Discussion Papers
Department of Economics
University of Copenhagen

No. 18-10

BOOTSTRAP INFERENCE ON THE BOUNDARY OF THE PARAMETER SPACE WITH APPLICATION TO CONDITIONAL VOLATILITY MODELS

Giuseppe Cavaliere, Heino Bohn Nielsen, Rasmus Søndergaard Pedersen
and Anders Rahbek

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

BOOTSTRAP INFERENCE ON THE BOUNDARY OF THE PARAMETER SPACE WITH APPLICATION TO CONDITIONAL VOLATILITY MODELS

Giuseppe Cavaliere*, Heino Bohn Nielsen†,
Rasmus Søndergaard Pedersen† and Anders Rahbek†

Abstract

It is a well-established fact that testing a null hypothesis on the boundary of the parameter space, with an unknown number of nuisance parameters at the boundary, is infeasible in practice in the sense that limiting distributions of standard test statistics are non-pivotal. In particular, likelihood ratio statistics have limiting distributions which can be characterized in terms of quadratic forms minimized over cones, where the shape of the cones depends on the unknown location of the (possibly mulitiple) model parameters not restricted by the null hypothesis. We propose to solve this inference problem by a novel bootstrap, which we show to be valid under general conditions, irrespective of the presence of (unknown) nuisance parameters on the boundary. That is, the new bootstrap replicates the unknown limiting distribution of the likelihood ratio statistic under the null hypothesis and is bounded (in probability) under the alternative. The new bootstrap approach, which is very simple to implement, is based on shrinkage of the parameter estimates used to generate the bootstrap sample toward the boundary of the parameter space at an appropriate rate. As an application of our general theory, we treat the problem of inference in finite-order ARCH models with coefficients subject to inequality constraints. Extensive Monte Carlo simulations illustrate that the proposed bootstrap has attractive finite sample properties both under the null and under the alternative hypothesis.

Keywords: Inference on the boundary; Nuisance parameters on the boundary; ARCH models; Bootstrap.

*Department of Economics, University of Bologna, Italy. †Department of Economics, University of Copenhagen, Denmark. Correspondence to: Giuseppe Cavaliere, Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; email: giuseppe.cavaliere@unibo.it. The authors thank Michael McAleer, Federico Bandi, Valentino Dardanoni, Michael Wolf, Christian Brownlees, Anders Kock, David Preinerstorfer and participants to the 2018 "New Developments in Econometrics and Time Series" workshop held in Copenhagen, September 2018. This research was supported by the Danish Council for Independent Research (DSF Grant 7015-00028B). Cavaliere also acklowledges partial financial support from the University of Bologna (Alma Idea 2017 grant).

1 Introduction

We consider (likelihood ratio-based) testing the null hypothesis that some of the parameters of a statistical model lie on the boundary of the parameter space. This is a non-standard testing problem which has been widely analyzed in the case where the parameters not restricted by the null hypothesis are in the interior of the parameter space, see Andrews (2001) and the references therein. However, the assumption that the only parameters which may lie on the boundary are those restricted by the null hypothesis excludes several important cases in empirical applications. A classic example, which we discuss in detail in the paper, is testing hypotheses in (G)ARCH models subject to non-negativity parameter constraints; see Francq and Zakoïan (2009). In this case, the practitioner may want to test whether some of the (G)ARCH parameters are zero, but (s)he is uncertain about the location of the remaining parameters.

This testing problem is particularly involved because the relevant null asymptotic distributions depend on whether the parameters not restricted by the null hypothesis – henceforth, 'nuisance parameters' – lie on the boundary or not. More specifically, likelihood ratio [LR] statistics have limiting distributions which can be characterized in terms of quadratic forms minimized over cones, where the shape of the cones depends on the unknown location of the (possibly multiple) nuisance parameters. The widely applied assumption that such parameters are not on the boundary (which corresponds to the assumption that the location of the parameters not restricted by the null hypothesis is known) is implausible in most testing problems, such as the aforementioned (G)ARCH case.

Attempts to deal with inference problems involving nuisance parameters potentially on the boundary of the parameter space are given in the literature; see e.g. Andrews and Guggenberger (2009), Elliott, Müller and Watson (2015), McCloskey (2017), Ketz (2018) and the reference therein. Here, however, we take a completely different route. Specifically we propose and analyze a novel bootstrap-based testing approach which can be applied to this testing problem.

Interestingly, the bootstrap is usually regarded as invalid when applied to testing whether some parameters are on the boundary of the parameter space, see e.g. Horowitz (2001). For instance, Andrews (2000) shows that in a simple location model with i.i.d. Gaussian errors the asymptotic distribution of the bootstrap maximum likelihood [ML] estimator of the location parameter is random in the limit, and hence fails to mimic the asymptotic distribution of the original ML estimator. The bootstrap in Andrews (2000) does not impose the null hypothesis on the bootstrap sample – that is, it is an example of the widely applied 'unrestricted bootstrap' – and this is crucial when interest is in testing that a parameter is on the boundary. In contrast, Cavaliere, Nielsen and Rahbek (2017) show that randomness of the limiting distribution can be avoided by applying a bootstrap scheme which imposes the null hypothesis on the bootstrap sample, that is, the 'restricted bootstrap', see also Davidson and MacKinnon (2006). However, the approach of Cavaliere, Nielsen and Rahbek (2017) requires that

all parameters not restricted by the null hypothesis are in the interior of the parameter space and, when this is not the case, also this bootstrap fails to replicate the correct asymptotic distribution, see the discussion in Section 3 below. An analog requirement is made in Francq and Zakoïan (2009) for testing that some coefficients in a general (G)ARCH model are equal to zero.

To overcome this drawback, we propose here a straightforward, bootstrap-based testing approach, which is very simple to implement and moreover delivers asymptotically correctly sized tests without losing the consistency property, irrespectively of the location of the parameters not restricted by the null hypothesis (the nuisance parameters). In particular, we show that a simple modification of either the restricted bootstrap, or the unrestricted bootstrap, delivers correct inference in large samples. Such modification is based on *shrinkage* of the original estimates of the parameters not restricted by the null hypothesis toward the boundary of the parameter space at an appropriate rate. A similar approach, which draws back to Beran (1997), is advocated in Andrews (2000, p. 403) for a one-parameter location model. As we demonstrate, this modification of the bootstrap scheme is able to eliminate the randomness in the limiting distribution of the bootstrap LR statistic. Consequently, we are able to provide high-level conditions on the data and bootstrap generating processes such that the bootstrap test allows control of the rejection probability under the null in large samples, irrespective of the presence of nuisance parameters on the boundary. We also discuss sufficient conditions for this novel modified bootstrap tests to be consistent under the alternative hypothesis.

As an application of our theory, in the paper we treat the problem of inference in finite-order ARCH models with coefficients subject to inequality (i.e. non-negativity) constraints. Using a fixed-volatility bootstrap scheme to illustrate, see Cavaliere, Pedersen and Rahbek (2018) and Beutner, Heinemann and Smeekes (2018), we show that our modified bootstrap LR test is asymptotically valid under the null and consistent under the alternative under standard regularity conditions.

We complete the paper by providing an extensive Monte Carlo experiment based on the ARCH model, where we show three important facts. First, we show that neglecting the presence of parameters on the boundary affects the size of asymptotic and bootstrap tests, which do not take into account the unknown location of the nuisance parameters. These tests may in general be either undersized or oversized, depending on the location of nuisance parameters and their implied correlation structure. Second, we show that even in samples of moderate size our modified bootstrap test has excellent properties under the null, while its power is indistinguishable to the power of asymptotic LR test based on the artificial assumption (see above for a discussion of this assumption) that all nuisance parameters are in the interior of the parameter space. Third, our Monte Carlo simulations show that the small sample properties of our modified bootstrap are extremely good irrespective of the bootstrap sample being based on restricted or unrestricted model parameter estimates.

The paper is organized as follows. Section 2 describes the general framework and

introduces the main assumptions on the estimators, the parameter space, the null hypothesis and the test statistics. The special case that will be considered throughout, namely the ARCH(q) model, is detailed here in Section 2.1. Section 3 presents the new bootstrap tests and analyzes their large sample properties, in particular by showing validity of the tests under the null and under the alternative hypotheses. The theory is applied to the ARCH(q) case in Section 4, while the small-sample properties are investigated by Monte Carlo simulation in Section 5. Section 6 concludes. All proofs are placed in the Appendix.

NOTATION. We make use of the following notation and definitions throughout. With $\mathbb{R}_+$ we denote the set of non-negative real numbers; with $\mathbb{I}(\cdot)$ we denote the indicator function, and '$x := y$' ('$x =: y$') indicates that x is defined by y (y is defined by x). We let $\{0\}^k := \{0\} \times \cdots \times \{0\}$ (k times), while $0_k = (0, ..., 0)'$ (of dimension k). We say that a set $A \subset \mathbb{R}^p$ is *locally equal* to a set $B \subset \mathbb{R}^p$ if there exists $C(0_p, \varepsilon)$ such that $A \cap C(0_p, \varepsilon) = B \cap C(0_p, \varepsilon)$, with $C(0_p, \varepsilon)$ an open cube in $\mathbb{R}^p$ centered at 0_p and with edge length 2ε, $\varepsilon > 0$. For any vector or matrix, x, $\|x\|$ denotes the usual Euclidean norm, $\|x\| := [tr(x'x)]^{1/2}$; moreover, the norm of a vector x with respect to a (square) matrix M is defined as $\|x\|_M^2 := x'Mx$ and $M > 0$ means the matrix M is positive definite.

Unless differently specified, limits are taken for $n \to \infty$. We use P^* and E^* respectively to denote probability and expectation, conditional on the original sample. With $\xrightarrow{w}$ and $\xrightarrow{p}$ we denote weak convergence and convergence in probability, respectively. For a given sequence X_n^* computed on the bootstrap data, $X_n^* = o_p^*(1)$, in probability, and $X_n^* \xrightarrow{p^*}_p X$ mean that for any $\epsilon > 0$, $P^*(||X_n^*|| > \epsilon) \xrightarrow{p} 0$ and $P^*(||X_n^* - X|| > \epsilon) \xrightarrow{p} 0$, respectively. Similarly, $X_n^* = O_p^*(1)$, in probability, means that, for every $\epsilon > 0$, there exists a constant $M > 0$ such that, for all large n, $P(P^*(||X_n^*|| > M) < \epsilon)$ is arbitrarily close to one. Finally, weak convergence (in probability) of X_n^* to a random variable X is denoted by $X_n^* \xrightarrow{w^*}_p X$.

2 THE SETTING

We address inference and testing in statistical models with parameters $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ where some of the parameters in θ are subject to an inequality constraint. Specifically, we look at the case where such parameters are restricted to be greater than or equal to zero, and test whether some of these parameters are indeed zero. Inference and testing is infeasible in practice, in the sense that it is not known whether the parameters in θ which are not restricted by the null hypothesis lie on the boundary or not. That is, the location of the nuisance parameters is unknown, and, as is detailed below, asymptotic inference is non-pivotal.

To address the issue, it is useful to partition the $d_\theta \times 1$ parameter vector θ as

$$\theta = (\gamma', \beta', \delta')',$$

where γ, β and δ are of dimension d_γ, d_β and d_δ respectively, with $d_\gamma + d_\beta + d_\delta = d_\theta$. The true parameter value is denoted by $\theta_0 = (\gamma_0', \beta_0', \delta_0')'$ and the null hypothesis H_0 we consider testing is given by

$$\mathsf{H}_0 : \gamma = 0_{d_\gamma}.$$

Thus, the parameters in θ are for simplicity, and without any loss of generality, partitioned into, or simply *labelled* as, *(i)* γ, the parameters of interest, which are the d_γ parameters restricted to zero under the hypothesis H_0, and; *(ii)* the remaining parameters β and δ. The parameters in δ are the d_δ parameters which are known *a priori* to have true values in the interior of the parameter space. The parameters in β – which we call 'nuisance parameters' in the following – can attain true values which are zero or not, and it is unknown a priori whether these are at the boundary or in the interior of the parameter space. Reflecting the partitioning of θ, the parameter space Θ is assumed to be given by

$$\Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta, \tag{1}$$

where $\gamma \in \Theta_\gamma := [0, \gamma_U]^{d_\gamma}$, $\gamma_U > 0$, $\beta \in \Theta_\beta := [0, \beta_U]^{d_\beta}$, $\beta_U > 0$, and $\delta \in \Theta_\delta \subset \mathbb{R}^{d_\delta}$ with Θ_δ compact. We emphasize that for the true value of the nuisance parameters β_0 and δ_0, the vector δ_0 is assumed to be an interior point in Θ_δ, and hence not on the boundary, while for β_0 it is *not* known whether parts of it are on the boundary (that is, equal to zero) or not.

We assume in addition that the statistical model is given by the variables $(x_t)_{t=1}^n$ together with a (quasi log-) likelihood function – or, more generally, an objective function – denoted here by $L_n(\theta)$. In particular, the unrestricted and restricted estimators of θ are given by

$$\hat{\theta}_n = (\hat{\gamma}_n', \hat{\beta}_n', \hat{\delta}_n')' := \arg\max_{\theta \in \Theta} L_n(\theta), \quad \tilde{\theta}_n = (\tilde{\gamma}_n', \tilde{\beta}_n', \tilde{\delta}_n')' := \arg\max_{\theta \in \Theta_{\mathsf{H}_0}} L_n(\theta), \tag{2}$$

where the optimization set under the null hypothesis is given by

$$\Theta_{\mathsf{H}_0} = \{\theta \in \Theta \ : \gamma = 0_{d_\gamma}\}, \tag{3}$$

such that $\tilde{\gamma}_n = 0_{d_\gamma}$. The (quasi-)likelihood ratio statistic for the hypothesis H_0 is given by

$$LR_n := -2(L_n(\tilde{\theta}_n) - L_n(\hat{\theta}_n)). \tag{4}$$

Andrews (2001) derives the limiting distribution of the likelihood ratio statistic for the null hypothesis H_0 under a set of standard regularity conditions, in addition to the conditions on the parameter space(s). The standard regularity conditions for the asymptotic theory are as follows.

ASSUMPTION 1 *Assume* (i), *that for* $\theta_0 \in \Theta$, *the unrestricted estimator* $\hat{\theta}_n$ *is consistent, that is,* $\hat{\theta}_n = \theta_0 + o_p(1)$, *and likewise for the restricted estimator* $\tilde{\theta}_n$, $\tilde{\theta}_n = \theta_0 + o_p(1)$ *under* H_0. *Furthermore, for* $\theta_0 \in \Theta_{\mathsf{H}_0}$:

(ii) $-n^{-1}\partial^2 L_n(\theta_0)/\partial\theta\partial\theta' \xrightarrow{p} \Omega > 0$, $\Omega^{-1}n^{-1/2}\partial L_n(\theta_0)/\partial\theta \xrightarrow{w} Z$,
(iii) $\max_{i,j,k=1,2,...,d_\theta} \sup_{\theta\in\Theta} |n^{-1}\partial^3 L_n(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k| \leq \kappa_n$,
where $\kappa_n = O_p(1)$ *and* Z *a* d_θ*-dimensional Gaussian random variable with covariance matrix* $\Omega^{-1}\Sigma\Omega^{-1}$, $\Sigma > 0$.

REMARK 2.1 Note that, as is standard, Assumption 1*(iii)* can be replaced by the requirement that a uniform law of large numbers applies to the second order derivative, $n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta'$, see also Andrews (1999) and Jensen and Rahbek (2004) for a discussion. □

For the parameter space Θ in (1), we denote by $k \in \{0, 1, .., d_\beta\}$ the unknown number of nuisance parameters which are (at their true value) on the boundary of the parameter space and we make the following assumption.

ASSUMPTION 2 *The shifted parameter space,* $\Theta - \theta_0$*, is locally equal to the cone* Λ *given by*

$$\Lambda := \Lambda_\gamma \times \Lambda_\beta \times \Lambda_\delta, \tag{5}$$

where $\Lambda_\gamma = \mathbb{R}_+^{d_\gamma}$, $\Lambda_\delta = \mathbb{R}^{d_\delta}$ *and* $\Lambda_\beta = \Lambda_1 \times .... \times \Lambda_{d_\beta}$, *with* k *of the* Λ_i*'s equal to* $\mathbb{R}_+$ *and the remaining* $d_\beta - k$ *equal to* $\mathbb{R}$.

REMARK 2.2 It is important to stress that the shape of the cone Λ in (5) varies depending on the unknown value k of nuisance parameters at the boundary. The above formulation of Λ allows, in particular, for any combination of nuisance parameters on the boundary. □

From Andrews (1999, 2001), it follows that under Assumptions 1 and 2, the test statistic LR_n in (4) converges in distribution to a non-standard, non-pivotal distribution, say $\mathcal{L}$. In general, $\mathcal{L}$ can be written as a difference between two quadratic forms minimized separately over cones which depend on the unknown k, or equivalently on the shape of the cone Λ_β, defined in Assumption 2. Specifically, if $\theta_0 = (\gamma', \beta', \delta') \in \Theta_{\mathsf{H}_0}$ with $\delta_0 \in \text{int}(\Theta_\delta)$ then as in Andrews (2001) we have that

$$LR_n \xrightarrow{w} \mathcal{L} := \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta} \|\lambda - HZ\|^2_{(H\Omega^{-1}H')^{-1}} - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta} \|\lambda - HZ\|^2_{(H\Omega^{-1}H')^{-1}} \tag{6}$$

with H a matrix of dimension $(d_\gamma + d_\beta) \times d_\theta$ such that $(\gamma', \beta')' = H\theta$.

Thus, the limiting distribution $\mathcal{L}$ depends on the unknown cone Λ_β, as well as on the covariances Ω and Σ, and we may write $\mathcal{L} = \mathcal{L}(\Omega, \Sigma, \Lambda_\beta)$. Hence, in general, the distribution of $\mathcal{L}$ is non-pivotal and asymptotic inference is infeasible. We propose a new bootstrap as detailed in Section 3 to circumvent this.

Before, we next briefly discuss the just presented theory in terms of the well-known ARCH(q) model.

2.1 The ARCH(q) model

Consider the finite-order linear ARCH(q) model with $q \geq 1$,

$$x_t = \sigma_t(\theta)\,\eta_t, \text{ with } \sigma_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 \tag{7}$$

for $t = 1, 2, ..., n$, $(x_{-q+1}, ..., x_0)$ fixed in the statistical analysis and η_t i.i.d.(0,1), where η_t has a Lebesgue density that is strictly positive in a neighborhood of zero.

We assume $\omega \in [\omega_L, \omega_U]$, with $\omega_U > \omega_L > 0$, and $(\alpha_1, ..., \alpha_q) \in [0, \alpha_U]^q$, $\alpha_U > 0$. The true values of the parameters are denoted by $\omega_0, \alpha_{10}, ..., \alpha_{q0}$. The setting above covers hypotheses such as any (non-empty) subset of $\{\alpha_1, ..., \alpha_q\}$ are equal to zero. However, to keep notation simple, we focus on the simple hypothesis $\mathsf{H}_0 : \alpha_q = 0$.

Notice that while we assume *a priori* that the true value ω_0 of the intercept term ω in the ARCH model is an interior point, importantly it is *unknown* whether the true values of the remaining ARCH nuisance parameters equal zero or not; that is, it is unknown whether $\alpha_{i0} = 0$ or $\alpha_{i0} > 0$ for $i = 1, ..., q-1$.

In terms of the notation introduced above, we make the following assumption on the parameter space for θ as well as on the true parameter θ_0.

Assumption 3 *Consider the ARCH(q) model given by (7). With $\theta = (\gamma, \beta', \delta)'$, where $\gamma = \alpha_q$, $\beta = (\alpha_1, ..., \alpha_{q-1})'$ and $\delta = \omega$, assume that*

$$\gamma \in \Theta_\gamma := [0, \alpha_U],\ \beta \in \Theta_\beta := [0, \alpha_U]^{q-1}, \text{ and } \quad \delta \in \Theta_\delta := [\omega_L, \omega_U], \tag{8}$$

with $\omega_U > \omega_L > 0$ and $\alpha_U > 0$. Moreover, assume that at the true parameter vector θ_0, $\theta_0 \in \Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$, with $\delta_0 \in \operatorname{int}\Theta_\delta$, the ARCH process $\{x_t\}$ is stationary and ergodic with $E[x_t^6] < \infty$.

With the Gaussian (quasi-) log-likelihood function given by

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta), \quad l_t(\theta) = -\tfrac{1}{2}(\log \sigma_t^2(\theta) + \tfrac{x_t^2}{\sigma_t^2(\theta)}), \tag{9}$$

we can define the unrestricted estimator of θ, $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\beta}_n', \hat{\delta}_n)'$ where $\hat{\gamma}_n = \hat{\alpha}_{q,n}$, $\hat{\beta}_n := (\hat{\alpha}_{1,n}, ..., \hat{\alpha}_{q-1,n})'$ and $\hat{\delta}_n = \hat{\omega}_n$, as the maximizer of (9) over Θ. Similarly, the restricted estimator of θ, denoted by $\tilde{\theta}_n = (0, \tilde{\beta}_n', \tilde{\delta}_n)'$ where $\tilde{\beta}_n := (\tilde{\alpha}_{1,n}, ..., \tilde{\alpha}_{q-1,n})'$ and $\tilde{\delta}_n = \tilde{\omega}_n$, is the maximizer of (9) over $\Theta_{\mathsf{H}_0} := \{0\} \times \Theta_\beta \times \Theta_\delta$.

It follows as in Andrews (2001), see also Francq and Zakoïan (2009), that under Assumption 3, and with $\theta_0 \in \Theta_{\mathsf{H}_0}$ defined in (3), $LR_n \to_w \mathcal{L}$ with $\mathcal{L}$ given in (6).

As emphasized earlier, the limiting distribution $\mathcal{L}$ in (6) is non-pivotal and an asymptotic test infeasible in practice, as it depends on the unknown number $k \in \{0, 1, 2, ..., , d_\beta = q-1\}$ of nuisance parameters on the boundary of the parameter space.

When (as done here) the null hypothesis restricts one parameter only, i.e. $d_\gamma = 1$, some remarks can be made about the distribution of $\mathcal{L}$ depending on the number k of nuisance parameters at zero.

For the case of $k = 0$, where there are no nuisance parameters on the boundary, the distribution of $\mathcal{L}$ reduces to the well-known mixture distribution $\mathcal{M} = \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, i.e. a mixture of a χ_1^2 and singular random variable with probability mass at zero, both with weights $\frac{1}{2}$; see e.g. Andrews (2001), Francq and Zakoïan (2009) and Cavaliere *et al.* (2017).

For the case of $k = 1$, where one (and only one) nuisance parameter is on the boundary, the distribution $\mathcal{L}$ can be characterized by a correlation parameter ρ of a bivariate Gaussian variable, Z_ρ. This can be seen by combining our proof of Theorem 1 below with the theory of Kopylev and Sinha (2010, 2011), defining $Z_\rho = H_1 Z$, with H_k defined in (A.2) in the Appendix. In particular, for $\rho \geq 0$, the distribution is a mixture of independent χ_0^2, χ_1^2, and χ_2^2 variables with mixture weights $(\frac{1}{2} - p, \frac{1}{2}, p)$, where $p := \sin^{-1}(\rho)/2\pi$. This distribution is shifted to the right compared to the mixture distribution $\mathcal{M}$, in the sense that $P(\mathcal{L} = 0) = \frac{1}{2} - p \leq \frac{1}{2} = P(\mathcal{M} = 0)$. For $\rho < 0$, the distribution is not a mixture of χ^2- distributed random variables. Interestingly, the distribution is shifted to the left compared to $\mathcal{M}$. That is, for $k = 1$ and $\rho < 0$, and $P(\mathcal{L} = 0) = \frac{1}{2} + \sin^{-1}(-\rho)/2\pi > P(\mathcal{M} = 0)$. Observe that for the ARCH(q) case, with $k = 1$, the correlation ρ is negative, $\rho < 0$. Hence, a test which neglects the presence of the nuisance parameter on the boundary and hence uses $\mathcal{M}$ as the reference null distribution may be undersized in large samples.

For the remaining cases, where $1 < k \leq d_\beta$, the distribution $\mathcal{L}$ cannot, to the best of our knowledge, be characterized by a mixture of χ^2-distributed random variables. However, we conjecture that it depends on the correlation structure of a Gaussian $(k+1)$-dimensional random vector, similarly to the $k = 1$ case.

Noticeably, for the ARCH(q) model where $k = d_\beta$; that is, with all nuisance parameters in β on the boundary, $\mathcal{L}$ is distributed as the mixture $\mathcal{M}$, since the matrix $H\Omega^{-1}H'$ in (6) is block-diagonal with respect to γ and β, as demonstrated by Demos and Sentana (1998, Appendix A); see also Francq and Zakoïan (2009, Section 7.1) and Pedersen and Rahbek (2018).

3 A NEW BOOTSTRAP

As detailed in Section 2, the limiting distribution $\mathcal{L}$ is non-pivotal, hence rendering asymptotic inference infeasible in general. As anticipated earlier, we propose here a new bootstrap which is based on shrinking the parameter estimators used to generate the bootstrap sample, see Andrews (2000) for a simple one-parameter location model. In this respect, our bootstrap involves the use of Hodges-Le Cam super-efficient type estimators, see *e.g.* Bickel, Klaassen, Ritov and Wellner (1998) and the references therein. We provide a full asymptotic theory for the validity of the new bootstrap,

and as a by-product we also discuss why conventional bootstrap methods – such as the standard, restricted or unrestricted bootstrap – do not work in the case where there are nuisance parameters possibly at the boundary.

The setup of the bootstrap we consider is as follows. As is standard, we consider bootstrap data $\{x_t^*\}_{t=1}^n$ with x_t^* generated (possibly recursively) as a function of: *(i)* the original data, $\{x_t\}_{t=1}^n$; *(ii)* possibly lagged x_t^*'s or exogenous variables, X_t^*, *(iii)* a bootstrap true parameter value θ_n^*, which is some function of $\{x_t\}_{t=1}^n$; *(iv)* a random vector of bootstrap shocks, independent of the original data, denoted here by π_n^*. That is,

$$x_t^* := f\left(\theta_n^*, \{x_t\}_{t=1}^n, X_t^*, \pi_n^*\right), \ t = 1, 2, ..., n. \tag{10}$$

REMARK 3.1 The bootstrap true parameter value θ_n^* in *(iii)* is crucial in defining the properties of the bootstrap. Usually θ_n^* is set equal to $\hat{\theta}_n$, the unrestricted estimator of θ_0, or to $\tilde{\theta}_n$, the estimator of θ_0 obtained with the null hypothesis imposed (see Davidson and MacKinnon, 2006), or a hybrid of the two (see e.g. Swensen, 2004, for an application to co-integration). For standard testing problems, the associated (un-)restricted bootstraps are often asymptotically valid, or consistent. For some non-testing problems, such as for inference on the number of unit roots (Cavaliere, Rahbek and Taylor, 2012) and in the presence of infinite variance innovations (Davidson and Flachaire, 2008), the restricted bootstrap based on $\tilde{\theta}_n$ is asymptotically valid even when the bootstrap based on $\hat{\theta}_n$ may fail. In the testing problem considered here, both the unrestricted and the restricted bootstraps fail, making the bootstrap unable to mimic the target distribution $\mathcal{L}$ under the null hypothesis, see Remark 3.4 below. The bootstrap discussed in the section circumvents this drawback.

REMARK 3.2 The role of π_n^* in *(iv)* is crucial, as it defines – along with the function $f(\cdot)$ – the bootstrap resampling scheme. For instance, for the usual i.i.d. bootstrap, $\pi_n^* := (\pi_{n1}^*, ..., \pi_{nn}^*)$ is the (random) number of times each of the original observations (or some residuals) are selected during the re-sampling process; for the wild bootstrap, π_n^* is the vector of bootstrap i.i.d. innovations used to rescale the original data (or residuals). □

Corresponding to the bootstrap data $\{x_t^*\}_{t=1}^n$ we introduce a bootstrap (quasi) log-likelihood, or criterion function, $L_n^*(\theta)$, and the associated bootstrap (unrestricted and restricted) estimators,

$$\hat{\theta}_n^* := \arg\max_{\theta\in\Theta} L_n^*(\theta), \text{ and } \tilde{\theta}_n^* := \arg\max_{\theta\in\Theta_{\mathsf{H}_0}} L_n^*(\theta). \tag{11}$$

The bootstrap (quasi-)likelihood ratio statistic for the hypothesis H_0 is given by

$$LR_n^* = -2(L_n^*(\tilde{\theta}_n^*) - L_n^*(\hat{\theta}_n^*)). \tag{12}$$

Importantly, as discussed in Remark 3.1, the bootstrap likelihood ratio statistic, LR_n^*, will not for the (un-)restricted bootstrap replicate the unknown non-pivotal distribution $\mathcal{L}$ in (6), even asymptotically.

Instead of the classical bootstraps, we propose here to choose θ_n^* differently: First, we impose the null hypothesis H_0 on θ_n^*, which corresponds to setting γ_n^* equal to zero. Second, and crucially, we impose a requirement on the rate of consistency for β_n^* in θ_n^*, which we refer to as 'shrinkage', for the reasons explained below. Formally, we make the following assumption about θ_n^*.

ASSUMPTION 4 *With $\theta_n^* = (\gamma_n^{*\prime}, \beta_n^{*\prime}, \delta_n^{*\prime})'$, let $\gamma_n^* = 0_{d_\gamma}$ and assume that*

$$\theta_n^* \xrightarrow{p} \theta_0^\dagger := (0'_{d_\gamma}, \beta_0^{\dagger\prime}, \delta_0^{\dagger\prime}) \in \Theta_{\mathsf{H}_0} \tag{13}$$

under H_0 as well as the alternative, where $\theta_0^\dagger = \theta_0$ under H_0. Furthermore, for $i = 1, ...d_\beta$, assume

$$\sqrt{n}(\beta_{n,i}^* - \beta_{0,i}^\dagger) = \begin{cases} o_p(1) & \text{if } \beta_{0,i}^\dagger = 0 \\ O_p(1) & \text{if } \beta_{0,i}^\dagger > 0 \end{cases}, \tag{14}$$

and $\sqrt{n}(\delta_n^* - \delta_0^\dagger) = O_p(1)$.

For comparison, the classical unrestricted bootstrap where $\theta_n^* = \hat{\theta}_n$ does satisfy (13) under H_0, but not under the alternative. Moreover, for the unrestricted bootstrap, $\gamma_n^* \neq 0_{d_\gamma}$ and, in addition, the convergence rates in (14) do not apply. For the restricted bootstrap, $\theta_n^* = \tilde{\theta}_n$, and satisfies by definition $\gamma_n^* = 0_{d_\gamma}$, but as for the unrestricted bootstrap, the convergence rates in (14) do not apply. As to (13), this follows under H_0, while under the alternative it is non-trivial for various models to establish if, or indeed if not, θ_n^* converges to some pseudo-true value $\theta_0^\dagger$.

A particular bootstrap scheme satisfying Assumption 4 is given by choosing $\delta_n^* = \hat{\delta}_n$ and

$$\beta_{n,i}^* = \hat{\beta}_{n,i}\mathbb{I}(\hat{\beta}_{n,i} > c_n) \quad i = 1, ..., d_\beta$$

with c_n a scalar sequence converging to zero at an appropriate rate, as seen in the following lemma.

LEMMA 1 *Under Assumption 1, and with the sequence $\{c_n\}_{n=1,2,....}$ satisfying*

$$c_n \to 0 \text{ and } \sqrt{n}c_n \to \infty \quad as \quad n \to \infty, \tag{15}$$

then θ_n^* defined by

$$\theta_n^* = (\gamma_n^{*\prime}, \beta_n^{*\prime}, \delta_n^{*\prime})' = (0'_{d_\gamma}, \{\hat{\beta}_{n,i}\mathbb{I}(\hat{\beta}_{n,i} > c_n)\}'_{i=1,..,d_\beta}, \hat{\delta}_n')', \tag{16}$$

satisfies Assumption 4 with $\theta_0^\dagger = (0'_{d_\gamma}, \beta_0', \delta_0')'$.

The proposed shrinkage in terms of the c_n sequence, or more generally, the requirement on β_n^* in (14) ensures that the bootstrap replicates the unknown limiting distribution $\mathcal{L}$ under the null, while being of order $O_p^*(1)$, in probability, under the alternative. That is, as is established in Theorem 1 below, the new bootstrap is consistent even though it is unknown if any of the nuisance parameters are on the boundary or not.

REMARK 3.3 Alternatively, in (16) the unrestricted estimators $\hat{\beta}_n$ and $\hat{\delta}_n$ could be replaced by the restricted estimators, $\tilde{\beta}_n$ and $\tilde{\delta}_n$. However, as already mentioned, in that case it may not be trivial to establish $\theta_n^* \to_p \theta_0^\dagger$ in Assumption 4 under the alternative.□

For the bootstrap we make the following assumption which need to be verified on a case by case basis depending on the model of interest. Assumption 5 is the bootstrap equivalent of Assumption 1.

ASSUMPTION 5 *Assume that* (i) $\hat{\theta}_n^*, \tilde{\theta}_n^* = \theta_0^\dagger + o_p^*(1)$, *in probability, for some* $\theta_0^\dagger = (0'_{d_\gamma}, \beta_0^{\dagger\prime}, \delta_0^{\dagger\prime})' \in \Theta_{\mathsf{H}_0}$. *Furthermore,*

$$\begin{aligned} &\text{(ii)} \quad -n^{-1}\partial^2 L_n^*(\theta_n^*)/\partial\theta\partial\theta' \xrightarrow{p^*}_p \Omega^* > 0, \text{ with } \Omega^* = \Omega \text{ under } \mathsf{H}_0, \\ &\qquad \text{and, } \Omega^{*-1} n^{-1/2} \partial L_n^*(\theta_n^*)/\partial\theta \xrightarrow{w^*}_p Z^*, \\ &\text{(iii)} \quad \max_{i,j,k=1,2,...,d_\theta} \sup_{\theta\in\Theta} |n^{-1}\partial^3 L_n^*(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k| \leq \kappa_n^*, \end{aligned}$$

where $\kappa_n^* = O_p^*(1)$, *in probability, and* Z^* *a* d_θ*-dimensional Gaussian random variable with positive definite covariance* $\Omega^{*-1}\Sigma^*\Omega^{*-1}$, *with* $Z^* \overset{d}{=} Z$ *under* H_0.

We can then state the following general result:

THEOREM 1 *Consider the model for* $\{x_t\}_{t=1}^n$ *with (quasi-)likelihood function* $L_n(\theta)$, *and assume that Assumptions 1 and 2 hold, such that the (quasi-)likelihood ratio statistic satisfies* $LR_n \to_w \mathcal{L}$, *with* $\mathcal{L}$ *defined in (6). Then, with the bootstrap data* $\{x_t^*\}_{t=1}^n$ *defined in (10) and the bootstrap (quasi-) likelihood ratio statistic* LR_n^* *in (12), under Assumptions 4 and 5, we have under* H_0,

$$LR_n^* \xrightarrow{w^*}_p \mathcal{L}.$$

Under the alternative and Assumptions 1, 2, 4 and 5, then $LR_n^* = O_p^*(1)$, *in probability, with* $LR_n^* \xrightarrow{w^*}_p \mathcal{L}^\dagger$ *defined in (A.3) in the appendix.*

REMARK 3.4 If we replace β_n^* by the unrestricted estimator $\hat{\beta}_n$ (or the restricted estimator $\tilde{\beta}_n$) in the construction of θ_n^*, then LR_n^* does not converge weakly (in probability) to $\mathcal{L}$, hence invalidating the consistency of the classic unrestricted and restricted bootstraps. To see this, note that in the proof of Theorem 1, it is used that by Assumption 4 the convergence rate of β_n^* should satisfy (14). With $\beta_n^* = \hat{\beta}_n, \tilde{\beta}_n$, that is in the case of no shrinkage, it only holds that $\sqrt{n}(\beta_{n,i}^* - \beta_{0,i}^\dagger) = O_p(1)$ for $i = 1, ...d_\beta$ and hence (14) does not apply. Furthermore, with U the weak limit of $\sqrt{n}(\beta_n^* - \beta_0^\dagger)$, it can be shown that the limiting distribution of LR_n^* in this case is given by (6), with HZ replaced by $H(Z^* + U)$, where Z^* under H_0 has the same distribution as Z. That is, while in (6), Z has mean zero, $Z^* + U$, conditional on U, has mean U. Alternatively, the limiting distribution of LR_n^* is given by (6), with Λ_β replaced by $\Lambda_\beta - HU$, that is Λ_β shifted stochastically, corresponding to the appropriate limit of $\sqrt{n}(H(\Theta - \theta_n^*))$. This is in line with the results in Cavaliere, Nielsen and Rahbek (2015), where in the context

of co-integration, an Ornstein-Uhlenbeck process with stochastic diffusion coefficient characterizes the limiting distribution of the bootstrap LR statistics. Obviously, when there are no nuisance parameters on the boundary, shrinkage is not required and the classic unrestricted or restricted bootstraps are asymptotically valid. □

4 Bootstrap theory applied to ARCH(q)

We consider here in detail bootstrap-based inference for the ARCH(q) model of Section 2.1, and establish that the proposed bootstrap indeed satisfies the regularity conditions for Theorem 1. That is, we show here that the proposed bootstrap is consistent in the ARCH(q) model case, in the sense that under the null hypothesis it replicates the limiting distribution $\mathcal{L}$, while under the alternative the bootstrap LR statistic converges in distribution to a random variable $\mathcal{L}^{\dagger}$ and hence is bounded, in probability.

When testing the simple hypothesis $\mathsf{H}_0 : \alpha_q = 0$, the bootstrap ARCH($q$) data are generated as

$$x_t^* = f\left(\theta_n^*, \{x_t\}_{t=1}^n, X_t^*, \pi_n^*\right) = \sigma_t\left(\theta_n^*\right)\eta_t^*, \text{ for } t = 1, ..., n, \tag{17}$$

with θ_n^* given as in (16):

$$\theta_n^* = (\gamma_n^*, \beta_n^{*\prime}, \delta_n^*)' = (0, \{\hat{\alpha}_{i,n}\mathbb{I}\left(\hat{\alpha}_{i,n} > c_n\right)\}_{i=1}^{q-1}, \hat{\omega}_n)', \tag{18}$$

where the $\hat{\alpha}_{i,n}$'s and $\hat{\omega}_n$ are unrestricted estimators of the ARCH parameters obtained on the original data, see Section 2.1. Here the bootstrap conditional volatility $\sigma_t^2\left(\theta_n^*\right)$ is given by

$$\sigma_t^2\left(\theta_n^*\right) = \delta_n^* + (\beta_n^{*\prime}, \gamma_n^*)' X_t^* \tag{19}$$

$$X_t^* = \left(x_{t-1}^2, ..., x_{t-q}^2\right)', \tag{20}$$

hence corresponding to a non-recursive, fixed volatility bootstrap as in Cavaliere *et al.* (2018) and Beutner *et al.* (2018), and $X_1^* = \left(x_0^2, ...x_{-q+1}^2\right)$ fixed.

As to the bootstrap resampling scheme, that is π_n^* in (17), we let $\pi_n^* = (\eta_1^*, ..., \eta_n^*)$ where the η_t^*'s are bootstrap innovations $\{\eta_t^*\}_{t=1}^n$ obtained by re-sampling with replacement from the normalized and re-scaled estimated residuals, $\{\hat{\eta}_t\}_{t=1}^n$, defined as $\hat{\eta}_t := x_t/\sigma_t(\hat{\theta}_n)$. That is, η_t^* is re-sampled with replacement from

$$\hat{\eta}_t^s := \frac{\hat{\eta}_t - \bar{\eta}_n}{\sqrt{n^{-1}\sum_{t=1}^n (\hat{\eta}_t - \bar{\eta}_n)^2}}, \; \bar{\eta}_n := n^{-1}\sum_{t=1}^n \hat{\eta}_t. \tag{21}$$

The Gaussian bootstrap criterion function is given by

$$L_n^*\left(\theta\right) = \sum_{t=1}^n l_t^*\left(\theta\right), \quad l_t^*\left(\theta\right) = -\tfrac{1}{2}(\log\sigma_t^2(\theta) + \tfrac{x_t^{*2}}{\sigma_t^2(\theta)}), \tag{22}$$

where $\hat{\theta}_n^*$ ($\tilde{\theta}_n^*$) maximizes $L_n^*(\theta)$ (under H_0). Finally, the bootstrap quasi-likelihood ratio statistic LR_n^* for the hypothesis $\mathsf{H}_0 : \alpha_q = 0$, is given by (12).

We establish next that the regularity conditions in Assumption 5 hold for this bootstrap, such that Theorem 1 holds, as formulated in the next proposition.

PROPOSITION 1 *Let Assumption 3 holds with $Ex_t^8 < \infty$, and consider the bootstrap data $\{x_t^*\}_{t=1}^n$ as generated by (17). Then, under H_0, the (quasi-)likelihood ratio statistic LR_n^* satisfies $LR_n^* \xrightarrow{w^*}_p \mathcal{L}$, provided the sequence $\{c_n\}$ satisfies (15). Under the alternative, $LR_n^* \xrightarrow{w^*}_p \mathcal{L}^\dagger$, with $\mathcal{L}^\dagger$ defined in (A.3) in the appendix.*

The proof of Proposition 1 follows in two steps. First, the above choice of θ_n^*, see (18), implies by Lemma 1 that Assumption 4 holds. Second, Lemmas 2–4 below imply that Assumption 5 holds, such that Theorem 1 applies and the desired result is obtained.

REMARK 4.1 As previously mentioned, although we focus here on a simple hypothesis such that $d_\gamma = 1$, all results generalize to the case of $d_\gamma > 1$.

REMARK 4.2 While the asymptotic theory for the standard LR statistic requires existence of 6^{th} order moments for x_t, our implementation of the fixed regressor bootstrap is based on the sufficient condition $Ex_t^8 < \infty$. This is needed to analyze the asymptotic behaviour of the third-order derivatives of the bootstrap quasi-log likelihood. Our simulation results, see section 5 below, suggest that this requirement may not be necessary.□

LEMMA 2 *Under Assumption 3, and with $\{x_t^*\}$ given by (17), it holds that the bootstrap unrestricted and restricted estimators $\hat{\theta}_n^*, \tilde{\theta}_n^*$ satisfy Assumption 5(i); that is,*

$$\hat{\theta}_n^*, \tilde{\theta}_n^* \xrightarrow{p^*}_p \theta_0^\dagger = (0, \{\alpha_{i,0}\}_{i=1}^{q-1}, \omega_0).$$

REMARK 4.3 To establish the result in Lemma 2 a non-standard asymptotic criterion function is introduced in the arguments under the alternative. □

LEMMA 3 *Under Assumption 3, and with $\{x_t^*\}$ given by (17), it holds that the bootstrap score and information satisfy Assumption 5(ii); that is,*

$$-n^{-1}\partial^2 L_n^*(\theta_n^*)/\partial\theta\partial\theta' \xrightarrow{p^*}_p \Omega^* > 0; \text{ and } \Omega^{*-1} n^{-1/2} \partial L_n^*(\theta_n^*)/\partial\theta \xrightarrow{w^*}_p Z^*,$$

with Z^* distributed as a $N(0, \Omega^{*-1}\Sigma^*\Omega^{*-1})$ random variable. Here $\Omega^* = \Omega$ under H_0, while $\Omega^* = E[\frac{1}{2}\sigma_t^{-4}(\theta_0^\dagger) z_t z_t']$ under the alternative, with z_t as defined in (B.5) and $\Sigma^* = E[(\partial l_t(\theta_0^\dagger)/\partial\theta)(\partial l_t(\theta_0^\dagger)/\partial\theta')]$.

LEMMA 4 *Under Assumption 3, with $\{x_t^*\}$ given by (17), and the additional assumption that $E[x_t^8] < \infty$, it holds that the bootstrap third order derivatives of the (quasi-) likelihood function satisfy Assumption 5(iii); that is,*

$$\max_{i,j,k=1,\ldots,q+1} \sup_{\theta\in\Theta} \left| n^{-1}\partial^3 L_n^*(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k \right| \leq \kappa_n^*, \text{ with } \kappa_n^* = O_p^*(1),$$

in probability.

REMARK 4.4 The fixed volatility bootstrap implemented in this section can be replaced by a recursive bootstrap scheme, see e.g. Hidalgo and Zaffaroni (2007), Corradi and Iglesias (2008) and Jeong (2017). This can be done by replacing X_t^* in (20) by $X_t^* := \left(x_{t-1}^{*2}, ..., x_{t-q}^{*2}\right)'$. Accordingly, the bootstrap criterion function changes to

$$L_n^*(\theta) = \sum_{t=1}^{n} l_t^*(\theta), \quad l_t^*(\theta) := -\tfrac{1}{2}(\log \sigma_t^{*2}(\theta) + \tfrac{x_t^{*2}}{\sigma_t^{*2}(\theta)})$$

with $\sigma_t^{*2}(\theta) = \omega + \sum_{i=1}^{q} \alpha_i (x_{t-i}^*)^2$. □

5 NUMERICAL RESULTS

In this section we illustrate the finite sample properties of the proposed bootstrap LR tests using a detailed simulation study based on an ARCH(q) model with $q = 5$. First, we aim at exploring the performance in terms of size and power of our new bootstrap test across different choices of the bootstrap true values and different volatility resampling schemes. Second, we aim at analyzing the robustness of the result over different choices of the shrinkage sequence $\{c_n\}$, and in particular to show that the test behaviour is not substantially affected by such choices. Third, we aim at providing evidence about the superiority of our bootstrap tests over existing techniques, such as the 'm out of n' bootstrap (see Hall and Yao, 2003, for some applications to ARCH-type models), a 'plain' restricted bootstrap and the asymptotic test based on the mixture $\mathcal{M} = \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ defined in Section 4. This section is organized as follows. In Section 5.1 we describe the model, the null hypothesis, the bootstrap and non-bootstrap test statistics and the design of the Monte Carlo experiments. In Section 5.2 we analyze the empirical rejection probabilities [ERP] of the tests under the null hypothesis. In Section 5.3 we analyse the behaviour of the test under the alternative hypothesis, in particular by discussing both raw and (pointwise) size-adjusted ERPs when the null hypothesis does not hold. In Section 5.4 we discuss the choice of the shrinkage sequence $\{c_n\}$ on our tests and compare with the choice of the length of the bootstrap samples for the 'm out of n' bootstrap.

5.1 MONTE CARLO DESIGN

The data generating process is

$$x_t = \sigma_t \eta_t, \quad \sigma_t^2 := \omega + \sum_{i=1}^{5} \alpha_i x_{t-i}^2 \ (t = 1, 2, ..., n)$$

with η_t i.i.d. $N(0,1)$ and initialized at $x_{1-q}, ..., x_0 = 0$. The null hypothesis of interest is univariate and of the form

$$\mathsf{H}_0 : \alpha_5 = 0.$$

The parameter vector can be written as $\theta := (\gamma, \beta', \delta')'$, where $\gamma := \alpha_5$, $\beta := (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, and $\delta := \omega$. All parameters are restricted to be non-negative. Observe that the number of nuisance parameters in β – that is, the parameters which may or may not be on the boundary of the parameter space – is $d_\beta = 4$. Hence, the number k of nuisance parameters on the boundary may take any value in the set $\{0, 1, 2, 3, 4\}$. Accordingly, in order to investigate properties of the proposed bootstrap test for different values of k, we consider five cases, denoted by $(\mathsf{C_k})_{k=0}^4$, and defined as follows:

$$(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = \begin{cases} (1, & 0.1, & 0.1, & 0.1, & 0.1, & \alpha_{5,0}) & (\mathsf{C_0}) \\ (1, & 0.133, & 0.133, & 0.133, & 0, & \alpha_{5,0}) & (\mathsf{C_1}) \\ (1, & 0.2, & 0.2, & 0, & 0, & \alpha_{5,0}) & (\mathsf{C_2}) \\ (1, & 0.4, & 0, & 0, & 0, & \alpha_{5,0}) & (\mathsf{C_3}) \\ (1, & 0, & 0, & 0, & 0, & \alpha_{5,0}) & (\mathsf{C_4}) \end{cases}$$

Thus, for the case $\mathsf{C_k}$ there are k nuisance parameters on the boundary (that is, equal to zero) and $d_\beta - k$ interior points. Notice that across cases we always have that $\alpha_{1,0} + ... + \alpha_{4,0} = 0.4$. As to the value of $\alpha_{5,0}$ we set $\alpha_{5,0} = 0$ under $\mathsf{H_0}$, and $\alpha_{5,0} > 0$ under the alternative.

We consider four different versions of the proposed bootstrap, depending on how the vector θ_n^* of bootstrap true values is chosen an on whether the fixed volatility bootstrap or the recursive bootstrap are selected. Specifically, we have:

(*i*) The proposed bootstrap (denoted as 'unrestricted, fixed vol.' in the following), with θ_n^* defined as

$$\theta_n^* := (0, \{\hat{\alpha}_{i,n} 1(\hat{\alpha}_{i,n} > c_n)\}_{i=1}^4, \hat{\omega}_n)',$$

see (18), and hence based on the unrestricted parameter estimates $\{\hat{\alpha}_{1,n}, ..., \hat{\alpha}_5, \hat{\omega}_n\}$; moreover, $\sigma_t^2(\theta_n^*)$ is as defined in (19) (fixed volatility bootstrap);

(*ii*) A recursive volatility version of the proposed bootstrap ('unrestricted, recursive vol.'), with θ_n^* as in (*i*), $X_t^* := \left(x_{t-1}^{*2}, ..., x_{t-5}^{*2}\right)'$ and conditional variance defined recursively, see Remark 4.4;

(*iii*) A restricted version of the proposed bootstrap ('restricted, fixed vol.'), see Remark 3.3, based on $\theta_n^* := (0, \{\tilde{\alpha}_{i,n} 1(\tilde{\alpha}_{i,n} > c_n)\}_{i=1}^4, \tilde{\omega}_n)'$, where $\tilde{\omega}_n$ and the $\tilde{\alpha}_{i,n}$'s are parameter estimates obtained with the null hypothesis imposed;

(*iv*) A recursive volatility version of (*iii*) ('restricted, recursive vol.').

For comparison, results are also reported for the classic restricted bootstrap (that is, without shrinkage), based on $\theta_n^* = \tilde{\theta}_n$, which is asymptotically valid only for $k = 0$ (no nuisance parameters on the boundary) or $k = 4$ (all nuisance parameters on the boundary), see Remark 3.4. Along with the restricted bootstrap we further consider the 'm out of n' bootstrap. We also report results for an 'infeasible' version of the asymptotic test ('infeasible asymptotic'), based on the unrealistic assumption that the

practitioner knows how many (and which) nuisance parameters are on the boundary[1]. Finally, we also report results for the an asymptotic test ('$\mathcal{M}$-based asymptotic') based on the quantiles of the $\mathcal{M}$ distribution discussed in Section 4, which is valid only for the cases where $k = 0$ or $k = q - 1 = 4$.

As to the choice of the shrinkage sequence c_n, we set $c_n = \nu n^{-\varepsilon}$, with $\varepsilon = 0.45$ and $\nu = 1.60$, such that $c_{100} = 0.195$, $c_{500} = 0.093$, and $c_{1000} = 0.068$. In this respect, we note e.g. that in case $\mathsf{C_4}$ for $n = 1000$, c_n corresponds to the approximate 98% quantile of the simulated distribution of $\hat{\alpha}_i$, for $i = 1, 2, 3, 4$. For the 'm out of n' bootstrap implementation, we set the size m_n of the bootstrap sample to $cn/\log(n)$, with $c = 1.5$. This implies that $m_n = 32$ for $n = 100$ and $m_n = 217$ for $n = 1000$. Different choices of c_n and m_n are discussed in Section 5.4.

Throughout, we use $10,000$ Monte Carlo replications while we use $B = 199$ bootstrap repetitions to approximate the distribution of the LR statistics[2,3]. Sample of size $n \in \{100, 500, 1000\}$ are considered throughout. All tests are run at the nominal 10% significance level.

5.2 Empirical rejection probabilities under the null

Table 1 reports the empirical rejection probabilities (as estimated on the $10,000$ Monte Carlo replications) under the null hypothesis, $\mathsf{H}_0 : \alpha_{5,0} = 0$, for the five cases $\mathsf{C_0}$-$\mathsf{C_4}$. As summary measures to compare the performance across cases and sample sizes, we also report the mean absolute deviation [MAD] and the root mean square error [RMSE] between the ERPs and the chosen 10% nominal level. In Table 1 we focus on the preferred versions of the shrinkage-based bootstrap and the 'm out of n' bootstrap, while results for additional cases are presented in Section 5.4, Table 4.

[Table 1 about here]

The following points can be made out of the analysis.

First, the ERPs of the different implementations of the shrinkage-based bootstrap are all remarkably close to the nominal level, even at the smaller sample sizes. Results do not change across different numbers of nuisance parameters on the boundary, *i.e.* across the five cases $(\mathsf{C_i})_{i=0}^4$.

Second, recursive bootstrap implementations of our tests perform slightly better than the corresponding fixed volatility bootstraps. This results is different from what

[1] For this asymptotic test, critical values are obtained by simulation based on samples of size $T = 20,000$.

[2] Unreported simulations show that varying the number B of bootstrap repetitions does not imply any changes in the results.

[3] Computations have been performed using Ox 8.0, see Doornik (2007). Code is available upon request.

reported in Cavaliere *et al.* (2018), where however no nuisance parameters on the boundary of the parameter space are allowed.

Third, there are no substantial differences in terms of which estimator is chosen in order to construct θ_n^*; that is, (shrinkage) unrestricted and restricted bootstraps have similar behaviour in terms of size control. While the choice of the bootstrap true parameters is indeed crucial in other testing problems (see *e.g.* Cavaliere *et al.*, 2012) and, in particular, restricted estimators tend to deliver better size control, for the testing problem considered here this is not the case.

Fourth, in terms of finite-sample size control, the proposed bootstrap tests are clearly superior to the 'm out of n' bootstrap, which is oversized for small values of k and undersized for larger values of k. Overall, the MAD and RMSE of the 'm out of n' bootstrap is approximately doubled compared to those of our shrinkage-based procedure. Similarly, the proposed bootstrap tests substantially outperform the infeasible asymptotic test based on the assumption that the limiting null distribution of LR_n is known in advance. This is an important result, as it clearly show that not only the proposed bootstrap estimates the correct limiting distribution LR_n, but it also delivers significant finite-sample refinements, even at the larger sample sizes.

Fifth, as expected, see the discussion in Remark 3.4, the standard restricted bootstrap performs well in case $\mathsf{C_0}$, where there are no nuisance parameters on the boundary. This is consistent with the theory in Cavaliere *et al.* (2017) and Cavaliere *et al.* (2018), where the parameters not restricted by the null hypothesis are all in the interior of the parameter space. Unfortunately, this bootstrap is not valid in the general case.

Sixth, for case $\mathsf{C_0}$, in terms of size our shrinkage bootstrap tests are again very similar to the restricted bootstrap tests, despite shrinkage is not required here. This shows that cost of shrinkage – when it is not needed – is actually very low. For cases $\mathsf{C_1}$-$\mathsf{C_4}$, the standard bootstrap is not mimicking the correct null distribution and its implementation leads to undersized tests[4].

Seventh, regarding the asymptotic test based on critical values from the $\mathcal{M} = \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ mixture, we observe that it tends to be somewhat conservative in small samples for the two cases where the test is asymptotically valid (case $\mathsf{C_0}$ and case $\mathsf{C_4}$). For case $\mathsf{C_1}$, with one nuisance parameter on the boundary of the parameter space, the asymptotic test becomes undersized, even for $n = 1000$, reflecting that the true limiting distribution shifts to the left in the ARCH case, see the discussion at the end of Section 2.1. In case $\mathsf{C_2}$ and case $\mathsf{C_3}$, the asymptotic test gets increasingly undersized, suggesting that the true limiting distribution also shifts to the left in these cases.

Overall, the proposed bootstrap procedure gives excellent size control, irrepectively of how many (if any) nuisance parameters are on the boundary of the parameter space.

[4]We conjecture that the fact that the standard bootstrap tests are undersized is a consequence of the correlation structure in the ARCH case. For other models with positive correlations, the standard bootstrap may be over-sized.

5.3 Empirical rejection probabilities under the alternative

We now investigate the ERPs for tests of $\mathsf{H}_0 : \alpha_5 = 0$ under the alternative $\mathsf{H}_1 : \alpha_5 = \bar{\alpha}$ where

$$\bar{\alpha} \in (0.025, 0.05, 0.1, 0.2, 0.3) .$$

The corresponding ERPs are reported in Table 2 for the five cases C_0–C_4 and for samples of size $n = 500$. In addition, in order to make the ERPs directly comparable, in Table 3 we report pointwise size-corrected rejection frequencies. These are constructed as follows: for each case under the null, $\alpha_{5,0} = 0$, we store the nominal level that would have given an ERP of 10%, and then use this nominal level for parameter combinations under the alternative, $\alpha_{5,0} > 0$. This type of size-correction is obviously infeasible in practice, but makes the ERP's directly comparable, see also Davidson and MacKinnon (2006) and Cavaliere *et al.* (2015). The last column in Table 3 corresponds to the (size-adjusted) power of the (infeasible) asymptotic test.

[Table 2 about here]

[Table 3 about here]

The following points can be made out of these tables.

First, and as expected, for all tests, power is monotonically increasing as the true $\alpha_{5,0}$ gets further away from the null hypothesis.

Second, the shrinkage device implemented in the proposed bootstrap tests does not seem to affect the power of the test. The behaviour in terms of (size-adjusted) ERPs of our tests matches the (size-adjusted) ERPs of the ($\mathcal{M}$-based and infeasible) asymptotic test. In particular, this is true even for the cases where shrinkage is not necessary (for instance, case C_0).

Third, there are no substantial power differences in terms of which estimator is chosen in order to construct θ_n^*: shrinkage with the unrestricted estimator and shrinkage with the restricted estimator deliver bootstrap tests with similar behaviour in terms of ERPs under the alternative hypothesis. While in other testing problems the use of unrestricted estimators tend to deliver better power, this is not the case here. A possible explanation is that for both our restricted and unrestricted bootstraps we set $\gamma = 0$ in the bootstrap DGP – that is, we impose the null hypothesis on the bootstrap sample. Hence, our shrinkage bootstrap based on $\hat{\theta}_n$ differs from a standard, unrestricted bootstrap, where $\gamma = \hat{\gamma}_n$ and the null hypothesis to be tested on the bootstrap sample is $\tilde{\mathsf{H}}_0 : \gamma = \hat{\gamma}$; *cf.* Hall (1992).

Fourth, in terms of ERFs under the alternative, recursive bootstrap implementations of our tests perform slightly better than the corresponding fixed volatility bootstraps. The gap between recursive and fixed volatility bootstraps is, however, rather marginal.

In summary, the new tests show excellent power properties, with ERFs almost identical to those of the infeasible LR test based on the unrealistic assumption that the practitioners knows which nuisance parameters are on the boundary of the parameter space.

5.4 Choice of the tuning parameters

We conclude this section with a brief analysis on the choice of the shrinkage sequence c_n used to construct the bootstrap true values. More specifically, in order to investigate the effect of the choice of c_n we set, as done earlier in this section, $c_n := \nu n^{-\varepsilon}$, with $\varepsilon = 0.45$. The tuning parameter ν is now chosen in the set $\mathcal{V} := \{0.2, 0.4, 0.8, 1.2, 1.6, 2.0\}$ (recall that the results in sections 5.2 and 5.3 are based on $\nu = 1.6$). With this choice of $\mathcal{V}$ we are able to cover quantiles of the distribution of $\hat{\alpha}_i$ for $i = 1, 2, 3, 4$ from approximately 60% to 99%.

We also consider the choice of the length of the bootstrap sample for the 'm out of n' bootstrap implementation. Here we set, as before, $m_n := cn/\log(n)$ with the tuning parameter c in the set $\mathcal{C} := \{1, 1.5, 2, 2.5, 3, 3.5\}$ (the results in sections 5.2 and 5.3 are based on $c = 1.5$). With this choice, m_{100} ranges from 21 to 76 while m_{1000} ranges from 144 to 506.

[Table 4 about here]

The most important point that can be made out of Table 4 is that the finite-sample behaviour of the shrinkage-based bootstrap tests under the null hypothesis is quite robust with respect to the choice of tuning parameter ν. In particular, for $v \geq 0.8$ we find no remarkable differences, for all the sample sizes n considered. For $n \leq 500$, smaller values of v implies that the cut-off point is such that virtually all bootstrap parameter values are not set to zero corresponding to no shrinkage. As a result, the tests tend to behave as the standard restricted bootstrap and therefore can be slightly undersized. In general, our bootstrap test tends to outperform the 'm out of n' bootstrap across different values of c and ν.

6 Conclusions

Testing whether a subset of the parameters lie on the boundary of the parameter space is a classic inference problem in statistics and econometrics. The 'parameter on the boundary problem' is particularly important for economics, where most models involve

parameters restricted by some inequality constraints; see e.g. Chernozhucov, Hong and Tamer (2007). Chernoff (1954) was the first to notice that Wilks' classical result about the χ^2-type asymptotic distribution of likelihood ratio statistics breaks down when the true parameter is a boundary point. Andrews (1999, 2001) provide a comprehensive framework for dealing with estimation with parameters on the boundary and testing that a subset of the parameters is on the boundary. While dealing with very general econometric models, parameter spaces and restrictions, a maintained assumption which is required in order to obtain feasible tests is that the parameters not restricted by the null hypothesis are indeed interior points (see Francq and Zakoïan, 2007, 2009). When this is not the case – as it is in most empirical applications – the asymptotic distributions of the test statistics depends on nuisance parameters which are unknown.

In this paper we have proposed a bootstrap-based approach to (LR) testing whether a subset of the parameter vector lie on the boundary of the parameter set, here defined thorough inequality constraints. The bootstrap just requires a simple, straightforward to implement, adjustment of the parameter values used to generate the bootstrap data. We have shown that our bootstrap consistently estimate the relevant asymptotic null distribution, irrespective of the number (and location) of nuisance parameters on the boundary. Under the alternative, the associated bootstrap statistics are bounded in probability, hence making the bootstrap test consistent.

Validity of the bootstrap for 'parameter on the boundary' problems is far from being expected. In particular, even in simple econometric models the classic (unrestricted) bootstrap fails to mimic the correct asymptotic distributions (Andrews, 2000). Other bootstraps such as the restricted bootstrap works only in the special case where there are no further parameters on the boundary (Cavaliere *et al.*, 2017, 2018). In this respect, our results unexpectedly show that the bootstrap may indeed be an extremely powerful device in econometric models featuring parameters on the boundary.

In the paper we have also shown how our results can be applied to the classic problem of inference in ARCH models subject to non-negativity parameter constraints; that is, testing significance of one ARCH coefficient when there is uncertainty about the nullity of the remaining parameters. There are many further open problems in the literature that may be analyzed in our framework.

In the application to ARCH we have focused on a single parameter constraint d_γ, but the analysis can be extended to tests on a general subvector of parameters. For instance, consider the ARCH(22) for daily returns $x_t := \eta_t \sqrt{\omega + \sum_{i=1}^{22} \alpha_i x_{t-i}^2}$, η_t i.i.d.(0,1) and, in the spirit of the HARCH model of Corsi (2007), suppose that interest is in the null hypothesis $\mathsf{H}_0 : \bigcup_{i \neq 1,5,22} \alpha_i = 0$, which implies that the only relevant ARCH parameters are those corresponding to the daily ($i = 1$), weekly ($i = 5$) and monthly ($i = 22$) frequencies. The asymptotic distribution of the LR test for H_0 depends on α_i, $i = 1, 5, 22$ being on the boundary or not. The implementation of our bootstrap test allow inference without prior knowledge of the location of these three parameters.

Another important application is within the PARX class of models of Agosto, Cav-

aliere, Kristensen and Rahbek (2016), which assumes that the behaviour of a count variable y_t over time can be described by a Poisson random variable, with intensity λ_t measurable with respect to the past information set and given by

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} + \sum_{k=1}^{r} \gamma_k x_{kt},$$

where the (exogenous) regressors x_{kt}'s, as well as the α_i's, β_j's and γ_k's are all non-negative. The outcome of an asymptotic test on *any* of the parameters depends on the location of the remaining parameters (and, in particular, on whether they are boundary points or not). Our bootstrap approach circumvents this problem and allow inference without making unrealistic assumption on the location of the unknown parameters.

There are obviously further extensions of our work which are left open for future research. For instance, we have here focused on parameters spaces defined through non-negativity constraints. The case of general linear and nonlinear restriction is indeed important and deserves further investigations. We conjecture that versions of the bootstrap defined here would apply to the general case.

References

Agosto A., G. Cavaliere, D. Kristensen and A. Rahbek (2016): "Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)", *Journal of Empirical Finance*, 38(B), 640–663.

Andrews, D.W.K. (1999): "Estimation when a Parameter is on a Boundary", *Econometrica*, 67(6), 1341–1383.

—— (2000): "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space", *Econometrica*, 68(2), 399–405.

—— (2001): "Testing when a Parameter is on the Boundary of the Maintained Hypothesis", *Econometrica*, 69(3), 683–734.

Andrews, D.W.K. and M. Buchinsky (2000): "A Three-Step Method for Choosing the Number of Bootstrap Repetitions", *Econometrica*, 68(1), 23–51.

Andrews, D.W.K. and P. Guggenberger (2009): "Hybrid and Size-Corrected Subsampling Methods", *Econometrica*, 77(3), 721–762.

Beran, R (1997): "Diagnosing Bootstrap Success", *Annals of the Institute of Statistical Mathematics,* 49, 1–24.

Beutner E., A. Heinemann and S. Smeekes (2018): "A Residual Bootstrap for Conditional Value-at-Risk", arXiv:1808.09125.

Bickel, Peter J.; C.A.J. Klaassen, Y. Ritov and J.A. Wellner (1998): "Efficient and Adaptive Estimation for Semiparametric Models". Springer: New York.

Cavaliere, G., H.B. Nielsen and A. Rahbek (2015): "Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models", *Econometrica,* 83(2), 813–831.

—— (2017): "On the Consistency of Bootstrap Testing for a Parameter on the Boundary of the Parameter Space", *Journal of Time Series Analysis* 38, 513–534.

Cavaliere, G., R.S. Pedersen and A. Rahbek (2018): "The Fixed Volatility Bootstrap for a Class of ARCH(q) Models", *Journal of Time Series Analysis,* 39, 920–941.

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2012): "Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models", *Econometrica,* 80, 1721–1740.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio", *Annals of Mathematical Statistics,* 54, 573–578.

Chernozhukov V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica,* 75, 1243–1284.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility", *Journal of Financial Econometrics,* 7, 174–196.

Corradi, V. and E.M. Iglesias (2008): "Bootstrap Refinements for QML Estimators of the GARCH(1,1) Parameters", *Journal of Econometrics,* 144, 500–510.

Davidson R. and E. Flachaire (2008): "The Wild Bootstrap, Tamed at Last", *Journal of Econometrics,* 146, 162–169.

Davidson R. and J. MacKinnon (2006): "The Power of Bootstrap and Asymptotic Tests", *Journal of Econometrics*, 133, 421–441.

Demos A. and E. Sentana (1998): "Testing for GARCH Effects: A One-Sided Approach", *Journal of Econometrics*, 86, 97—127.

Doornik J. (2007): *Object Oriented Matrix Programming Using Ox*, 3rd Edition London: Timberlake Consultants Press.

Elliott, G., U. Müller, and M. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis", *Econometrica*, 83(2), 771–811.

FRANCQ C. AND J.M. ZAKOÏAN (2007): "Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero", *Stochastic Processes and their Applications*, 117(9), 1165–1372.

FRANCQ C. AND J.M. ZAKOÏAN (2009): "Testing the Nullity of GARCH Coefficients: Correction of the Standard Tests and Relative Efficiency Comparisons", *Journal of American Statistical Association*, 104, 313–324.

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*. Springer: Berlin.

HALL, P. AND Q. YAO (2003): "Inference in ARCH and GARCH Models with Heavy-Tailed Errors", *Econometrica*, 71, 285–317.

HIDALGO, J. AND P. ZAFFARONI (2007) "A goodness-of-fit Test for ARCH(∞) Models", *Journal of Econometrics*, 141, 973–1013.

HOROWITZ, J. (2001): "The bootstrap", In Heckman J.J. and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, ch. 52, 3159–3228.

JEONG, M. (2017): "Residual-Based GARCH Boostrap and Second Order Asymptotic Refinement", *Econometric Theory*, 33, 779–790.

JENSEN S.T AND A. RAHBEK (2004): "Asympotic Inference for Non-Stationary GARCH", *Econometric Theory*, 30, 1203–1226.

KETZ P. (2018): "Subvector Inference When the True Parameter Vector May Be Near or At the Boundary", *Journal of Econometrics,* First published online on 05 September 2018.

KOPYLEV L. AND B. SINHA (2010): "On the Asymptotic Distribution of Likelihood Ratio Test when Parameters lie on the Boundary", Technical Report, Department of Mathematics and Statistics, UMBC.

—— (2011): "On the Asymptotic Distribution of Likelihood Ratio Test when Parameters lie on the Boundary", *Sankhyā B,* 73, 20–41.

MCCLOSKEY A. (2017): "Bonferroni-Based Size-Correction for Nonstandard Testing Problems", *Journal of Econometrics,* 200, 17–35.

PEDERSEN, R.S. AND A. RAHBEK (2018): "Testing GARCH-X Type Models", *Econometric Theory,* First published online on 18 October 2018.

SILVAPULLE M.J. AND P.K. SEN (2005): *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. John Wiley & Sons: New York.

SWENSEN A.R. (2004): "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models", *Econometrica*, 74, 1699–1714.

Appendix

This appendix is organized as follows. In Section A we present the proofs of our main general results; that is Theorem 1 and the related Lemma 1. In Section B we provide the proofs of the lemmas used to prove bootstrap validity for the ARCH model.

A Proofs of general results

A.1 Proof of Theorem 1

By definition, the bootstrap likelihood ratio statistic is given by

$$LR_n^* = 2(L_n^*(\hat{\theta}_n^*) - L_n^*(\tilde{\theta}_n^*)) = 2(L_n^*(\hat{\theta}_n^*) - L_n^*(\theta_n^*) - (L_n^*(\tilde{\theta}_n^*) - L_n^*(\theta_n^*)).$$

Next, as in Andrews (2001, eq.(3.3)) expand the bootstrap likelihood function as follows,

$$\begin{aligned} L_n^*(\theta) - L_n^*(\theta_n^*) &= \frac{\partial L_n^*(\theta_n^*)}{\partial \theta'}(\theta - \theta_n^*) + \frac{1}{2}(\theta - \theta_n^*)' \frac{\partial^2 L_n^*(\theta_n^*)}{\partial \theta \partial \theta'}(\theta - \theta_n^*) + R_n^*(\theta) \\ &= \frac{1}{2} Z_n^{*\prime} J_n^* Z_n^* - \frac{1}{2} q_n^*[n^{1/2}(\theta - \theta_n^*)] + R_n^*(\theta), \end{aligned}$$

with $J_n^* := -n^{-1}\frac{\partial^2 L_n^*(\theta_n^*)}{\partial \theta \partial \theta'}$, $Z_n^* := n^{-1/2}(J_n^*)^{-1}\frac{\partial L_n^*(\theta_n^*)}{\partial \theta}$ and

$$q_n^*[\lambda] := (\lambda - Z_n^*)' J_n^* (\lambda - Z_n^*).$$

Furthermore, due to Assumption 5, it holds as in Andrews (2001, Lemma 1), that $\sqrt{n}(\hat{\theta}_n^* - \theta_n^*)$ and$\sqrt{n}(\tilde{\theta}_n^* - \theta_n^*)$ are $O_p^*(1)$, in probability. This together with Assumption 5(*iii)*, implies that $R_n^*(\theta) = o_p^*(1)$, in probability, for $\theta = \hat{\theta}_n^*, \tilde{\theta}_n^*$.

By Assumption 5, it follows by Andrews (2001, proof of Theorem 4(a)) that the bootstrap likelihood ratio statistic is given by,

$$LR_n^* = q_n^*[n^{1/2}(\tilde{\theta}_n^* - \theta_n^*)] - q_n^*[n^{1/2}(\hat{\theta}_n^* - \theta_n^*)] + o_{p^*}(1),$$

in probability. Next,

$$\begin{aligned} q_n^*[n^{1/2}(\hat{\theta}_n^* - \theta_n^*)] &= \left\| n^{1/2}(\hat{\theta}_n^* - \theta_n^*) - Z_n^* \right\|_{J_n^*}^2 = \inf_{\theta \in \Theta} \left\| n^{1/2}(\theta - \theta_n^*) - Z_n^* \right\|_{J_n^*}^2 + o_p^*(1) \\ &= n \inf_{\theta \in \Theta} \left\| \theta - \theta_n^* - Z_n^* n^{-1/2} \right\|_{J_n^*}^2 + o_p^*(1) \\ &= n \inf_{\lambda \in \Theta - \theta_0^\dagger} \left\| \lambda - W_n^* n^{-1/2} \right\|_{J_n^*}^2 + o_p^*(1), \end{aligned}$$

in probability, where

$$W_n^* = n^{1/2}(\theta_n^* - \theta_0^\dagger) + Z_n^*. \tag{A.1}$$

Let $\Lambda^\dagger$ be defined as $\Lambda^\dagger := \Lambda_\gamma \times \Lambda_\beta^\dagger \times \Lambda_\delta$, where $\Lambda_\beta^\dagger = \Lambda_1^\dagger \times .... \times \Lambda_{d_\beta}^\dagger$, with $k^\dagger$ of the $\Lambda_i^\dagger$'s equal to $\mathbb{R}_+$ and the remaining $d_\beta - k^\dagger$ equal to $\mathbb{R}$. Under H_0, $k^\dagger = k$ and $\Lambda^\dagger = \Lambda$ of Assumption 2.

By Silvapulle and Sen (2005, Corollary 4.7.5), Assumption 2 and Lemma A.1*(i)*, which we give at the end of this section, then as in Andrews (2001, Lemma 7),

$$\inf_{\lambda \in \Theta - \theta_0^\dagger} \left\| \lambda - W_n^* n^{-1/2} \right\|_{J_n^*}^2 = \inf_{\lambda \in \Lambda^\dagger} \left\| \lambda - W_n^* n^{-1/2} \right\|_{J_n^*}^2 + o_p^*(n^{-1}),$$

in probability, such that

$$q_n^*[n^{1/2}(\hat{\theta}_n^* - \theta_n^*)] = \inf_{\lambda \in \Lambda^\dagger} \|\lambda - W_n^*\|_{J_n^*}^2 + o_p^*(1),$$

in probability.

For any given $k^\dagger$, without loss of generality consider the partition $\beta_0^\dagger = (\beta_0^{(k^\dagger)\prime}, \beta_0^{(d_\beta - k^\dagger)\prime})'$ with $\beta_{0,i}^{(k^\dagger)} = 0$ for $i = 1, ..., k^\dagger$ and $\beta_{0,i}^{(d_\beta - k^\dagger)} > 0$ for $i = k^\dagger + 1, ..., d_\beta$. Likewise, let $\Lambda_{\beta_k^\dagger} = \mathbb{R}_+^{k^\dagger}$ denote the part of the cone $\Lambda_\beta^\dagger$ corresponding to the boundary points $\beta_0^{(k^\dagger)}$, and let $H_{k^\dagger}$ denote the selection matrix of dimension $(d_\gamma + k^\dagger) \times d_\theta$ such that

$$H_{k^\dagger}\theta = (\gamma', \beta^{(k^\dagger)\prime})'. \tag{A.2}$$

Then, as in Andrews (2001, proof of Theorem 2(b)), it holds that, in probability,

$$\begin{aligned} q_n^*[n^{1/2}(\hat{\theta}_n^* - \theta_n^*)] &= \inf_{\lambda \in \Lambda^\dagger} \|\lambda - W_n^*\|_{J_n^*}^2 + o_p^*(1) \\ &= \inf_{\lambda \in \Lambda_\gamma \times \Lambda_{\beta_k}^\dagger} \|\lambda - H_{k^\dagger} W_n^*\|_{(H_{k^\dagger} J_n^{*-1} H_{k^\dagger}')^{-1}}^2 + o_p^*(1). \end{aligned}$$

Analogously, for the restricted bootstrap estimator we have that

$$q_n^*[n^{1/2}(\tilde{\theta}_n^* - \theta_n^*)] = \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_{\beta_k^\dagger}^\dagger} \|\lambda - H_{k^\dagger} W_n^*\|_{(H_{k^\dagger} J_n^{*-1} H_{k^\dagger}')^{-1}}^2 + o_p^*(1),$$

in probatility.

Collecting terms, it holds by Lemma A.1*(ii)* that, in probability,

$$\begin{aligned} LR_n^* &= q_n^*[n^{1/2}(\tilde{\theta}_n^* - \theta_n^*)] - q_n^*[n^{1/2}(\hat{\theta}_n^* - \theta_n^*)] + o_{p^*}(1) \\ &= \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_{\beta_k^\dagger}^\dagger} \|\lambda - H_{k^\dagger} W_n^*\|_{(H_{k^\dagger} J_n^{*-1} H_{k^\dagger}')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_{\beta_k}^\dagger} \|\lambda - H_{k^\dagger} W_n^*\|_{(H_{k^\dagger} J_n^{*-1} H_{k^\dagger}')^{-1}}^2 + o_p^*(1) \\ &= \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_{\beta_k^\dagger}^\dagger} \|\lambda - H_{k^\dagger} Z^*\|_{(H_{k^\dagger} \Omega^{*-1} H_{k^\dagger}')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_{\beta_k^\dagger}^\dagger} \|\lambda - H_{k^\dagger} Z^*\|_{(H_{k^\dagger} \Omega^{*-1} H_{k^\dagger}')^{-1}}^2 + o_p^*(1). \end{aligned}$$

Under H_0, it follows that $LR_n^* \xrightarrow{w^*}_p \mathcal{L}$ as claimed, since under H_0, $k^\dagger = k$, $\Lambda^\dagger = \Lambda$, $\Omega^* = \Omega, \Sigma^* = \Sigma$ and, by an application of the just given arguments from Andrews (2001, Theorem 2(b)),

$$\mathcal{L} = \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_\beta} \|\lambda - HZ\|_{(H\Omega^{-1}H')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_\beta} \|\lambda - HZ\|_{(H\Omega^{-1}H')^{-1}}^2$$

$$= \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_{\beta_k}} \|\lambda - H_k Z\|^2_{\left(H_k \Omega^{-1} H_k'\right)^{-1}} - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_{\beta_k}} \|\lambda - H_k Z\|^2_{\left(H_k \Omega^{-1} H_k'\right)^{-1}}.$$

Finally under the alternative, $LR_n^* \xrightarrow{w^*}_p \mathcal{L}^\dagger$, with

$$\mathcal{L}^\dagger = \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_{\beta_k^\dagger}} \|\lambda - H_{k^\dagger} Z^*\|^2_{(H_{k^\dagger} \Omega^{*-1} H_{k^\dagger}')^{-1}} - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_{\beta_k^\dagger}} \|\lambda - H_{k^\dagger} Z^*\|^2_{(H_{k^\dagger} \Omega^{*-1} H_{k^\dagger}')^{-1}}. \tag{A.3}$$

This completes the proof. □

LEMMA A.1 *With W_n^* defined in (A.1), θ_n^* satisfying Assumption 4, then under Assumption 5, (i) $W_n^* = O_p^*(1)$, in probability, and (ii) $H_{k^\dagger} W_n^* \xrightarrow{w^*}_p H_{k^\dagger} Z^*$, where $H_{k^\dagger}$ is given by (A.2).*

PROOF. Recall that $W_n^* = n^{1/2}(\theta_n^* - \theta_0^\dagger) + Z_n^*$. The result in *(i)* follows as, by Assumption 4, $n^{1/2}(\theta_n^* - \theta_0^\dagger) = O_p(1)$ and, by Assumption 5, $Z_n^* = O_p^*(1)$, in probability. Turning to *(ii)*, $H_{k^\dagger} W_n^* = H_{k^\dagger} n^{1/2}(\theta_n^* - \theta_0^\dagger) + H_{k^\dagger} Z_n^*$, where by definition of $H_{k^\dagger}$ and Assumption 4, it follows that $H_{k^\dagger} n^{1/2}(\theta_n^* - \theta_0^\dagger) = o_p(1)$. By Assumption 5, $H_{k^\dagger} Z_n^* \xrightarrow{w^*}_p H_{k^\dagger} Z^*$, and the result holds by an application of a bootstrap version of Slutzky's Lemma. □

A.2 PROOF OF LEMMA 1

By Assumption 1, it follows by Andrews (2001, Lemma 1) that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. As in this case $\theta_0^\dagger = (\gamma_0^\dagger, \beta_0^{\dagger\prime}, \delta_0^{\dagger\prime})' = (0_{d_\gamma}', \beta_0', \delta_0')'$, we have $\sqrt{n}(\delta_n^* - \delta_0^\dagger) = O_p(1)$, while $\sqrt{n}(\hat{\beta}_{n,i} - \beta_{0,i}^\dagger) = O_p(1)$, for $i = 1, ..., d_\beta$. It remains to show that (14) in Assumption 4 holds.

Suppose first that $\beta_{0,i}^\dagger = 0$. By (15), $\hat{\beta}_{n,i} c_n^{-1} = n^{1/2} \hat{\beta}_{n,i} (n^{1/2} c_n)^{-1} = O_p(1) o(1) = o_p(1)$. Hence for any $\varepsilon > 0$,

$$P(\mathbb{I}(\hat{\beta}_{n,i} > c_n) > \varepsilon) \leq P(\mathbb{I}(\hat{\beta}_{n,i} > c_n) = 1) = P(\hat{\beta}_{n,i} > c_n) = P(\hat{\beta}_{n,i}/c_n > 1) \to 0,$$

and we have that $\mathbb{I}(\hat{\beta}_{n,i} > c_n) = o_p(1)$. Hence, for $\beta_{0,i}^\dagger = 0$,

$$\sqrt{n}(\beta_{n,i}^* - \beta_{0,i}^\dagger) = \sqrt{n}\beta_{n,i}^* = \sqrt{n}\hat{\beta}_{n,i}\mathbb{I}(\hat{\beta}_{n,i} > c_n) = O_p(1) o_p(1) = o_p(1).$$

Suppose next that $\beta_{0,i}^\dagger > 0$, and note that

$$\sqrt{n}(\beta_{n,i}^* - \beta_{0,i}^\dagger) = \sqrt{n}(\hat{\beta}_{n,i} - \beta_{0,i}^\dagger)\mathbb{I}(\hat{\beta}_{n,i} > c_n) - \sqrt{n}\beta_{0,i}^\dagger \mathbb{I}(\hat{\beta}_{n,i} \leq c_n). \tag{A.4}$$

It holds that $n^{1/2}(\hat{\beta}_{n,i} - \beta_{0,i}^\dagger)/(n^{1/2} c_n) = o_p(1)$ and $\beta_{0,i}^\dagger / c_n \to \infty$, such that for any $\varepsilon > 0$

$$P(n^{1/2}(\hat{\beta}_{n,i} - \beta_{0,i}^\dagger)(n^{1/2} c_n)^{-1} + \beta_{0,i}^\dagger / c_n > \varepsilon) \to 1,$$

i.e. $n^{1/2}(\hat{\beta}_{n,i} - \beta^{\dagger}_{0,i})(n^{1/2}c_n)^{-1} + \beta^{\dagger}_{0,i}/c_n$ diverges to ∞. Hence, for any $\varepsilon > 0$

$$\begin{aligned} P((1 - \mathbb{I}(\hat{\beta}_{n,i} > c_n)) > \varepsilon) &= P(\mathbb{I}(\hat{\beta}_{n,i} \leq c_n) > \varepsilon) \leq P(\hat{\beta}_{n,i} \leq c_n) \\ &= P(n^{1/2}(\hat{\beta}_{n,i} - \beta^{\dagger}_{0,i})(n^{1/2}c_n)^{-1} \leq 1 - \beta^{\dagger}_{0,i}/c_n) \\ &= P(n^{1/2}(\hat{\beta}_{n,i} - \beta^{\dagger}_{0,i})(n^{1/2}c_n)^{-1} + \beta^{\dagger}_{0,i}/c_n \leq 1) \\ &= 1 - P(n^{1/2}(\hat{\beta}_{n,i} - \beta^{\dagger}_{0,i})(n^{1/2}c_n)^{-1} + \beta^{\dagger}_{0,i}/c_n > 1) \to 0, \end{aligned}$$

and we have that $\mathbb{I}(\hat{\beta}_{n,i} > c_n) - 1 = o_p(1)$. We conclude that $\sqrt{n}(\hat{\beta}_{n,i} - \beta^{\dagger}_{0,i})\mathbb{I}(\hat{\beta}_{n,i} > c_n) = O_p(1)$, so in light of (A.4), it remains to show that $\sqrt{n}\beta^{\dagger}_{0,i}\mathbb{I}(\hat{\beta}_{n,i} \leq c_n) = O_p(1)$. Note that for any $\varepsilon > 0$, by similar arguments as above,

$$P(\sqrt{n}\mathbb{I}(\hat{\beta}_{n,i} \leq c_n) > \varepsilon) \leq P(\mathbb{I}(\hat{\beta}_{n,i} \leq c_n) = 1) = P(\hat{\beta}_{n,i} \leq c_n) \to 0,$$

and we have that $\sqrt{n}\mathbb{I}(\hat{\beta}_{n,i} \leq c_n) = o_p(1)$. We conclude that for $\beta^{\dagger}_{0,i} > 0$, $\sqrt{n}(\beta^{*}_{n,i} - \beta^{\dagger}_{0,i}) = O_p(1)$. □

B Proofs of Lemmata 2-4

Throughout this section, we make use of the following notation and results. First, we let

$$z_t := (x^2_{t-q}, x^2_{t-1}, ..., x^2_{t-(q-1)}, 1)', \tag{B.5}$$

such that with $k > 0$, $E[\|z_t\|^k] < \infty$ if $E[x^{2k}_t] < \infty$. Second, with $\sigma^2_t(\theta) := \theta' z_t$, and for any $\theta, \tilde{\theta} \in \Theta$, it holds that

$$\frac{\sigma^2_t(\theta)}{\sigma^2_t(\tilde{\theta})} \leq \omega_L^{-1}\|\theta\|\|z_t\|.$$

Finally, suppose that $E[x^{2k}_t] < \infty$ with $k > 0$. Then,

$$E\left(\frac{\sigma^2_t(\theta)}{\sigma^2_t(\tilde{\theta})}\right)^k < \infty$$

for any $\theta, \tilde{\theta} \in \Theta$.

B.1 Proof of Lemma 2

We initially consider the convergence of $\hat{\theta}^{*}_n$. Recall that the bootstrap true value is given by $\theta^{*}_n = (0, \{\hat{\alpha}_{i,n}1(\hat{\alpha}_{i,n} > c_n)\}^{q-1}_{i=1}, \hat{\omega}_n)$. Under Assumption 3 and the stated condition on $\{c_n\}$, Lemma 1 applies such that

$$\theta^{*}_n \to_p \theta^{\dagger}_0 = (0, \{\alpha_{i,0}\}^{q-1}_{i=1}, \omega_0) \tag{B.6}$$

under H_0 and under the alternative, where, in particular, $\theta^{\dagger}_0 = \theta_0$ under H_0.

We now prove that the bootstrap unrestricted estimator is consistent for $\theta^{\dagger}_0$; that is, $\hat{\theta}^{*}_n \xrightarrow{p^*}_p \theta^{\dagger}_0$. Consistency of the bootstrap restricted estimator $\tilde{\theta}^{*}_n$ follows using similar

arguments but with Θ replaced by Θ_{H_0} in (3) with $d_\gamma = 1$. The proof consists of two steps. First, we show the uniform convergence result

$$\sup_{\theta\in\Theta} |n^{-1}L_n^*(\theta) - M(\theta)| \xrightarrow{p^*}_p 0,$$

where $M(\cdot)$ is an asymptotic estimating function given by

$$M(\theta) = -\tfrac{1}{2}E\left[\log\sigma_t^2(\theta) + \tfrac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)}\right].$$

Second, we show that identification in terms of $M(\cdot)$ applies; that is, for any $\theta \in \Theta$, we have that $M(\theta_0^\dagger) \geq M(\theta)$, with equality if and only if $\theta = \theta_0^\dagger$.

UNIFORM CONVERGENCE. Consider the following inequality

$$\begin{aligned}\sup_{\theta\in\Theta} |n^{-1}L_n^*(\theta) - M(\theta)| &\leq \sup_{\theta\in\Theta} |n^{-1}L_n^*(\theta) - E^*[n^{-1}L_n^*(\theta)]| + \sup_{\theta\in\Theta} |E^*[n^{-1}L_n^*(\theta)] - M(\theta)| \\ &=: T_{1,n}^* + T_{2,n},\end{aligned}$$

with $T_{1,n}^*$,$T_{2,n}$ implicitly defined. We have

$$\begin{aligned}T_{1,n}^* &= \sup_{\theta\in\Theta} |n^{-1}L_n^*(\theta) - E^*[n^{-1}L_n^*(\theta)]| \\ &= \sup_{\theta\in\Theta} \left| n^{-1}\textstyle\sum_{t=1}^n \frac{1}{2}\left(\log\sigma_t^2(\theta) + \frac{x_t^{*2}}{\sigma_t^2(\theta)}\right) - n^{-1}\sum_{t=1}^n \frac{1}{2}\left(\log\sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)}\right)\right| = \sup_{\theta\in\Theta} |G_n^*(\theta)|\end{aligned}$$

where $2G_n^*(\theta) := n^{-1}\sum_{t=1}^n(\eta_t^{*2} - 1)\sigma_t^2(\theta_n^*)\sigma_t^{-2}(\theta)$. In order to show that $T_{1,n}^* \xrightarrow{p^*}_p 0$, we apply Lemma B.4 of Cavaliere, Nielsen, and Rahbek (2017) which requires establishing that, for all $\theta, \tilde{\theta} \in \Theta$,

$$G_n^*(\theta) \xrightarrow{p^*}_p 0,\ |G_n^*(\theta) - G_n^*(\tilde{\theta})| \leq B_n^*||\theta - \tilde{\theta}||, \tag{B.7}$$

where B_n^* does not depend on θ and $\tilde{\theta}$ and satisfies $E^*[B_n^*] = O_p(1)$.

Consider the first term in (B.7). By Chebychev inequality and using that $E^*[(\eta_t^{*2} - 1)(\eta_s^{*2} - 1)] = 0$ for $t \neq s$, for any $\theta \in \Theta$,

$$\begin{aligned}P^*(|G_n^*(\theta)| > \varepsilon) &\leq Cn^{-2}E^*\left[\textstyle\sum_{t=1}^n \left(\frac{(\eta_t^{*2}-1)\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)}\right)^2\right] \\ &\quad + Cn^{-2}E^*\left[\textstyle\sum_{t=1,s=1,t\neq s}^n \left(\frac{(\eta_t^{*2}-1)\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)}\right)\left(\frac{(\eta_s^{*2}-1)\sigma_s^2(\theta_n^*)}{\sigma_s^2(\theta)}\right)\right] \\ &= Cn^{-2}E^*\left[(\eta_t^{*2} - 1)^2\right]\textstyle\sum_{t=1}^n \left(\frac{\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)}\right)^2 \\ &\leq C\omega_L^{-2}n^{-2}E^*\left[(\eta_t^{*2} - 1)^2\right]\|\theta_n^*\|^2\sum_{t=1}^n \|z_t\|^2 \\ &= Cn^{-2}O_p(1)O_p(1)O_p(n) = O_p(n^{-1}),\end{aligned}$$

where the last equality holds by Lemma B.5 together with the fact that $\{x_t\}$ is ergodic with $E[x_t^4] < \infty$. Consider now the second term in (B.7). We have

$$\begin{aligned}
|G_n^*(\theta) - G_n^*(\tilde{\theta})| &= \left| \tfrac{1}{2} n^{-1} \textstyle\sum_{t=1}^n (\eta_t^{*2} - 1) \sigma_t^2(\theta_n^*) \left(\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\sigma_t^2(\tilde{\theta})} \right) \right| \\
&= \left| \tfrac{1}{2} n^{-1} \textstyle\sum_{t=1}^n (\eta_t^{*2} - 1) \sigma_t^2(\theta_n^*) \left(\frac{z_t'(\tilde{\theta} - \theta)}{\sigma_t^2(\theta)\sigma_t^2(\tilde{\theta})} \right) \right| \\
&\leq ||\theta - \tilde{\theta}|| \left\| \tfrac{1}{2} n^{-1} \textstyle\sum_{t=1}^n (\eta_t^{*2} - 1) \sigma_t^2(\theta_n^*) \left(\frac{1}{\sigma_t^2(\theta)\sigma_t^2(\tilde{\theta})} \right) z_t \right\| \\
&\leq ||\theta - \tilde{\theta}|| \tfrac{1}{2} \omega_L^{-2} n^{-1} \textstyle\sum_{t=1}^n \left\| (\eta_t^{*2} - 1) \sigma_t^2(\theta_n^*) z_t \right\| =: ||\theta - \tilde{\theta}|| B_n^*,
\end{aligned}$$

and it is straightforward to show that $E^*[B_n^*] = O_p(1)$, using again Lemma B.5, $\theta_n^* = O_p(1)$, and that $\{x_t\}$ is ergodic with $Ex_t^4 < \infty$.

Next, consider $T_{2,n}$. We have that

$$\begin{aligned}
T_{2,n} &= \sup_{\theta \in \Theta} |E^*[n^{-1} L_n^*(\theta)] - M(\theta)| \\
&= \sup_{\theta \in \Theta} \left| -n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)} \right) - M(\theta) \right| \\
&\leq \sup_{\theta \in \Theta} \left| -n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right) - M(\theta) \right| \\
&+ \sup_{\theta \in \Theta} \left| n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)} \right) - n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right) \right| \\
&= \sup_{\theta \in \Theta} \left| -n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right) - M(\theta) \right| \\
&+ \sup_{\theta \in \Theta} \left| n^{-1} \textstyle\sum_{t=1}^n \frac{1}{2} \left(\frac{\sigma_t^2(\theta_n^*) - \sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right) \right|,
\end{aligned}$$

where the first term tends to zero in probability by the ULLN (since $Ex_t^2 < \infty$), and the second term is bounded by

$$||\theta_n^* - \theta_0^\dagger|| \tfrac{1}{2\omega_L} n^{-1} \textstyle\sum_{t=1}^n ||z_t|| = o_p(1).$$

We conclude that $T_{2,n} \xrightarrow{p} 0$, and hence the desired result holds.

IDENTIFICATION. First, note that $M(\theta_0^\dagger) - M(\theta)$ is well-defined on Θ since $E[x_t^2] < \infty$. Then

$$\begin{aligned}
M(\theta_0^\dagger) - M(\theta) &= -\tfrac{1}{2} E \left[\log \sigma_t^2(\theta_0^\dagger) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta_0^\dagger)} \right] + \tfrac{1}{2} E \left[\log \sigma_t^2(\theta) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right] \\
&= \tfrac{1}{2} E \left[-\log \left(\frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} \right) + \frac{\sigma_t^2(\theta_0^\dagger)}{\sigma_t^2(\theta)} - 1 \right] \geq 0,
\end{aligned}$$

with equality if and only if $\sigma_t^2(\theta_0^\dagger) = \sigma_t^2(\theta)$ with probability one, which by standard arguments is true if and only if $\theta = \theta_0^\dagger$. This completes the proof. □

B.2 Proof of Lemma 3

The proof follows by Lemmas B.2 and B.3 below together with an application of the bootstrap version of Slutzky's Lemma.

Lemma B.2 *Suppose that Assumption 3 holds. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} s_t^*(\theta_n^*) \xrightarrow{w^*}_p N(0,\Sigma^*),$$

where $s_t^*(\theta_n^*) = \partial l_t^*(\theta_n^*)/\partial\theta$ *and* $\Sigma^* = E[(\partial l_t(\theta_0^\dagger)/\partial\theta)(\partial l_t(\theta_0^\dagger)/\partial\theta')] > 0$.

Proof: Recall that

$$n^{-1}L_n^*(\theta) = -\tfrac{1}{2n}\textstyle\sum_{t=1}^n \left\{\log\sigma_t^2(\theta) + \frac{x_t^{*2}}{\sigma_t^2(\theta)}\right\}, \quad \sigma_t^2(\theta) = \theta' z_t.$$

and hence that $s_t^*(\theta_n^*) = -\frac{1}{2}\left\{1-\eta_t^{*2}\right\}\sigma_t^{-2}(\theta_n^*)z_t$.

Similar to Cavaliere *et al.* (2018) the result holds by verifying, with $\mathcal{F}_t^* = \sigma(x_s^* : s = 0,...,t)$:

(i) $E^*[s_t^*(\theta_n^*)|\mathcal{F}_{t-1}^*] = 0$,

(ii) $n^{-1}\sum_{t=1}^n E^*[s_t^*(\theta_n^*)s_t^*(\theta_n^*)'|\mathcal{F}_{t-1}^*] \xrightarrow{p^*}_p \Sigma^* > 0$,

(iii) $n^{-1}\sum_{t=1}^n E^*[(\lambda' s_t^*(\theta_n^*))^2 1_{(|\lambda' s_t^*(\theta_n^*)|>\epsilon n^{1/2})}] \xrightarrow{p} 0$,

for any $\lambda \in \mathbb{R}^k$ and any $\epsilon > 0$. Condition (i) is immediate, since $E^*[\eta_t^{*2}|\mathcal{F}_{t-1}^*] = E^*[\eta_t^{*2}] = 1$. For (ii), note that $s_t^*(\theta_n^*)$ is, conditionally on the data, an independent process, and hence

$$\begin{aligned} n^{-1}\textstyle\sum_{t=1}^n E^*[s_t^*(\theta_n^*)s_t^*(\theta_n^*)'|\mathcal{F}_{t-1}^*] &= n^{-1}\textstyle\sum_{t=1}^n E^*[s_t^*(\theta_n^*)s_t^*(\theta_n^*)'] \\ &= n^{-1}\textstyle\sum_{t=1}^n \left(2\sigma_t^2(\theta_n^*)\right)^{-2} z_t z_t' E^*[\left(1-\eta_t^{*2}\right)^2] \\ &= \tfrac{1}{4}\left(n^{-1}\textstyle\sum_{t=1}^n \sigma_t^{-4}(\theta_n^*)z_t z_t'\right) E^*[\left(1-\eta_t^{*2}\right)^2]. \end{aligned}$$

By Lemma B.5, $E^*[(1-\eta_t^{*2})^2] = O_p(1)$. Moreover, as $Ex_t^6 < \infty$, $E\left[\sup_{\theta\in\Theta}\|\sigma_t^{-4}(\theta)z_t z_t'\|\right] < \infty$ and, by the ULLN for stationary and ergodic processes,

$$\sup_{\theta\in\Theta}\left\|n^{-1}\textstyle\sum_{t=1}^n \sigma_t^{-4}(\theta)z_t z_t' - E\left[\sigma_t^{-4}(\theta)z_t z_t'\right]\right\| \xrightarrow{p} 0. \tag{B.8}$$

Using (B.8), $\theta_n^* \xrightarrow{p} \theta_0^\dagger$, compactness of Θ, and continuity of $E\left[\sigma_t^{-4}(\theta)z_t z_t'\right]$ at $\theta_0^\dagger$, we have that

$$n^{-1}\textstyle\sum_{t=1}^n \sigma_t^{-4}(\theta_n^*)z_t z_t' \xrightarrow{p} E\left[\sigma_t^{-4}(\theta_0^\dagger)z_t z_t'\right],$$

and hence that (ii) holds with $\Sigma^* > 0$, since $\lambda' z_t \neq 0$ with probability one for any $\lambda \in R^{q+1}$.

Turning to (iii), $\lambda' s_t^*(\theta_n^*) = (1-\eta_t^{*2})\,\sigma_t^{-2}(\theta_n^*)\lambda' z_t$, such that

$$\begin{aligned}
&E^*\left[(\lambda' s_t^*(\theta_n^*))^2\,\mathbb{I}(|\lambda' s_t^*(\theta_n^*)| > \epsilon n^{1/2})\right] \leq Cn^{-1/2}E^*\left[(\lambda' s_t^*(\theta_n^*))^3\right] \\
&= Cn^{-1/2}\left(\sigma_t^{-2}(\theta_n^*)\lambda' z_t\right)^3 E^*\left[\left(1-\eta_t^{*2}\right)^3\right] \\
&\leq Cn^{-1/2}\sup_{\theta\in\Theta}|\sigma_t^{-2}(\theta)\lambda' z_t|^3 E^*\left[\left(1-\eta_t^{*2}\right)^3\right] \\
&\leq C|\lambda' z_t|^3 n^{-1/2}E^*\left[\left(1-\eta_t^{*2}\right)^3\right] = n^{-1/2}|\lambda' z_t|^3 O_p(1),
\end{aligned}$$

where we have used Lemma B.5. Hence, (iii) holds as $Ex_t^6 < \infty$. □

LEMMA B.3 *Suppose that Assumption 3 holds and, in addition, that* $E[x_t^8] < \infty$. *Then,*

$$-n^{-1}\partial^2 L_n^*(\theta_n^*)/\partial\theta\partial\theta' \xrightarrow{p^*}_p \Omega^*$$

where $\Omega^* = \frac{1}{2}E[\sigma_t^{-4}(\theta_0^\dagger)z_t z_t']$ *with* z_t *defined in (B.5), and* $\Omega^* = \Omega$ *under* H_0.

PROOF: Define $\Omega^* := \frac{1}{2}E[\sigma_t^{-4}(\theta_0^\dagger)z_t z_t']$, $\Omega_n^* := \frac{1}{2}n^{-1}\sum_{t=1}^n \sigma_t^{-4}(\theta_n^*)z_t z_t'$ and $J_n^*(\theta) := -n^{-1}\partial^2 L_n^*(\theta)/\partial\theta\partial\theta'$. We have

$$\|J_n^*(\theta_n^*) - \Omega^*\| \leq \|J_n^*(\theta_n^*) - \Omega_n^*\| + \|\Omega_n^* - \Omega^*\|, \tag{B.9}$$

where the second term tends to zero in probability by standard arguments using that $E[x_t^6] < \infty$ and $\theta_n^* - \theta_0^\dagger = o_p(1)$. To see that first term tends to zero, note that the result holds if for all $i, j = 1, ..., q+1$,

$$\left|-n^{-1}\partial^2 L_n^*(\theta_n^*)/\partial\theta_i\partial\theta_j - (\Omega_n^*)_{ij}\right| \xrightarrow{p^*}_p 0.$$

By definition,

$$\begin{aligned}
&-n^{-1}\partial^2 L_n^*(\theta_n^*)/\partial\theta_i\partial\theta_j - (\Omega_n^*)_{ij} \\
&= n^{-1}\textstyle\sum_{t=1}^n(\eta_t^{*2} - \frac{1}{2})\sigma_t^{-4}\left(\theta_n^*\right)[z_t z_t']_{i,j} - \frac{1}{2}n^{-1}\sum_{t=1}^n\sigma_t^{-4}\left(\theta_n^*\right)[z_t z_t']_{i,j} \\
&= n^{-1}\textstyle\sum_{t=1}^n\left(\eta_t^{*2} - 1\right)\sigma_t^{-4}\left(\theta_n^*\right)[z_t z_t']_{i,j},
\end{aligned}$$

and by arguments similar to the ones given in Cavaliere *et al.* (2018, Proof of Lemma A.8), using that $Ex_t^8 < \infty$, we have that $\|J_n^*(\theta_n^*) - \Omega_n^*\|$ is $o_{p^*}(1)$, in probability, and hence $\|J_n^*(\theta_n^*) - \Omega^*\| = o_{p^*}(1)$, in probability. Finally, observe that under H_0, $\theta_0^\dagger = \theta_0$, such that $\Omega = -E\left[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'\right] = \frac{1}{2}E[\sigma_t^{-4}\left(\theta_0\right)z_t z_t'] = \Omega^*$.

□

B.3 PROOF OF LEMMA 4

It holds that for any $i, j, k = 1, ..., q+1$,

$$n^{-1}\frac{\partial^3 L_n^*\left(\theta\right)}{\partial\theta_i\partial\theta_j\partial\theta_k} = n^{-1}\sum_{t=1}^n\left(3\frac{x_t^{*2}}{\sigma_t^2(\theta)} - 1\right)\frac{z_{t,i}z_{t,j}z_{t,k}}{\sigma_t^6(\theta)}.$$

Using that $x_t^{*2} = \eta_t^{*2}\sigma_t^2(\theta_n^*) = \eta_t^{*2}\theta_n^{*\prime}z_t$ with z_t defined in (B.5), for any $i,j,k = 1,...,q+1$,

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\frac{\partial^3 L_n^*(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| &\leq n^{-1}\sum_{t=1}^n\left(3\eta_t^{*2}\frac{\sigma_t^2(\theta_n^*)}{\sigma_t^2(\theta)}+1\right)\frac{z_{t,i}z_{t,j}z_{t,k}}{\sigma_t^6(\theta)}\\ &\leq n^{-1}\sum_{t=1}^n\left(3\omega_L^{-1}\eta_t^{*2}\|\theta_n^*\|\|z_t\|+1\right)\omega_L^{-3}z_{t,i}z_{t,j}z_{t,k}\\ &=: K_n^*,\end{aligned}$$

and it suffices to show that $K_n^* = O_p^*(1)$, in probability. Given the the stated moment conditions, this follows by an application of Markov's inequality conditionally on the original data.

B.4 Auxiliary lemmata

Lemma B.4 *Suppose that Assumption 3 holds. If $E|x_t|^k < \infty$ for some $k \geq 1$, then $n^{-1}\sum_{t=1}^n \hat{\eta}_t^k \xrightarrow{p} E\eta_t^k$.*

Proof: Note that, with z_t defined in (B.5),

$$\hat{\eta}_t^k = \frac{x_t^k}{(\sigma_t^2(\hat{\theta}_n))^{k/2}} = \frac{x_t^k}{(\hat{\theta}_n'z_t)^{k/2}} = \eta_t^k + \eta_t^k\left[\left(\frac{\theta_0'z_t}{\hat{\theta}_n'z_t}\right)^{k/2}-1\right] = \eta_t^k + g_t(\hat{\theta}_n),$$

where we set $g_t(\theta) := \eta_t^k((\theta_0'z_t(\theta'z_t)^{-1})^{k/2}-1)$, such that $g_t(\theta_0) = 0$. By the LLN for i.i.d. processes, $n^{-1}\sum_{t=1}^n \eta_t^k \xrightarrow{p} E\eta_t^k$. To show that

$$n^{-1}\textstyle\sum_{t=1}^n \eta_t^k g_t(\hat{\theta}_n) = o_p(1), \tag{B.10}$$

and, for any $\theta\in\Theta$,

$$|g_t(\theta)| \leq |\eta_t|^k + |\eta_t|^k|\theta_0'z_t|^{k/2}\omega_L^{-k/2}.$$

Hence,

$$E\sup_{\theta\in\Theta}\|g_t(\theta)\| \leq C + CE[|x_t|^k] < \infty,$$

with C denoting a generic positive constant. By the ULLN for ergodic processes,

$$\sup_{\theta\in\Theta}\left|n^{-1}\textstyle\sum_{t=1}^n g_t(\theta) - Eg_t(\theta)\right| = o_p(1)$$

and, using that $\|\hat{\theta}_n - \theta_0\| = o_p(1)$, Θ is compact, and $E[g_t(\theta)]$ is continuous at θ_0, we have that $n^{-1}\sum_{t=1}^n g_t(\hat{\theta}_n) - E[g_t(\theta_0)] = o_p(1)$, which implies (B.10). □

Lemma B.5 *Suppose that Assumption 3 holds. Then, $E^*[\eta_t^{*k}] \xrightarrow{p} E[\eta_t^k]$, for $k\in[1,6]$.*

Proof: The result follows by Lemma B.4 and the arguments given in the proof of Lemma A.11 in Cavaliere, Pedersen, and Rahbek (2018). □

TABLE 1: Empirical Size for the ARCH(5) Example.

	Shrinking-based bootstrap					Standard	Asymptotic	
	Unrestricted		Restricted		'm out of n'	restricted		
	fixed vol.	recursive	fixed vol.	recursive	bootstrap	bootstrap	$\mathcal{M}$-based	Infeasible
n	$\mathsf{C_0}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.1, 0.1, 0.1, 0.1, 0)$							
100	10.0	11.4	9.7	11.2	11.4	9.5	6.7	7.2
500	11.0	11.3	10.7	11.3	12.1	10.1	8.4	9.0
1000	10.8	11.3	10.6	11.1	12.2	10.5	9.3	9.8
	$\mathsf{C_1}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.133, 0.133, 0.133, 0, 0)$							
100	9.3	10.8	9.0	10.7	10.7	8.9	6.4	7.3
500	10.2	10.7	10.0	10.8	11.1	9.0	7.6	9.2
1000	10.1	10.5	9.9	10.4	11.0	9.1	8.2	9.7
	$\mathsf{C_2}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.2, 0.2, 0, 0, 0)$							
100	8.2	9.7	8.1	9.6	9.4	7.7	5.2	7.5
500	9.1	10.1	9.1	10.0	9.2	7.2	6.2	8.9
1000	10.0	10.3	9.9	10.4	9.0	7.6	6.9	9.5
	$\mathsf{C_3}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.4, 0, 0, 0, 0)$							
100	8.1	10.2	7.9	10.0	8.9	7.0	5.2	8.0
500	9.2	10.0	9.2	10.0	7.8	6.5	5.9	8.8
1000	10.1	10.4	10.2	10.4	8.2	6.8	6.6	9.5
	$\mathsf{C_4}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0, 0, 0, 0, 0)$							
100	8.3	10.9	7.9	10.7	9.3	7.4	7.2	7.5
500	9.1	10.1	9.1	10.0	8.5	7.1	8.2	8.7
1000	10.0	10.3	9.9	10.3	8.2	7.7	9.2	9.7
MAD	0.73	0.57	0.78	0.52	1.32	1.94	2.85	1.31
RMSE	0.98	0.73	1.05	0.67	1.44	2.22	3.12	1.60

NOTES: *Empirical rejection frequencies under the null hypothesis, $\alpha_{5,0} = 0$. The nominal level is 10%. Bootstrap p-values are based on 199 bootstrap replications and the simulation is based on 10000 Monte Carlo replications. The shrinking-based bootstrap uses $c_n = 1.6n^{-0.45}$ and the m-out-of-n bootstrap uses $m_n = 1.5n/log(n)$. The feasible asymptotic test uses the distribution $\mathcal{M}$ for all cases. The infeasible asymptotic test uses critical values simulated for each case with $T = 20000$. MAD and RMSE measure the overall deviation from the nominal level across cases and sample sizes.*

TABLE 2: Empirical Unadjusted Power for the ARCH(5) Example.

	Shrinking-based bootstrap					Standard	Asymptotic	
	Unrestricted		Restricted		'm out of n'	restricted		
	fixed vol.	recursive	fixed vol.	recursive	bootstrap	bootstrap	$\mathcal{M}$-based	Infeasible
$\alpha_{5,0}$	$\mathsf{C_0}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.1, 0.1, 0.1, 0.1, 0)$							
0	11.0	11.3	10.7	11.3	12.1	10.1	8.4	9.0
0.025	25.1	25.6	24.3	25.6	26.8	23.8	20.8	21.8
0.05	41.9	42.6	41.2	42.7	44.2	40.6	37.1	38.2
0.1	73.3	73.9	72.8	73.8	74.9	72.1	69.6	70.5
0.2	97.0	97.3	97.0	97.3	97.6	97.1	96.5	96.7
0.3	99.8	99.9	99.8	99.8	99.9	99.8	99.7	99.8
	$\mathsf{C_1}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.133, 0.133, 0.133, 0, 0)$							
0	10.2	10.7	10.0	10.8	11.1	9.0	7.6	9.2
0.025	24.0	25.2	23.6	25.2	25.8	21.9	20.1	22.6
0.05	41.6	43.0	41.1	43.0	43.5	39.4	36.9	40.1
0.1	73.5	74.5	73.1	74.4	75.0	71.9	70.0	72.3
0.2	97.3	97.5	97.3	97.5	97.7	97.2	96.7	97.1
0.3	99.8	99.9	99.8	99.9	99.9	99.8	99.8	99.8
	$\mathsf{C_2}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.2, 0.2, 0, 0, 0)$							
0	9.1	10.1	9.1	10.0	9.2	7.2	6.2	8.9
0.025	25.3	26.5	25.3	26.4	25.1	21.2	19.5	24.6
0.05	44.7	46.6	44.5	46.7	44.9	40.5	38.0	44.2
0.1	77.4	78.5	77.3	78.4	77.5	74.4	72.6	77.3
0.2	98.2	98.3	98.2	98.3	98.3	97.7	97.5	98.2
0.3	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
	$\mathsf{C_3}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.4, 0, 0, 0, 0)$							
0	9.2	10.0	9.2	10.0	7.8	6.5	5.9	8.8
0.025	29.0	30.2	28.8	30.3	26.7	23.0	21.8	28.1
0.05	51.6	52.7	51.4	52.9	49.2	44.4	43.5	50.7
0.1	82.5	83.7	82.5	83.6	81.4	78.3	78.0	82.5
0.2	98.9	99.0	98.8	99.0	98.8	98.4	98.4	98.9
0.3	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
	$\mathsf{C_4}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0, 0, 0, 0, 0)$							
0	9.1	10.1	9.1	10.0	8.5	7.1	8.2	8.7
0.025	21.9	23.5	21.6	23.4	20.1	17.7	20.9	21.5
0.05	38.3	40.0	38.0	39.9	36.2	33.1	36.6	37.7
0.1	70.1	72.3	70.0	72.3	68.0	65.4	69.3	70.2
0.2	96.8	97.1	96.8	97.1	96.3	95.9	96.7	96.9
0.3	99.8	99.8	99.8	99.8	99.7	99.6	99.8	99.8

NOTES: *Empirical rejection frequencies under the alternative hypothesis, $\alpha_{5,0} > 0$, for $n = 500$. The nominal level is 10%. Bootstrap p-values are based on 199 bootstrap replications and the simulation is based on 10000 Monte Carlo replications. The shrinking-based bootstrap uses $c_n = 1.6n^{-0.45}$ and the m-out-of-n bootstrap uses $m_n = 1.5n/log(n)$.*

TABLE 3: Empirical Size-adjusted Power for the ARCH(5) Example.

	Shrinking-based bootstrap					Standard	Asymptotic
	Unrestricted		Restricted		'm *out of* n'	restricted	Infeasible
	fixed vol.	recursive	fixed vol.	recursive	bootstrap	bootstrap	$\mathcal{M}$-based
$\alpha_{5,0}$	$\mathsf{C_0}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.1, 0.1, 0.1, 0.1, 0)$						
0	10.1	10.3	10.3	10.3	10.3	10.1	10.0
0.025	23.2	23.9	23.4	23.7	24.0	23.8	23.4
0.05	39.7	40.5	40.2	40.5	41.2	40.6	40.3
0.1	71.5	72.2	71.8	72.1	72.2	72.1	72.2
0.2	96.7	97.0	96.8	97.0	97.1	97.1	97.0
0.3	99.8	99.8	99.8	99.8	99.8	99.8	99.8
	$\mathsf{C_1}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.133, 0.133, 0.133, 0, 0)$						
0	10.2	10.2	10.0	10.3	10.0	10.4	10.0
0.025	24.0	24.3	23.6	24.2	24.0	24.4	23.7
0.05	41.6	42.0	41.1	42.0	41.7	42.2	41.5
0.1	73.5	73.8	73.1	73.6	73.5	74.2	73.5
0.2	97.3	97.4	97.3	97.4	97.3	97.5	97.4
0.3	99.8	99.8	99.8	99.9	99.9	99.9	99.9
	$\mathsf{C_2}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.2, 0.2, 0, 0, 0)$						
0	10.0	10.1	10.0	10.0	10.1	10.1	10.0
0.025	27.0	26.5	26.8	26.4	26.7	26.7	26.5
0.05	46.7	46.6	46.7	46.7	46.7	47.1	46.4
0.1	78.7	78.5	78.6	78.4	78.7	78.4	78.6
0.2	98.4	98.3	98.4	98.3	98.4	98.4	98.4
0.3	99.9	99.9	99.9	99.9	99.9	99.9	99.9
	$\mathsf{C_3}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.4, 0, 0, 0, 0)$						
0	10.2	10.5	10.2	10.0	10.1	10.1	10.0
0.025	30.9	31.3	30.7	30.3	30.6	30.7	30.4
0.05	53.3	53.8	53.0	52.9	53.4	53.0	53.2
0.1	83.6	84.2	83.5	83.6	83.7	83.6	83.8
0.2	99.0	99.0	98.9	99.0	99.0	99.0	99.1
0.3	99.9	99.9	99.9	99.9	100.0	100.0	100.0
	$\mathsf{C_4}$: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0, 0, 0, 0, 0)$						
0	10.2	10.1	10.1	10.0	10.0	10.1	10.0
0.025	23.4	23.5	23.2	23.4	22.9	23.2	23.8
0.05	40.1	40.0	40.1	39.9	39.7	39.4	40.4
0.1	71.8	72.3	71.8	72.3	71.3	71.3	72.4
0.2	97.1	97.1	97.1	97.1	96.9	97.0	97.1
0.3	99.8	99.8	99.8	99.8	99.8	99.8	99.8

NOTES: *Pointwise size-adjusted rejection frequencies under the alternative hypothesis,* $\alpha_{5,0} > 0$*, for* $n = 500$*. The nominal level is 10%. Bootstrap p-values are based on 199 bootstrap replications and the simulation is based on 10000 Monte Carlo replications. The shrinking-based bootstrap uses* $c_n = 1.6n^{-0.45}$ *and the m-out-of-n bootstrap uses* $m_n = 1.5n/log(n)$*.*

TABLE 4: Empirical Size for the ARCH(5) Example with Varying c_n and m_n.

	Shrinking-based bootstrap Unrestricted, fixed volatility, $c_n = \nu n^{-0.45}$						'm out of n' bootstrap $m_n = cn/\log(n)$				
	$\nu = 0.2$	$\nu = 0.4$	$\nu = 0.8$	$\nu = 1.2$	$\nu = 1.6$	$\nu = 2.0$	$m = 1$	$m = 1.5$	$m = 2$	$m = 2.5$	$m = 3$
n	C_0: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.1, 0.1, 0.1, 0.1, 0)$										
100	8.5	8.6	9.2	9.6	10.0	10.2	12.0	11.4	10.8	10.0	9.7
500	9.5	9.7	10.1	10.6	11.0	11.7	13.0	12.1	11.5	11.1	10.8
1000	9.8	9.8	10.0	10.3	10.8	11.3	13.0	12.2	11.9	11.3	11.1
	C_1: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.133, 0.133, 0.133, 0, 0)$										
100	8.2	8.3	8.7	9.1	9.3	9.6	11.3	10.7	10.0	9.5	9.3
500	9.1	9.2	9.5	9.9	10.2	10.5	11.9	11.1	10.5	10.0	9.8
1000	9.6	9.7	9.9	10.0	10.1	10.3	11.6	11.0	10.4	9.8	9.7
	C_2: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.2, 0.2, 0, 0, 0)$										
100	7.4	7.6	7.7	8.1	8.2	8.5	9.6	9.4	8.9	8.2	8.0
500	8.5	8.7	8.9	9.1	9.1	9.2	10.0	9.2	8.6	8.2	7.8
1000	9.3	9.4	9.7	9.9	10.0	10.0	9.8	9.0	8.7	8.2	8.1
	C_3: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0.4, 0, 0, 0, 0)$										
100	7.6	7.8	7.9	8.0	8.1	8.1	9.5	8.9	8.6	7.6	7.5
500	8.6	8.7	9.0	9.2	9.2	9.2	8.8	7.8	7.5	7.1	6.8
1000	9.4	9.6	9.9	10.1	10.1	10.2	8.6	8.2	7.7	7.4	7.4
	C_4: $(\omega_0, \alpha_{1,0}, \alpha_{2,0}, \alpha_{3,0}, \alpha_{4,0}, \alpha_{5,0}) = (1, 0, 0, 0, 0, 0)$										
100	8.0	8.1	8.2	8.3	8.3	8.3	10.5	9.3	8.4	7.8	7.6
500	9.1	9.1	9.0	9.1	9.1	9.2	9.0	8.5	8.2	7.6	7.5
1000	9.9	9.9	10.0	10.0	10.0	9.9	8.8	8.2	8.1	7.8	7.7
MAD	1.17	1.05	0.82	0.72	0.73	0.82	1.29	1.32	1.37	1.55	1.66
RMSE	1.40	1.28	1.11	0.98	0.97	1.03	1.56	1.44	1.54	1.81	1.92

NOTES: *Empirical rejection frequencies under the null hypothesis, $\alpha_{5,0} = 0$. The nominal level is 10%. Bootstrap p-values are based on 199 bootstrap replications and the simulation is based on 10000 Monte Carlo replications. MAD and RMSE measure the overall deviation from the nominal level across cases and sample sizes.*