

Bootstrap Quasi-Likelihood Ratio Tests for Nested Models

Patrice Bertail¹ and Pascal Lavergne²

¹Université Paris 10 and CREST

²Toulouse School of Economics

February 13, 2016
(*Preliminary Draft*)

Abstract

We consider *quasi-likelihood ratio* (QLR) tests for testing restrictions on parameters, based on the difference of M-estimation objective functions of the two models to be compared. We show that for an univariate restriction the QLR statistic can be rendered asymptotically pivotal. We then propose a simple bootstrap procedure that yields asymptotically valid critical values even if the model is misspecified. The method modifies the bootstrap objective function to mimic what happens when the null hypothesis is true. A Monte-Carlo study illustrates that bootstrap QLR tests provide a viable and competitive alternative to standard bootstrap Wald tests.

Keywords: Hypothesis Testing, Likelihood Ratio, Bootstrap, Misspecified Models.

JEL classification: Primary C12; Secondary C15.

Address correspondence: Pascal Lavergne, Toulouse School of Economics, Université Toulouse 1, 21 Allées de Brienne, 31000 Toulouse FRANCE. Email: pascal.lavergne@ut-capitole.fr

1 Introduction

Testing restrictions on parameters, including testing for significance, is very common in applied economics, whether to evaluate an economic hypothesis, to check a modeling assumption, or to build confidence intervals. In a maximum likelihood context, the well-known trinity of tests include the Wald test, the score or Lagrange multiplier test, and the likelihood ratio test. Each one can be generalized for more general models and estimation methods. The most popular testing approach among practitioners is the Wald one. Standard t-tests are reported by all statistical and econometric software, and in most cases robust versions are available. Bulgin on Eicker (1963), White (1980) has shown how the Wald statistic can be rendered robust to heteroscedasticity in a linear regression and White (1982) has further developed a more general way to obtain Wald and score statistics robust to potential misspecification in a maximum likelihood context. This has motivated further work on the precise properties of these robust statistics and their small sample behavior, which has been shown to be pretty sensitive to the implementation details and the design of the data, see Mackinnon and White (1985), Chesher and Jewitt (1987), and Chesher and Austin (1991) among others. Bootstrap methods have been investigated in the aim of obtaining better approximations, see in particular Wu (1986), Hall and Horowitz (1996), Horowitz (1997), Davidson and MacKinnon (1999), Andrews (2002), and Goncalves and White (2004). Kline and Santos (2012) have recently pointed out that in a regression model the popular wild bootstrap does not improve over the asymptotic approximation for the Wald statistic when the model is misspecified. By contrast, the naive bootstrap of Efron (1979) does improve on asymptotics in that situation, as shown by Hall and Horowitz (1996).

Our interest in this work focuses on likelihood ratio tests and their generalizations, labeled *quasi-likelihood ratio* (QLR) tests, based on the objective functions of M-estimators, among which maximum likelihood (ML). The QLR test does not appear to be as widely used in empirical work as Wald is. A potential explanation is that it requires to estimate both the general model and the model under the considered restrictions on parameters,

while a Wald test requires only unrestricted estimation. However, the QLR test has its own merits. In particular, it is by construction invariant to (nonlinear) transformations of the restrictions under test. By contrast, one can obtain any desired value for Wald test statistic by choosing an appropriate transformation, and the adequacy of the usual asymptotic approximation may then substantially vary, see Gregory and Veall (1985), Lafontaine and White (1986), and Phillips and Park (1988). The QLR statistic has a limiting chi-square distribution under the null hypothesis in an ML context when the parametric model is correctly specified, see Wilks (1938). But in general, and in particular under misspecification, it behaves asymptotically as a weighted sum of independent chi-squares, as show by Foutz and Srivastava (1977) and Kent (1982) for unconditional ML, Vuong (1989) for conditional ML, and Marcellino and Rossi (2008) for M-estimators. Hence, it is necessary to evaluate critical values on a case-by-case basis by numerical simulation.

This work has two main contributions. First, we explain how to recover asymptotic pivotalness for the QLR statistic when testing an *univariate* restriction. This case is of key interest in practice, as when testing for the significance of a specific parameter component, or for building a confidence interval by inverting the test. We show that this is as easy as it is to render the Wald statistic robust to misspecification. The new *adjusted* QLR used in conjunction with standard critical values from a chi-square distribution yields an asymptotically valid test. Second, we design a simple general bootstrap procedure for obtaining critical values for QLR tests that avoid reliance on the asymptotic approximation and the computational burden of simulating the critical values. Bootstrapping is a viable all purpose alternative that can perform well, though we cannot expect bootstrap to provide asymptotic refinements if the test statistic is not asymptotically pivotal. For testing univariate restrictions, bootstrapping the adjusted statistic may be expected to provide critical values that are more accurate than asymptotic ones.

It seems at first unclear how to bootstrap a QLR statistic. In a ML context, the distribution, if correctly specified, could be used together with estimated parameters to generate data following the null hypothesis. If however the model is potentially misspecified, it is dubious that generating data from an incorrect conditional distribution can be helpful.

An alternative possibility would be to use subsampling, which is valid under fairly weak assumptions, see Politis and Romano (1994), but this would inevitably entail some power loss. By contrast, we use the nonparametric (or naive) bootstrap, but to abide to the first guideline of bootstrap, namely that the resampled statistic should reflect what happens under the null hypothesis, even when it does not hold true, we rely on a simple device. We define the bootstrap estimator as the one that equates the score vector of the bootstrap objective function to the score vector of the constrained objective function for the initial sample. This way we constrain the bootstrap likelihood to mimic what happens when we impose restrictions on the original data. In essence, this is the recentering method proposed by Hall and Horowitz (1996) for t-tests based on Generalized Method of Moments estimators. These authors point out that one needs to recenter the moment conditions in overidentified models because there is generally no solution to the whole set of moment conditions. Without recentering, the bootstrap would implement a moment condition that does not hold in the population of the bootstrap samples. What is happening in our setup is pretty similar: restrictions on parameters provide overidentifying restrictions, so that the score vector evaluated at the restricted estimator is typically not zero. Therefore, we recenter the bootstrap score vector by the original score vector evaluated at the restricted estimator to mimic what happens under the null hypothesis. Hence we label our method as *bootstrap under the null hypothesis*.

While some recent work has investigated QLR tests and bootstrap methods, the connection of the recentering method with restrictions on parameters appears to have been mostly overlooked. A method explored by several authors is to compute a bootstrap statistic that does not test the original hypothesis $H_0 : r(\theta) = 0$, but rather $r(\theta) = r(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the initial estimator. Along these lines, Camponovo (2015) establishes higher-order improvements of a block bootstrap test in a dependent data context, and Spokoiny and Zhilova (2015) study the conditions of validity of the multiplier bootstrap. In a general semiparametric context, Chen and Pouzo (2012, 2015) apply this method to obtain asymptotically valid critical values when testing restrictions on parameters. Using a different approach, Kim (2003) proposes to bootstrap a LR test by using the estimated

parametric restricted model to generate bootstrap samples and establishes higher-order improvements for a well specified model. Shi (2015) focuses instead on LR and QLR tests for comparing nonnested or overlapping models.

Our paper is organized as follows. In Section 2, we recall the main properties of QLR tests for restrictions on parameters. We also explain how an asymptotically pivotal statistic can be readily obtained when testing an univariate restriction, which is pretty common in empirical work. We label the new statistic the *adjusted* QLR. In Section 3, we detail our bootstrap method and we show that it yields asymptotically valid critical values, whether the QLR statistic is adjusted or not. Section 4 contains the results of a Monte-Carlo study on the comparative small sample performances of bootstrap QLR tests. Our experiments reveal that (i) the finite sample behavior of the QLR test may not be well approximated by asymptotics under misspecification, (ii) bootstrapping is helpful in reaching a rejection probability that is close to the desired level, and (iii) the bootstrap adjusted QLR test is competitive with a robust to misspecification bootstrap Wald test.

2 Quasi-Likelihood Ratio Tests for Nested Models

Let $z \in \mathbb{R}^k$ be a random vector. and $Q(\theta) = \mathbb{E}q(z, \theta)$ a function from \mathbb{R}^p to \mathbb{R} that admits a unique minimizer θ_* on the parameter space Θ denoted as θ_* , i.e.

$$\theta_* = \arg \min_{\Theta} Q(\theta) \quad Q(\theta) = \mathbb{E}q(z, \theta).$$

For a random sample $\{z_t, t = 1, \dots, n\}$ from z , a consistent estimator of θ_* is defined as

$$\hat{\theta}_n = \arg \min_{\Theta} Q_n(\theta) \quad Q_n(\theta) = \mathbb{E}_n q(z, \theta) = n^{-1} \sum_{t=1}^n q(z_t^*, \theta).$$

When the objective function is the expected conditional log-likelihood, i.e. $q(z, \theta) = \log f(y|x, \theta)$, Gourieroux, Monfort, and Trognon (1984) show that the parameters of the conditional mean can be estimated consistently whenever it is well specified even if the likelihood itself is not. If the objective function is ordinary least-squares, i.e.

$q(z, \theta) = (y - x'\theta)^2$, then similarly one can estimate θ consistently even in the presence of heteroscedasticity of unknown form. If the conditional mean $x'\theta$ is itself misspecified, so that the linear form is only an approximation of the true conditional expectation, one can still rely on least-squares to estimate the approximation parameters and be interested in making inference on them, see White (1980).

Let $\Gamma \subset \mathbb{R}^q$, $q < p$, be the restricted parameter space and $\varphi(\cdot)$ an injective mapping from Γ to Θ . The null hypothesis of interest is

$$H_0 : \theta_* = \varphi(\gamma_*), \quad \gamma_* \in \Gamma \quad \text{against} \quad H_A : \theta_* \notin \varphi(\Gamma).$$

We focus here on restrictions in explicit form for convenience, but as is well known it is generally possible to rewrite restrictions in implicit form, see e.g. Gourieroux and Monfort (1995). Because $Q(\cdot)$ is uniquely minimized at θ_* , this yields the equivalent formulation

$$H_0 : Q(\theta_*) - Q(\varphi(\gamma_*)) = 0 \quad \gamma_* \in \Gamma \quad \text{against} \quad H_A : \min_{\gamma} Q(\theta_*) - Q(\varphi(\gamma)) > 0.$$

The M-estimator subject to the constraint that θ belongs to $\varphi(\Gamma)$ is

$$\hat{\theta}_n^0 = \arg \min_{\varphi(\Gamma)} Q_n(\theta) = \varphi(\hat{\gamma}_n) \quad \text{with} \quad \hat{\gamma}_n = \arg \min_{\Gamma} Q_n(\varphi(\gamma)).$$

The Quasi-Likelihood Ratio (QLR) statistic for testing H_0 is

$$\text{QLR}_n = 2n \left[Q_n(\hat{\theta}_n^0) - Q_n(\hat{\theta}_n) \right].$$

In a conditional ML context, the parametric likelihood ratio asymptotically follows a chi-square under H_0 for a well-specified model. Under misspecification, Vuong (1989) establishes that it generally tends to a weighted sum of independent chi-squares. Marcellino and Rossi (2008) generalize Vuong's result to the general case of M-estimation with dependent data. We here state the asymptotic results for iid data, and provide a short proof of these results.

Assumption A (a) Θ is compact, (b) $q(z, \theta)$ is twice continuously differentiable in θ , (c) $\mathbb{E} \sup_{\Theta} |q(z, \theta)| < \infty$

Assumption B (a) There exists a unique minimizer θ_* of $Q(\theta) = \mathbb{E}q(z, \theta)$ on the parameter space Θ and θ_* is interior to Θ . There exists a unique minimizer γ_* of $Q(\varphi(\gamma))$ on the parameter space Γ and γ_* is interior to Γ . (b) each element of $\nabla_{\theta}q(z, \theta)$ is uniformly bounded on a neighborhood of θ_* by a function with finite squared expectation. (c) each element of $\nabla_{\theta\theta'}q(z, \theta)$ is uniformly bounded on a neighborhood of θ_* by a function with finite expectation. (d) $A(\theta_*) = \nabla_{\theta\theta'}Q(\theta_*)$ is non singular.

Assumption C $\varphi(\cdot)$ is continuously differentiable with $H = \frac{\partial}{\partial \gamma'}\varphi(\gamma_*)$ of full rank.

We will also need the following definition.

Definition 1 Let $U = (U_1, \dots, U_m)'$ be a vector of m independent standard normal variables and $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$. The random variable $\sum_{i=1}^m \lambda_i U_i^2$ is distributed as a weighted sum of chi-squares with parameters (m, λ) and its (cumulative) distribution is denoted by $\mathcal{M}_m(\cdot, \lambda)$.

Theorem 2.1 Under Assumption C, A, and B,

(a) under H_0 ,

$$n^{1/2} \begin{bmatrix} \hat{\theta}_n - \theta_* \\ \hat{\gamma}_n - \gamma_* \end{bmatrix} = - \begin{bmatrix} A^{-1}(\theta_*) \\ (H' A(\theta_*) H)^{-1} H' \end{bmatrix} \Delta_n + o_p(1),$$

where $\Delta_n \xrightarrow{d} N(0, B(\theta_*))$, $B(\theta) = \text{Var } q(z, \theta)$.

(b) Under H_0 , $\text{QLR}_n \xrightarrow{d} \mathcal{M}_p(\cdot : \lambda_*)$, where λ_* is the vector of p eigenvalues of

$$W = B(\theta_*) \left[A^{-1}(\theta_*) - H (H' A(\theta_*) H)^{-1} H' \right].$$

(c) Under H_A , $\text{QLR}_n \xrightarrow{p} +\infty$.

Proof. (a) Under Assumption A, $\sup_{\Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$, so that $\widehat{\theta}_n - \theta_* = o_p(1)$. Similarly, $\widehat{\theta}_n^0 - \varphi(\gamma_*) = o_p(1)$. By a Taylor expansion, we have under H_0

$$\begin{aligned}\sqrt{n} \left(\widehat{\theta}_n - \theta_* \right) &= - [\nabla_{\theta\theta'} Q(\theta_*)]^{-1} \sqrt{n} \nabla_{\theta} Q_n(\theta_*) + o_p(1) \\ \sqrt{n} \left(\widehat{\gamma}_n - \gamma_* \right) &= - [H' \nabla_{\theta\theta'} Q(\theta_*) H]^{-1} \sqrt{n} H' \nabla_{\theta} Q_n(\theta_*) + o_p(1).\end{aligned}$$

Moreover, $\sqrt{n} \nabla_{\theta} Q_n(\theta_*) \xrightarrow{d} N(0, B(\theta_*))$.

(b) Using a Taylor expansion of $Q_n(\theta_*)$ and (a),

$$\begin{aligned}Q_n(\widehat{\theta}_n) &= Q_n(\theta_*) - \frac{1}{2} \left(\widehat{\theta}_n - \theta_* \right)' A(\theta_*) \left(\widehat{\theta}_n - \theta_* \right) + o_p(1/n), \\ Q_n(\varphi(\widehat{\gamma})) &= Q_n(\varphi(\gamma_*)) - \frac{1}{2} \left(\widehat{\gamma}_n - \gamma_* \right)' H' A(\theta_*) H \left(\widehat{\gamma}_n - \gamma_* \right) + o_p(1/n).\end{aligned}$$

Hence,

$$\text{QLR}_n = n \left(\widehat{\theta}_n - \theta_* \right)' A(\theta_*) \left(\widehat{\theta}_n - \theta_* \right) - n \left(\widehat{\gamma}_n - \gamma_* \right)' H' A(\theta_*) H \left(\widehat{\gamma}_n - \gamma_* \right),$$

which is a quadratic form in the asymptotically normal vector in (a), and thus converges to a weighted sum of chi-squares with parameters (p, λ_*) . From Vuong (1989), λ_* is the vector of eigenvalues of the matrix

$$\begin{bmatrix} A(\theta_*) & \mathbf{0} \\ \mathbf{0} & -H' A(\theta_*) H \end{bmatrix} \begin{bmatrix} A^{-1}(\theta_*) \\ (H' A(\theta_*) H)^{-1} H' \end{bmatrix} B(\theta_*) \begin{bmatrix} A^{-1}(\theta_*) \\ H (H' A(\theta_*) H)^{-1} \end{bmatrix},$$

which are also the ones of

$$B(\theta_*) \begin{bmatrix} A^{-1}(\theta_*) \\ H (H' A(\theta_*) H)^{-1} \end{bmatrix} \begin{bmatrix} A(\theta_*) & \mathbf{0} \\ \mathbf{0} & -H' A(\theta_*) H \end{bmatrix} \begin{bmatrix} A^{-1}(\theta_*) \\ (H' A(\theta_*) H)^{-1} H' \end{bmatrix}.$$

This equals the desired matrix.

(c) Under H_A , we have $\widehat{\theta}_n \xrightarrow{p} \theta_*$ and $\widehat{\gamma}_n \xrightarrow{p} \gamma_*$, so that

$$\text{QLR}_n = 2n [Q_n(\varphi(\gamma_*)) - Q_n(\theta_*)] + o_p(n) \xrightarrow{p} +\infty$$

under H_A . ■

When the information matrix equality $B(\theta_*) = A(\theta_*)$ holds, which happens for a correctly

likelihood model, then there exactly $r = p - q$ ones in λ_* and q zeros. Hence the LR statistic asymptotically follows a centered chi-square distribution with r degrees of freedom. Also in a linear regression model with correctly specified mean, i.e. $q(z, \theta) = (y - x'\theta)^2$, we have $B(\theta_*) = \sigma^2 A(\theta_*)$ whenever errors are homoscedastic, where $\sigma^2 = \mathbb{E} q^2(z, \theta)$, so that we can easily render QLR asymptotically pivotal. In general however, the QLR statistic has a more involved asymptotic distribution under H_0 . We now derive an alternative formula for λ_* that appear to be new.

Corollary 2.2 *Under the assumptions of Theorem 2.1, λ_* has q elements equal to zero and the remaining $r = p - q$ non-zero eigenvalues are the eigenvalues of*

$$(d' A^{-1}(\theta_*) d)^{-1} d' A^{-1}(\theta_*) B(\theta_*) A^{-1}(\theta_*),$$

where d is any $p \times r$ matrix orthogonal to H such that $[d \ H]$ is full-rank.

Proof. Since

$$\begin{aligned} W &= B(\theta_*) A^{-1/2}(\theta_*) \left[\mathbf{I} - A^{1/2}(\theta_*) H (H' A(\theta_*) H)^{-1} A^{1/2}(\theta_*) H' \right] A^{-1/2}(\theta_*) \\ &= B(\theta_*) A^{-1/2}(\theta_*) M A^{-1/2}(\theta_*), \end{aligned}$$

it has the same eigenvalues as $M A_f^{-1/2} B_f A_f^{-1/2} M$, as M is idempotent. Also

$$C = \begin{bmatrix} (H' A(\theta_*) H)^{-1/2} H' A^{1/2}(\theta_*) & \mathbf{0} \\ \mathbf{0} & (d' A^{-1}(\theta_*) d)^{-1/2} d' A^{-1/2}(\theta_*) \end{bmatrix}$$

is an orthogonal matrix. Hence W has the same eigenvalues as $C M A_f^{-1/2} B_f A_f^{-1/2} M C'$, whose non-zero eigenvalues are the ones of

$$(d' A^{-1}(\theta_*) d)^{-1/2} d' A^{-1}(\theta_*) B(\theta_*) A^{-1}(\theta_*) (d' A^{-1}(\theta_*) d)^{-1/2},$$

as M is the orthogonal projection onto the span of $A^{-1/2}(\theta_*) H$. Finally, the eigenvalues are unchanged by premultiplying by $(d' A^{-1}(\theta_*) d)^{-1/2}$ and postmultiplying by $(d' A^{-1}(\theta_*) d)^{1/2}$.

■

The result implies that we need only to determine the $r = p - q$ non-zero eigenvalues. Computing the quantiles of the QLR statistic under H_0 can be entertained by simulation on a case-by-case basis: it entails the determination of the eigenvalues of a consistent estimator of W , that can be obtained from empirical analogs of $A(\theta)$ and $B(\theta)$ evaluated at the quasi ML estimators. The resulting test is an asymptotically valid one. However the process may be tedious, and I failed to locate any use of such a procedure in applications. Hence bootstrapping appears as a natural method to determine critical values.

The case of a univariate restriction is of particular empirical relevance, as when one tests for the significance of a specific parameter component. Moreover, inverting a test on a scalar parameter can be used to obtain a confidence interval. For an univariate restriction, λ_* is scalar, and specifically

$$\lambda_* = \frac{d' A^{-1}(\theta_*) B(\theta_*) A^{-1}(\theta_*) d}{d' A_f^{-1}(\theta_*) d}.$$

The QLR statistic is then asymptotically distributed under H_0 as $\lambda_* \chi_1^2$, and it can easily be rendered asymptotically pivotal. Given a consistent estimator $\widehat{\lambda}_n$, the adjusted QLR statistic is

$$\text{AQLR}_n = \frac{\text{QLR}_n}{\widehat{\lambda}_n} \xrightarrow{d} \chi_1^2 \quad \text{under } H_0,$$

and a test can be easily entertained using standard critical values. Estimating λ_* requires consistent estimation of $A^{-1}(\theta_*)$ and $A^{-1}(\theta_*) B(\theta_*) A^{-1}(\theta_*)$. The latter happens to be the asymptotic variance of the ML estimator $\widehat{\theta}_n$, that should be estimated to build the robust version of the Wald statistic, see White (1982). Hence it is as easy to obtain the adjusted QLR statistic than it is to render the Wald one robust to misspecification. Consistent estimation of $A(\theta_*)$ and $B(\theta_*)$ can be achieved by using their empirical equivalent evaluated at the ML estimator, that is

$$\widehat{A}_n = n^{-1} \sum_{t=1}^n \nabla_{\theta\theta'} q(z_t, \widehat{\theta}_n) \quad \widehat{B}_n = n^{-1} \sum_{t=1}^n \nabla_{\theta} q(z_t, \widehat{\theta}_n) \nabla_{\theta'} q(z_t, \widehat{\theta}_n).$$

For the particular case of a linear regression

$$y = x' \beta + \varepsilon = x'_1 \beta_1 + x'_2 \beta_2 + \varepsilon \quad \varepsilon | x \sim \mathcal{N}(0, \sigma^2), \quad (2.1)$$

Lien and Vuong (1987, Lemma 2) provide an alternative characterization of the r non-zero eigenvalues when testing for r linear restrictions. Our corollary implies that, when using a normal likelihood for the error term and testing $H_0 : \beta_1 = \beta_{1,0}$, with $\dim \beta_1 = r$, λ_* can be estimated as the vector of eigenvalues of

$$[I_1' \hat{\sigma}^2 (X'X)^{-1} I_1]^{-1} I_1' (X'X)^{-1} X' \hat{\Omega} X (X'X)^{-1} I_1, \quad \text{where} \quad I_1 = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$\hat{\Omega}$ is an estimator of $\text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, $\sigma_i^2 = \mathbb{E}(\varepsilon_i^2 | x_i)$, and $\hat{\sigma}^2$ is consistent for $\sigma^2 = \mathbb{E} \varepsilon^2$. For an univariate restriction,

$$\hat{\lambda}_n = \frac{e_1' (X'X)^{-1} X' \hat{\Omega} X (X'X)^{-1} e_1}{e_1' \hat{\sigma}^2 (X'X)^{-1} e_1} \quad e_1 = (1, 0, \dots, 0)'$$

The denominator is exactly the heteroscedasticity-robust variance estimator used for the robust Wald statistic. When one uses the least-squares criterion instead of the normal log-likelihood, the estimated eigenvalue is $\hat{\sigma}^2 \hat{\lambda}_n$.

3 Bootstrap test

We now propose a simple bootstrap procedure for the QLR test that serves as an alternative to the case-by case determination by simulation of the quantiles of the QLR statistic. Moreover, since we can easily adjust the statistic when testing an univariate restriction, we could expect a good control of level of the bootstrap resulting test. We rely on a nonparametric (naive) bootstrap, that is we generate bootstrap samples drawn with replacement from the original data. We then entertain constrained and unconstrained estimation, but we modify the objective function to make it behave as if the restrictions were true. Specifically, let $\{(z_t^*), t = 1, \dots, n\}$ be a naive bootstrap sample and $Q_n^*(\theta) = n^{-1} \sum_{t=1}^n q(z_t^*, \theta)$ the criterion based on this bootstrap sample. We consider a penalized criterion, that is

$$\hat{\theta}_n^* = \arg \min_{\Theta} \tilde{Q}_n^*(\theta) = \arg \min_{\Theta} Q_n^*(\theta) - \frac{\partial}{\partial \theta'} Q_n(\hat{\theta}_n^0) (\theta - \hat{\theta}_n^0). \quad (3.2)$$

Similarly, the constrained bootstrap estimator is

$$\widehat{\theta}_n^{0*} = \arg \min_{\varphi(\Gamma)} \widetilde{Q}_n^*(\theta) = \arg \min_{\varphi(\Gamma)} Q_n^*(\theta) - \frac{\partial}{\partial \theta'} Q_n(\widehat{\theta}_n^0) (\theta - \widehat{\theta}_n^0) . \quad (3.3)$$

If constraints are linear, and since the score $\frac{\partial}{\partial \gamma} Q_n(\varphi(\widehat{\gamma}_n))$ with respect to γ is indeed zero, we can equivalently solve $\widehat{\theta}_n^{0*} = \varphi(\widehat{\gamma}_n^*)$, where $\widehat{\gamma}_n^* = \arg \min_{\Gamma} Q_n^*(\varphi(\gamma))$. Our bootstrap QLR statistic is

$$\begin{aligned} \text{QLR}_n^* &= 2n \left[\widetilde{Q}_n^*(\widehat{\theta}_n^{0*}) - \widetilde{Q}_n^*(\widehat{\theta}_n^*) \right] \\ &= 2n \left[Q_n^*(\widehat{\theta}_n^{0*}) - Q_n^*(\widehat{\theta}_n^*) - \frac{\partial}{\partial \theta'} Q_n(\widehat{\theta}_n^0) (\widehat{\theta}_n^{0*} - \widehat{\theta}_n^*) \right] . \end{aligned}$$

When testing an univariate restriction, we rely on the formula of Corollary 2.2, evaluated using the same quantities based on resampled data, to compute an adjusted statistic AQLR_n^* . The bootstrap and adjusted bootstrap test are defined as $\mathbb{I}(\text{QLR}_n^* > c_{1-\alpha}^*)$ and $\mathbb{I}(\text{QLR}_n^* > c_{1-\alpha}^{a*})$, where $c_{1-\alpha}^*$ and $c_{1-\alpha}^{a*}$ are the $1 - \alpha$ quantile of QLR_n^* and AQLR_n^* .

To understand the use of the penalized criterion, there are two main intuitions. First, let us think about the criterion itself. It is clear that $Q_n^*(\theta)$ converges to $Q_n(\theta)$ conditionally on the initial sample. Hence at the limit (3.2) becomes

$$\min_{\theta} Q_n(\theta) - \frac{\partial}{\partial \theta'} Q_n(\widehat{\theta}_n^0) (\theta - \varphi(\widehat{\gamma}_n)) .$$

Remembering that the constrained estimator $\widehat{\theta}_n^0$ is obtained solving

$$\max_{\theta, \lambda, \gamma} Q_n(\theta) - \mu (\theta - \varphi(\gamma)) ,$$

and that at the optimum $\gamma = \widehat{\gamma}_n$ and $\mu = \frac{\partial}{\partial \theta'} Q_n(\widehat{\theta}_n^0)$, the solution of the limiting problem is $\widehat{\theta}_n^0$. So, conditionally on the initial sample, $\widehat{\theta}_n^*$ estimates $\widehat{\theta}_n^0$. Andrews (2002), who studies the nonparametric block bootstrap with dependent data, uses a similar modification of the optimization problem for building a bootstrap Wald statistic. In his setup, penalization accounts for the fact that the average score evaluated using block-bootstrapped data can be different from zero. Here we use it to account for the fact that the score evaluated at the constrained estimator is not zero.

A second way to understand the use of the penalized criterion is to consider the first-order conditions of (3.2),

$$\frac{\partial}{\partial \theta} Q_n^*(\hat{\theta}_n^*) = \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n^0). \quad (3.4)$$

Our device appears as a recentering method, as proposed by Hall and Horowitz (1996). While they use it to account for the non nullity of the empirical moments in GMM overidentified models, recentering is here used to ensure that the score of the bootstrap estimator equals the score of the restricted estimator, and so the bootstrap estimator mimics what happens if the restrictions were true. We hence label our method *bootstrap under the null hypothesis*.

Theorem 3.1 *Under Assumption C, A, and B, the tests $\mathbb{I}(\text{QLR}_n > c_{1-\alpha}^*)$ and $\mathbb{I}(\text{QLR}_n > c_{1-\alpha}^{a*})$ have asymptotically level α and are consistent against H_A .*

Proof. Under Assumption A, $\sup_{\Theta} |Q_n^*(\theta) - Q(\theta)| \xrightarrow{p} 0$ and $\sup_{\Theta} \|\nabla_{\theta} Q_n^*(\theta) - \nabla_{\theta} Q(\theta)\| \xrightarrow{p} 0$. Since $\hat{\theta}_n^0 \xrightarrow{p} \varphi(\gamma_*)$, we have $\sup_{\Theta} |\tilde{Q}_n^*(\theta) - Q(\theta) + \nabla_{\theta} Q(\theta)(\theta - \varphi(\gamma_*))| \xrightarrow{p} 0$, so that $\hat{\theta}_n^* - \varphi(\gamma_*)$ and then $\hat{\theta}_n^* - \hat{\theta}_n^0$ are both $o_p(1)$. By the same arguments, $\hat{\theta}_n^{0*} - \hat{\theta}_n^0 = \varphi(\hat{\gamma}_n^*) - \varphi(\hat{\gamma}_n) = o_p(1)$. By a Taylor expansion,

$$\begin{aligned} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n^0) &= -[\nabla_{\theta\theta'} Q(\varphi(\gamma_*))]^{-1} \sqrt{n} [\nabla_{\theta} Q_n^*(\varphi(\gamma_*)) - \nabla_{\theta} Q_n(\varphi(\gamma_*))] + o_p(1) \\ \sqrt{n} (\hat{\gamma}_n^* - \hat{\gamma}_n) &= -[H' \nabla_{\theta\theta'} Q(\varphi(\gamma_*)) H]^{-1} H' \sqrt{n} [\nabla_{\theta} Q_n^*(\varphi(\gamma_*)) - \nabla_{\theta} Q_n(\varphi(\gamma_*))] + o_p(1). \end{aligned}$$

Under Assumptions A and B,

$$\sqrt{n} [\nabla_{\theta} Q_n^*(\varphi(\gamma_*)) - \nabla_{\theta} Q_n(\varphi(\gamma_*))] = \frac{1}{\sqrt{n}} \sum_{t=1}^n q(z_t^*, \varphi(\gamma_*)) - q(z_t, \varphi(\gamma_*))$$

has an asymptotic distribution $N(0, B(\varphi(\gamma_*)))$ conditionally on the initial sample, see e.g. Gine and Zinn (1989). This yields that the vector

$$n^{1/2} \begin{bmatrix} \hat{\theta}_n^* - \hat{\theta}_n^0 \\ \hat{\gamma}_n^* - \hat{\gamma}_n \end{bmatrix}$$

has conditionally on the initial sample the same distribution as the vector in Theorem 2.1 (a) under H_0 .

By a Taylor expansion of $\tilde{Q}_n^*(\hat{\theta}_n^0)$ around $\hat{\theta}_n^*$ and $\hat{\theta}_n^{0*}$,

$$\begin{aligned}\tilde{Q}_n^*(\hat{\theta}_n^*) &= Q_n^*(\hat{\theta}_n^0) - \frac{1}{2} \left(\hat{\theta}_n^* - \hat{\theta}_n^0 \right)' A(\varphi(\gamma_*)) \left(\hat{\theta}_n^* - \hat{\theta}_n^0 \right) + o_p(n^{-1}) \\ \tilde{Q}_n^*(\hat{\theta}_n^{0*}) &= Q_n^*(\hat{\theta}_n^0) - \frac{1}{2} (\hat{\gamma}_n^* - \hat{\gamma}_n)' H' A(\varphi(\gamma_*)) H (\hat{\gamma}_n^* - \hat{\gamma}_n) + o_p(n^{-1}) \\ \text{QLR}_n^* &= n \left(\hat{\theta}_n^* - \hat{\theta}_n^0 \right)' A(\varphi(\gamma_*)) \left(\hat{\theta}_n^* - \hat{\theta}_n^0 \right) - (\hat{\gamma}_n^* - \hat{\gamma}_n)' H' A(\varphi(\gamma_*)) H (\hat{\gamma}_n^* - \hat{\gamma}_n) + o_p(1).\end{aligned}$$

The asymptotic conditional behavior of the latter statistic is thus the same as the one of QLR_n under H_0 . ■

Practically, bootstrapping involves generating B bootstrap statistics to approximate accurately enough the quantiles of the statistic's distribution. As solving (3.2) and its constrained counterpart a large number of times may be computationally demanding, one can use approximations to bypass this task. First note that both $\hat{\theta}_n^*$ and $\hat{\theta}_n^{0*}$ are estimating $\hat{\theta}_n^0$, so that one can focus on a neighborhood of the latter quantity. Moreover, a Taylor expansion of (3.4) yields

$$\hat{\theta}_n^* - \hat{\theta}_n^0 = - \left[\frac{\partial}{\partial \theta \partial \theta'} Q_n^*(\hat{\theta}_n^0) \right]^{-1} \left[\frac{\partial}{\partial \theta} Q_n^*(\hat{\theta}_n^0) - \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n^0) \right] + o_p(n^{-1/2}).$$

Hence one can use this approximation and take a number of Newton-Raphson steps from $\hat{\theta}_n^0$ to obtain an accurate approximation of the bootstrap estimators values. Specifically, one can use the rule

$$\begin{aligned}\theta_{n,0}^* &= \hat{\theta}_n^0, \\ \theta_{n,j}^* &= \theta_{n,j-1}^* - A_{n,j-1}^{*-1} \left[\frac{\partial}{\partial \theta} Q_n^*(\theta_{n,j-1}^*) - \frac{\partial}{\partial \theta} Q_n(\theta_{n,j-1}^*) \right] \quad j = 1, \dots, k,\end{aligned}$$

where $A_{n,j-1}^* = \frac{\partial}{\partial \theta \partial \theta'} Q_n^*(\theta_{n,j-1}^*)$. Alternatively, $A_{n,j-1}^*$ can be any consistent estimator of $\frac{\partial}{\partial \theta \partial \theta'} \mathbb{E} Q_n(\cdot)$ evaluated at $\theta_{n,j-1}^*$. A similar iterative device can be used to approximate $\hat{\theta}_n^{0*}$. Results from Andrews (2002) could be adapted here to ensure that the k -step bootstrap estimators are close enough to the exact bootstrap estimators.

4 Monte-Carlo

We ran some experiments to assess the finite sample performance of our bootstrap QLR test and compare it to bootstrap Wald tests. The setup is similar to the one used by Kline and Santos (2012), where heteroscedasticity and misspecification can both be present. Specifically, we generated the variable y according to the linear model

$$y_i = x_{1i} + x_{2i} + x_{3i} + \psi x_{1i}x_{2i} + (1 + \lambda x_{1i}) \eta,$$

where (v_i, x_{2i}, x_{3i}) are distributed as a trivariate standard normal with zero correlation and $x_{1i} = \exp(v_i) - \mathbb{E}[\exp(v_i)] / (\text{Var}(\exp v_i))^{1/2}$. The log-normal specification of x_1 generate observations with high leverage, which can create serious obstacles to heteroscedasticity robust inference as shown by Chesher and Jewitt (1987). Additionally, the log-normal specification provides us with an asymmetric covariate, which is helpful in avoiding overly optimistic results, see Chesher (1995). The error η follows a centered and standardized exponential distribution and is independent of the regressors. Though all of its moments exist, it exhibits substantial skewness. The parameter λ allows to introduce heteroscedasticity and the parameter ψ captures misspecification of the conditional mean. We focus on inference on the coefficient β_1 in the assumed linear regression model

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2).$$

In our simulations, $\alpha = 0$ and each of the beta parameters equals 1. When $\psi = 0$, the conditional mean is properly specified. Otherwise, it is misspecified, and while the resulting ε is uncorrelated with regressors, it will have non-zero conditional mean. Due to the regressors' independence, β_1 is unchanged under misspecification of the conditional mean, which is convenient when studying the behavior of the tests under misspecification. The Wald statistic uses an heteroscedasticity-robust covariance estimator computed with the HC_3 method as recommended by Mackinnon and White (1985). We consider two bootstrap versions of the test, one using the naive bootstrap following Hall and Horowitz (1996), and one using a wild residual bootstrap using the two-point distribution of Mammen (1993). For building QLR statistics, we consider either a likelihood criterion that

assumes a normal heteroscedastic distribution for ε , or the least-squares (LS) criterion. The QLR statistic based on LS is $n \frac{\widehat{\sigma}_n^{0,2} - \widehat{\sigma}_n^2}{\widehat{\sigma}_n^2}$, where $\widehat{\sigma}_n^2$ and $\widehat{\sigma}_n^{0,2}$ are the unconstrained and constrained estimators of the error's variance. For a linear conditional mean and homoscedastic errors, both QLR statistics are asymptotically chi-square under the null hypothesis. For a misspecified model, each can be adjusted by $\widehat{\lambda}_n$ to obtain the same asymptotic behavior. The latter is built on the empirical equivalent of $A(\theta)$ and $B(\theta)$ evaluated at the unconstrained estimator.

We examined the ability of tests to deliver an accurate level. Our experiments used 399 bootstrap repetitions and 10,000 Monte Carlo replications. We computed the errors in rejection probability (ERP), that is the differences between empirical rejection rates and the nominal level. We report our results in a set of figures that allow to compare the ERP of the different tests.

In Figure 1, we compare the asymptotic and bootstrap robust Wald tests with the asymptotic tests based on adjusted QLR statistics. Note that the adjusted LS-based QLR statistic is equal to the robust Wald test statistic for our linear model. All Wald tests are oversized, and in some cases severely so if there is misspecification in mean or variance. The wild bootstrap Wald test never improves upon the asymptotic one, while the naive bootstrap test corrects part of the error in rejection probability for misspecified models. The asymptotic adjusted LR (ALR) test mostly lies in between the asymptotic and naive bootstrap test, showing that bootstrap may be useful.

In Figure 2, we compare the bootstrap robust Wald tests with the bootstrap QLR tests based on the misspecified likelihood. We look at the LR test using our bootstrap method under the null (LR-b0), its adjusted version (ALR-b0), as well as the LR test bootstrapped following Camponovo (2015) (LR-b). The latter is based on the difference between $\min_{\Theta} Q_n^*(\theta)$ and $\min_{\Theta \cup H_{A_n}} Q_n^*(\theta)$, where H_{A_n} imposes that $\beta_1 = \widehat{\beta}_{1n}$. Our bootstrap test LR-b0 performs consistently much better than LR-b, which is due solely to the bootstrap method. In misspecified cases, the latter always performs worse than the wild bootstrap Wald test. The unadjusted bootstrap LR test is mostly worse than the naive bootstrap Wald, and even worse than the wild bootstrap Wald when both mean

and variance are misspecified. This is not surprising given that we are bootstrapping an asymptotically non pivotal statistic. Our bootstrap ALR test outperforms its best competitor, the naive bootstrap Wald test, in all cases, and often by a wide margin. Note also that Wald tests always overreject the null hypothesis, which should yield confidence intervals with coverage inferior to the required confidence level. By contrast, our bootstrap LR test is not systematically oversized in our experiment.

Figure 3 concerns the behavior of bootstrap tests based on least-squares. The LS test bootstrapped following Camponovo (2015) (LS-b) exhibits large overrejection in all cases. The best performing test is our bootstrap adjusted LS test, who always has a better level than the naive bootstrap Wald test. More surprising is that the unadjusted bootstrap test does as well or better than the naive bootstrap Wald test.

Figure 4 compares the empirical power of our three best sized tests, namely naive bootstrap Wald, adjusted bootstrap LR and LS tests, based on 2000 replications for each considered value of β_1 . Power is lowered by misspecification in mean or variance, but the power curves of the three tests are similar in shape. Though the Wald test is always oversized under the null hypothesis, it can be outperformed under some alternatives by our bootstrap tests. To sum up, our results suggest that adjusting and bootstrapping QLR tests with our bootstrap under the null method provides a viable alternative to standard bootstrap Wald tests, with a better control of level and at least comparable power.

References

- ANDREWS, D. W. K. (2002): “Higher-Order Improvements of a Computationally Attractive K-Step Bootstrap for Extremum Estimators,” *Econometrica*, 70, 119–162.
- CAMPONOVO, L. (2015): “Asymptotic Refinements of Nonparametric Bootstrap for Quasi-Likelihood Ratio Tests for Classes of Extremum Estimators,” Tech. rep., University of St.Gallen.
- CHEN, X. AND D. POUZO (2012): “Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals,” *Econometrica*, 80, 277–321.
- (2015): “Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models,” *Econometrica*, 83, 1013–1079.

- CHESHER, A. (1995): "A Mirror-Image Invariance for M-Estimators," *Econometrica*, 63, 207–211.
- CHESHER, A. AND G. AUSTIN (1991): "The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics," *Journal of Econometrics*, 47, 153–173.
- CHESHER, A. AND I. JEWITT (1987): "The Bias of a Heteroskedasticity Consistent Covariance-Matrix Estimator," *Econometrica*, 55, 1217–1222.
- DAVIDSON, R. AND J. G. MACKINNON (1999): "Bootstrap Testing in Nonlinear Models." *International Economic Review*, 40, 487.
- EFRON, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *Ann. Statist.*, 7, 1–26.
- EICKER, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Ann. Math. Statist.*, 34, 447–456.
- FOUTZ, R. V. AND R. C. SRIVASTAVA (1977): "Performance of Likelihood Ratio Test When Model is Incorrect," *Annals of Statistics*, 5, 1183–1194.
- GINE, E. AND J. ZINN (1989): "Necessary Conditions for the Bootstrap of the Mean," *Ann. Statist.*, 17, 684–691.
- GONCALVES, S. AND H. WHITE (2004): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199 – 219.
- GOURIEROUX, C. AND A. MONFORT (1995): *Statistics and Econometric Models*, Themes in Modern Econometrics, Cambridge University Press.
- GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681–700.
- GREGORY, A. W. AND M. R. VEALL (1985): "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 53, pp. 1465–1468.
- HALL, P. AND J. L. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-Of-Moments Estimators," *Econometrica*, 64, 891–916.
- HOROWITZ, J. (1997): "Bootstrap Methods in Econometrics: Theory and Numerical Performance," in *Advances in Economics and Econometrics: Theory and Applications, 7th World Congress, vol. 3*, ed. by D. M. Kreps and K. F. Wallis, Cambridge University Press.

- KENT, J. T. (1982): “Robust Properties of Likelihood Ratio Tests,” *Biometrika*, 69, 19–27.
- KIM, J. H. (2003): “Higher-Order Improvements of the Restricted Parametric Bootstrap for Tests,” National University of Singapore.
- KLINE, P. AND A. SANTOS (2012): “Higher Order Properties of the Wild Bootstrap Under Misspecification,” *Journal of Econometrics*, 171, 54–70.
- LAFONTAINE, F. AND K. J. WHITE (1986): “Obtaining Any Wald Statistic You Want,” *Economics Letters*, 21, 35–40.
- LIEN, D. AND Q. H. VUONG (1987): “Selecting the Best Linear Regression Model: A Classical Approach,” *Journal of Econometrics*, 35, 3 – 23.
- MACKINNON, J. G. AND H. WHITE (1985): “Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties,” *Journal of Econometrics*, 29, 305–325.
- MAMMEN, E. (1993): “Bootstrap and Wild Bootstrap for High-Dimensional Linear-Models,” *Annals of Statistics*, 21, 255–285.
- MARCELLINO, M. AND B. ROSSI (2008): “Model Selection for Nested and Overlapping Nonlinear, Dynamic and Possibly Mis-specified Models*,” *Oxford Bulletin of Economics and Statistics*, 70, 867–893.
- PHILLIPS, P. C. B. AND J. Y. PARK (1988): “On the Formulation of Wald Tests of Nonlinear Restrictions,” *Econometrica*, 56, 1065–1083.
- POLITIS, D. N. AND J. P. ROMANO (1994): “Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions,” *Ann. Statist.*, 22, 2031–2050.
- SHI, X. (2015): “A Nondegenerate Vuong Test,” *Quantitative Economics*, 6, 85–121.
- SPOKOINY, V. AND M. ZHILOVA (2015): “Bootstrap Confidence Sets Under Model Misspecification,” *Ann. Statist.*, 43, 2653–2675.
- VUONG, Q. H. (1989): “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses,” *Econometrica*, 57, 307–333.
- WHITE, H. (1980): “Using Least Squares to Approximate Unknown Regression Functions,” *International Economic Review*, 21, 149–70.

- (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, pp. 1–25.
- WILKS, S. S. (1938): “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,” *Annals of Mathematical Statistics*, 9, 60–62.
- WU, C. F. J. (1986): “Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion,” *Annals of Statistics*, 14, 1261–1295.

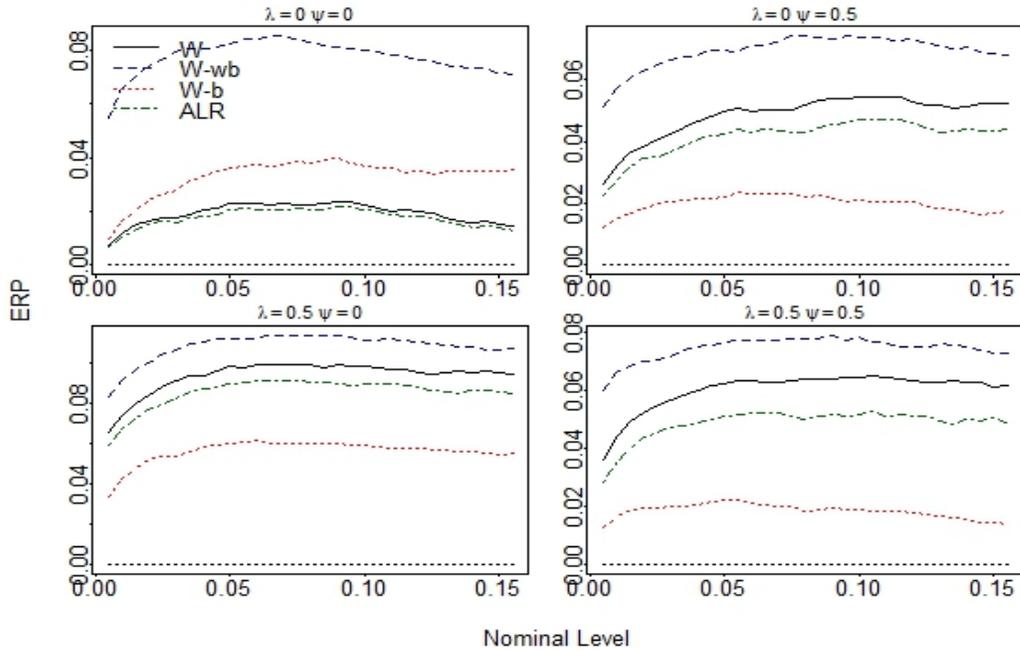


Figure 1: Errors in rejection probability for $n = 100$: Asymptotic Wald (W) and ALR tests, Wild bootstrap (wb) and Naive bootstrap (b) Wald tests

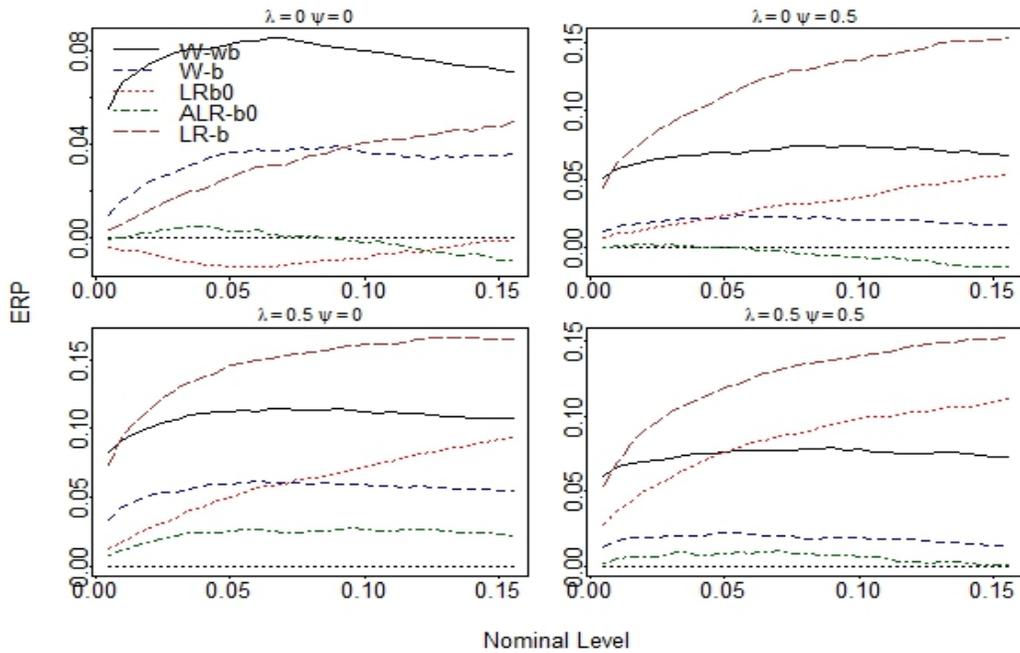


Figure 2: Errors in rejection probability for $n = 100$: Wild bootstrap (wb) and Naive bootstrap (b) Wald tests, Bootstrap LR and adjusted LR under H_0 (b0), Naive Bootstrap LR

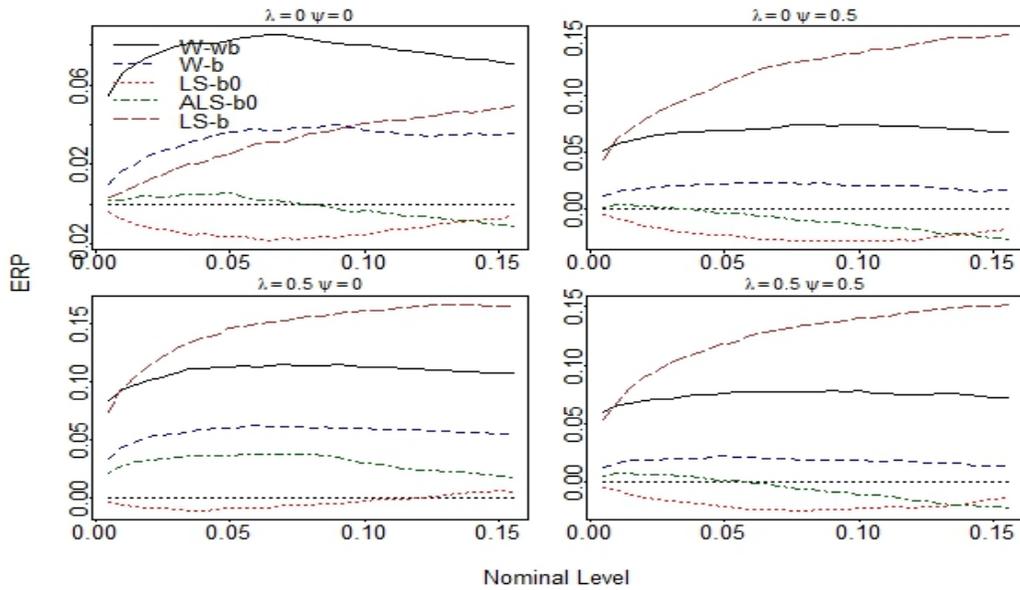


Figure 3: Errors in rejection probability for $n = 100$: Wild bootstrap (wb) and Naive bootstrap (b) Wald tests, Bootstrap LS and adjusted LS under H_0 (b0), Naive Bootstrap LS

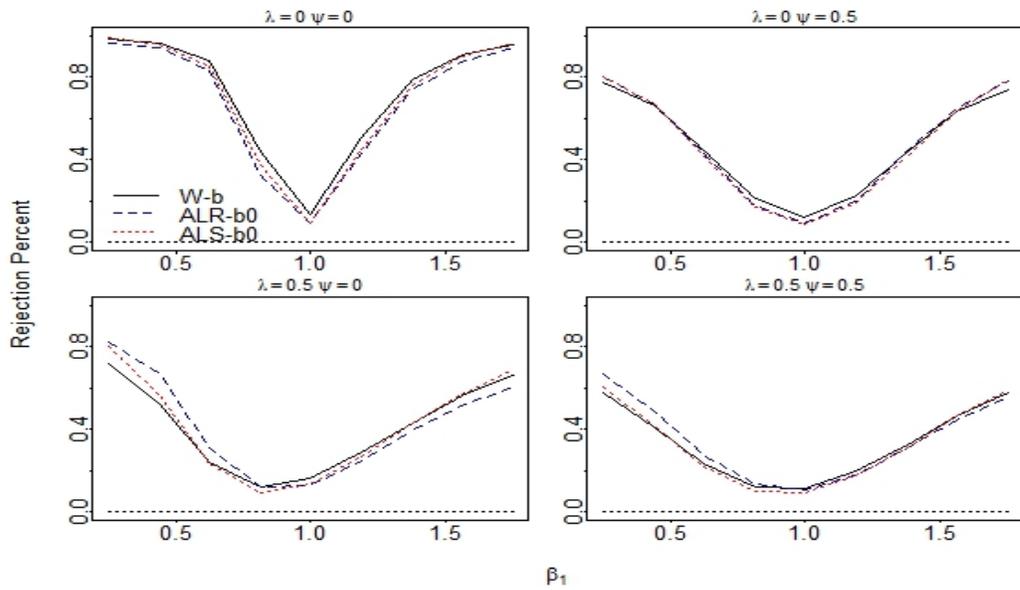


Figure 4: Rejection frequency for $n = 100$: Naive bootstrap (b) Wald tests, adjusted LR and LS under H_0 (b0)