty^{d_\beta})$. However, such a choice of CVs leads to a very conservative test. Alternatively, one may use the $1 - \alpha$ quantile of $\mathcal{L}_\infty(b^*, b^*)$, where b^* is the maximizer across $[0, \infty]^{d_\beta}$ of all $1 - \alpha$ quantiles of $\mathcal{L}_\infty(b, b)$, in Andrews and Guggenberger (2009). However, also this choice of CV can lead to a conservative test with lower power. Perhaps more importantly, b^* is burdensome to compute when the number of nuisance parameters $d_\beta > 2$. To reduce the conservative nature of these CVs, McCloskey (2017) suggests to use the $1 - \alpha + \eta$ quantile of $\mathcal{L}_\infty(\tilde{b}, \tilde{b})$ for some $\eta \in (0, \alpha)$, where $\tilde{b}$ is the maximizer across a $(1 - \eta)$-level confidence set for b. Unfortunately, this proposal suffers a similar computational drawback to that of Andrews and Guggenberger (2009) when $d_\beta > 2$.

We finally note that the shrinkage-based bootstrap of Cavaliere, Nielsen, Pedersen and Rahbek (2022) essentially seeks to choose between $\mathcal{L}_\infty(0, 0)$ or $\mathcal{L}_\infty(\infty, \infty)$ (component-wise) in a data-driven way. However, it fails to control size uniformly since there may exist values $b \in (0, \infty)$ such that $\mathcal{L}_\infty(b, b) \geq \mathcal{L}_\infty(0, 0)$.

3.2 Prerequisites and additional assumptions

In order to construct computationally-feasible and uniformly-valid CVs, we require two main ingredients, which are very simple to satisfy in applications. First, we need consistent estimators of the covariance matrices Ω_0 and Σ_0, see Assumptions 3–4. Second, we need the construction of a consistent and asymptotically Gaussian estimator $\check{\beta}_n$ of β_n with asymptotic covariance matrix given by $\Sigma_\beta = H_\beta \Omega_0^{-1} \Sigma_0 \Omega_0^{-1} H_\beta'$ where Ω_0 is defined in Assumption 3 and H_β is the selection matrix satisfying $H_\beta \theta = \beta$. Moreover, we require a consistent estimator $\check{\Sigma}_{\beta,n}$ of the covariance matrix Σ_β. We formalize these requirements through the following assumptions, which hold for any sequence $\{\psi_n\} \subset \Psi_0$ satisfying (2.9).

Assumption 5 *There exists a matrix $\hat{\Sigma}_n$ such that $\hat{\Sigma}_n \xrightarrow{p} \Sigma_0$.*

Assumption 6 *There exist estimators $\check{\beta}_n$ and $\check{\Sigma}_{\beta,n}$ which satisfy (i) $\sqrt{n}(\check{\beta}_n - \beta_n) \xrightarrow{d} N(0, \Sigma_\beta)$, with $\Sigma_\beta = H_\beta \Omega_0^{-1} \Sigma_0 \Omega_0^{-1} H_\beta'$, and (ii) $\check{\Sigma}_{\beta,n} \xrightarrow{p} \Sigma_\beta$.*

Assumption 7 *Fix $\alpha \in (0,1)$ and $\eta \in (0, \alpha)$. Under every sequence $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n \to \psi_0$ and (2.9), with corresponding limits b, Ω_0 and Σ_0, the limiting LR statistic satisfies*

$$\mathbb{P}\left(\mathcal{L}_\infty(b,b) = 0\right) < 1 - \alpha + \eta.$$

Assumption 7 is an upper-quantile regularity condition. Lemma A.4 in Supplemental Appendix A shows that the only possible atom of $\mathcal{L}_\infty(b,b)$ is at zero and that its distribution function is continuous and strictly increasing above zero. Assumption 7 therefore places the $1-\alpha+\eta$ quantile used by Algorithm 1 in the continuous part of the distribution and yields precisely the quantile regularity needed below. The condition is a property of the limiting Gaussian experiment and must be verified over the relevant collection of limiting experiments in a given application. Assumption 7 holds whenever the parameter of interest γ is not sign-restricted ($s = d_\gamma$), since then $\mathcal{L}_\infty(b,b)$ has no atom at zero. When some components of γ are sign-restricted the condition has content. In the leading scalar case of Section 1, the atom probability admits a closed form, is monotone in b with the sign of the correlation between Y_γ and Y_b, and is bounded above by 3/4 uniformly in $b \geq 0$ and in the correlation; see Proposition A.1 in Supplemental Appendix A. Consequently, in this model Assumption 7 holds for *every* positive definite Σ and *every* $b \in [0,\infty]$ as soon as $\alpha - \eta \leq 1/4$, and therefore at conventional significance levels. We note that a similar condition arises in the standard case with no nuisance parameter, where the limiting distribution is given by $\mathcal{L}_\infty(\infty,\infty) = (\max\{N(0,1),0\})^2$. This random variable is zero with probability 1/2, and, consequently, one must choose a level $\alpha < 1/2$ in order to have a strictly positive critical value.

As a simple example of an estimator satisfying Assumption 6 in general, consider the one-step iterated Newton-Raphson estimator for θ:

$$\check{\theta}_n = \hat{\theta}_n - \left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial L_n(\hat{\theta}_n)}{\partial\theta}. \tag{3.1}$$

As observed by Ketz (2018), by definition, $\check{\theta}_n$ may not belong to the parameter space Θ, and hence $L_n(\check{\theta}_n)$ may not even be well-defined. We have the following lemma due to Ketz (2018).

Lemma 3.2 *Under any sequence* $\{\psi_n\} \subset \Psi_0$ *satisfying (2.9) and Assumptions 1–4, with* $\check{\theta}_n$ *given by (3.1),*

$$\sqrt{n}(\check{\theta}_n - \theta_n) \xrightarrow{d} N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1}).$$

In particular, we have that $\check{\beta}_n = H_\beta\check{\theta}_n$ *is asymptotically normal,*

$$\sqrt{n}(\check{\beta}_n - \beta_n) \xrightarrow{d} N(0, \Sigma_\beta), \tag{3.2}$$

where $\Sigma_\beta = H_\beta\Omega_0^{-1}\Sigma_0\Omega_0^{-1}H_\beta'$. *If, in addition, Assumption 5 holds, then*

$$\check{\Sigma}_{\beta,n} = H_\beta \left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} \hat{\Sigma}_n \left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} H_\beta' \xrightarrow{p} \Sigma_\beta. \tag{3.3}$$

Hence, for the one-step iterated Newton-Raphson estimator $\check{\theta}_n$ of (3.1) it holds that Assumptions 1–5 imply Assumption 6. We illustrate this in terms of the two running examples.

Running example: Regression. To verify Assumptions 5 and 6 for the linear regression example, we make the following additional assumptions.

Assumption LinIID 2 *For all* $\psi \in \Psi_0$ *and some constants* $\nu, c > 0$, $\mathbb{E}_\psi[|x_{i,t}x_{j,t}x_{k,t}x_{l,t}|^{1+\nu}] \leq c$ *for* $i,j,k,l = 1,\dots,d_\beta + 1$.

The following Proposition states that Assumptions 5–6 hold for the linear regression example when choosing $\hat{\Sigma}_n$ as the Eicker-White heteroskedasticity-robust estimator[2],

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 x_t x_t', \quad \hat{\varepsilon}_t = y_t - x_t'\hat{\theta}_{LS}. \tag{3.4}$$

[2]Note that one could alternatively define the estimator in terms of the constrained residuals, that is, $\hat{\varepsilon}_t = y_t - x_t'\hat{\theta}_n$.

Proposition 3.1 *Suppose* $\check{\beta}_n = \hat{\beta}_{LS}$ *and* $\check{\Sigma}_{\beta,n}$ *is equal to the upper* $d_\beta \times d_\beta$ *block of* $S_{xx}^{-1}\hat{\Sigma}_n S_{xx}^{-1}$ *with* $\hat{\Sigma}_n$ *given by (3.4). Then Assumptions LinIID 1–2 imply Assumptions 5–6.*

Remark 3.1 *The linear regression example can easily be extended to time series data and triangular arrays. In particular, under appropriate strong mixing conditions on* $\{W_t\}_{t\in\mathbb{Z}}$ *and replacing* $\hat{\Sigma}_n$ *with a HAC estimator (Newey and West, 1987), it is possible to verify Assumptions 1–5.* □

Running example: ARCH. Under Assumptions ARCH 1-4 it holds that $\Sigma_0 = (\kappa/2)\Omega_0$; hence, we can estimate Σ_0 using $\hat{\Sigma}_n = (\hat{\kappa}_n/2)\hat{\Omega}_n$ with

$$\hat{\Omega}_n = -n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}, \qquad \hat{\kappa}_n = n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^4 - 1)$$

where $\hat{\varepsilon}_t = y_t/\hat{\sigma}_t(\hat{\theta}_n)$. Moreover, let $\check{\beta}_n = H_\beta\check{\theta}_n$ with $\check{\theta}_n$ given by (3.1), and finally let

$$\check{\Sigma}_{\beta,n} = (\hat{\kappa}_n/2)H_\beta\hat{\Omega}_n^{-1}H_\beta'. \tag{3.5}$$

Proposition 3.2 *With* $\check{\beta}_n$ *and* $\check{\Sigma}_{\beta,n}$ *as defined above, Assumptions ARCH 1-4 imply Assumptions 5–6.* □

3.3 Uniformly valid CV

We can now finally state our main algorithm to construct uniformly valid CVs, which are denoted as $\mathsf{CV}_{\alpha,n}$, where $\alpha \in (0,1)$ denotes the nominal significance level. As anticipated, it relies on the asymptotically normal estimator $\check{\beta}_n$ and the consistent covariance matrix estimator $\check{\Sigma}_{\beta,n}$ in Assumption 6.

Algorithm 1

With $\hat{\theta}_n$, $\hat{\Omega}_n = -n^{-1}\partial^2 L_n(\hat{\theta}_n)/\partial\theta\partial\theta'$, $\hat{\Sigma}_n$, $\check{\beta}_n$, $\check{\Sigma}_{\beta,n}$ *and the nominal significance level* $\alpha \in (0,1)$ *as inputs:*

1. *Choose some* $\eta \in (0,\alpha)$*;*
2. *Compute* $\check{\Omega}_{\beta,n} = \mathrm{diag}(\check{\Sigma}_{\beta,n})^{-1/2}\check{\Sigma}_{\beta,n}\,\mathrm{diag}(\check{\Sigma}_{\beta,n})^{-1/2}$*;*

3. *Compute* $\check{q}_{1-\eta,n}$ *as the* $(1-\eta)$*-quantile of* $\max_{i=1,\ldots,d_\beta} |\check{Z}_{\beta,i}|$ *with* $\check{Z}_\beta \overset{d}{=} N(0, \check{\Omega}_{\beta,n})$*;*

4. *Compute* $b_{L,n} = (b_{L,n,1}, \ldots, b_{L,n,d_\beta})'$ *and* $b_{U,n} = (b_{U,n,1}, \ldots, b_{U,n,d_\beta})'$ *by setting* $b_{L,n,i} = \max\{0, \check{b}_{L,n,i}\}$ *and* $b_{U,n,i} = \max\{0, \check{b}_{U,n,i}\}$ *for* $i = 1, \ldots, d_\beta$*, where*

$$\check{b}_{L,n} = \sqrt{n}\check{\beta}_n - \check{q}_{1-\eta,n}\,\mathrm{diagv}(\check{\Sigma}_{\beta,n})^{1/2},\ \check{b}_{U,n} = \sqrt{n}\check{\beta}_n + \check{q}_{1-\eta,n}\,\mathrm{diagv}(\check{\Sigma}_{\beta,n})^{1/2};$$

5. *Compute* $\mathsf{CV}_{\alpha,n}$ *as the* $1-\alpha+\eta$ *quantile of* $\mathcal{L}_{\infty,n}(b_{L,n}, b_{U,n})$*, where* $\mathcal{L}_{\infty,n}(\cdot,\cdot)$ *is defined as* $\mathcal{L}_\infty(\cdot,\cdot)$ *in (2.12) with* Ω_0 *and* Σ_0 *replaced by* $\hat{\Omega}_n$ *and* $\hat{\Sigma}_n$*, respectively.*

Notice that the quantiles in Steps 3 and 5 of Algorithm 1 can be computed straightforwardly by simulation. Notice also that the elements of $b_{L,n}$ and $b_{U,n}$ are maximized against zero in order to ensure that $b_{L,n}$ and $b_{U,n}$ lie in $[0,\infty)^{d_\beta}$. In contrast to existing CV construction approaches that require one to compute the maximal quantile of $\mathcal{L}_{\infty,n}(b,b)$ either over the entire parameter space $b \in [0,\infty]^{d_\beta}$ (e.g., the least favorable approach of Andrews and Guggenberger, 2009) or a confidence set inside of it (e.g., McCloskey, 2017), our approach remains computationally tractable even when d_β is not small because it only requires computing a quantile of $\mathcal{L}_{\infty,n}(\cdot,\cdot)$ at the single point $(b_{L,n}, b_{U,n})$. See Step 5 of Algorithm 1. Like the CLR test of Ketz (2018), this quantile computation at a single point simply involves repeatedly simulating the value of two quadratic programming problems with linear constraints because each component of (2.12) is equal to the value of a quadratic programming problem with linear constraints.

We can now state our main theorem, where we establish the uniform asymptotic size control of the LR test based on our proposed CV, $\mathsf{CV}_{\alpha,n}$.

Theorem 3.3 *Fix* $\alpha \in (0,1)$ *and* $\eta \in (0,\alpha)$*, and let* $\mathsf{CV}_{\alpha,n}$ *be constructed as in Algorithm 1. Then under Assumptions 1–7,*

$$\limsup_{n\to\infty} \sup_{\psi\in\Psi_0} \mathbb{P}_\psi\left(\mathsf{LR}_n > \mathsf{CV}_{\alpha,n}\right) \leq \alpha.$$

We emphasize that the underlying assumptions of Theorem 3.3 are intuitive and typically possible to verify for a given statistical model. Propositions 3.1 and 3.2 verify Assumptions 1–6 for the linear regression and ARCH models, respectively. Assumption

7 is a separate property of the resulting limiting Gaussian experiment and provides the upper-quantile regularity required for the strict-rejection test.

Remark 3.2 *Theorem 3.3 states that Algorithm 1 provides asymptotically valid CVs for any choice of the user-chosen parameter* $\eta \in (0, \alpha)$. *We note that* $1 - \eta$ *(approximately) quantifies the probability of the event that* b *lies inside of the confidence hyper-rectangle with lower corner* $b_{L,n}$ *and upper corner* $b_{U,n}$ *so that a larger value of* η *corresponds to a (pointwise) larger value of* $b_{L,n}$ *and smaller value of* $b_{U,n}$. *Given the monotonicity of* $\mathcal{L}_\infty(x, y)$ *described in Lemma 3.1, this implies that a larger value of* η *leads to* $\mathsf{CV}_{\alpha,n}$ *being constructed from a stochastically smaller random variable,* $\mathcal{L}_{\infty,n}(b_{L,n}, b_{U,n})$. *However, since* $\mathsf{CV}_{\alpha,n}$ *is equal to the* $1 - \alpha + \eta$ *quantile of this random variable, a larger value of* η *corresponds to a larger quantile, entailing an opposing effect on the value of* $\mathsf{CV}_{\alpha,n}$. *In practice, we recommend a small value of* $\eta = \alpha/10$ *because it performs well across a wide range of simulation designs (see Section 4.3) and yields a value of* $\mathsf{CV}_{\alpha,n}$ *that is close to the standard CV that ignores the boundary, the* $1 - \alpha$ *quantile of* $\mathcal{L}_{\infty,n}(\infty, \infty)$, *when the nuisance parameter* β *is far from its boundary. The latter fact can be seen by noting that when the components of* β *are far from zero, Step 4 of Algorithm 1 produces very large values of both* $b_{L,n}$ *and* $b_{U,n}$ *so that* $\mathsf{CV}_{\alpha,n}$ *is approximately equal to the* $1-\alpha+\eta$ *quantile of* $\mathcal{L}_{\infty,n}(\infty, \infty)$. *This recommendation is also consistent with the existing related literature on CVs constructed using Bonferroni bounds (e.g., McCloskey, 2017).*

4 Simulation experiments

In this section, we begin by examining the performance of our proposed uniform LR test in the context of the ARCH model, for which our test is the only test we are aware of with proven uniform asymptotic size control. After this, we examine its performance in several Gaussian shift experiments.

4.1 ARCH with one-dimensional parameter of interest

We consider testing for the presence of an explanatory covariate when another covariate may be present. Specifically, consider the model

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, \ldots, n,$$
$$\sigma_t^2 = \delta_1 + \delta_2 y_{t-1}^2 + \gamma x_{t-1,1} + \beta x_{t-1,2},$$

where $\delta_1 > 0$, $\delta_2, \gamma, \beta \geq 0$, and $\{\varepsilon_t\}_{t=1}^n$ is an i.i.d. process with $\varepsilon_t \sim N(0,1)$. We consider testing $\mathsf{H}_0 : \gamma = 0$ against $\mathsf{H}_1 : \gamma > 0$. We set $\delta_1 = 0.1$ and $\delta_2 = 0.5$ and assume that the researcher knows that the true values of the ARCH coefficients δ_1 and δ_2 are not near their boundaries of zero, so that the only nuisance parameter that potentially causes a discontinuity in the null distribution is β. In terms of the covariates, we follow the simulation design in Nielsen, Pedersen, Rahbek and Thorsen (2025) and let $(x_{t,1}, x_{t,2}) = (\exp(v_{t,1}), \exp(v_{t,2}))$ with $V_t = (v_{1,t}, v_{2,t})'$ a bivariate autoregression with correlated innovations satisfying $V_t = aV_{t-1} + \epsilon_t$, $t = 1, \ldots, n$. Here $\{\epsilon_t\}_{t=1}^n$ is an i.i.d. $N_2(0, \Sigma)$ process with

$$\Sigma = c\left(1 - a^2\right)\begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$$

independent of $\{\varepsilon_t\}_{t=1}^n$, $\rho_{12} \in (-1, 1)$, $a = 0.9$ and $c = 0.5$. The simulated realizations of the processes for y_t and $(x_{t,1}, x_{t,2})$ use a burn-in period of 1000 observations and the sample size is $n = 1000$. The results are reported for 10,000 Monte Carlo replications and $\eta = \alpha/10$.

We compare rejection frequencies of our uniform LR test (labeled Unif) with the standard LR test (Std), the CLR test, and the least favorable (LF) approach.[3] We report rejection frequencies for parameter values $\gamma \in \{0, 0.01, 0.05, 0.1, 0.25\}$, corresponding to H_0 and increasingly distant alternatives in H_1, $\beta \in \{0, 0.1, 0.25, 0.5\}$, corresponding to a nuisance parameter at and increasingly away from the boundary, and various correlations $\rho_{12} \in \{-0.95, -0.75, -0.5, 0, 0.5, 0.75, 0.95\}$. These DGPs imply an asymptotic correlation ρ between any estimators of γ and β satisfying Assumption 6. To see this, let H

[3] We describe the construction and logic behind the CLR test in detail in Supplemental Appendix B.

denote the selection matrix such that $H\theta = (\gamma, \beta)'$. Then $\Omega_0^{-1}|_H = H\Omega_0^{-1}H'$ is the 2×2 sub-block of the inverse information matrix corresponding to γ and β, and

$$\Sigma^\dagger = \mathrm{Diag}\big(\Omega_0^{-1}\big|_H\big)^{-1/2} \Omega_0^{-1}\big|_H \ \mathrm{Diag}\big(\Omega_0^{-1}\big|_H\big)^{-1/2} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

is the implied population correlation matrix. This defines a mapping from the covariate correlation ρ_{12} to the correlation between Y_γ and Y_β in the limiting experiment[4]:

ρ_{12}	-0.95	-0.75	-0.50	0.00	0.50	0.75	0.95
ρ	0.58	0.48	0.34	0.00	-0.43	-0.69	-0.93

Tables 1 and 2 report the results for a nominal level of $\alpha = 0.05$ with Table 1 corresponding to β being at and near its boundary and Table 2 corresponding to β being further from its boundary. We note that our uniform LR test has rejection frequencies that are well-controlled at the nominal level across all parameter values under H_0. The standard LR test overrejects for $\beta = 0$ and negative values of ρ_{12} under H_0 (e.g., up to 7.8% for $\rho_{12} = -0.95$). The CLR test slightly overrejects at $\beta = 0$ for positive ρ_{12} under H_0 but appears to have roughly correct size in this setting. The least favorable LR test has rejection frequencies below or close to α for all configurations. We note that, in this design, the nuisance parameter β is scalar and non-negative. When ρ is non-negative, this is consistent with the conditions of Corollary 2 in Ketz (2018), which guarantee the asymptotic similarity of the CLR test when performing one-sided inference in the presence of a scalar parameter of interest. Our simulation results further support the asymptotic similarity of the CLR test in this ARCH context since the CLR test has null rejection frequencies close to α for most configurations with $\beta > 0$, with the overrejection at $\beta = 0$ (or small) and positive ρ_{12} likely being a finite-sample phenomenon.

[Tables 1 and 2 around here]

The last row of Table 1 reports average computation times per replication, obtained using parallelization across 60 cores on a machine with dual AMD EPYC 9635 processors (2.25 GHz, 768 GB RAM). The uniform LR test has computation times comparable to

[4]The correlations ρ are computed numerically.

those of the CLR test, while the LF approach is roughly 20–30 times slower due to its grid search over nuisance parameter values. This computational difference substantially widens as the dimension of the nuisance parameter grows, making the LF approach computationally infeasible for higher-dimensional nuisance parameters.

In terms of power, for $\gamma \in \{0.01, 0.05, 0.1, 0.25\}$, our uniform LR test shows notable power advantages over the CLR test for negative values of ρ_{12}, which correspond to nuisance parameter estimators that are positively correlated with the estimator of the parameter of interest. For instance, with $\beta = 0$ and $\rho_{12} = -0.95$, the uniform LR test achieves a rejection frequency of 96.2% at $\gamma = 0.05$, compared to 80.3% for the CLR test. This is also true of the least favorable LR test, which can display even higher power in these scenarios. But since this test becomes very computationally cumbersome/infeasible when the nuisance parameter grows in dimension, it is difficult to recommend in practice beyond the one-dimensional case. For positive ρ_{12}, the CLR test is typically more powerful. Moreover, the power of the uniform LR and CLR tests is quite close as β moves away from its boundary, across all values of ρ_{12}. We further explore the power properties of the uniform LR test in relation to the CLR test under various simulation designs in the next two sections to investigate whether the power advantage of the uniform LR test under positive correlation when the nuisance parameter is at the boundary appears to generalize.

4.2 ARCH with two-dimensional parameter of interest

We now extend the ARCH simulation experiment of Section 4.1 to a setting where γ is two-dimensional. Now consider the model

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, \ldots, n,$$
$$\sigma_t^2 = \delta_1 + \delta_2 y_{t-1}^2 + \gamma_1 x_{1,t-1} + \gamma_2 x_{2,t-1} + \beta x_{3,t-1},$$

where $\delta_1 > 0$, $\delta_2, \gamma_1, \gamma_2, \beta \geq 0$, and $\{\varepsilon_t\}_{t=1}^n$ is an i.i.d. process with $\varepsilon_t \sim N(0,1)$. We consider testing $\mathsf{H}_0 : \gamma = (\gamma_1, \gamma_2)' = 0$ against $\mathsf{H}_1 : \max\{\gamma_1, \gamma_2\} > 0$. We again set $\delta_1 = 0.1$ and $\delta_2 = 0.5$ and assume the researcher knows that δ_1 and δ_2 are away from their boundaries so that β remains the only nuisance parameter potentially causing a discontinuity in the null distribution. Similar to the setting in Nielsen et al. (2025,

Appendix B), the covariates $x_{1,t}$, $x_{2,t}$, $x_{3,t}$ are non-negative and generated via a Gaussian copula with Gamma marginals. Specifically, let $\{z_t\}$ be an i.i.d. process, independent of $\{\varepsilon_t\}$, with $z_t \sim N(0, C)$, where C is a 3×3 correlation matrix, and define $u_{i,t} = F_N(z_{i,t})$, where $F_N(\cdot)$ is the standard normal CDF. With $F_{G(a,b)}$ denoting the CDF of the Gamma distribution with shape parameter a and scale parameter b, the covariates are

$$x_{i,t} = \frac{F^{-1}_{G(a_i,b)}(u_{i,t})}{a_i b}, \quad i = 1, 2, 3,$$

where the normalization ensures $\mathbb{E}[x_{i,t}] = 1$. The Gamma shape parameters are $a_1 = 3$, $a_2 = 2$, $a_3 = 10$, and the common scale parameter is $b = 10$. The correlation matrix C is parametrized using the Generalized Fisher Transformation (GFT) of Archakov and Hansen (2021): for an unrestricted vector $g \in \mathbb{R}^3$, the GFT inverse mapping produces a valid correlation matrix $C(g)$. We consider here three different copula correlation structures, determined by the vectors $g_1 = (0, 0, 0)'$, $g_2 = (0, -1, 0)'$, and $g_3 = (-0.8, -0.65, -0.3)'$. Let H denote the selection matrix such that $H\theta = (\gamma_1, \gamma_2, \beta)'$. Then $\Omega_0^{-1}|_H = H\Omega_0^{-1}H'$ is the 3×3 sub-block of the inverse information matrix corresponding to γ and β, and

$$\Sigma^\dagger = \mathrm{Diag}\big(\Omega_0^{-1}\big|_H\big)^{-1/2}\,\Omega_0^{-1}\big|_H\,\mathrm{Diag}\big(\Omega_0^{-1}\big|_H\big)^{-1/2}$$

is the implied population correlation matrix, which we evaluate numerically. The three configurations yield the following copula and population correlation matrices:

$$g_1: \quad C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_1^\dagger = \begin{pmatrix} 1 & 0.011 & -0.002 \\ 0.011 & 1 & -0.020 \\ -0.002 & -0.020 & 1 \end{pmatrix},$$

$$g_2: \quad C_2 = \begin{pmatrix} 1 & 0 & -0.762 \\ 0 & 1 & 0 \\ -0.762 & 0 & 1 \end{pmatrix}, \quad \Sigma_2^\dagger = \begin{pmatrix} 1 & 0.002 & 0.708 \\ 0.002 & 1 & -0.007 \\ 0.708 & -0.007 & 1 \end{pmatrix},$$

$$g_3: \quad C_3 = \begin{pmatrix} 1 & -0.584 & -0.466 \\ -0.584 & 1 & -0.070 \\ -0.466 & -0.070 & 1 \end{pmatrix}, \quad \Sigma_3^\dagger = \begin{pmatrix} 1 & 0.594 & 0.540 \\ 0.594 & 1 & 0.371 \\ 0.540 & 0.371 & 1 \end{pmatrix}.$$

We compare the rejection frequencies of our uniform LR test (Unif) with the standard LR test (Std) and the CLR test. The standard LR test uses the population correlation $\Sigma^{\dagger}_{\gamma\gamma}$ (the 2×2 sub-block of $\Sigma^{\dagger}$ corresponding to γ), computed once from a large pilot sample, to determine the chi-bar-squared weights. We now report rejection frequencies for parameter values $\gamma_1, \gamma_2 \in \{0, 0.05, 0.1, 0.25\}$ and $\beta \in \{0, 0.1, 0.25, 0.5\}$. As before, we set $n = 1000$, use a burn-in period of 1000 and 10,000 Monte Carlo replications, and set $\eta = \alpha/10$.

Table 3 reports the results for a nominal level of $\alpha = 0.05$. Under the null ($\gamma_1 = \gamma_2 = 0$), the standard LR test over-rejects when $\beta = 0$ and the correlation between γ and β estimates is substantial (configurations C_2, C_3), confirming that ignoring the nuisance parameter on the boundary distorts the size. The uniform LR and CLR tests control size well across all configurations. Under the alternatives, our uniform LR test shows substantial power gains relative to the CLR test when there is positive correlation between the parameter of interest and the nuisance parameters, corresponding to C_2 and C_3, and β is at the boundary of its parameter space ($\beta = 0$). For example, under C_2 with $\beta = 0$, the uniform LR test has a rejection frequency of 78% while that of the CLR test is 58.5%. When there is very little correlation between the parameter of interest and the nuisance parameters, corresponding to C_1, the CLR test is more powerful. These relative power advantages are broadly in line with the results for a one-dimensional parameter of interest in the previous section: when there is positive correlation between the estimates of the parameters of interest and the nuisance parameters, the uniform LR test is more powerful at and, by continuity, very close to the boundary of the nuisance parameter space. We also note that as the nuisance parameter moves away from its boundary, the uniform LR test's power advantage disappears. However, the relative power of the CLR and uniform LR test as the nuisance parameter moves away from the boundary is complicated in the ARCH setting by the fact that as β grows larger, the variance of the parameter estimates grows quickly, resulting in power declines for all tests that can be seen throughout Tables 1–3. In order to better isolate the effect of moving the nuisance parameter away from its boundary while fixing the variances of the parameter estimates, we next turn to analyzing the Gaussian shift experiment under various correlation structures.

[Table 3 around here]

4.3 Gaussian shift with two-dimensional parameter of interest

Not only does examining the Gaussian shift experiment allow us to fix the variance of the parameter estimates but it also allows us to make more general (asymptotic) comparisons since it pertains to the asymptotic representation of common regular estimators across a wide variety of settings. Consider the multivariate location model

$$y_t = \theta + \varepsilon_t, \quad t = 1, \ldots, n, \tag{4.1}$$

where $\theta = (\gamma', \beta)' \in [0,\infty)^2 \times [0,\infty)$ and $\{\varepsilon_t\}_{t=1}^n$ is an i.i.d. process with $\varepsilon_t \sim N(0, \Sigma)$ for Σ a *known* 3×3 covariance matrix. Here $\gamma = (\gamma_1, \gamma_2)'$ is two-dimensional, β is a scalar nuisance parameter, and as in the previous section we wish to test $\mathsf{H}_0 : \gamma = 0$ against $\mathsf{H}_1 : \max\{\gamma_1, \gamma_2\} > 0$. For the log-likelihood function $L_n(\theta) = -\frac{1}{2}\sum_{t=1}^n (y_t - \theta)'\Sigma^{-1}(y_t - \theta)$, we have $\hat{\Omega}_n = \Sigma^{-1}$, $\hat{\Sigma}_n = \Sigma^{-1}$, and sandwich estimator $\hat{V}_n = \hat{\Omega}_n^{-1}\hat{\Sigma}_n\hat{\Omega}_n^{-1} = \Sigma$. Note that the Newton–Raphson one-step estimator of θ in (3.1) is the sample mean $\check{\theta}_n = (\check{\gamma}_n', \check{\beta}_n)' = n^{-1}\sum_{t=1}^n y_t$. We now set the number of observations to $n = 500$, which only serves as a scaling for the covariance matrix of the estimator $\check{\theta}_n$ in this setting, and the number of Monte Carlo replications to 10,000.

We again compare the rejection frequencies of our uniform LR test (Unif) with the standard LR test (Std) and the CLR test across various covariance matrices which we can directly control in this setting. We now consider parameter values $\gamma_1, \gamma_2 \in \{0, 0.1\}$ and $\beta \in \{0, 0.1, 0.25\}$ and also look at the performance of the uniform LR test across a range of values $\eta \in \{\alpha/20, \alpha/10, \alpha/2, 9\alpha/10\}$. To further investigate the generality of the good relative power performance of the uniform LR test under positive correlations between the estimators of the parameter of interest and the nuisance parameters, we consider six covariance matrices spanning a range of such positive correlations:

$$\Sigma_1 = \begin{pmatrix} 1 & 0.34 & 0.73 \\ 0.34 & 1 & 0.73 \\ 0.73 & 0.73 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & -0.82 & 0.28 \\ -0.82 & 1 & 0.28 \\ 0.28 & 0.28 & 1 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & -0.26 & 0.54 \\ -0.26 & 1 & 0.54 \\ 0.54 & 0.54 & 1 \end{pmatrix},$$

$$\Sigma_4 = \begin{pmatrix} 1 & 0.86 & 0.86 \\ 0.86 & 1 & 0.86 \\ 0.86 & 0.86 & 1 \end{pmatrix}, \quad \Sigma_5 = \begin{pmatrix} 1 & -0.46 & 0.46 \\ -0.46 & 1 & 0.46 \\ 0.46 & 0.46 & 1 \end{pmatrix}, \quad \Sigma_6 = \begin{pmatrix} 1 & -0.86 & 0.86 \\ -0.86 & 1 & -0.86 \\ 0.86 & -0.86 & 1 \end{pmatrix}.$$

Tables 4–6 report the results for a nominal level of $\alpha = 0.05$ with Table 4 corresponding to β at its boundary ($\beta = 0$), Table 6 corresponding to β away from its boundary ($\beta = 0.25$) and Table 5 corresponding to an intermediate value of β ($\beta = 0.1$). We again see that our uniform LR test and the CLR test appear to maintain good size control across all configurations while the size of the standard LR test is distorted, especially at the boundary of the parameter space. As in the ARCH simulations, for small values of η we can see that the power of the uniform LR test is very competitive in the presence of positive correlations between the estimators of the parameters of interest and the nuisance parameter, especially when β is at its boundary of 0.

[Tables 4–6 around here]

We also note two other features of the uniform LR test's power relative to the CLR's which were present in the ARCH simulations but are easier to see in the present setting in which the parameter estimators' variances do not increase with the value of β: (i) the relative power gain of the uniform LR test vs the CLR test disappears at the intermediate value of β and (ii) the power of the two tests nearly coincides when β is far from its boundary. Regarding (i), when the estimators are positively correlated, the draws with large estimates of the parameter of interest that drive rejection also tend to produce large estimates of the nuisance parameter. This causes the uniform LR CV to be evaluated further from the boundary, thereby increasing the probability of rejection. The CLR test removes this favorable co-movement by conditioning on

$$X_n = \check{\beta}_n - \hat{V}_{n,\beta\gamma}\hat{V}_{n,\gamma\gamma}^{-1}\check{\gamma}_n,$$

that is, the nuisance parameter estimator after the estimator of the parameter of interest has been projected out (see Supplemental Appendix B for more details). The coefficient $\hat{V}_{n,\beta\gamma}\hat{V}_{n,\gamma\gamma}^{-1}$ is chosen precisely so that X_n and $\check{\gamma}_n$ are asymptotically uncorrelated and hence, by the joint asymptotic normality of the one-step estimator, asymptotically independent; the conditional CV therefore cannot inherit the upward co-movement of $\check{\beta}_n$ with $\check{\gamma}_n$ that benefits the uniform LR test. Moreover, writing $V_0 = \Omega_0^{-1}\Sigma_0\Omega_0^{-1}$ for the probability limit of $\hat{V}_n$, the statistic X_n is centered at $\beta - V_{0,\beta\gamma}V_{0,\gamma\gamma}^{-1}\gamma$, so that under positive correlation the very alternatives $\gamma > 0$ that drive rejection shift the conditioning variable *downward*, relative to β. When $\beta = 0$ this shift moves X_n to or below the boundary, resulting in a relatively large conditional CV. When $\beta = 0.1$, the positive nuisance parameter value offsets that downward shift, so the conditional CV falls and the uniform LR test can lose its relative power advantage.

Regarding (ii), as β moves away from its boundary, so do the endpoints of the confidence hyper-rectangle for b computed in Step 4 of Algorithm 1, $b_{L,n}$ and $b_{U,n}$. Consequently, the CV $\mathsf{CV}_{\alpha,n}$ computed in Step 5 is evaluated as a quantile of the asymptotic distribution of the LR statistic at a value far from the boundary. The CLR test's CV is similarly computed away from the boundary as β grows large (see Supplemental Appendix B), leading to tests with similar power in these cases. Finally, we note that smaller values of the user-chosen parameter η yield better power for the uniform LR test, with power more-or-less settling down once η is chosen to be small enough at $\eta = \alpha/10$.

As a further investigation into the generality of our conclusions on the relative power performance of the uniform LR and CLR test, we next consider a variant of the location model (4.1) in which $\gamma = (\gamma_1, \gamma_2)' \in \mathbb{R}^2$ is unrestricted, while $\beta \in [0, \infty)$ remains non-negative. We now test $\mathsf{H}_0 : \gamma = 0$ against $\mathsf{H}_1 : \gamma \neq 0$ with $\theta = (\gamma', \beta)' \in \mathbb{R}^2 \times [0, \infty)$.[5] We find that the relative power strengths between the CLR and uniform LR test continue to hold when the parameter of interest is no longer restricted to have non-negative entries. This can be seen, for example, in Table 7 which reports the rejection frequencies of the two tests under the exact same parameter values as Table 4 after changing the parameter space

[5]In this setting the standard LR test uses the $\chi^2(2)$ CV, which ignores the constraint $\beta \geq 0$.

for γ to $\mathbb{R}^2$. Analogous conclusions regarding the choice of η hold for these designs as well.

[Table 7 around here]

As a final investigation into the generality of our conclusions on the relative power performance of the uniform LR and CLR test, we increase the dimension of the nuisance parameter β to three. We continue to test $\mathsf{H}_0 : \gamma = 0$ against $\mathsf{H}_1 : \max\{\gamma_1, \gamma_2\} > 0$, now with $\theta = (\gamma', \beta')' \in [0, \infty)^2 \times [0, \infty)^3$ so that $\gamma = (\gamma_1, \gamma_2)'$ is two-dimensional and $\beta = (\beta_1, \beta_2, \beta_3)'$ is three-dimensional, all constrained to be non-negative. We test H_0 with Σ a known 5×5 covariance matrix. We consider three covariance matrices entailing positive correlations between the estimators of the parameter of interest and the nuisance parameters, generated using the GFT parametrization:

$$\Sigma_7 = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & \cdot & \cdot \\ 0.641 & 0 & 1 & \cdot & \cdot \\ 0 & 0 & 0 & 1 & \cdot \\ 0.641 & 0 & 0.243 & 0 & 1 \end{pmatrix}, \qquad \Sigma_8 = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ 0.230 & 1 & \cdot & \cdot & \cdot \\ 0.528 & 0.537 & 1 & \cdot & \cdot \\ 0.528 & 0.537 & 0.602 & 1 & \cdot \\ 0.443 & 0.037 & 0.121 & 0.121 & 1 \end{pmatrix},$$

$$\Sigma_9 = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ 0.386 & 1 & \cdot & \cdot & \cdot \\ -0.367 & 0.107 & 1 & \cdot & \cdot \\ -0.159 & 0.162 & -0.434 & 1 & \cdot \\ 0.386 & 0.403 & 0.107 & 0.162 & 1 \end{pmatrix}.$$

The standard LR test uses the chi-bar-squared CV, ignoring the constraint $\beta \geq 0$. We again find qualitatively similar results to those for the other simulation designs, e.g., in Table 8, where $\beta = 0$. As in other cases, the negative correlations present in Σ_9 lead to worse power performance of the uniform LR test relative to the CLR test whereas purely positive correlations make the uniform LR test much more competitive. It is also notable that the higher-dimensional nuisance parameter in these simulations entails more severe over-rejection for the standard LR test, e.g., 12.9% for Σ_8.

[Table 8 around here]

To conclude this simulation section, we gather a few practical takeaway recommendations from our simulation experiments. First, the standard LR test does not appear to control size while the uniform LR and CLR tests do. We therefore recommend the latter two tests over the former. Second, the LF approach to CV construction is significantly more computationally intensive than the uniform LR and CLR approaches and loses computational tractability quickly as the dimension of the boundary constrained-nuisance parameter grows beyond one. We therefore again recommend the latter two tests over the former, especially in the case of a multi-dimensional nuisance parameter. Finally, when deciding between the uniform LR and CLR tests, we recommend that the researcher first estimate the asymptotic correlation(s) between the estimators in Assumption 6 of the parameter(s) of interest and the boundary-constrained nuisance parameter(s) and if these correlation(s) appear to be mostly positive, and especially of a larger magnitude, we recommend the use of the uniform LR test over the CLR test. Otherwise, we recommend the CLR test. The logic behind this final recommendation is twofold: (i) the uniform LR test appears to have better power properties under positive correlations when the nuisance parameter is at or very close to the boundary, which is presumably the region of the parameter space of most empirical interest in problems for which the user is concerned about nuisance parameter boundary-robustness; (ii) the aforementioned correlation matrix is typically consistently estimable so that this recommended decision rule should not produce pretest size distortions.

5 Empirical illustration

In this section we consider an empirical illustration of our proposed test. We consider ARCH models for daily returns of various stock indices and test for the presence of ARCH effects and spillovers from the U.S. stock market. Inspired by the HAR model of Corsi (2009), and similar to Nielsen et al. (2025, Section 5), we include lagged Realized Volatility (RV) covariates to account for potential high persistence in the conditional variance of the index returns. With the *daily* index return given by y_t in (2.5), let

$$\sigma_t^2 = \delta + \beta_{ARCH} y_{t-1}^2 + \beta_{RV} RV_{t-1} + \beta_W RV_{W,t-1} + \beta_M RV_{M,t-1} + \beta_{SPX} SPX_{t-1}^2,$$

with $\delta > 0$ and $\beta_{ARCH}, \beta_{RV}, \beta_W, \beta_M, \beta_{SPX} \geq 0$, where δ is assumed to be bounded away from its boundary but not $\beta_{ARCH}, \beta_{RV}, \beta_W, \beta_M, \beta_{SPX}$. The variable RV_t is the RV of the index based on 5-minute intraday returns at day t, $RV_{W,t} = 5^{-1} \sum_{i=0}^{4} RV_{t-i}$ is the weekly average RV, and $RV_{M,t} = 22^{-1} \sum_{i=0}^{21} RV_{t-i}$ is the monthly average RV. Lastly, the variable SPX_t is the (continuously compounded, close-to-close) return on the S&P 500 index at day t. Based on data retrieved from the Oxford Man Realized Library covering the period 3 January 2000 to 27 June 2018, we analyze the following indices: Australian ASX All Ordinaries (AORD), Belgian BEL 20 (BFX), Spanish IBEX 35 (IBEX), IPC Mexico (MXX), Indian NIFTY50 (NSEI) and Danish OMX C20 (OMX) (notice that OMX C20 begins on 3 October 2005). For each of these indices, we test the hypothesis of no ARCH, $\mathsf{H}_{ARCH} : \beta_{ARCH} = 0$, and the hypothesis of no spillovers from the U.S. stock market, $\mathsf{H}_{SPX} : \beta_{SPX} = 0$.

Table 9 contains point estimates of the model parameters for each index series. Moreover, it contains the values of the LR and CLR statistics along with their corresponding CVs based on 5% and 10% nominal levels. The uniform LR and CLR tests both fail to reject H_{ARCH} for all indices except the Mexican MXX and Indian NSEI indices, for which they both reject at the 5% level. By contrast, both tests reject H_{SPX} at the 5% level for all indices except for the Danish OMX index for which the results are mixed with the uniform LR test failing to reject and the CLR test rejecting at the 5% level. But even for the CLR test the rejection appears to be weakest for the Danish OMX index, leading to the conclusion that spillovers from the U.S. stock market appear to be weakest for the Danish OMX index.

[Table 9 around here]

In short, we find little evidence of ARCH effects in most of the index return series, whereas most of the series appear subject to spillovers from the U.S. stock market.

6 Conclusions

This paper proposes a novel and computationally efficient method for constructing CVs for LR tests that achieve uniform asymptotic size control, even when nuisance parameters lie on or near the boundary of their parameter space. The key innovation lies in using confidence bounds for an asymptotically Gaussian approximation of a nuisance parameter estimator and exploiting the monotonicity properties of the two components of the LR statistic's asymptotic distribution. This allows for the formation of valid CVs via straightforward Monte Carlo simulation, without the need for intensive optimization or conservative LF approaches. The method remains tractable in settings for which the nuisance parameter is not low-dimensional and is broadly applicable, including to models such as constrained regressions and ARCH specifications, offering the first proven uniformly valid test for the latter.

Acknowledgements

We thank Yanqin Fan (Co-Editor), an Associate Editor, and two anonymous referees for important comments and suggestions on a previous draft of the paper. We also thank Pascal Lavergne, Frank Kleibergen, and participants at the 2022 CIREQ Montreal Econometrics Conference, the 2023 Robust Econometric Methods in Financial Econometrics Workshop in Copenhagen, the 2023 ICEEE Conference in Cagliari, and the 2024 Toulouse Financial Econometrics Conference for comments.

McCloskey acknowledges support from the National Science Foundation under Grant SES-2341730. Rahbek and Cavaliere gratefully acknowledge support from the Independent Research Fund Denmark (DFF Grant 7015-00028) and the Italian Ministry of University and Research (PRIN 2020 Grant 2020B2AKFW). Pedersen acknowledges support from Independent Research Fund Denmark ("Robust Econometric Methods in Empirical and Quantitative Finance", FSE 0133-00162B).

References

Andrews, D. W. K. (2001): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica*, vol. 69, 683–734.

Andrews, D. W. K. and Cheng, X. (2012): "Estimation and inference with weak, semi-strong, and strong identification", *Econometrica*, vol. 80, 2153–2211.

Andrews, D. W. K. and Guggenberger, P. (2009): “Hybrid and size-corrected subsampling methods”, *Econometrica*, vol. 77, 721–762.

Archakov, I. and Hansen, P. R. (2021): “A new parametrization of correlation matrices”, *Econometrica*, vol. 89, 1699–1715.

Berger, R. L. and Boos, D. D. (1994): “P values maximized over a confidence set for the nuisance parameter”, *Journal of the American Statistical Association*, vol. 89, 1012–1016.

Berry, S., Levinsohn, J., and Pakes, A. (1995): “Automobile prices in market equilibrium”, *Econometrica*, vol. 63, 841–890.

Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2022): “Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models”, *Journal of Econometrics*, vol. 227, 241–263.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2017): “On the consistency of bootstrap testing for a parameter on the boundary of the parameter space”, *Journal of Time Series Analysis*, vol. 38, 513–534.

Cavaliere, G., Perera, I., and Rahbek, A. (2024): “Specification tests for GARCH processes”, *Journal of Business and Economic Statistics*, vol. 42, 197–214.

Corsi, F. (2009): “A simple approximate long-memory model of realized volatility”, *Journal of Financial Econometrics*, vol. 7, 174–196.

Drton, M. (2009): “Likelihood ratio tests and singularities”, *Annals of Statistics*, vol. 37, 979–1012.

Fan, Y. and Shi, X. (2023): “Wald, QLR, and score tests when the parameters are subject to linear inequality constraints”, *Journal of Econometrics*, vol. 235, 2005–2026.

Francq, C. and Thieu, L. Q. (2019): “QML inference for volatility models with covariates”, *Econometric Theory*, vol. 35, 37–72.

Francq, C., and Zakoïan, J.-M. (2009), “Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons”, *Journal of the American Statistical Association*, vol. 104, 313–324.

Geyer, C. J. (1994): “On the asymptotics of constrained M-estimators”, *The Annals of Statistics*, vol. 22, 1993–2010.

Hong, H. and Li, J. (2020): “The numerical bootstrap”, *The Annals of Statistics*, vol. 48, 397–412.

Ketz, P. (2018): “Subvector inference when the true parameter vector may be near or at the boundary”, *Journal of Econometrics*, vol. 207, 285–306.

Ketz, P. and McCloskey, A. (2025): “Short and simple confidence intervals when the directions of some effects are known”, *Review of Economics and Statistics*, vol. 107, 820–834

Li, J. (2025): “The proximal bootstrap for constrained estimators”, *Journal of Statistical Planning and Inference*, vol. 236, 106245.

McCloskey, A. (2017): “Bonferroni-based size-correction for nonstandard testing problems”, *Journal of Econometrics*, vol. 200, 17–35.

McCloskey, A. (2020): “Asymptotically uniform tests after consistent model selection in the linear regression model”, *Journal of Business and Economic Statistics*, vol. 38, 810–825.

Mitchell, J. D., Allman, E. S., and Rhodes, J. A. (2019): “Hypothesis testing near singularities and boundaries”, *Electronic Journal of Statistics*, vol. 13, 2150–2193.

Newey, W. K. and McFadden, D. L. (1994): "Large sample estimation and hypothesis testing". In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, vol. 4, 2111–2245.

Newey, W. K. and West, K. D. (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, vol. 55, 703–708.

Nielsen, H. B., Pedersen, R. S., Rahbek, A., and Thorsen, S.N. (2025): "Testing in GARCH-X models: Boundary, correlations and bootstrap theory", *Journal of Time Series Analysis*, vol. 46, 1175–1204.

Romano, J. P., Shaikh, A., and Wolf, M.. (2014): "A practical two-step method for testing moment inequalities", *Econometrica*, vol. 82, 1979–2002.

Self, S. G. and Liang, K.-Y. (1987): "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, vol. 82, 605–610.

Shapiro, A. (1989): "Asymptotic properties of statistical estimators in stochastic programming", *The Annals of Statistics*, vol. 17, 841–858.

Silvapulle, M. J. (1996): "A test in the presence of nuisance parameters", *Journal of the American Statistical Association*, vol. 91, 1690–1693.

Silvapulle, M. J. and Silvapulle, P. (1995): "A score test against one-sided alternatives", *Journal of the American Statistical Association*, vol. 90, 342–349.

Supplemental Appendix

A Proofs

A.1 Proofs of main results

Proof of Lemma 2.1

We first introduce some notation. Let $b^{(1)} = (b_1^{(1)}, \dots, b_{d_\beta}^{(1)})'$ and $b^{(2)} = (b_1^{(2)}, \dots, b_{d_\beta}^{(2)})'$ satisfy

$$b_i^{(1)} = \mathbb{I}(b_i < \infty) b_i, \ \ b_i^{(2)} = \mathbb{I}(b_i = \infty) b_i$$

for $i = 1, \dots, d_\beta$, and where we work under the convention that $0 \times \infty = 0$. Note that $b = b^{(1)} + b^{(2)}$. Likewise, let $\beta_n^{(1)} = (\beta_{n,1}^{(1)}, \dots, \beta_{n,d_\beta}^{(1)})'$ and $\beta_n^{(2)} = (\beta_{n,1}^{(2)}, \dots, \beta_{n,d_\beta}^{(2)})'$ satisfy

$$\beta_{n,i}^{(1)} = \mathbb{I}(b_i < \infty)\beta_{n,i}, \quad \beta_{n,i}^{(2)} = \mathbb{I}(b_i = \infty)\beta_{n,i}$$

for $i = 1, \dots, d_\beta$. Define

$$\theta_n^{(1)} = (0'_{d_\gamma}, \beta_n^{(1)\prime}, 0'_{d_\delta})' \quad \text{and} \quad \theta_n^{(2)} = (\gamma_0', \beta_n^{(2)\prime}, \delta_n')'. \tag{A.1}$$

In particular, $\theta_n = \theta_n^{(1)} + \theta_n^{(2)}$, and we have that under any $\{\psi_n\}$ sequence,

$$\sqrt{n}\theta_n^{(1)} \to (0'_{d_\gamma}, b^{(1)}, 0'_{d_\delta})' \equiv \tau \tag{A.2}$$

and

$$\sqrt{n}(\Theta - \theta_n^{(2)}) \cap \mathbb{R}^{d_\theta} \to \tilde{\Lambda} \quad \text{and} \quad \sqrt{n}(\Theta_{\mathsf{H}_0} - \theta_n^{(2)}) \cap \mathbb{R}^{d_\theta} \to \tilde{\Lambda}_0 \tag{A.3}$$

where $\Theta_{\mathsf{H}_0} \equiv \{\theta \in \Theta : \gamma = \gamma_0\}$ and $\tilde{\Lambda}$ and $\tilde{\Lambda}_0$ are convex cones (with zero vertex) given respectively by

$$\tilde{\Lambda} = \Lambda_\gamma \times \tilde{\Lambda}_\beta \times \Lambda_\delta \quad \text{and} \quad \tilde{\Lambda}_0 = \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta \times \Lambda_\delta, \tag{A.4}$$

for $\Lambda_\gamma = \mathbb{R}^s \times [0,\infty)^{d_\gamma - s}$, $\Lambda_\delta = \mathbb{R}^{d_\delta}$ and $\tilde{\Lambda}_\beta = \tilde{\Lambda}_{\beta,1} \times \cdots \times \tilde{\Lambda}_{\beta,d_\beta}$, with

$$\tilde{\Lambda}_{\beta,i} = \begin{cases} \mathbb{R}_+ & \text{if } b_i < \infty \\ \mathbb{R} & \text{if } b_i = \infty \end{cases}, \quad i = 1, \dots, d_\beta.$$

As in Andrews (1999), we note that

$$L_n(\theta) = L_n(\theta_n) + \frac{1}{2} Z_n' \Omega_n Z_n - \frac{1}{2} \| \sqrt{n}(\theta - \theta_n) - Z_n \|_{\Omega_n}^2 + R_n(\theta),$$

with

$$\Omega_n \equiv -n^{-1} \frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'}, \tag{A.5}$$

$$Z_n = \Omega_n^{-1} n^{-1/2} \frac{\partial L_n(\theta_n)}{\partial\theta}, \tag{A.6}$$

$$\|\lambda - Z_n\|_{\Omega_n}^2 = (\lambda - Z_n)' \, \Omega_n (\lambda - Z_n), \quad \lambda \in \mathbb{R}^{d_\theta}. \tag{A.7}$$

Hence,

$$\begin{aligned}
\mathsf{LR}_n &= 2(L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)) \\
&= \|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 - \|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 + 2\left(R_n(\hat{\theta}_n) - R_n(\tilde{\theta}_n)\right) \\
&= \|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 - \|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 + o_p(1),
\end{aligned}$$

where the last equality follows by Assumptions 1–2. We seek to show that, jointly,

$$\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}_0} \|\lambda - (Z + \tau)\|_{\Omega_0}^2$$

and

$$\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|_{\Omega_0}^2.$$

Given this convergence, the limiting distribution, $\mathcal{L}_\infty(b, b)$, of LR_n is then derived by standard arguments, using the structure of $\tilde{\Lambda}$ and $\tilde{\Lambda}_0$ and that $\Lambda_\delta = \mathbb{R}^{d_\delta}$:

$$\begin{aligned}
&\inf_{\lambda \in \tilde{\Lambda}_0} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 - \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 \\
&= \inf_{\lambda \in \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta \times \Lambda_\delta} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 - \inf_{\lambda \in \Lambda_\gamma \times \tilde{\Lambda}_\beta \times \Lambda_\delta} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 \\
&= \inf_{\lambda \in \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta} \|\lambda - H(Z + \tau)\|_{(H\Omega_0^{-1}H')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \tilde{\Lambda}_\beta} \|\lambda - H(Z + \tau)\|_{(H\Omega_0^{-1}H')^{-1}}^2 \\
&= \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_\beta(b)} \|\lambda - HZ\|_{(H\Omega_0^{-1}H')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_\beta(b)} \|\lambda - HZ\|_{(H\Omega_0^{-1}H')^{-1}}^2 \\
&= \mathcal{L}_\infty(b, b).
\end{aligned}$$

Let us focus on the convergence of $\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2$, noting that the convergence of $\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ follows by similar arguments, and that the convergence holds jointly, as $\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ and $\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ are functions of the same data $\{W_t\}$. It suffices to show the following three properties:

1. With $\hat{\theta}_{q,n}$ satisfying $\|\sqrt{n}(\hat{\theta}_{q,n} - \theta_n) - Z_n\|^2_{\Omega_n} = \inf_{\theta\in\Theta} \|\sqrt{n}(\theta - \theta_n) - Z_n\|^2_{\Omega_n}$, it holds that
$$\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n} = \|\sqrt{n}(\hat{\theta}_{q,n} - \theta_n) - Z_n\|^2_{\Omega_n} + o_p(1).$$

2. It holds that
$$\|\sqrt{n}(\hat{\theta}_{q,n}-\theta_n)-Z_n\|^2_{\Omega_n} = \inf_{\theta\in\Theta} \|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n} = \inf_{\lambda\in\tilde{\Lambda}} \|\lambda-\sqrt{n}\theta_n^{(1)}-Z_n\|^2_{\Omega_n}+o_p(1).$$

3. It holds that
$$\inf_{\lambda\in\tilde{\Lambda}} \|\lambda - \sqrt{n}\theta_n^{(1)} - Z_n\|^2_{\Omega_n} \xrightarrow{d} \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z+\tau)\|^2_{\Omega_0}.$$

These properties follow from Lemmas A.1–A.3. □

Proof of Lemma 3.1

With $\underline{b} \leq b \leq \bar{b}$, it holds that $\Lambda_\beta(\underline{b}) \subset \Lambda_\beta(b) \subset \Lambda_\beta(\bar{b})$. This implies that

$$\inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(\underline{b})} Q(\lambda) \geq \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(b)} Q(\lambda),$$

and

$$\inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(b)} Q(\lambda) \geq \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(\bar{b})} Q(\lambda).$$

Hence,

$$\begin{aligned}\mathcal{L}_\infty(b,b) &= \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(b)} Q(\lambda) - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(b)} Q(\lambda) \leq \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(\underline{b})} Q(\lambda) - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(\bar{b})} Q(\lambda)\\ &= \mathcal{L}_\infty(\underline{b},\bar{b}),\end{aligned}$$

as required. □

Proof of Lemma 3.2

By a Taylor-type expansion at θ_n,

$$\sqrt{n}\left(\check{\theta}_n-\theta_n\right)=\left[I_{d_\theta}-\left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'}\right]\sqrt{n}(\hat{\theta}_n-\theta_n)$$
$$-\left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}n^{-1/2}\frac{\partial L_n(\theta_n)}{\partial\theta}+o_p\left(1\right)\xrightarrow{d}N(0,\Omega_0^{-1}\Sigma_0\Omega_0^{-1})$$

where we have used Assumptions 1–4 and Remark 2.2. This proves (3.2). In addition, (3.3) follows directly from Assumptions 1–3 and 5. □

Proof of Theorem 3.3

Let $\mathsf{CV}_{q,n}(x,y)$ denote the qth quantile of $\mathcal{L}_{\infty,n}(x,y)$ and now let $\mathsf{CV}_q(x,y)$ denote the qth quantile of $\mathcal{L}_\infty(x,y)$. For the parameter space Ψ_0, standard subsequencing arguments (e.g., Andrews and Guggenberger, 2009, and McCloskey, 2017) provide that showing

$$\lim_{n\to\infty}\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\leq\alpha \tag{A.8}$$

under all $\{\psi_n\}$ sequences in Ψ_0 satisfying $\psi_n\to\psi_0$ and (2.9) is sufficient for proving the statement of the theorem. Consider any such sequence; then, we have

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&=\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})>\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\right)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\geq\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\geq\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\leq\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n>\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\right) \qquad\qquad (A.9)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right).
\end{aligned}$$

Note the following:

(a) when evaluated at any continuity point of the distribution function of $\mathcal{L}_\infty(b,b)$, the distribution function of $\mathcal{L}_{\infty,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)$ converges in probability to the distribution function of $\mathcal{L}_\infty(b,b)$ by Assumptions 1–3 and 5 and the continuous mapping theorem;

(b) $\mathsf{LR}_n \xrightarrow{d} \mathcal{L}_\infty(b,b)$ by Lemma 2.1;

(c) by Assumption 7, $\mathbb{P}(\mathcal{L}_\infty(b,b) \le 0) < 1-\alpha+\eta$ so that Lemma A.4 implies

$$\lim_{\epsilon\uparrow 0} \mathbb{P}(\mathcal{L}_\infty(b,b) \le \mathsf{CV}_{1-\alpha+\eta}(b,b)+\epsilon) = \mathbb{P}(\mathcal{L}_\infty(b,b) \le \mathsf{CV}_{1-\alpha+\eta}(b,b)) = 1-\alpha+\eta$$

and $\mathbb{P}(\mathcal{L}_\infty(b,b) \le \mathsf{CV}_{1-\alpha+\eta}(b,b)+\epsilon) > 1-\alpha+\eta$ for all $\epsilon > 0$.

Therefore, Lemma 5(ii) of Andrews and Guggenberger (2010) provides that

$$\lim_{n\to\infty} \mathbb{P}_{\psi_n}\left(\mathsf{LR}_n > \mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n, \sqrt{n}\beta_n)\right) = \alpha-\eta. \tag{A.10}$$

For $q_{1-\eta}$ equal to the $(1-\eta)$-quantile of $\max_{i=1,\ldots,d_\beta}|Z_{\beta,i}|$ with $Z_\beta \overset{d}{=} N(0,\Omega_\beta)$ and $\Omega_\beta = \operatorname{diag}(\Sigma_\beta)^{-1/2}\Sigma_\beta \operatorname{diag}(\Sigma_\beta)^{-1/2}$,

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n) > \mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n) \le \mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\le 1-\mathbb{P}_{\psi_n}\left(b_{L,n} \le \sqrt{n}\beta_n \le b_{U,n}\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\check{b}_{L,n} \le \sqrt{n}\beta_n \le b_{U,n}\right)\\
&\le 1-\mathbb{P}_{\psi_n}\left(\check{b}_{L,n} \le \sqrt{n}\beta_n \le \check{b}_{U,n}\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(-\check{q}_{1-\eta,n}\operatorname{diagv}(\check{\Sigma}_{\beta,n})^{1/2} \le \sqrt{n}(\check{\beta}_n-\beta_n) \le \check{q}_{1-\eta,n}\operatorname{diagv}(\check{\Sigma}_{\beta,n})^{1/2})\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(-\check{q}_{1-\eta,n} \le \frac{\sqrt{n}(\check{\beta}_{n,i}-\beta_{n,i})}{\sqrt{\check{\Sigma}_{\beta,n,ii}}} \le \check{q}_{1-\eta,n} \text{ for all } i=1,\ldots,d_\beta\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\max_{i=1,\ldots,d_\beta}\left|\frac{\sqrt{n}(\check{\beta}_{n,i}-\beta_{n,i})}{\sqrt{\check{\Sigma}_{\beta,n,ii}}}\right| \le \check{q}_{1-\eta,n}\right)\\
&\to 1-\mathbb{P}\left(\max_{i=1,\ldots,d_\beta}|Z_{\beta,i}| \le q_{1-\eta}\right) = \eta,
\end{aligned} \tag{A.11}$$

where the inequalities inside of probabilities are evaluated element-wise across vectors, the first inequality follows from Lemma 3.1, the second equality follows from $\beta_n \ge 0$ and the convergence follows from Assumption 6 and the continuity of the distribution of the maximum of the absolute values of a nondegenerate Gaussian random vector.

Together, (A.9)–(A.11) imply (A.8), and therefore the statement of the theorem. $\square$

Proposition A.1 *Consider the model of Section 1 with* $d_\gamma = d_\beta = 1$, $\gamma \geq 0$, $Z \sim \mathrm{N}(0, \Sigma)$, *and* $\mathcal{L}(b, b)$ *as in* (1.1). *Let* Σ *be a correlation matrix with correlation parameter* $\rho \in (-1, 1)$. *Write* $\sigma = \sqrt{1-\rho^2}$, *let* Φ *denote the standard normal distribution function, and let* $\Phi_2(\,\cdot\,, \cdot\,; r)$ *denote the standard bivariate normal distribution function with correlation* r. *Then*

$$\mathbb{P}\left(\mathcal{L}(b,b)=0\right) = \tfrac{1}{2}\Phi\left(\frac{b}{\sigma}\right) + \Phi_2\left(\frac{\rho b}{\sigma}, -\frac{b}{\sigma}; -\rho\right) \equiv f(b/\sigma). \tag{A.12}$$

Moreover, with $\phi(\cdot)$ *the pdf of the standard normal distribution,* $f'(t) = \rho\,\phi(\rho t)\,\Phi(-t\sigma)$ *for* $t \geq 0$, *so that* $\mathbb{P}(\mathcal{L}(b,b)=0)$ *is nonincreasing in* b *when* $\rho \leq 0$ *and nondecreasing in* b *when* $\rho \geq 0$, *with endpoint values*

$$\mathbb{P}\left(\mathcal{L}(0,0)=0\right) = \frac{1}{4} + \frac{\arccos\rho}{2\pi} \qquad \textit{and} \qquad \mathbb{P}\left(\mathcal{L}(\infty,\infty)=0\right) = \frac{1}{2}.$$

Consequently,

$$\sup_{b\in[0,\infty]} \mathbb{P}\left(\mathcal{L}(b,b)=0\right) = \max\left\{\frac{1}{2},\ \frac{1}{4} + \frac{\arccos\rho}{2\pi}\right\} < \frac{3}{4}. \tag{A.13}$$

Proof. Write $A = \Sigma^{-1}$, $C_b = [0,\infty] \times [-b,\infty]$ and $F_b = \{0\} \times [-b,\infty]$, so that, as in (1.1), $\mathcal{L}(b,b) = \inf_{\lambda\in F_b} Q(\lambda) - \inf_{\lambda\in C_b} Q(\lambda)$ with $Q(\lambda) = (\lambda - Z)'A(\lambda - Z)$. Since $Q(\lambda) \to \infty$ as $\|\lambda\| \to \infty$, both infima are unchanged if C_b and F_b are replaced by their intersections with $\mathbb{R}^2$,

$$C_b^\circ = [0,\infty) \times [-b,\infty) \qquad \text{and} \qquad F_b^\circ = \{0\} \times [-b,\infty),$$

which are closed and convex, and we work with the latter two sets throughout. As Q is strictly convex, the infimum over C_b° is attained at a unique point $\hat\lambda = (\hat\lambda_\gamma, \hat\lambda_b)'$, and likewise over F_b°. Note that

$$\{\mathcal{L}(b,b)=0\} = \{\hat\lambda \in F_b^\circ\} = \{\hat\lambda_\gamma = 0\}.$$

Write $(X, Y) = (Z_\gamma, Z_b)$, a standard bivariate normal pair with correlation ρ, and index the entries of A by γ and b as well. Partition $\{\hat\lambda_\gamma = 0\}$ according to whether the constraint on the b-coordinate is slack. Let $b < \infty$.

Case (a): $\hat\lambda_\gamma = 0$ and $\hat\lambda_b > -b$. It holds that $[A(Z - \hat\lambda)]_b = 0$, i.e. $\hat\lambda_b = Y + (A_{b\gamma}/A_{bb})X =$

$Y - \rho X$, while optimality in the γ-coordinate requires $[A(Z - \hat{\lambda})]_\gamma \le 0$. Substituting the expression for $\hat{\lambda}_b$,

$$[A(Z - \hat{\lambda})]_\gamma = A_{\gamma\gamma} X - \frac{A_{\gamma b} A_{b\gamma}}{A_{bb}} X = \frac{\det A}{A_{bb}} X,$$

so that, $\det A > 0$ and $A_{bb} > 0$, this optimality condition is equivalent to $X \le 0$, while the slackness condition $\hat{\lambda}_b > -b$ reads $Y - \rho X > -b$. Hence case **(a)** is the event $\{X \le 0\} \cap \{Y - \rho X > -b\}$. Because $Y - \rho X \sim \mathrm{N}(0, 1 - \rho^2)$ is independent of X,

$$\mathbb{P}\left(\textbf{(a)}\right) = \mathbb{P}(X \le 0)\, \mathbb{P}(Y - \rho X > -b) = \tfrac{1}{2} \Phi(b/\sigma).$$

Case (b): $\hat{\lambda}_\gamma = 0$ and $\hat{\lambda}_b = -b$. Here $\hat{\lambda} = (0, -b)'$, which is optimal if and only if $[A(Z - v)]_\gamma \le 0$ and $[A(Z - v)]_b \le 0$. Using $A = \Sigma^{-1}$, these two inequalities read

$$X \le \rho(Y + b) \qquad \text{and} \qquad Y + b \le \rho X.$$

Set $W_1 = \rho(Y + b) - X$ and $W_2 = \rho X - (Y + b)$. Then (W_1, W_2) is bivariate normal with $\mathbb{E} W_1 = \rho b$, $\mathbb{E} W_2 = -b$, $\operatorname{Var} W_1 = \operatorname{Var} W_2 = 1 - \rho^2$ and $\operatorname{Cov}(W_1, W_2) = \rho^3 - \rho = -\rho(1 - \rho^2)$, so that their correlation equals $-\rho$. Therefore

$$\mathbb{P}\left(\textbf{(b)}\right) = \mathbb{P}(W_1 \ge 0, W_2 \ge 0) = \Phi_2\!\left(\frac{\rho b}{\sigma}, -\frac{b}{\sigma}; -\rho\right).$$

Events **(a)** and **(b)** are disjoint and exhaust $\{\hat{\lambda}_\gamma = 0\}$, which gives (A.12). When $b = \infty$, only case **(a)** above arises and $\mathbb{P}(\mathcal{L}(\infty, \infty) = 0) = \mathbb{P}(X \le 0) = 1/2$, in agreement with (A.12).

For the derivative, write $t = b/\sigma$ and recall that, for $|r| < 1$, differentiating Φ_2 in each of its arguments and factoring the bivariate normal density into the marginal of one coordinate and the conditional $N(r\,\cdot\,, 1 - r^2)$-distribution of the other gives

$$\partial_a \Phi_2(a, c; r) = \phi(a)\, \Phi\!\left(\frac{c - ra}{\sqrt{1 - r^2}}\right) \qquad \text{and} \qquad \partial_c \Phi_2(a, c; r) = \phi(c)\, \Phi\!\left(\frac{a - rc}{\sqrt{1 - r^2}}\right).$$

Both partial derivatives are continuous, so $\Phi_2(\,\cdot\,, \cdot\,; r)$ is totally differentiable and the chain rule applies along the curve $t \mapsto (a(t), c(t))$. With $a(t) = \rho t$, $c(t) = -t$, $r = -\rho$ and $\sqrt{1 - r^2} = \sigma$, so that $a'(t) = \rho$ and $c'(t) = -1$,

$$f'(t) = \tfrac{1}{2}\phi(t) + \rho\, \phi(\rho t)\, \Phi\!\left(\frac{-t + \rho^2 t}{\sigma}\right) - \phi(t)\, \Phi\!\left(\frac{\rho t - \rho t}{\sigma}\right) = \rho\, \phi(\rho t)\, \Phi(-t\sigma),$$

where the second term is $a'(t)\,\partial_a\Phi_2$ and the third is $c'(t)\,\partial_c\Phi_2$, and where we have used $(-t+\rho^2 t)/\sigma = -t(1-\rho^2)/\sigma = -t\sigma$, $\phi(-t)=\phi(t)$ and $\Phi(0)=1/2$. Thus f' has the sign of ρ. Finally, $f(0)=\frac{1}{4}+\Phi_2(0,0;-\rho)=\frac{1}{2}-\arcsin(\rho)/(2\pi)=\frac{1}{4}+\arccos(\rho)/(2\pi)$, using that $\Phi_2(0,0;r)=1/4+\arcsin(r)/(2\pi)$, and $f(\infty)=1/2+0=1/2$. Combining the monotonicity with the endpoint values yields (A.13); the strict inequality follows since $\arccos\rho<\pi$ for $\rho\in(-1,1)$. $\square$

A.2 Technical lemmas for proving main results

Lemma A.1 *Let $\hat\theta_{q,n}$ satisfy $\|\sqrt{n}(\hat\theta_{q,n}-\theta_n)-Z_n\|^2_{\Omega_n}=\inf_{\theta\in\Theta}\|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n}$, with Ω_n given by (A.5) and Z_n given by (A.6). Under Assumptions 1-4 and any sequence $\{\psi_n\}\subset\Psi_0$ satisfying $\psi_n\to\psi_0$ and (2.9),*

$$\|\sqrt{n}(\hat\theta_n-\theta_n)-Z_n\|^2_{\Omega_n}=\|\sqrt{n}(\hat\theta_{q,n}-\theta_n)-Z_n\|^2_{\Omega_n}+o_p(1).$$

Proof. The result follows directly from arguments given in Andrews (1999, proof of Theorem 2) with θ_0 replaced by θ_n. $\square$

Lemma A.2 *Under Assumptions 3-4 and any sequence $\{\psi_n\}\subset\Psi_0$ satisfying $\psi_n\to\psi_0$ and (2.9), it holds that*

$$\inf_{\theta\in\Theta}\|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n}=\inf_{\lambda\in\tilde\Lambda}\|\lambda-\sqrt{n}\theta_n^{(1)}-Z_n\|^2_{\Omega_n}+o_p(1),$$

where Ω_n given by (A.5), Z_n given by (A.6), $\tilde\Lambda$ is given by (A.4) and $\theta_n^{(1)}$ is given by (A.1).

Proof. By definition, using (A.1),

$$\|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n}=\|\sqrt{n}(\theta-\theta_n^{(2)})-(Z_n+\sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n}.$$

Hence,

$$\begin{aligned}\inf_{\theta\in\Theta}\|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n}&=\inf_{\theta\in\Theta}\|\sqrt{n}(\theta-\theta_n^{(2)})-(Z_n+\sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n}\\&=\inf_{\lambda\in\sqrt{n}(\Theta-\theta_n^{(2)})}\|\lambda-(Z_n+\sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n}\\&=\inf_{\lambda\in\sqrt{n}(\Theta-\theta_n^{(2)})\cap\mathbb{R}^{d_\theta}}\|\lambda-(Z_n+\sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n},\end{aligned}$$

where the last equality follows by the fact that the quadratic form diverges for non-finite λ since Ω_n is positive definite. By Assumptions 3-4 and (A.2), we have that $Z_n + \sqrt{n}\theta_n^{(1)} = O_p(1)$. Hence, the result now follows using the fact that $\sqrt{n}(\Theta - \theta_n^{(2)}) \cap \mathbb{R}^{d_\theta}$ contains zero for all $n \geq 1$, the set convergence in (A.3), and Silvapulle and Sen (2005, Corollary 4.7.5.1 and the comments on p. 194). $\square$

LEMMA A.3 *Under Assumptions 3-4 and any sequence $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n \to \psi_0$ and (2.9), it holds that*

$$\inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n} \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|^2_{\Omega_0},$$

where Ω_n given by (A.5), Z_n given by (A.6), $\tilde{\Lambda}$ is given by (A.4), $\theta_n^{(1)}$ is given by (A.1), and τ is defined in (A.2).

PROOF. Since $(Z_n + \sqrt{n}\theta_n^{(1)}) = O_p(1)$ and $\Omega_n = \Omega_0 + o_p(1)$ by Assumptions 3-4, we have by Silvapulle and Sen (2005, Lemma 4.10.2.2) that

$$\inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n} = \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_0} + o_p(1).$$

By Silvapulle and Sen (2005, Corollary 4.7.5.2) and the fact that $Z_n + \sqrt{n}\theta_n^{(1)} \xrightarrow{d} Z + \tau$, we have that $\inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_0} \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|^2_{\Omega_0}$. $\square$

LEMMA A.4 *For any $b \in [0, \infty]^{d_\beta}$, the distribution of $\mathcal{L}_\infty(b, b)$ has no atoms on $(0, \infty)$, and*

$$\mathbb{P}\left(x < \mathcal{L}_\infty(b, b) < y\right) > 0 \quad \textit{for every } 0 \leq x < y < \infty.$$

Consequently, its only possible atom is at zero. Writing F_b for the distribution function of $\mathcal{L}_\infty(b, b)$, for every $q \in (F_b(0), 1)$ the qth quantile $c_q(b)$ is strictly positive and

$$\lim_{\epsilon \uparrow 0} F_b\left(c_q(b) + \epsilon\right) = F_b\left(c_q(b)\right) = q, \qquad F_b\left(c_q(b) + \varepsilon\right) > q \quad \textit{for every } \varepsilon > 0. \tag{A.14}$$

PROOF. Let

$$A = (H\Omega_0^{-1}H')^{-1}, \qquad V = HZ,$$

and define the nested polyhedra

$$C_0(b) = \{0\}^{d_\gamma} \times \Lambda_\beta(b), \qquad C_1(b) = \Lambda_\gamma \times \Lambda_\beta(b).$$

For a closed convex set C, write

$$d_A^2(v, C) = \inf_{\lambda \in C} \|v - \lambda\|_A^2 \quad \text{and} \quad \Pi_C^A(v) = \arg\min_{\lambda \in C} \|v - \lambda\|_A^2.$$

Then

$$\mathcal{L}_\infty(b, b) = \ell_b(V), \qquad \ell_b(v) = d_A^2(v, C_0(b)) - d_A^2(v, C_1(b)).$$

The covariance matrix of V is positive definite by Assumptions 3–4; hence, V has an everywhere-positive Lebesgue density.

Because $C_0(b) \subseteq C_1(b)$, ℓ_b is continuous and nonnegative, and $\ell_b(0) = 0$. It is also unbounded above. To see this, choose any nonzero $g \in \Lambda_\gamma$ and let $v_t = (tg', 0'_{d_\beta})'$ for $t \geq 0$. Then $v_t \in C_1(b)$, so $d_A^2(v_t, C_1(b)) = 0$. Partitioning A conformably with $(\gamma', \beta')'$ gives

$$\begin{aligned} \ell_b(v_t) &= \inf_{u \in \Lambda_\beta(b)} \begin{pmatrix} tg \\ -u \end{pmatrix}' A \begin{pmatrix} tg \\ -u \end{pmatrix} \\ &\geq t^2 g' \left(A_{\gamma\gamma} - A_{\gamma\beta} A_{\beta\beta}^{-1} A_{\beta\gamma} \right) g \longrightarrow \infty, \end{aligned}$$

where the Schur complement in parentheses is positive definite. Since $\mathbb{R}^{d_\gamma + d_\beta}$ is connected, the continuous image of this space under ℓ_b is an interval. The preceding observations imply that the range of ℓ_b is $[0, \infty)$.

It remains to rule out positive atoms. The KKT conditions for projection onto a polyhedron imply that $\Pi_{C_0(b)}^A$ and $\Pi_{C_1(b)}^A$ are affine on each cell of a finite polyhedral partition of $\mathbb{R}^{d_\gamma + d_\beta}$ corresponding to the finitely many possible active sets. Thus, ℓ_b is a quadratic polynomial on each full-dimensional cell. If, for some $t > 0$, the level set $\{v : \ell_b(v) = t\}$ had positive Lebesgue measure, then on at least one full-dimensional cell the relevant quadratic polynomial would equal t on a set of positive measure and hence would be identically equal to t on that cell. Its gradient would therefore vanish there. On the interior of any such cell,

$$\nabla \ell_b(v) = 2A \left(\Pi_{C_1(b)}^A(v) - \Pi_{C_0(b)}^A(v) \right).$$

Since A is nonsingular, a zero gradient implies that the two projections coincide, in which case $\ell_b(v) = 0$, a contradiction. The cell boundaries have Lebesgue measure zero, so

$\mathbb{P}(\mathcal{L}_\infty(b,b)=t)=0$ for every $t>0$.

Finally, for any $0 \le x < y < \infty$, the range result supplies a point v with $\ell_b(v) \in (x,y)$. By continuity, the inverse image of (x,y) contains a nonempty open set, which has positive probability because V has an everywhere-positive density. This proves the first two claims. If $q > F_b(0)$, the qth quantile lies in the continuous, strictly increasing part of F_b, which yields (A.14). $\square$

A.3 Proofs and lemmas related to linear regression example

Proof of Proposition 2.1

Starting with Assumption 1, note that Assumptions LinIID 1.1–3 and a weak LLN for row-wise i.i.d. random variables imply that $\|S_{xx}-\mathbb{E}_{\psi_n}[x_t x_t']\| \xrightarrow{p} 0$ and $\|S_{x\varepsilon}-\mathbb{E}_{\psi_n}[x_t\varepsilon_t]\| \xrightarrow{p} 0$ (all convergence statements that follow are thus understood to be under any sequence $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n \to \psi_0$ and (2.9)). Since x_t under ψ_n converges in distribution to x_t under ψ_0 and $\max_j E_\psi[|x_{j,t}|^{2+\nu}] \le c$ for all $\psi \in \Psi_0$, we have that $\mathbb{E}_{\psi_n}[x_t x_t'] \to \mathbb{E}_{\psi_0}[x_t x_t'] = \Omega_0$, such that $S_{xx} \xrightarrow{p} \Omega_0$. Likewise, $\mathbb{E}_\psi[x_t\varepsilon_t] = 0$ for all $\psi \in \Psi_0$, such that $S_{x\varepsilon} \xrightarrow{p} 0$. Clearly, these convergence properties imply that S_{xx} is invertible with probability approaching one, such that

$$\hat{\theta}_{LS} - \theta_n = S_{xx}^{-1} S_{x\varepsilon} = o_p(1), \tag{A.15}$$

and

$$\|\hat{\theta}_{LS}-\theta_n\|^2_{S_{xx}} = (\hat{\theta}_{LS}-\theta_n)' S_{xx} (\hat{\theta}_{LS}-\theta_n) = o_p(1).$$

We then have by the triangle inequality

$$\|\hat{\theta}_n-\theta_n\|_{S_{xx}} \le \|\hat{\theta}_n-\hat{\theta}_{LS}\|_{S_{xx}} + \|\hat{\theta}_{LS}-\theta_n\|_{S_{xx}} \le 2\|\hat{\theta}_{LS}-\theta_n\|_{S_{xx}},$$

where the second equality follows by noting that $\theta_n \in \Theta$ and $\|\hat{\theta}_n-\hat{\theta}_{LS}\|_{S_{xx}} = \min_{\theta\in\Theta} \|\theta-\hat{\theta}_{LS}\|_{S_{xx}}$. We conclude that $\|\hat{\theta}_n-\theta_n\|_{S_{xx}} = o_p(1)$, and hence that $\hat{\theta}_n-\theta_n = o_p(1)$. By similar arguments we have that $\tilde{\theta}_n-\theta_n = o_p(1)$.

Moving now to Assumption 2, 1. clearly holds by the definition of $L_n(\cdot)$ and 2. holds trivially since $-n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta' = S_{xx}$ in this example. Assumption 3 also holds by the fact that $-n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta' = S_{xx}$ and that $S_{xx} \xrightarrow{p} \Omega_0$. Finally, Assumption 4 holds by Assumptions LinIID 1.1–3, the Cramér Wold device and a Liapounov central limit theorem (CLT) for row-wise i.i.d. random variables since $n^{-1/2}\partial L_n(\theta_n)/\partial\theta = n^{1/2}S_{x\varepsilon}$ in this example. $\square$

Proof of Proposition 3.1

The proof that Assumption 5 holds is nearly identical to those of White (1980, proof of Theorem 1). For Assumption 6, note that under any sequence $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n \to \psi_0$ and (2.9),

$$\sqrt{n}(\hat{\theta}_{LS} - \theta_n) = S_{xx}^{-1}\sqrt{n}S_{x\varepsilon} \xrightarrow{d} N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1})$$

by the continuous mapping theorem and the fact that Assumptions 3–4 hold. Furthermore, given that Assumptions 3 and 5 hold, the continuous mapping theorem implies $\hat{\Sigma}_{\beta,n} \xrightarrow{p} \Sigma_\beta$. $\square$

Lemma A.5 *Suppose that S_{xx} is invertible. For any set $\Lambda \subset \mathbb{R}^{1+d_\beta}$,*

$$\arg\inf_{\theta\in\Lambda}\sum_{t=1}^{n}(y_t - x_t'\theta)^2 = \arg\inf_{\theta\in\Lambda}(\theta - \hat{\theta}_{LS})' S_{xx}(\theta - \hat{\theta}_{LS}).$$

Proof. After noting that for any θ,

$$\begin{aligned}
\sum_{t=1}^{n}(y_t - x_t'\theta)^2 &= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS} - x_t'(\theta - \hat{\theta}_{LS}))^2 \\
&= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS})^2 + \sum_{t=1}^{n}(x_t'(\theta - \hat{\theta}_{LS}))^2 \\
&\quad - 2(\theta - \hat{\theta}_{LS})'\underbrace{\sum_{t=1}^{n} x_t(y_t - x_t'\hat{\theta}_{LS})}_{=0} \\
&= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS})^2 + n(\theta - \hat{\theta}_{LS})' S_{xx}(\theta - \hat{\theta}_{LS}),
\end{aligned}$$

the result follows immediately. $\square$

A.4 Proofs related to the ARCH example

Below we prove that Assumptions 1-6 hold under Assumptions ARCH 1-4. Throughout, we make use of

$$W_t = (y_t^2, F_{t-1}')' \in \mathcal{W} = \mathbb{R}_+ \times \{1\} \times \mathbb{R}_+^{d_\beta + d_\gamma}.$$

Proof that Assumption 1 holds

Under any sequence $\{\psi_n\} \subset \Psi_0$ satisfying $\psi_n \to \psi_0$ and (2.9), and using the compactness of Θ, to prove the convergence of $\hat{\theta}_n$, it suffices to show that

$$\sup_{\theta \in \Theta} |n^{-1} L_n(\theta) - \mathcal{L}(\theta)| \xrightarrow{p} 0, \tag{A.16}$$

with

$$\mathcal{L}(\theta) = -\frac{1}{2}\mathbb{E}_{\psi_0}\left[\log \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right] \tag{A.17}$$

and

$$\mathcal{L}(\theta) \leq \mathcal{L}(\theta_0) \text{ for any } \theta \in \Theta \text{ with equality if and only if } \theta = \theta_0. \tag{A.18}$$

We start out by showing that (A.16) by applying Lemma 11.3 of Andrews and Cheng (2013)[6]. Recall that

$$l_t(\theta) = -\frac{1}{2}\left(\log \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right) = -\frac{1}{2}\left(\log g(F_{t-1}, \theta) + \frac{W_{t,1}}{g(F_{t-1}, \theta)}\right),$$

with $g(F_{t-1}, \theta) := F_{t-1}'\theta$ and $W_{t,1} = y_t^2$, the first entry of W_t. For any $w \in \mathcal{W}$, let w_1 denote the first entry of w and w_2 the column vector of the remaining entries, that is, $w = (w_1, w_2')'$. Let

$$s(w, \theta) = \log g(w, \theta) + \frac{w_1}{g(w_2, \theta)} = \log (w_2'\theta) + \frac{w_1}{w_2'\theta}.$$

For any $\theta_1, \theta_2 \in \Theta$, a mean value expansion gives

$$\log (w_2'\theta_1) = \log (w_2'\theta_2) + \frac{1}{w_2'\theta^*} w_2'(\theta_1 - \theta_2),$$

[6] A careful inspection of the proof of that lemma shows that only strong mixing conditions as the ones stated in Assumption ARCH 1 are needed. In particular, the proof makes use of a weak LLN for triangular arrays of strongly mixing processes, which does not impose any rate of decay on the mixing coefficients.

with $\theta^* \in \Theta$ between θ_1 and θ_2. Likewise,

$$\frac{w_1}{w_2'\theta_1} = \frac{w_1}{w_2'\theta_2} - \frac{w_1}{(w_2'\theta^{**})^2} w_2'(\theta_1 - \theta_2),$$

with $\theta^{**} \in \Theta$ between θ_1 and θ_2. It holds that $w_2'\theta \geq \delta_L$ uniformly on $\mathcal{W} \times \Theta$. Consequently, for all $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned} |s(w,\theta_1) - s(w,\theta_2)| &= \left| \frac{1}{w_2'\theta^*} w_2'(\theta_1 - \theta_2) - \frac{w_1}{(w_2'\theta^{**})^2} w_2'(\theta_1 - \theta_2) \right| \\ &\leq \left(\delta_L^{-1} + \delta_L^{-2} w_1\right) \|w_2\| \, \|\theta_1 - \theta_2\| \, . \end{aligned}$$

With $M_1(w) := \left(\delta_L^{-1} + \delta_L^{-2} w_1\right) \|w_2\|$, we conclude that for any $\eta > 0$

$$|s(w,\theta_1) - s(w,\theta_2)| \leq M_1(w)\eta \tag{A.19}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4, it holds that

$$\begin{aligned} \mathbb{E}_\psi[M_1(W_t)] &= \mathbb{E}_\psi\left[\left(\delta_L^{-1} + \delta_L^{-2} y_t^2\right) \|F_{t-1}\|\right] \\ &\leq \delta_L^{-1}\mathbb{E}_\psi\left[\|F_{t-1}\|\right] + \delta_L^{-2}\mathbb{E}_\psi\left[y_t^2 \|F_{t-1}\|\right] \leq \tilde{c} \end{aligned}$$

for some constant $\tilde{c} \in (0,\infty)$ for all $\psi \in \Psi_0$. Moreover, we have that for $\theta \in \Theta$

$$|s(W_t,\theta)| \leq |\log(\delta_L)| + \|F_{t-1}\| d_\theta(\delta_U + \beta_U + \gamma_U) + \frac{y_t^2}{\delta_L},$$

so using Assumption ARCH 4 again, we have that

$$\mathbb{E}_\psi[\sup_{\theta\in\Theta} |s(W_t,\theta)|^{1+\nu}] \leq \tilde{c}$$

for some constants $\nu, \tilde{c} \in (0,\infty)$ for all $\psi \in \Psi_0$. We conclude that

$$\mathbb{E}_\psi[\sup_{\theta\in\Theta} |s(W_t,\theta)|^{1+\nu}] + \mathbb{E}_\psi[M_1(W_t)] \leq \bar{C}, \tag{A.20}$$

for some constants $\nu, \tilde{c} \in (0,\infty)$ for all $\psi \in \Psi_0$. Using the fact that $l_t(\theta) = -s(W_t,\theta)/2$ together with (A.19) and (A.20), we have that (A.16) holds by Lemma 11.3 of Andrews and Cheng (2013). Condition (A.18) holds by standard arguments and Assumption 2. The properties (A.16)-(A.18) imply the convergence of $\hat{\theta}_n$. By identical arguments (under H_0) we can prove that $\tilde{\theta}_n$ converges, replacing Θ by Θ_{H_0}. $\square$

Proof that Assumption 2 holds

Note that

$$\begin{aligned}
&\sup_{\theta\in\Theta:\|\theta-\theta_n\|\le\epsilon_n}\left\|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}-n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'}\right\| \\
&\le 2\sup_{\theta\in\Theta}\left\|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}-\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]\right\| \\
&+\sup_{\theta\in\Theta:\|\theta-\theta_n\|\le\epsilon_n}\left\|\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]-\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_n)}{\partial\theta\partial\theta'}\right]\right\|.
\end{aligned}$$

Hence Assumption 2 holds provided

$$\sup_{\theta\in\Theta}\left\|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}-\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]\right\|=o_p(1), \tag{A.21}$$

and that $\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta)/\partial\theta\partial\theta^{'}]$ is continuous. Both conditions are shown by an application of Lemma 11.3 of Andrews and Cheng (2013), and the proof follows closely the arguments given in the previous proof. Note initially, that (A.21) holds provided that for any $i,j=1,\ldots,d_\theta$

$$\sup_{\theta\in\Theta}\left|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta_i\partial\theta_j}-\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right]\right|=o_p(1). \tag{A.22}$$

It holds that for any $i,j=1,\ldots,d_\theta$,

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}&=-\frac{1}{2}\left[2\frac{y_t^2}{\sigma_t^6(\theta)}-\frac{1}{\sigma_t^4(\theta)}\right]\left(\frac{\partial\sigma_t^2(\theta)}{\partial\theta_i}\right)\left(\frac{\partial\sigma_t^2(\theta)}{\partial\theta_j}\right)\\
&=-\frac{1}{2}\left[2\frac{y_t^2}{(F_{t-1}'\theta)^3}-\frac{1}{(F_{t-1}'\theta)^2}\right]F_{t-1,i}F_{t-1,j}=s_{ij}(W_t,\theta),
\end{aligned}$$

with

$$s_{ij}(w,\theta)=-\frac{1}{2}\left[2\frac{w_1}{(w_2'\theta)^3}-\frac{1}{(w_2'\theta)^2}\right]w_{2,i}w_{2,j},\quad w=(w_1,w_2')'\in\mathcal{W}.$$

For any $\theta_1,\theta_2\in\Theta$, a mean value expansion gives that

$$\frac{w_1}{(w_2'\theta_1)^3}=\frac{w_1}{(w_2'\theta_2)^3}-3\frac{w_1}{(w_2'\theta^*)^4}w_2'(\theta_1-\theta_2),$$

and

$$\frac{1}{(w_2'\theta_1)^2}=\frac{1}{(w_2'\theta_2)^2}-2\frac{1}{(w_2'\theta^{**})^3}w_2'(\theta_1-\theta_2)$$

with $\theta^*,\theta^{**}\in\Theta$ between θ_1 and θ_2. Consequently, for any $\theta_1,\theta_2\in\Theta$,

$$s_{ij}(w,\theta_1)-s_{ij}(w,\theta_2)=-\frac{1}{2}\left[2\frac{w_1}{(w_2'\theta_1)^3}-\frac{1}{(w_2'\theta_1)^2}-\left(2\frac{w_1}{(w_2'\theta_2)^3}-\frac{1}{(w_2'\theta_2)^2}\right)\right]w_{2,i}w_{2,j}$$

$$= \left[\frac{3w_1}{(w_2'\theta^*)^4} - \frac{1}{(w_2'\theta^{**})^3} \right] w_{2,i} w_{2,j} w_2'(\theta_1 - \theta_2),$$

such that

$$|s_{ij}(w,\theta_1) - s_{ij}(w,\theta_2)| \leq \underbrace{\left(\frac{3w_1}{\delta_L^4} + \frac{1}{\delta_L^3} \right) w_{2,i} w_{2,j} \|w_2\|}_{:= M_{ij}(w)} \|\theta_1 - \theta_2\|.$$

We conclude that for any $\eta > 0$

$$|s_{ij}(w,\theta_1) - s_{ij}(w,\theta_2)| \leq M_{ij}(w)\eta \tag{A.23}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4, it holds that

$$\begin{aligned} \mathbb{E}_\psi[M_{ij}(W_t)] &= \mathbb{E}_\psi \left[\left(\frac{3y_t^2}{\delta_L^4} + \frac{1}{\delta_L^3} \right) F_{t-1,i} F_{t-1,j} \|F_{t-1}\| \right] \\ &\leq \frac{3}{\delta_L^4} \mathbb{E}_\psi \left[y_t^2 \|F_{t-1}\|^3 \right] + \delta_L^{-3} \mathbb{E}_\psi \left[\|F_{t-1}\|^3 \right] \leq \tilde{c} \end{aligned}$$

for some constant $\tilde{c} \in (0,\infty)$ for all $\psi \in \Psi_0$. Moreover, we have that for $\theta \in \Theta$

$$|s_{ij}(W_t,\theta)| \leq \frac{1}{2} \left[2\frac{y_t^2}{\delta_L^3} + \frac{1}{\delta_L^2} \right] F_{t-1,i} F_{t-1,j},$$

so applying Assumption ARCH 4 again, we have that

$$\mathbb{E}_\psi \sup_{\theta\in\Theta} [|s_{ij}(W_t,\theta)|^{1+\nu}] \leq c^*$$

for some constants $\nu, c^* \in (0,\infty)$ for all $\psi \in \Psi_0$. We conclude that for any $i,j = 1,\ldots,d_\theta$,

$$\mathbb{E}_\psi \sup_{\theta\in\Theta} [|s_{ij}(W_t,\theta)|^{1+\nu}] + \mathbb{E}_\psi[M_{ij}(W_t)] \leq c, \tag{A.24}$$

for some constant $c \in (0,\infty)$ for all $\psi \in \Psi_0$. Using (A.23) and (A.24) together with Lemma 11.3 of Andrews and Cheng (2013), we have that (A.22) holds and, hence, that (A.21) holds. Moreover, this lemma ensures that $\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j]$ is uniformly continuous on Θ for all $\psi_0 \in \Psi_0$. □

Proof that Assumption 3 holds

Recall that $\Omega_0 = -\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta']$ and note that the matrix is positive definite for all $\psi_0 \in \Psi_0$ under Assumptions ARCH 1-4, by standard arguments. It holds that

$$\begin{aligned}\left\| -n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} - \Omega_0 \right\| &= \left\| n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} - \mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]\right\| \\ &\leq \sup_{\theta\in\Theta}\left\| n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - \mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]\right\| \\ &+ \left\|\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right] - \mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_n)}{\partial\theta\partial\theta'}\right]\right\|,\end{aligned}$$

where the first term is $o_p(1)$ for all $\psi_0 \in \Psi_0$, by the arguments given in the proof of Assumption *2*. Moreover, from that proof, it holds that the second term is $o(1)$ for all $\psi_0 \in \Psi_0$. □

Proof that Assumption 4 holds

For a given ψ_n, the (scaled) score is given by

$$S_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_n)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} -\frac{1}{2}(\varepsilon_t^2 - 1)\frac{1}{\sigma_t^2(\theta_n)}F_{t-1},$$

with F_t being $\mathcal{F}_{t,n}$-measurable. For any non-zero constant vector $k \in \mathbb{R}^{d_\theta}$, let $s_t = -k'F_{t-1}\sigma_t^{-2}(\theta_n)(\varepsilon_t^2 - 1)/2$, such that $k'S_n = n^{-1/2}\sum_{t=1}^n s_t$, and note that by Assumption ARCH 3, $\mathbb{E}_{\psi_n}[s_t|\mathcal{F}_{t-1,n}] = 0$ almost surely. The result follows by an application of the Lindeberg CLT for martingale difference arrays combined with an application of the Cramér-Wold Theorem. Note that by Assumptions ARCH 3 and ARCH 4,

$$\begin{aligned}\mathbb{E}_{\psi_n}[s_t^2] &= \mathbb{E}_{\psi_n}[(\varepsilon_t^2 - 1)^2/4]k'\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}']k \\ &= (\kappa/4)k'\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}']k.\end{aligned}$$

The convergence of parameters under the drifting sequence induces convergence in distribution of (F_{t-1}, θ_n) under ψ_n to (F_{t-1}, θ_0) under ψ_0. Consequently, by the continuous mapping theorem $\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}' = (\theta_n' F_{t-1})^{-2}F_{t-1}F_{t-1}'$ converges in distribution to $(\theta_0' F_{t-1})^{-2}F_{t-1}F_{t-1}'$ under ψ_0. Assumption ARCH 4 implies that $\|F_{t-1}\sigma_t^{-2}(\theta_n)\|^2 \leq \delta_L^{-2}\|F_{t-1}\|^2$ is uniformly integrable, and consequently, we have that $\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}'] \to$

$\mathbb{E}_{\psi_0}[\sigma_t^{-4}(\theta_0)F_{t-1}F'_{t-1}]$. Hence,

$$\mathbb{E}_{\psi_n}[s_t^2] \to k'\Sigma_0 k,$$

with

$$\Sigma_0 := (\kappa/4)\,\mathbb{E}_{\psi_0}[\sigma_t^{-4}(\theta_0)F_{t-1}F'_{t-1}].$$

The matrix Σ_0 is positive definite by standard arguments and Assumption ARCH 2. It remains to show that for any constant $\epsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^{n} s_t^2 \mathbb{I}(s_t^2 > \sqrt{n}\epsilon) \xrightarrow{p} 0, \tag{A.25}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[s_t^2|\mathcal{F}_{t-1,n}\right] - \mathbb{E}_{\psi_n}[s_t^2] \xrightarrow{p} 0. \tag{A.26}$$

To show (A.25), note that by Assumptions ARCH 3 and ARCH 4 and Minkowski's inequality there exist constants a $\upsilon, c \in (0,\infty)$ (with c depending on k) such that for any $\eta > 0$,

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\frac{1}{n}\sum_{t=1}^{n} s_t^2\mathbb{I}(s_t^2 > \sqrt{n}\epsilon) > \eta\right) \leq \eta^{-1}\mathbb{E}_{\psi_n}\left[s_t^2\mathbb{I}(s_t^2 > \sqrt{n}\epsilon)\right] \\
&\leq \frac{1}{\eta\left(\sqrt{n}\epsilon\right)^{\upsilon}}\mathbb{E}_{\psi_n}\left[|s_t|^{2+\upsilon}\right] = \frac{1}{\eta\left(\sqrt{n}\epsilon\right)^{\upsilon}}\mathbb{E}_{\psi_n}\left[|(\varepsilon_t^2-1)\sigma_t^{-2}(\theta_n)k'F_{t-1}/2|^{2+\upsilon}\right] \\
&= \frac{1}{2^{2+\upsilon}\eta\left(\sqrt{n}\epsilon\right)^{\upsilon}}\mathbb{E}_{\psi_n}\left[|\varepsilon_t^2-1|^{2+\upsilon}\right]\mathbb{E}_{\psi_n}\left[|\sigma_t^{-2}(\theta_n)k'F_{t-1}|^{2+\upsilon}\right] \\
&\leq \frac{1}{2^{2+\upsilon}\eta\left(\sqrt{n}\epsilon\right)^{\upsilon}}\mathbb{E}_{\psi_n}\left[|\varepsilon_t^2-1|^{2+\upsilon}\right]\delta_L^{-(2+\upsilon)}\left(\sum_{i=1}^{d_\theta}|k_i|(\mathbb{E}_{\psi_n}\left[\|F_{t-1}\|^{2+\upsilon}\right])^{1/(2+\upsilon)}\right)^{2+\upsilon} \\
&\leq \frac{1}{2^{2+\upsilon}\eta\left(\sqrt{n}\epsilon\right)^{\upsilon}}c \to 0,
\end{aligned}$$

and we conclude that (A.25) holds. The convergence in (A.26) follows by an application of the (weak) LLN for row-wise stationary and strongly mixing triangular arrays (Andrews, 1988, p. 462), using that $\mathbb{E}_{\psi_n}[s_t^2|\mathcal{F}_{t-1,n}] = (\kappa/4)(\sigma_t^{-2}(\theta_n)k'F_{t-1})^2$ is uniformly integrable under Assumption ARCH 4. $\square$

Proof that Assumption 5 holds

First, note that by Assumptions 1–3, $\hat{\Omega}_n \xrightarrow{p} \Omega_0$. Consequently, it remains to show that

$$\hat{\kappa}_n \xrightarrow{p} \kappa. \tag{A.27}$$

We have that

$$\hat{\kappa}_n = n^{-1} \sum_{t=1}^{n} \left(\frac{y_t^4}{(\hat{\theta}_n' F_{t-1})^2} - 1 \right).$$

Using the same notation as before, let $f : \mathcal{W} \times \Theta \to \mathbb{R}$ be given by

$$f(w, \theta) = \frac{w_1^2}{(\theta' w_2)^2},$$

such that $\hat{\kappa}_n = n^{-1} \sum_{t=1}^{n} f(W_t, \hat{\theta}_n)$. With $\theta_1, \theta_2 \in \Theta$, a mean-value expansion gives that

$$f(w, \theta_1) - f(w, \theta_2) = -2 \frac{w_1^2}{(\theta^{*\prime} w_2)^3} w_2' (\theta_1 - \theta_2).$$

It holds that

$$| -2 \frac{w_1^2}{(\theta^{*\prime} w_2)^3} w_2' (\theta_1 - \theta_2)| \leq \frac{2 w_1^2}{\delta_L^3} \|w_2\| \|\theta_1 - \theta_2\|$$

such that with $M : \mathcal{W} \to \mathbb{R}$ given by

$$M(w) = \frac{2 w_1^2}{\delta_L^3} \|w_2\|,$$

$$|f(w, \theta_1) - f(w, \theta_2)| \leq M(w) \|\theta_1 - \theta_2\|.$$

Consequently, for any $\eta > 0$

$$|f(w, \theta_1) - f(w, \theta_2)| \leq M(w) \eta \tag{A.28}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4,

$$\mathbb{E}_\psi [M(W_t)] = \frac{2}{\delta_L^3} \mathbb{E}_\psi \left[y_t^2 \|F_{t-1}\| \right] \leq \tilde{c},$$

for some constant $\tilde{c} \in (0, \infty)$ for all $\psi \in \Psi_0$. Likewise, noting that for any $\theta \in \Theta$, $|f(W_t, \theta)| \leq y_t^4 / \delta_L^2$, we have by Assumption ARCH 4,

$$\mathbb{E}_\psi \left[\sup_{\theta \in \Theta} |f(W_t, \theta)|^{1+\varepsilon} \right] \leq c^*$$

for some constants $\varepsilon, c^* \in (0, \infty)$ for all $\psi \in \Psi_0$. Consequently,

$$\mathbb{E}_\psi \left[\sup_{\theta \in \Theta} |f(W_t, \theta)|^{1+\varepsilon} \right] + \mathbb{E}_\psi [M(W_t)] \leq \bar{c}, \tag{A.29}$$

for some constant $\bar{c} \in (0, \infty)$ for all $\psi \in \Psi_0$. Using (A.28) and (A.29) together with Lemma 11.3 of Andrews and Cheng (2013), we have that

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{t=1}^{n} f(W_t, \theta) - \mathbb{E}_{\psi_0}[f(W_t, \theta)] \right| = o_p(1)$$

and that $\mathbb{E}_{\psi_0}[f(W_t, \theta)]$ is uniformly continuous on Θ for all $\psi_0 \in \Psi_0$. Using that $\hat{\theta}_n - \theta_0 = o_p(1)$, we then have that

$$n^{-1} \sum_{t=1}^{n} f(W_t, \hat{\theta}_n) - \mathbb{E}_{\psi_0}[f(W_t, \theta_0)] = o_p(1),$$

or, equivalently, (A.27) holds. □

Proof that Assumption 6 holds

As $\check{\beta}_n$ is given in terms of the Newton-Raphson estimator, Assumption 6 holds by Lemma 3.2. □

B The conditional likelihood ratio test

This section outlines the CLR test of Ketz (2018), adapted to our setting. Recall from Section 2 that the parameter vector is $\theta = (\gamma', \beta', \delta)'$, and the parameter space $\Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$ satisfies $\Theta_\gamma = [\gamma_L, \gamma_U]^s \times [0, \gamma_U]^{d_\gamma - s}$, $\Theta_\beta = [0, \beta_U]^{d_\beta}$, with $-\infty \leq \gamma_L < 0 < \gamma_U \leq \infty$, $0 < \beta_U \leq \infty$, and $\Theta_\delta \subset (-\infty, \infty)^{d_\delta}$ is compact. We test $\mathsf{H}_0 : \gamma = 0$ against $\gamma \neq 0$, $\gamma \in \Theta_\gamma$. The construction of the CLR test proceeds in three steps: (i) the Newton–Raphson one-step estimator and its orthogonal decomposition, (ii) the CLR statistic, and (iii) the conditional CV.

Step 1: One-step estimator and projection residual. Let $\hat{\theta}_n$ denote the estimator given in (2.2). Let $\hat{V}_n = \hat{\Omega}_n^{-1} \hat{\Sigma}_n \hat{\Omega}_n^{-1}$ denote the sandwich covariance estimator, with $\hat{\Omega}_n$ and $\hat{\Sigma}_n$ estimators of the matrices Ω_0 and Σ_0 defined in Assumptions 3 and 4, respectively. Partition the sub-block of $\hat{V}_n$ corresponding to the parameters $(\gamma', \beta')'$ as

$$\hat{V}_{n,S} = \begin{bmatrix} \hat{V}_{n,\gamma\gamma} & \hat{V}_{n,\gamma\beta} \\ \hat{V}_{n,\beta\gamma} & \hat{V}_{n,\beta\beta} \end{bmatrix},$$

where $\hat{V}_{n,\gamma\gamma}$ is $d_\gamma \times d_\gamma$ and $\hat{V}_{n,\beta\beta}$ is $d_\beta \times d_\beta$.

Recall that Newton–Raphson one-step estimator starting from $\hat{\theta}_n$ is

$$\check{\theta}_n = \hat{\theta}_n - \left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial L_n(\hat{\theta}_n)}{\partial\theta}.$$

Write $\check{\theta}_n = (\check{\gamma}_n', \check{\beta}_n', \check{\delta}_n)'$. The key idea of Ketz (2018) is to decompose $\check{\theta}_n$ into two asymptotically independent components: the subvector $\check{\gamma}_n$ pertaining to the parameter of interest, and the projection residual

$$X_n = \check{\beta}_n - \hat{V}_{n,\beta\gamma}\hat{V}_{n,\gamma\gamma}^{-1}\check{\gamma}_n, \tag{B.1}$$

which is asymptotically sufficient for the nuisance parameter β and asymptotically independent of $\check{\gamma}_n$. Conditioning on X_n renders the distribution of the CLR statistic approximately free of the unknown β.

Step 2: The CLR statistic. Following Ketz (2018), define the extended parameter spaces G^e and B^e as the Cartesian products obtained by replacing each component interval of Θ_γ and Θ_β, respectively, with $[0,\infty)$ if the component has a boundary at zero, or $(-\infty,\infty)$ if it is unrestricted. For example, if $\Theta_\gamma = [0,\gamma_U] \times [\gamma_L,\gamma_U]$ then $G^e = [0,\infty) \times (-\infty,\infty)$.

Define

$$\beta_n(\gamma) = X_n + \hat{V}_{n,\beta\gamma}\hat{V}_{n,\gamma\gamma}^{-1}\gamma,$$

which reconstructs the implied β estimate from X_n and a given γ. The CLR statistic is

$$\begin{aligned}&\mathrm{CLR}(\gamma_0, \check{\gamma}_n, X_n, n^{-1}\hat{V}_{n,S}, G^e, B^e)\\ &= \inf_{\beta\in B^e}\left\|\begin{pmatrix}\check{\gamma}_n-\gamma_0\\ \beta_n(\check{\gamma}_n)-\beta\end{pmatrix}\right\|^2_{(n^{-1}\hat{V}_{n,S})^{-1}} - \inf_{\gamma\in G^e,\, \beta\in B^e}\left\|\begin{pmatrix}\check{\gamma}_n-\gamma\\ \beta_n(\check{\gamma}_n)-\beta\end{pmatrix}\right\|^2_{(n^{-1}\hat{V}_{n,S})^{-1}},\end{aligned} \tag{B.2}$$

where $\gamma_0 = 0$ under H_0 and $\|v\|_A^2 = v'Av$. Note that $\beta_n(\check{\gamma}_n) = \check{\beta}_n$ by (B.1), so the statistic in (B.2) is evaluated at the one-step estimates $(\check{\gamma}_n', \check{\beta}_n')'$. The infima are taken over the extended spaces G^e and B^e rather than the original compact parameter spaces Θ_γ and Θ_β; this ensures that only boundaries at zero are taken into account. Both infima are quadratic programs. Note that the CLR statistic uses the sandwich covariance $\hat{V}_{n,S}$ rather than the LR statistic in (2.3), and in particular does not require the computation of the restricted estimator $\tilde{\theta}_n$.

Step 3: Conditional critical value. Conditional on X_n and $\hat{V}_{n,S}$, the CV $\text{cv}_{1-\alpha}$ is the $(1-\alpha)$-quantile of

$$c_n = \text{CLR}\Big(\gamma_0, \left(n^{-1}\hat{V}_{n,\gamma\gamma}\right)^{1/2} Z + \gamma_0,\ X_n,\ n^{-1}\hat{V}_{n,S},\ G^e, B^e\Big), \tag{B.3}$$

where $Z|X_n, \hat{V}_{n,S} \sim N(0, I_{d_\gamma})$, and where $\check{\gamma}_n$ in (B.2) is replaced by the simulated draw $(n^{-1}\hat{V}_{n,\gamma\gamma})^{1/2}Z + \gamma_0$ and $\check{\beta}_n$ by $\beta_n\Big((n^{-1}\hat{V}_{n,\gamma\gamma})^{1/2}Z + \gamma_0\Big)$. That is, (B.3) replaces the observed one-step estimates with a draw from the (conditional) null distribution, while holding X_n and $\hat{V}_{n,S}$ fixed. The quantile is computed by simulation: draw $Z^{(1)}, \ldots, Z^{(B)}$ independently from $N(0, I_{d_\gamma})$, compute $c_n^{(j)}$ for each draw, and set $\text{cv}_{1-\alpha}$ to the empirical $(1-\alpha)$-quantile of $\{c_n^{(1)}, \ldots, c_n^{(B)}\}$. In the simulations and the empirical application, we use $B = 10{,}000$.

References

Andrews, D.W.K. (1988): "Laws of large numbers for dependent non-identically distributed random variables", *Econometric Theory*, vol. 4, 458–467.

Andrews, D.W.K. (1999): "Estimation when a parameter is on a boundary", *Econometrica*, vol. 67, 1341–1383.

Andrews, D.W.K. and Cheng, X. (2013),"Maximum likelihood estimation and uniform inference with sporadic identification failure", *Journal of Econometrics*, vol. 173, 36–56, Supplemental Appendices.

Andrews, D.W.K. and Guggenberger, P. (2009): "Hybrid and size-corrected subsampling methods", *Econometrica*, vol. 77, 721–762.

Andrews, D.W.K. and Guggenberger, P. (2010): "Asymptotic size and a problem with subsampling and with the m out of n bootstrap", *Econometric Theory*, vol. 26, 426–468.

McCloskey, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems", *Journal of Econometrics*, vol. 200, 17–35.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*. John Wiley & Sons, Hoboken, NJ.

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

γ	$\beta = 0$ Std	CLR	Unif	LF	$\beta = 0.1$ Std	CLR	Unif	LF
				$\rho_{12} = -0.95$				
0	0.0780	0.0363	0.0330	0.0568	0.0498	0.0372	0.0539	0.0362
0.01	0.4387	0.2173	0.2900	0.3663	0.1713	0.1330	0.1671	0.1354
0.05	0.9872	0.8032	0.9625	0.9734	0.6682	0.6058	0.5665	0.6008
0.1	0.9997	0.9165	0.9953	0.9959	0.8997	0.8670	0.8257	0.8670
0.25	1.0000	0.9578	0.9822	0.9825	0.9957	0.9891	0.9890	0.9932
				$\rho_{12} = -0.75$				
0	0.0706	0.0362	0.0300	0.0545	0.0460	0.0377	0.0496	0.0396
0.01	0.4326	0.2604	0.2842	0.3772	0.1900	0.1582	0.1844	0.1626
0.05	0.9876	0.8957	0.9652	0.9771	0.7640	0.7230	0.6918	0.7225
0.1	0.9997	0.9703	0.9970	0.9974	0.9553	0.9440	0.9158	0.9433
0.25	1.0000	0.9840	0.9915	0.9916		0.9976	0.9963	0.9980
				$\rho_{12} = -0.5$				
0	0.0613	0.0368	0.0264	0.0549	0.0466	0.0374	0.0470	0.0426
0.01	0.4195	0.2948	0.2776	0.3824	0.2029	0.1705	0.1966	0.1874
0.05	0.9869	0.9514	0.9673	0.9826	0.8229	0.7892	0.7779	0.8005
0.1	0.9998	0.9917	0.9981	0.9991	0.9767	0.9711	0.9605	0.9728
0.25	1.0000	0.9952	0.9966	0.9970	0.9996	0.9994	0.9987	0.9994
				$\rho_{12} = 0$				
0	0.0480	0.0384	0.0209	0.0531	0.0478	0.0384	0.0466	0.0526
0.01	0.3818	0.3350	0.2460	0.3894	0.1918	0.1606	0.1817	0.2017
0.05	0.9868	0.9810	0.9674	0.9861	0.8328	0.7995	0.8023	0.8348
0.1	0.9996	0.9994	0.9983	0.9997	0.9849	0.9800	0.9770	0.9841
0.25	1.0000	0.9996	0.9992	0.9998	0.9998	0.9998	0.9997	0.9998
				$\rho_{12} = 0.5$				
0	0.0297	0.0523	0.0125	0.0320	0.0484	0.0420	0.0397	0.0537
0.01	0.2772	0.3027	0.1660	0.2885	0.1420	0.1222	0.1175	0.1506
0.05	0.9608	0.9548	0.9189	0.9619	0.6943	0.6586	0.6400	0.7018
0.1	0.9985	0.9981	0.9945	0.9985	0.9470	0.9349	0.9247	0.9491
0.25	1.0000	1.0000	0.9999	1.0000	0.9995	0.9987	0.9983	0.9992
				$\rho_{12} = 0.75$				
0	0.0160	0.0605	0.0073	0.0178	0.0504	0.0463	0.0250	0.0552
0.01	0.1734	0.2358	0.0890	0.1801	0.1080	0.0978	0.0590	0.1155
0.05	0.8602	0.8504	0.7554	0.8657	0.4859	0.4552	0.3613	0.4950
0.1	0.9792	0.9743	0.9526	0.9787	0.8137	0.7915	0.7248	0.8180
0.25	0.9990	0.9990	0.9973	0.9992	0.9876	0.9845	0.9760	0.9872
				$\rho_{12} = 0.95$				
0	0.0026	0.0648	0.0017	0.0019	0.0509	0.0527	0.0195	0.0547
0.01	0.0280	0.1252	0.0135	0.0285	0.0733	0.0750	0.0278	0.0793
0.05	0.3522	0.3868	0.1821	0.3620	0.1820	0.1842	0.0940	0.1932
0.1	0.5853	0.5846	0.4068	0.5926	0.3304	0.3262	0.1953	0.3388
0.25	0.8057	0.7996	0.6683	0.8086	0.6180	0.6142	0.4554	0.6267
Time (s)	3.22e-05	0.209	0.184	5.232				

Table 1: Rejection frequencies corresponding to the ARCH model with a one-dimensional parameter of interest under the null hypothesis, $\gamma = 0$, and various alternatives, $\gamma > 0$, for $n = 1000$ ($\alpha = 0.05$) and the standard LR test (Std), the CLR test, the uniform LR test (Unif) and the least favorable test (LF). The last row reports average computation time per replication (seconds) across all configurations.

γ	$\beta = 0.25$ Std	CLR	Unif	LF	$\beta = 0.5$ Std	CLR	Unif	LF
				$\rho_{12} = -0.95$				
0	0.0446	0.0436	0.0511	0.0329	0.0285	0.0506	0.0392	0.0219
0.01	0.1276	0.1042	0.1334	0.0959	0.0804	0.0938	0.0921	0.0613
0.05	0.4792	0.4129	0.4698	0.4154	0.3292	0.2840	0.3369	0.2740
0.1	0.7459	0.6909	0.6978	0.6869	0.5637	0.5015	0.5602	0.4986
0.25	0.9580	0.9388	0.9180	0.9377	0.8673	0.8321	0.8264	0.8279
				$\rho_{12} = -0.75$				
0	0.0458	0.0410	0.0507	0.0405	0.0382	0.0481	0.0447	0.0329
0.01	0.1371	0.1147	0.1437	0.1168	0.0962	0.0939	0.1032	0.0814
0.05	0.5668	0.5153	0.5591	0.5192	0.3873	0.3440	0.3939	0.3477
0.1	0.8413	0.8102	0.8198	0.8092	0.6683	0.6169	0.6616	0.6227
0.25	0.9870	0.9817	0.9748	0.9807	0.9419	0.9278	0.9287	0.9262
				$\rho_{12} = -0.5$				
0	0.0464	0.0385	0.0492	0.0437	0.0418	0.0410	0.0461	0.0401
0.01	0.1394	0.1160	0.1446	0.1309	0.0965	0.0874	0.1028	0.0923
0.05	0.6168	0.5663	0.6070	0.5898	0.4166	0.3720	0.4226	0.3970
0.1	0.8948	0.8673	0.8811	0.8765	0.7235	0.6810	0.7205	0.7026
0.25	0.9952	0.9930	0.9915	0.9933	0.9713	0.9624	0.9668	0.9659
				$\rho_{12} = 0$				
0	0.0462	0.0377	0.0486	0.0521	0.0454	0.0374	0.0459	0.0508
0.01	0.1231	0.1013	0.1246	0.1349	0.0878	0.0745	0.0916	0.0979
0.05	0.5915	0.5428	0.5873	0.6028	0.3739	0.3312	0.3733	0.3869
0.1	0.8989	0.8747	0.8932	0.9023	0.7011	0.6542	0.6955	0.7097
0.25	0.9969	0.9960	0.9963	0.9970	0.9758	0.9683	0.9739	0.9768
				$\rho_{12} = 0.5$				
0	0.0480	0.0403	0.0481	0.0536	0.0463	0.0399	0.0469	0.0517
0.01	0.0948	0.0812	0.0933	0.1029	0.0739	0.0622	0.0722	0.0804
0.05	0.4160	0.3765	0.4055	0.4292	0.2395	0.2119	0.2363	0.2533
0.1	0.7666	0.7337	0.7533	0.7754	0.5003	0.4601	0.4940	0.5143
0.25	0.9881	0.9837	0.9850	0.9870	0.9149	0.8972	0.9078	0.9172
				$\rho_{12} = 0.75$				
0	0.0501	0.0446	0.0404	0.0535	0.0486	0.0444	0.0450	0.0508
0.01	0.0798	0.0703	0.0646	0.0853	0.0636	0.0570	0.0597	0.0704
0.05	0.2664	0.2431	0.2325	0.2775	0.1583	0.1444	0.1498	0.1665
0.1	0.5409	0.5131	0.4952	0.5508	0.3164	0.2945	0.3014	0.3267
0.25	0.9209	0.9091	0.8943	0.9210	0.7351	0.7124	0.7169	0.7424
				$\rho_{12} = 0.95$				
0	0.0517	0.0519	0.0193	0.0551	0.0514	0.0517	0.0190	0.0552
0.01	0.0633	0.0634	0.0243	0.0677	0.0586	0.0584	0.0214	0.0622
0.05	0.1162	0.1164	0.0547	0.1231	0.0887	0.0864	0.0359	0.0926
0.1	0.1966	0.1947	0.1031	0.2084	0.1277	0.1275	0.0615	0.1346
0.25	0.4281	0.4277	0.2780	0.4401	0.2768	0.2736	0.1534	0.2861

Table 2: Rejection frequencies corresponding to the ARCH model with a one-dimensional parameter of interest under the null hypothesis, $\gamma = 0$, and various alternatives, $\gamma > 0$, for $n = 1000$ ($\alpha = 0.05$) and the standard LR test (Std), the CLR test, the uniform LR test (Unif) and the least favorable test (LF).

γ_1	γ_2	β	C_1 Std	C_1 CLR	C_1 Unif	C_2 Std	C_2 CLR	C_2 Unif	C_3 Std	C_3 CLR	C_3 Unif
0	0	0	0.0530	0.0480	0.0269	0.0712	0.0464	0.0381	0.0693	0.0435	0.0385
		0.1	0.0530	0.0488	0.0337	0.0432	0.0490	0.0261	0.0457	0.0453	0.0294
		0.25	0.0452	0.0471	0.0347	0.0309	0.0500	0.0203	0.0308	0.0477	0.0238
		0.5	0.0384	0.0472	0.0316	0.0221	0.0510	0.0160	0.0224	0.0503	0.0175
0.05	0	0	0.8418	0.8165	0.7585	0.8566	0.5852	0.7804	0.8371	0.6993	0.7543
		0.1	0.4703	0.4395	0.3643	0.3025	0.2820	0.2075	0.3768	0.3495	0.2820
		0.25	0.2405	0.2289	0.1877	0.1467	0.1626	0.1045	0.1707	0.1817	0.1289
		0.5	0.1247	0.1312	0.1033	0.0696	0.1101	0.0531	0.0731	0.1104	0.0594
0	0.05	0	0.9421	0.9239	0.8993	0.9465	0.9176	0.9072	0.9356	0.8910	0.8925
		0.1	0.6197	0.5832	0.5180	0.6162	0.5810	0.5015	0.5814	0.5337	0.4802
		0.25	0.3284	0.3009	0.2687	0.3151	0.3028	0.2359	0.2869	0.2701	0.2206
		0.5	0.1634	0.1618	0.1364	0.1442	0.1628	0.1035	0.1227	0.1426	0.0953
0.05	0.05	0	0.9560	0.9450	0.9199	0.9617	0.8810	0.9268	0.7635	0.6082	0.6708
		0.1	0.7284	0.7011	0.6220	0.6454	0.6083	0.5276	0.3809	0.3482	0.2844
		0.25	0.4446	0.4225	0.3693	0.3737	0.3646	0.2823	0.2067	0.2060	0.1487
		0.5	0.2441	0.2401	0.2007	0.1897	0.2135	0.1409	0.1032	0.1343	0.0813
0.1	0	0	0.9909	0.9865	0.9784	0.9926	0.8765	0.9829	0.9916	0.9547	0.9786
		0.1	0.8465	0.8249	0.7715	0.6560	0.5765	0.5440	0.7548	0.7110	0.6598
		0.25	0.5494	0.5240	0.4638	0.3415	0.3370	0.2488	0.4417	0.4254	0.3543
		0.5	0.2847	0.2797	0.2385	0.1637	0.1956	0.1257	0.1980	0.2260	0.1585
0	0.1	0	0.9991	0.9986	0.9976	0.9992	0.9982	0.9978	0.9983	0.9958	0.9958
		0.1	0.9454	0.9296	0.9075	0.9479	0.9305	0.9060	0.9344	0.9098	0.8893
		0.25	0.7099	0.6758	0.6324	0.7062	0.6735	0.5991	0.6712	0.6322	0.5810
		0.5	0.3962	0.3770	0.3428	0.3773	0.3754	0.2954	0.3474	0.3407	0.2812
0.1	0.1	0	0.9975	0.9958	0.9939	0.9981	0.9823	0.9946	0.9362	0.8321	0.8885
		0.1	0.9577	0.9503	0.9227	0.9267	0.8919	0.8709	0.6695	0.6150	0.5670
		0.25	0.7975	0.7764	0.7176	0.7052	0.6780	0.5975	0.4218	0.4003	0.3232
		0.5	0.5189	0.5006	0.4536	0.4386	0.4360	0.3434	0.2328	0.2451	0.1778
0.25	0	0	1.0000	1.0000	0.9998	1.0000	0.9874	0.9995	1.0000	0.9991	0.9992
		0.1	0.9983	0.9966	0.9947	0.9876	0.9340	0.9724	0.9941	0.9833	0.9846
		0.25	0.9589	0.9509	0.9315	0.8132	0.7502	0.7200	0.9054	0.8799	0.8447
		0.5	0.7709	0.7518	0.7068	0.5133	0.5009	0.4029	0.6547	0.6343	0.5614
0	0.25	0	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.9999	0.9996
		0.1	1.0000	0.9999	0.9998	0.9999	0.9999	0.9994	0.9997	0.9995	0.9988
		0.25	0.9915	0.9882	0.9838	0.9914	0.9880	0.9807	0.9886	0.9834	0.9770
		0.5	0.8978	0.8802	0.8608	0.8961	0.8816	0.8356	0.8800	0.8531	0.8216
0.25	0.25	0	1.0000	1.0000	1.0000	1.0000	0.9993	0.9999	0.9899	0.9495	0.9779
		0.1	0.9993	0.9989	0.9976	0.9984	0.9925	0.9958	0.9361	0.8800	0.8903
		0.25	0.9896	0.9874	0.9795	0.9777	0.9612	0.9533	0.7873	0.7420	0.6970
		0.5	0.9264	0.9157	0.8822	0.8630	0.8431	0.7840	0.5705	0.5491	0.4687

Table 3: Rejection frequencies corresponding to the ARCH model with a two-dimensional parameter of interest under the null hypothesis, $\gamma_1 = \gamma_2 = 0$, and various alternatives, $\max\{\gamma_1, \gamma_2\} > 0$, for $n = 1000$ ($\alpha = 0.05$) and the standard LR test (Std), the CLR test and the uniform LR test (Unif). Left: C_1. Middle: C_2. Right: C_3.

(γ_1, γ_2)	Σ	Std	CLR	η_1	η_2	η_3	η_4
			Null: $\gamma_1 = \gamma_2 = 0$				
$(0,0)$	Σ_1	0.089	0.049	0.045	0.042	0.025	0.006
$(0,0)$	Σ_2	0.070	0.049	0.042	0.040	0.022	0.006
$(0,0)$	Σ_3	0.081	0.049	0.044	0.042	0.023	0.006
$(0,0)$	Σ_4	0.098	0.050	0.043	0.041	0.025	0.006
$(0,0)$	Σ_5	0.076	0.048	0.043	0.041	0.022	0.006
$(0,0)$	Σ_6	0.036	0.047	0.025	0.025	0.013	0.003
			Alternative: $(\gamma_1, \gamma_2) = (0, 0.1)$				
$(0,0.1)$	Σ_1	0.962	0.645	0.918	0.914	0.866	0.708
$(0,0.1)$	Σ_2	1.000	0.989	1.000	1.000	1.000	1.000
$(0,0.1)$	Σ_3	0.992	0.663	0.985	0.984	0.969	0.897
$(0,0.1)$	Σ_4	1.000	0.987	0.997	0.997	0.993	0.972
$(0,0.1)$	Σ_5	0.999	0.752	0.998	0.998	0.995	0.981
$(0,0.1)$	Σ_6	0.994	0.997	0.992	0.991	0.981	0.934
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0)$				
$(0.1,0)$	Σ_1	0.960	0.643	0.916	0.913	0.870	0.703
$(0.1,0)$	Σ_2	1.000	0.986	1.000	1.000	1.000	1.000
$(0.1,0)$	Σ_3	0.993	0.666	0.983	0.981	0.967	0.900
$(0.1,0)$	Σ_4	0.999	0.986	0.997	0.997	0.993	0.970
$(0.1,0)$	Σ_5	0.999	0.754	0.998	0.998	0.996	0.977
$(0.1,0)$	Σ_6	0.998	0.998	0.998	0.998	0.996	0.980
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0.1)$				
$(0.1,0.1)$	Σ_1	1.000	0.838	1.000	1.000	1.000	0.999
$(0.1,0.1)$	Σ_2	1.000	1.000	1.000	1.000	1.000	1.000
$(0.1,0.1)$	Σ_3	1.000	0.977	1.000	1.000	1.000	1.000
$(0.1,0.1)$	Σ_4	1.000	0.698	0.998	0.998	0.995	0.981
$(0.1,0.1)$	Σ_5	1.000	0.996	1.000	1.000	1.000	1.000
$(0.1,0.1)$	Σ_6	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Rejection frequencies corresponding to the multivariate location model with a two-dimensional parameter of interest $\gamma \geq 0$ under the null hypothesis and various alternatives for $n = 500$ ($\alpha = 0.05$) and $\beta = 0$. The columns Std and CLR correspond to the standard LR test and the CLR test and the columns $\eta_1, \ldots, \eta_4$ correspond to the uniform LR test with $\eta = \alpha/20, \alpha/10, \alpha/2, 9\alpha/10$, respectively.

(γ_1,γ_2)	Σ	Std	CLR	η_1	η_2	η_3	η_4
			Null: $\gamma_1=\gamma_2=0$				
$(0,0)$	Σ_1	0.057	0.049	0.040	0.040	0.025	0.006
$(0,0)$	Σ_2	0.038	0.049	0.044	0.043	0.023	0.005
$(0,0)$	Σ_3	0.050	0.049	0.041	0.041	0.025	0.006
$(0,0)$	Σ_4	0.063	0.050	0.037	0.037	0.022	0.006
$(0,0)$	Σ_5	0.045	0.048	0.042	0.041	0.025	0.005
$(0,0)$	Σ_6	0.031	0.049	0.027	0.027	0.015	0.003
			Alternative: $(\gamma_1,\gamma_2)=(0,0.1)$				
$(0,0.1)$	Σ_1	0.678	0.650	0.551	0.548	0.480	0.286
$(0,0.1)$	Σ_2	1.000	0.988	1.000	1.000	1.000	0.996
$(0,0.1)$	Σ_3	0.669	0.664	0.553	0.548	0.488	0.302
$(0,0.1)$	Σ_4	0.994	0.990	0.980	0.979	0.965	0.906
$(0,0.1)$	Σ_5	0.752	0.750	0.650	0.642	0.575	0.390
$(0,0.1)$	Σ_6	0.994	0.997	0.993	0.993	0.988	0.953
			Alternative: $(\gamma_1,\gamma_2)=(0.1,0)$				
$(0.1,0)$	Σ_1	0.675	0.647	0.541	0.538	0.472	0.275
$(0.1,0)$	Σ_2	1.000	0.986	1.000	1.000	1.000	0.995
$(0.1,0)$	Σ_3	0.673	0.667	0.550	0.544	0.479	0.293
$(0.1,0)$	Σ_4	0.993	0.990	0.979	0.977	0.965	0.899
$(0.1,0)$	Σ_5	0.755	0.753	0.650	0.641	0.572	0.382
$(0.1,0)$	Σ_6	0.994	0.997	0.991	0.991	0.985	0.944
			Alternative: $(\gamma_1,\gamma_2)=(0.1,0.1)$				
$(0.1,0.1)$	Σ_1	0.871	0.805	0.764	0.757	0.686	0.495
$(0.1,0.1)$	Σ_2	1.000	1.000	1.000	1.000	1.000	1.000
$(0.1,0.1)$	Σ_3	0.997	0.969	0.990	0.990	0.981	0.930
$(0.1,0.1)$	Σ_4	0.701	0.633	0.546	0.540	0.478	0.296
$(0.1,0.1)$	Σ_5	1.000	0.996	1.000	1.000	1.000	0.998
$(0.1,0.1)$	Σ_6	1.000	1.000	1.000	1.000	1.000	1.000

Table 5: Rejection frequencies corresponding to the multivariate location model with a two-dimensional parameter of interest $\gamma \geq 0$ under the null hypothesis and various alternatives for $n = 500$ ($\alpha = 0.05$) and $\beta = 0.1$. The columns Std and CLR correspond to the standard LR test and the CLR test and the columns $\eta_1, \dots, \eta_4$ correspond to the uniform LR test with $\eta = \alpha/20, \alpha/10, \alpha/2, 9\alpha/10$, respectively.

(γ_1, γ_2)	Σ	Std	CLR	η_1	η_2	η_3	η_4
Null: $\gamma_1 = \gamma_2 = 0$							
$(0, 0)$	Σ_1	0.057	0.049	0.047	0.044	0.025	0.006
$(0, 0)$	Σ_2	0.038	0.049	0.046	0.044	0.023	0.005
$(0, 0)$	Σ_3	0.050	0.049	0.047	0.044	0.026	0.006
$(0, 0)$	Σ_4	0.063	0.050	0.048	0.046	0.024	0.006
$(0, 0)$	Σ_5	0.045	0.048	0.046	0.044	0.026	0.005
$(0, 0)$	Σ_6	0.031	0.049	0.045	0.043	0.025	0.005
Alternative: $(\gamma_1, \gamma_2) = (0, 0.1)$							
$(0, 0.1)$	Σ_1	0.678	0.650	0.643	0.635	0.538	0.314
$(0, 0.1)$	Σ_2	0.984	0.988	0.985	0.986	0.974	0.901
$(0, 0.1)$	Σ_3	0.666	0.664	0.657	0.649	0.553	0.320
$(0, 0.1)$	Σ_4	0.994	0.990	0.989	0.989	0.978	0.929
$(0, 0.1)$	Σ_5	0.741	0.750	0.744	0.736	0.646	0.413
$(0, 0.1)$	Σ_6	0.994	0.997	0.997	0.996	0.993	0.962
Alternative: $(\gamma_1, \gamma_2) = (0.1, 0)$							
$(0.1, 0)$	Σ_1	0.675	0.647	0.635	0.626	0.530	0.299
$(0.1, 0)$	Σ_2	0.983	0.986	0.983	0.983	0.970	0.901
$(0.1, 0)$	Σ_3	0.669	0.667	0.658	0.646	0.547	0.311
$(0.1, 0)$	Σ_4	0.993	0.990	0.990	0.989	0.977	0.925
$(0.1, 0)$	Σ_5	0.743	0.753	0.745	0.737	0.646	0.404
$(0.1, 0)$	Σ_6	0.994	0.997	0.997	0.997	0.992	0.965
Alternative: $(\gamma_1, \gamma_2) = (0.1, 0.1)$							
$(0.1, 0.1)$	Σ_1	0.825	0.805	0.796	0.791	0.702	0.478
$(0.1, 0.1)$	Σ_2	1.000	1.000	1.000	1.000	1.000	1.000
$(0.1, 0.1)$	Σ_3	0.970	0.969	0.966	0.965	0.941	0.830
$(0.1, 0.1)$	Σ_4	0.679	0.633	0.622	0.614	0.516	0.292
$(0.1, 0.1)$	Σ_5	0.995	0.996	0.993	0.994	0.989	0.948
$(0.1, 0.1)$	Σ_6	1.000	1.000	1.000	1.000	1.000	1.000

Table 6: Rejection frequencies corresponding to the multivariate location model with a two-dimensional parameter of interest $\gamma \geq 0$ under the null hypothesis and various alternatives for $n = 500$ ($\alpha = 0.05$) and $\beta = 0.25$. The columns Std and CLR correspond to the standard LR test and the CLR test and the columns $\eta_1, \ldots, \eta_4$ correspond to the uniform LR test with $\eta = \alpha/20, \alpha/10, \alpha/2, 9\alpha/10$, respectively.

(γ_1, γ_2)	Σ	Std	CLR	η_1	η_2	η_3	η_4
			Null: $\gamma_1 = \gamma_2 = 0$				
$(0, 0)$	Σ_1	0.050	0.049	0.029	0.029	0.018	0.005
$(0, 0)$	Σ_2	0.051	0.047	0.029	0.028	0.018	0.004
$(0, 0)$	Σ_3	0.050	0.048	0.029	0.028	0.017	0.004
$(0, 0)$	Σ_4	0.050	0.048	0.029	0.029	0.018	0.005
$(0, 0)$	Σ_5	0.049	0.048	0.029	0.028	0.017	0.004
$(0, 0)$	Σ_6	0.046	0.047	0.027	0.027	0.016	0.004
			Alternative: $(\gamma_1, \gamma_2) = (0, 0.1)$				
$(0, 0.1)$	Σ_1	0.916	0.647	0.871	0.868	0.817	0.655
$(0, 0.1)$	Σ_2	1.000	0.988	1.000	1.000	1.000	1.000
$(0, 0.1)$	Σ_3	0.982	0.656	0.967	0.966	0.943	0.852
$(0, 0.1)$	Σ_4	0.998	0.987	0.995	0.994	0.990	0.964
$(0, 0.1)$	Σ_5	0.998	0.741	0.994	0.994	0.990	0.964
$(0, 0.1)$	Σ_6	0.985	0.987	0.971	0.970	0.952	0.871
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0)$				
$(0.1, 0)$	Σ_1	0.915	0.640	0.876	0.871	0.817	0.644
$(0.1, 0)$	Σ_2	1.000	0.986	1.000	1.000	1.000	1.000
$(0.1, 0)$	Σ_3	0.979	0.653	0.965	0.964	0.945	0.855
$(0.1, 0)$	Σ_4	0.998	0.986	0.996	0.995	0.989	0.961
$(0.1, 0)$	Σ_5	0.997	0.743	0.994	0.994	0.989	0.964
$(0.1, 0)$	Σ_6	0.997	0.987	0.994	0.993	0.989	0.961
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0.1)$				
$(0.1, 0.1)$	Σ_1	1.000	0.838	1.000	1.000	0.999	0.997
$(0.1, 0.1)$	Σ_2	1.000	1.000	1.000	1.000	1.000	1.000
$(0.1, 0.1)$	Σ_3	1.000	0.977	1.000	1.000	1.000	1.000
$(0.1, 0.1)$	Σ_4	0.999	0.698	0.997	0.996	0.993	0.974
$(0.1, 0.1)$	Σ_5	1.000	0.996	1.000	1.000	1.000	1.000
$(0.1, 0.1)$	Σ_6	1.000	1.000	1.000	1.000	1.000	1.000

Table 7: Rejection frequencies corresponding to the multivariate location model with an unrestricted two-dimensional parameter of interest γ under the null hypothesis and various alternatives for $n = 500$ ($\alpha = 0.05$) and $\beta = 0$. The columns Std and CLR correspond to the standard LR test and the CLR test and the columns $\eta_1, \dots, \eta_4$ correspond to the uniform LR test with $\eta = \alpha/20, \alpha/10, \alpha/2, 9\alpha/10$, respectively.

(γ_1, γ_2)	Σ	Std	CLR	η_1	η_2	η_3	η_4
			Null: $\gamma_1 = \gamma_2 = 0$				
$(0,0)$	Σ_7	0.095	0.049	0.017	0.016	0.009	0.002
$(0,0)$	Σ_8	0.129	0.047	0.025	0.023	0.012	0.002
$(0,0)$	Σ_9	0.060	0.048	0.010	0.009	0.004	0.001
			Alternative: $(\gamma_1, \gamma_2) = (0, 0.1)$				
$(0,0.1)$	Σ_7	0.667	0.588	0.373	0.366	0.281	0.133
$(0,0.1)$	Σ_8	0.889	0.602	0.671	0.664	0.573	0.348
$(0,0.1)$	Σ_9	0.863	0.767	0.635	0.628	0.547	0.337
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0)$				
$(0.1,0)$	Σ_7	0.981	0.690	0.902	0.898	0.853	0.692
$(0.1,0)$	Σ_8	0.940	0.656	0.785	0.777	0.699	0.486
$(0.1,0)$	Σ_9	0.710	0.624	0.417	0.411	0.322	0.156
			Alternative: $(\gamma_1, \gamma_2) = (0.1, 0.1)$				
$(0.1,0.1)$	Σ_7	0.998	0.908	0.976	0.975	0.961	0.873
$(0.1,0.1)$	Σ_8	0.999	0.865	0.989	0.988	0.979	0.923
$(0.1,0.1)$	Σ_9	0.882	0.796	0.672	0.666	0.579	0.364

Table 8: Rejection frequencies corresponding to the multivariate location model with a two-dimensional parameter of interest $\gamma \geq 0$ and a three-dimensional nuisance parameter $\beta \geq 0$ under the null hypothesis and various alternatives for $n = 500$ ($\alpha = 0.05$) and $\beta = 0$. The columns Std and CLR correspond to the standard LR test and the CLR test and the columns $\eta_1, \ldots, \eta_4$ correspond to the uniform LR test with $\eta = \alpha/20, \alpha/10, \alpha/2, 9\alpha/10$, respectively.

	AORD	BFX	IBEX	MXX	NSEI	OMX
δ	0.03	0.05	0.08	0.25	0.35	0.27
β_{ARCH}	0.01	0.01	0.00	0.08	0.16	0.00
β_{RV}	0.09	0.63	0.62	0.30	0.39	0.49
β_{W}	0.69	0.79	0.52	0.37	0.38	0.58
β_{M}	0.34	0.00	0.17	0.72	0.22	0.05
β_{SPX}	0.19	0.07	0.09	0.07	0.15	0.06
LR(H_{ARCH})	0.66	0.66	0.00	16.78	106.42	0.00
Uniform CV (5%)	6.61	6.70	8.44	16.77	20.38	18.44
Uniform CV (10%)	4.56	4.66	5.65	13.03	14.92	14.75
CLR(H_{ARCH})	0.59	0.44	0.00	7.31	23.02	0.00
CLR CV (5%)	2.70	2.74	2.71	2.72	2.65	2.68
CLR CV (10%)	1.63	1.64	1.63	1.66	1.60	1.63
LR(H_{SPX})	441.02	23.24	21.79	22.07	59.99	6.05
Uniform CV (5%)	7.54	7.35	8.50	16.91	20.22	18.46
Uniform CV (10%)	5.52	5.52	6.22	13.18	14.47	14.98
CLR(H_{SPX})	182.37	12.89	9.98	5.09	17.47	3.66
CLR CV (5%)	2.73	2.75	2.41	2.71	2.71	2.47
CLR CV (10%)	1.68	1.67	1.41	1.63	1.65	1.47
Observations	4757	4802	4772	4729	4670	3259

Table 9: Estimation results and individual tests for the empirical illustration. The first block of rows contains the point estimates of the ARCH model parameters. The second block of rows contains the values of the LR statistic LR and its corresponding uniform CVs at the 5% and 10% nominal levels as well as the values of the CLR statistic and its corresponding CVs, all for the null hypothesis $\mathsf{H}_{ARCH} : \beta_{ARCH} = 0$ across different indices. The third block of rows contains the same values as the second but for the null hypothesis $\mathsf{H}_{SPX} : \beta_{SPX} = 0$