TI 2001-119/4
Tinbergen Institute Discussion Paper

How to Implement the Bootstrap in Static or Stable Dynamic Regression Models

Noud P.A. van Giersbergen
Jan F. Kiviet

Department of Quantitative Economics, Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

How to Implement the Bootstrap in Static or Stable Dynamic Regression Models: Test Statistic versus Confidence Region Approach*

Noud P.A. van Giersbergen[a] and Jan F. Kiviet[b,*]
[a,b]Department of Quantitative Economics,
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands

November 29, 2001

Abstract

By combining two alternative formulations of a test statistic with two alternative resampling schemes we obtain four different bootstrap tests. In the context of static linear regression models two of these are shown to have serious size and power problems, whereas the remaining two are adequate and in fact equivalent. The equivalence between the two valid implementations is shown to break down in dynamic regression models. Then the procedure based on the test statistic approach performs best, at least in the AR(1)-model. Similar finite-sample phenomena are illustrated in the ARMA(1,1)-model through a small-scale Monte Carlo study and an empirical example.

JEL classification: C12; C15; C22

Keywords: Asymptotic rejection probabilities; Autoregressive models; Bootstrap; Hypothesis testing; Resampling schemes

* This research was sponsored by the Economics Research Foundation (ECOZOEK), which is part of the Netherlands Organization for Scientific Research (NWO). This paper is a completely revised version of the discussion paper "How to Implement Bootstrap Hypothesis Testing in Static and Dynamic Regression Models", which was first presented at the EC^2 meeting in Oxford (16-17 december 1993, UK). We are grateful to the editor Richard Blundell, an associate editor, and two anonymous referees for their helpful suggestions and for their patience during the revision of this manuscript.
*Corresponding author. Tel.: +31-20-5254224; fax: +31-20-5254349. *E-mail address:* jfk@fee.uva.nl (J. Kiviet)

1 Introduction

There exists a well-known correspondence between confidence sets and hypothesis testing. A confidence set with confidence coefficient $1 - \alpha$ entails the set of parameter values which are not rejected by a hypothesis test performed at significance level α. This duality is exploited in this paper to define and examine several bootstrap testing procedures in static and stable dynamic regression models. See Hansen (1999) for an analysis of bootstrap procedures which exploit this duality in unstable autoregressive models. For comprehensive overviews on the bootstrap see the books by Davison and Hinkley (1997), Efron and Tibshirani (1993) and Hall (1992), while the papers of Li and Maddala (1996), Horowitz (1997) and Berkowitz and Kilian (2000) provide recent reviews from an econometric (time-series) perspective. The bootstrap procedures to be examined here differ in their choice of the type of estimator employed in the resampling scheme and in the form of the test statistic used. More specifically, a resampling scheme can use either a restricted or an unrestricted estimator in defining bootstrap observables. In addition, a bootstrap test statistic can be centred around either a restricted or an unrestricted estimator. The use of a restricted estimator in both the resampling scheme and the test statistic is typical for the test statistic approach (in which one tries to assess the null distribution), while the use of an unrestricted estimator is characteristic for the confidence region approach (which does not adhere to just one specific null hypothesis); see Beran (1986) for more details about the two approaches in the context of independent and identically distributed (i.i.d.) random vectors. We shall examine these two approaches and their cross-fertilizations in a regression context.

To illustrate the basic idea, consider the simple regression model

$$y = \mu + \delta x + \varepsilon, \tag{1}$$

where y and x are $n \times 1$ vectors and ε is a $n \times 1$ random vector whose components are i.i.d. with unknown distribution F_ε with mean 0 and finite variance σ^2, *i.e.* $\varepsilon_i \sim F_\varepsilon(0, \sigma^2)$ for $i = 1, ..., n$. Suppose that we are interested in testing the null hypothesis $H_0 : \delta = \delta_0$ against the one-sided alternative $H_1 : \delta < \delta_0$ for some given value of δ_0. In general, the finite-sample distribution of the test statistic used, say T, is unknown since it depends on some nuisance parameters. The bootstrap tackles this problem by replacing the nuisance parameters by their empirical analogues. For example, the non-parametric bootstrap replaces the unknown distribution function F_ε by the empirical distribution function (EDF) of the residuals. In general, however, the bootstrap distribution of the test statistic does not possess a closed-form expression. In practice, fortunately, any desired characteristic of the bootstrap distribution can be approximated by a bootstrap simulation. If T_b^* denotes the b-th bootstrap realization of the test statistic T, then the bootstrap uses the empirical distribution of $\{T_b^*\}_{b=1}^B$ to approximate the actual distribution of T. Below, we shall describe in some detail how to test the hypothesis $H_0 : \delta = \delta_0$ using the bootstrap based on either the test statistic or on the confidence region approach.

The test statistic approach. Let

$$\tau(\delta_0) = (\hat{\delta} - \delta_0)/s_{\hat{\delta}} \tag{2}$$

denote the well-known t-statistic, which would have a Student t-distribution if the disturbances ε_i were normally distributed and $\delta = \delta_0$. In addition, let y^{*^r} denote the bootstrap observables defined by the 'restricted' resampling scheme $(*^r)$

$$y^{*^r} = \hat{\mu}_r + \delta_0 x + \varepsilon^*, \tag{3}$$

where $\hat{\mu}_r$ denotes the estimator of μ under the null hypothesis and ε_i^* $(i = 1, ..., n)$ is drawn randomly with replacement from the (un)restricted residuals. The bootstrap analogue of the t-statistic $\tau(\delta_0)$ based on the bootstrap sample $\mathcal{X}^{*^r} = (y^{*^r}\ x)$ is given by

$$\tau_r^{*^r} \equiv \tau^{*^r}(\delta_0) = (\hat{\delta}^* - \delta_0)/s_{\hat{\delta}^*}, \tag{4}$$

where the bootstrap estimators $\hat{\delta}^*$ and $s_{\hat{\delta}^*}$ will be defined explicitly in the next section. Let $\tau_{r,\alpha}^{*^r}$ denote the α-quantile of the bootstrap distribution of the test statistic $\tau^{*^r}(\delta_0)$, *i.e.* $\mathbb{P}_*[\tau_r^{*^r} \leq \tau_{r,\alpha}^{*^r}] = \alpha$, where the super-index $*^r$ denotes that the resampling scheme is based on restricted estimators while the sub-index r indicates that restricted (r) estimators have been used in the test statistic. The notation $\mathbb{P}_*[\cdot]$ stresses the fact that the probability is defined conditional on the data and with respect to resampling scheme (3). The *bootstrap test procedure* φ^* rejects the restriction $\delta = \delta_0$ if the observed t-statistic $\tau(\delta_0)$ given in (2) is smaller than the α-quantile of the bootstrap distribution of $\tau^{*^r}(\delta_0)$, *i.e.* the hypothesis $H_0 : \delta = \delta_0$ is rejected against the alternative $H_1 : \delta < \delta_0$ if $\tau(\delta_0) \leq \tau_{r,\alpha}^{*^r}$. From the analysis in Beran (1986), it follows that, under the null hypothesis, the rejection probability of the bootstrap test φ^* based on the quantiles of the theoretical bootstrap distribution (*i.e.* for $B = \infty$) converges to α, *i.e.* $\mathbb{P}[\tau(\delta_0) \leq \tau_{r,\alpha}^{*^r} | H_0] \to \alpha$ as $n \to \infty$[1].

The confidence region approach. An asymptotically pivotal root for the regression parameter δ is given by

$$R(\delta) = (\hat{\delta} - \delta)/s_{\hat{\delta}}. \tag{5}$$

A root is a function of both the parameter of interest and its estimator, and can be used to construct a confidence interval; see Beran (1987) for more details on roots. In the confidence region approach, the bootstrap sample $\mathcal{X}^{*^u} = (y^{*^u}\ x)$ is based on the 'unrestricted' resampling scheme $(*^u)$

$$y^{*^u} = \hat{\mu} + \hat{\delta} x + \varepsilon^*,$$

[1] In the analysis, the population distribution function F_ε is considered as a *fixed* nuisance parameter.

where ε_i^* is drawn randomly with replacement from the residuals. Let

$$R_u^{*^u} \equiv R^{*^u}(\hat{\delta}) = (\hat{\delta}^* - \hat{\delta})/s_{\hat{\delta}^*} \tag{6}$$

denote the bootstrap analogue of the root $R(\delta)$ based on the bootstrap sample $\mathcal{X}^{*^u}$. Note that in the bootstrapped root $R^{*^u}(\hat{\delta})$, the estimator $\hat{\delta}$ takes over the pivoting role of the unknown parameter δ. Let $R_{u,\alpha}^{*^u}$ denote the α-quantile of the bootstrap distribution of the root $R^{u^*}(\hat{\delta})$, *i.e.* $\mathbb{P}_*[R_u^{*^u} \leq R_{u,\alpha}^{*^u}] = \alpha$, where the super-index $*^u$ denotes that the bootstrap data is based on the 'unrestricted' resampling scheme while the sub-index u indicates that the root is based on the unrestricted (u) estimators. A one-sided $(1-\alpha)$-level percentile-t confidence interval for δ is given by

$$I_{1,t}^* = (-\infty, \hat{\delta} - s_{\hat{\delta}} R_{u,\alpha}^{*^u}), \tag{7}$$

which equals the set of parameters δ defined by $\{\delta : R(\delta) > R_{u,\alpha}^{*^u}\}$. Consider the situation that the parameter value under test is δ_0. The parameter value δ_0 lies outside the bootstrap confidence interval $I_{1,t}^*$ if the condition $R(\delta_0) \leq R_{u,\alpha}^{*^u}$ holds. Since the root $R(\delta_0)$ equals the t-statistic $\tau(\delta_0)$, the corresponding *bootstrap test procedure* ψ^* rejects the restriction $\delta = \delta_0$ if the t-statistic $\tau(\delta_0)$ is smaller than the α-quantile of the bootstrap distribution of $R^{*^u}(\hat{\delta})$, *i.e.* the hypothesis $H_0 : \delta = \delta_0$ is rejected against the alternative $H_1 : \delta < \delta_0$ if $\tau(\delta_0) \leq R_{u,\alpha}^{*^u}$. Under certain regularity conditions, Freedman (1981) has shown that the bootstrap test ψ^* also has the correct asymptotic rejection probability under the null hypothesis.

To summarize, we have distinguished two different bootstrap testing procedures. For testing the null hypothesis $H_0 : \delta = \delta_0$ against the alternative $H_1 : \delta < \delta_0$ using the t-statistic $\tau(\delta_0)$, the bootstrap test φ^* rejects H_0 if $\tau(\delta_0) \leq \tau_{r,\alpha}^{*^r}$ whereas the bootstrap test ψ^* rejects H_0 if $\tau(\delta_0) \leq R_{u,\alpha}^{*^u}$. In Section 3, it will be shown that in static regression models the tests φ^* and ψ^* are equivalent in finite samples since the α-quantile of the bootstrap distribution based on the test statistic approach coincides with the α-quantile of the bootstrap distribution based on the confidence region approach, *i.e.* $\tau_{r,\alpha}^{*^r} = R_{u,\alpha}^{*^u}$. In the case where the model contains only an intercept, this equivalence has been noticed before by, for instance, Beran (1986, Example 2) and Tibshirani (1992).

Instead of combining (un)restricted estimators in both the resampling scheme and the bootstrap test statistic, one could also opt for a crosswise combination. However, Hall and Wilson (1991) warn against the use of the procedure based on the bootstrap distribution of the t-statistic $\tau^*(\delta_0)$ and the 'unrestricted' resampling scheme $y^{*^u} = \hat{\mu} + \hat{\delta}x + \varepsilon^*$, because this latter implementation is said to lead to low test power. Yet another implementation results by combining the root $R^*(\hat{\delta})$ with the 'restricted' resampling scheme $y^{*^r} = \hat{\mu}_r + \delta_0 x + \varepsilon^*$. The latter two procedures will be referred to as hybrid implementations since they mix up the test statistic with the confidence region approach. So, by combining two formulations of a test statistic, *viz.* $\tau^*(\delta_0)$ and $\tau^*(\hat{\delta})$, with two ways to construct the bootstrap observables, four different bootstrap test procedures are obtained. The aim of this paper is to investigate the differences and similarities between these various implementations in both static and dynamic multiple regression models. The papers by Carpenter (1999), DiCiccio and Romano (1990) and Hansen (1999) also exploit the test statistic approach to construct confidence intervals, although their focus is quite distinct from ours. Moreover, none of these papers considers the hybrid implementations.

The paper is organised as follows. Section 2 takes a closer look at bootstrap hypothesis testing and defines the various test statistics in linear regression models for both single and joint hypotheses. In Section 3, we investigate the various bootstrap implementations and examine whether the one-to-one correspondence between the test statistic and confidence region approach already found in models with just an intercept continues to hold in finite samples of multiple regressions when testing one or several parameters jointly. Furthermore, we derive the asymptotic rejection probabilities for each testing procedure and we demonstrate that bootstrap tests based on any of the two hybrid implementations will produce zero rejections asymptotically, irrespective whether the null hypothesis is true or false. The only exception is the case where restricted estimators are used in resampling and the statistic is centred around the unrestricted estimator and the alternative hypothesis is one-sided. Then there is a serious overrejection problem. We provide an intuitive explanation for this failure of the two hybrid implementations in the case where a single coefficient is under test. In Section 3, we also take a closer look at the issue of using restricted or unrestricted residuals in the resampling schemes. In Section 4, the finite-sample performance of the two asymptotically valid implementations is investigated in stable dynamic models. We observe that in dynamic regression models, the test statistic and confidence region approach lead to different findings in finite samples. In a small-scale Monte Carlo study, finite-sample inference based on the two approaches is examined in an ARMA(1,1) model. In Section 5, the two approaches are compared on the basis of an empirical example. The final section discusses the major findings.

2 Various Bootstrap Regression Test Procedures

In the standard –not necessarily Gaussian– linear multiple regression model

$$y = X\beta + \varepsilon, \tag{8}$$

y is a $n \times 1$ vector, X is a fixed $n \times k$ matrix, β is a $k \times 1$ vector of unknown parameters and ε is a $n \times 1$ random vector with i.i.d. components $\varepsilon_i \sim F_\varepsilon(0, \sigma^2)$. In addition, we assume that the number of coefficients k is fixed, only $\mathcal{X} = (y\, X)$ is observable and the matrix X has rank k. For the asymptotic analysis we assume that $X'X/n \to Q$ as $n \to \infty$, where Q has finite elements and is positive definite. Of course, these conditions are rather restrictive but they allow us to apply the asymptotic results obtained by Freedman (1981).

For relevant literature on the bootstrap in this class of models see Wu (1986) who gives an extensive overview of various resampling techniques. It appears that the bootstrap still works when the dimension k of the model increases as the sample size $n \to \infty$; see Bickel and Freedman (1983) and Mammen (1993). Hall (1989) and Navidi (1989) show that only when an asymptotically pivotal (test) statistic is bootstrapped, the bootstrap is capable of achieving an Edgeworth correction, which loosely speaking means that the bootstrap approximation to the distribution of interest is better than the first-order normal approximation; see also Davidson and MacKinnon (1999) for an explanation of the higher-order refinement provided by bootstrap tests. In the econometric literature on the bootstrap, the use of asymptotically pivotal statistics, whose asymptotic distributions are independent of unknown parameters, has been advocated by Horowitz (1994).

Let $\hat{\beta} = (X'X)^{-1}X'y$ be the ordinary least squares (OLS) estimator of β and denote the OLS residuals by $\hat{\varepsilon} = y - X\hat{\beta} = M_x\varepsilon$, where $M_x = I - X(X'X)^{-1}X'$. The (co)variance matrix of $\hat{\beta}$ is estimated by $s^2_{\hat{\beta}} = s^2(X'X)^{-1}$, where $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-k)$. The i-th diagonal component of $s^2_{\hat{\beta}}$ is indicated by $s^2_{\hat{\beta}_i}$. For testing a set of $m(\leq k)$ independent linear restrictions on the vector of coefficients β, we consider the hypotheses

$$H_0\colon R\beta = r_0 \qquad \text{against} \qquad H_1\colon R\beta \neq r_0,$$

where R and r_0 are known matrices of dimension $m \times k$ and $m \times 1$ respectively. Let $\hat{\beta}_r$ denote the restricted OLS estimator under the restriction $R\beta = r_0$, then

$$\hat{\beta}_r = \hat{\beta} + (X'X)^{-1}R'W^{-1}(r_0 - R\hat{\beta}), \tag{9}$$

with $W = R(X'X)^{-1}R'$. Obviously $R\hat{\beta}_r = r_0$, *i.e.* the restricted estimators obey the linear restrictions. Note that when testing a single coefficient restriction $\beta_i = \beta_0$, the i-th component of $\hat{\beta}_r$ is equal to β_0. The restricted residuals are denoted by $\hat{\varepsilon}_r = y - X\hat{\beta}_r$. The traditional F-statistic can be written as

$$\lambda(\hat{\beta}_r) = \frac{1}{m}(\hat{\beta} - \hat{\beta}_r)'R'W^{-1}R(\hat{\beta} - \hat{\beta}_r)/s^2. \tag{10}$$

This test statistic has an exact F-distribution in finite samples under the null hypothesis in the Gaussian linear fixed-regressors model. For testing whether the i-th element of β equals a given scalar β_0, we consider the hypotheses

$$H_0\colon \beta_i = \beta_0 \qquad \text{against} \qquad H_2\colon \beta_i < \beta_0 \quad \text{or} \quad H_3\colon \beta_i > \beta_0,$$

where $i \in \{1, ..., k\}$. The habitual t-statistic is denoted as

$$\tau(\beta_0) = (\hat{\beta}_i - \beta_0)/s_{\hat{\beta}_i}. \tag{11}$$

For testing the hypothesis $H_0 : \beta_i = \beta_0$ against the two-sided alternative $H_1 : \beta_i \neq \beta_0$, we may use the squared t-statistic $\tau(\beta_0)^2$ which equals the corresponding F-statistic.

Bootstrap inference can be obtained as follows. Let $\hat{F}_\varepsilon$ denote some estimator of the underlying population distribution function. In the non-parametric bootstrap, $\hat{F}_\varepsilon$ usually equals the empirical distribution function (EDF) of the demeaned unrestricted residuals $\hat{\varepsilon}$. In case of hypothesis testing, $\hat{F}_\varepsilon$ is occasionally taken as the EDF of the demeaned restricted residuals $\hat{\varepsilon}_r$. The use of the unrestricted residuals, however, ensures that the EDF $\hat{F}_\varepsilon$ converges to the population distribution F_ε in a suitable metric even when the null hypothesis is false; we shall elaborate on this issue later on. For the moment, we shall stick to using unrestricted residuals in all resampling schemes considered. Conditional on the data $\mathcal{X}$, let the components of $\varepsilon^* = (\varepsilon_1^*, ..., \varepsilon_n^*)$ be i.i.d. with common distribution $\hat{F}_\varepsilon$. The bootstrap observations y^* can be generated according to one of the following two resampling schemes:

$$\text{Scheme I } (*^u): \qquad y^{*^u} = X\hat{\beta} + \varepsilon^*,$$

$$\text{Scheme II } (*^r): \qquad y^{*^r} = X\hat{\beta}_r + \varepsilon^*.$$

So, these schemes use the initial unrestricted (u) or the restricted (r) OLS estimator respectively for constructing a bootstrap sample.

The bootstrap analogue of $\hat{\beta}$ is defined as $\hat{\beta}^* = (X'X)^{-1}X'y^*$ under both resampling schemes ($y^* \in \{y^{*^u}, y^{*^r}\}$). The (co)variance matrix of $\hat{\beta}^*$ is estimated by $s^2_{\hat{\beta}^*} = s^{2*}(X'X)^{-1}$, where $s^{2*} = \hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/(n-k)$ and $\hat{\varepsilon}^* = y^* - X\hat{\beta}^*$ are the bootstrap residuals.

In addition to the two options regarding the resampling schemes, the bootstrap test statistics may be centred around either the unrestricted or the restricted estimator as indicated in the introduction. Centring the bootstrap test statistic around the unrestricted estimator seems reasonable from a confidence-region point of view. Following the bootstrap principle that population parameters are replaced by their estimators, the bootstrap test statistic may be based on a studentised version of the quantity $\hat{\beta}_i^* - \hat{\beta}_i$. However, the test statistic approach suggests to base the bootstrap test statistic on a studentised version of the quantity $\hat{\beta}_i^* - \beta_0$, since under the null the mean of $\hat{\beta}_i^*$ is known to be β_0. The combination of resampling schemes and test statistics leads to four feasible test procedures. For $j \in \{u, r\}$, the bootstrapped F-statistics are defined as

$$\lambda_u^{*^j} \equiv \lambda^{*^j}(\hat{\beta}) = \frac{1}{m}(\hat{\beta}^* - \hat{\beta})'R'W^{-1}R(\hat{\beta}^* - \hat{\beta})/s^{2*} \qquad \text{with } (\hat{\beta}^*, s^{2*}) \text{ based on } \mathcal{X}^{*^j} \tag{12}$$

and

$$\lambda_r^{*^j} \equiv \lambda^{*^j}(\hat{\beta}_r) = \frac{1}{m}(\hat{\beta}^* - \hat{\beta}_r)'R'W^{-1}R(\hat{\beta}^* - \hat{\beta}_r)/s^{2*} \qquad \text{with } (\hat{\beta}^*, s^{2*}) \text{ based on } \mathcal{X}^{*^j}, \tag{13}$$

where $\mathcal{X}^{*^u} = (y^{*^u} X)$ and $\mathcal{X}^{*^r} = (y^{*^r} X)$. Hence, the super-index denotes which resampling scheme is used and the sub-index indicates which type of estimator is used for centring. Similarly, the bootstrapped t-statistics are denoted as $\tau_u^{*^u}, \tau_r^{*^u}$, $\tau_u^{*^r}$ and $\tau_r^{*^r}$.

The α-quantiles of the bootstrap distributions of the test statistics λ^* and τ^* are denoted by λ^*_α and τ^*_α respectively, *i.e.*

$$\mathbb{P}_*[\lambda^* \leq \lambda^*_\alpha] = \alpha \qquad \text{and} \qquad \mathbb{P}_*[\tau^* \leq \tau^*_\alpha] = \alpha. \tag{14}$$

To distinguish the quantiles of the various bootstrap distributions, let $\lambda^{*^u}_{r,\alpha}$ denote the α-quantile of the bootstrap distribution of the test statistic $\lambda^{*^u}_r \equiv \lambda^{*^u}(\hat{\beta}_r)$, so that the quantiles use the same index notation as the test statistics. In this way, the quantiles based on the other three procedures are denoted by $\lambda^{*^u}_{u,\alpha}$, $\lambda^{*^r}_{u,\alpha}$ and $\lambda^{*^r}_{r,\alpha}$ respectively. Similarly, $\tau^{*^j}_{l,\alpha}$ $(j, l \in \{u, r\})$ refers to the α-quantile of the bootstrap distribution of the appropriate t-statistic τ^*.

In order to indicate the rejection probability of each procedure, let $\alpha_1(\lambda^{*^j}_l)$ $(j, l \in \{u, r\})$ denote the rejection probability of $H_0\colon R\beta = r_0$ against $H_1\colon R\beta \neq r_0$ at nominal significance level α, *i.e.*

$$\alpha_1(\lambda^{*^j}_l) = \mathbb{P}[\lambda \geq \lambda^{*^j}_{l,1-\alpha}], \qquad (j, l \in \{u, r\}). \tag{15}$$

In addition, the rejection probabilities of $H_0 : \beta_i = \beta_0$ against $H_2 : \beta_i < \beta_0$ and $H_3 : \beta_i > \beta_0$ are denoted by

$$\alpha_2(\tau^{*^j}_l) = \mathbb{P}[\tau \leq \tau^{*^j}_{l,\alpha}] \qquad \text{and} \qquad \alpha_3(\tau^{*^j}_l) = \mathbb{P}[\tau \geq \tau^{*^j}_{l,1-\alpha}]. \tag{16}$$

In the next section, we shall show that two of the four bootstrap test procedures have size and power problems and that the other two are appropriate and in fact equivalent in the context of static linear regression models. Further on, we shall show that the latter situation is different in dynamic regression models.

3 Properties of the Various Bootstrap Test Procedures

We begin this section by showing in Proposition 1 that the F-statistic obtained by the confidence region approach is algebraically equivalent to the F-statistic obtained by the test statistic approach, *i.e.* $\lambda^{*^u}(\hat{\beta}) = \lambda^{*^r}(\hat{\beta}_r)$. Self evidently, the same result holds for the t-statistics $\tau^{*^u}_u$ and $\tau^{*^r}_r$.

Proposition 1 *In the linear fixed-regressors model (8), the F-statistics $\lambda^{*^u}(\hat{\beta})$ and $\lambda^{*^r}(\hat{\beta}_r)$ defined in (12) and (13) are equivalent.*

Proof. Consider the OLS estimator $\hat{\beta}^*_b = (X'X)^{-1}X'y^*_b$ based on the resampling scheme

$$y^*_b = Xb + \varepsilon^*, \tag{17}$$

where b denotes some arbitrary $k \times 1$ vector and ε^* denotes an $n \times 1$ vector of bootstrap disturbances. Observe that resampling schemes I and II result from (17) for $b = \hat{\beta}$ and $b = \hat{\beta}_r$ respectively. Using the identities $(\hat{\beta}^*_b - b) = (X'X)^{-1}X'\varepsilon^*$ and $\hat{\varepsilon}^* = M_x\varepsilon^*$, straightforward derivations show that

$$\lambda^{*^{u}}(\hat{\beta}) = \frac{1}{m}\frac{\varepsilon^{*\prime}X(X'X)^{-1}R'W^{-1}R(X'X)^{-1}X'\varepsilon^*}{\varepsilon^{*\prime}M_x\varepsilon^*/(n-k)} = \lambda^{*^{r}}(\hat{\beta}_r). \tag{18}$$

The expression in the middle of equation (18) reveals that the bootstrap distribution of both test statistics depends only on the fixed quantities (R, X, k, m, n) and on the bootstrap disturbances ε^*. □

The result stated in Proposition 1 is intuitively clear since the t and F-statistics are invariant to location changes in the linear regression model. The equivalence of the test statistics as stated in Proposition 1 holds for any realization ε^* as long as both test statistics are evaluated at the same realization of ε^*. Therefore, the equivalence holds irrespective whether in resampling the unrestricted or the restricted residuals are employed. Note that the test statistic approach is based on the critical values $\lambda^{*^r}_{r,\alpha}$, whereas the confidence region approach is based on the critical values $\lambda^{*^u}_{u,\alpha}$. The result in (18) implies that $\lambda^{*^r}_{r,\alpha} = \lambda^{*^u}_{u,\alpha}$, *i.e.* the α-quantiles of the bootstrap distributions of the F-statistics $\lambda^{*^u}(\hat{\beta})$ and $\lambda^{*^r}(\hat{\beta}_r)$ are identical. Likewise, we find $\tau^{*^u}_{u,\alpha} = \tau^{*^r}_{r,\alpha}$, so that in the linear fixed-regressors model, there exists a one-to-one correspondence between the test statistic and confidence region approach in finite samples.

Next, we shall derive the asymptotic rejection frequencies for each procedure. We consider the asymptotic properties of the F-statistics first and subsequently look at the asymptotic properties of the t-statistics. From, *e.g.*, Mammen (1993), it follows that both the test statistic and confidence region approach lead to correct asymptotic rejection probabilities under the null as stated in Proposition 2. In fact, since the F and t-statistics are asymptotically pivotal, these two approaches lead to bootstrap inference which is second-order correct as noted earlier.

Proposition 2 *Under $H_0 : R\beta = r$, tests based on the distributions of $\lambda^{*^u}(\hat{\beta})$ and $\lambda^{*^r}(\hat{\beta}_r)$ have asymptotic rejection probability α at nominal significance level α.*

Proof. The result for $\lambda^{*^u}(\hat{\beta})$ follows from Theorem 5 contained in Mammen (1993). Since $\lambda^{*^r}(\hat{\beta}_r) = \lambda^{*^u}(\hat{\beta})$, it also applies to $\lambda^{*^r}(\hat{\beta}_r)$. □

We now turn to the asymptotic rejection probabilities of the bootstrapped F-statistics for the two hybrid implementations. After some algebra we obtain

$$\lambda^{*^u}(\hat{\beta}_r) = \frac{s^2}{s^{2*}}\lambda(\hat{\beta}_r) + \lambda^{*^u}(\hat{\beta}) + \gamma^* \tag{19}$$

and

$$\lambda^{*^r}(\hat{\beta}) = \frac{s^2}{s^{2*}}\lambda(\hat{\beta}_r) + \lambda^{*^r}(\hat{\beta}_r) - \gamma^*, \tag{20}$$

where $\gamma^* = \frac{2}{m}\left[R(X'X)^{-1}X'\varepsilon + R\beta - r_0\right]' W^{-1}R(X'X)^{-1}X'\varepsilon^*/s^{2*}$. Note that the original F-statistic $\lambda(\hat{\beta}_r)$ appears in the expressions of these two test statistics next to the appropriate bootstrap F-statistics $\lambda^{*''}(\hat{\beta})$ or $\lambda^{*'}(\hat{\beta}_r)$. Moreover, there is also the appearance of another term γ^* which depends on both the original and bootstrap disturbances. The next proposition, which generalizes the result of Hall and Wilson (1991), shows that tests based on the bootstrap distributions of $\lambda^{*''}(\hat{\beta}_r)$ or $\lambda^{*'}(\hat{\beta})$ are inconsistent tests with asymptotic rejection probability zero for any true value of β.

Proposition 3 *Irrespective of the true value of β, tests based on the distributions of $\lambda^{*''}(\hat{\beta}_r)$ and $\lambda^{*'}(\hat{\beta})$ given in (12) and (13) have asymptotic rejection probability zero at nominal significance level α, provided that $\alpha < 0.5$.*

Proof. First, we show that $\alpha_1(\lambda_r^{*''}) \to 0$ as $n \to \infty$. We find from (15) and by substitution of (19) that

$$\begin{aligned}
\alpha_1(\lambda_r^{*''}) &= \mathbb{P}[\lambda(\hat{\beta}_r) \geq \lambda_{r,1-\alpha}^{*''}] \\
&= \mathbb{P}[\lambda(\hat{\beta}_r) \geq \lambda(\hat{\beta}_r) + \{\lambda^{*''}(\hat{\beta}) + \gamma^*\}_{1-\alpha} + o_p(1)] \\
&= \mathbb{P}[\{\lambda^{*''}(\hat{\beta}) + \gamma^*\}_{1-\alpha} \leq 0] + o_p(1) \to 0,
\end{aligned}$$

where $\{\lambda^{*''}(\hat{\beta}) + \gamma^*\}_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the convolution of the distributions of the F-statistic $\lambda^{*''}(\hat{\beta})$ and the term γ^*. The convergence result in the last equation is obtained as follows. Suppose that $\{\lambda^{*''}(\hat{\beta}) + \gamma^*\}_{1-\alpha} \leq 0$. This would imply that

$$\mathbb{P}[\lambda^{*''}(\hat{\beta}) + \gamma^* \leq 0] > \tfrac{1}{2}. \tag{21}$$

However, Lemma 1 in the Appendix shows that the probability in (21) is less than or equal to 0.5 asymptotically leading to a contradiction. Hence, the probability that the $(1-\alpha)$-quantile is less than zero tends to zero. The proof for $\alpha_1(\lambda_u^{*'})$ proceeds along the same lines. □

Note that the results obtained in Proposition 3 imply that tests based on the bootstrap distributions of $\left[\tau^{*''}(\beta_0)\right]^2$ and $\left[\tau^{*'}(\hat{\beta}_i)\right]^2$ have asymptotic rejection probability zero when testing a single ($m = 1$) restriction $H_0: \beta_i = \beta_0$ in model (8) against the two-sided alternative $H_1: \beta_i \neq \beta_0$.

We now turn to examining the asymptotic rejection probabilities of the various bootstrapped t-statistics when testing $H_0: \beta_i = \beta_0$ against one-sided alternatives. From Proposition 2, it follows that bootstrap inference based on either the test statistic or the confidence region approach has correct size asymptotically against one-sided alternatives. The two hybrid implementations lead again to asymptotically incorrect sizes, but one leads to underrejection and the other to overrejection. It is easily shown that the asymptotic rejection probability of the test based on the distribution of $\tau^{*''}(\beta_0)$ is zero for a nominal level $\alpha < 0.5$ irrespective of the true value of β_i, *i.e.*

$$\mathbb{P}[\tau(\beta_0) \leq \tau_{r,\alpha}^{*''}] \to 0 \qquad \text{for } \alpha < 0.5. \tag{22}$$

A similar result holds for inference based on the critical value $\tau^{*''}_{r,1-\alpha}$ for the other-sided alternative. Hence, we see that the test based on the distribution of $\tau^{*''}(\beta_0)$ does not just have low power (rejection probability) as noticed by Hall (1992, Section 3.12) in the i.i.d. setup (regression with an unknown intercept only), but it also has asymptotic size zero, which disqualifies this implementation of the bootstrap procedure more fundamentally.

The inadequacy of the procedure based on the distribution of $\tau^{*''}(\beta_0)$ can be explained as follows. After some straightforward manipulations, we obtain

$$\tau^{*''}(\beta_0) = \frac{s}{s^*}\tau(\beta_0) + \tau^{*''}(\hat{\beta}_i). \tag{23}$$

Hence under the null, $\tau^{*''}(\beta_0)$ is approximately distributed around the observed test statistic $\tau(\beta_0)$, *i.e.* $\mathbb{E}_*[\tau^{*''}(\beta_0)] \to \tau(\beta_0)$, because $\tau^{*''}(\hat{\beta}_i)|\mathcal{X} \xrightarrow{d} \mathcal{N}(0, 1)$ and $s/s^* \xrightarrow{p} 1$. Note that the distribution of $\tau^{*''}(\beta_0)$ depends on the realization of $\tau(\beta_0)$, which is to be considered fixed in the bootstrap procedure. So, the bootstrap distribution will vary with each realization of $\tau(\beta_0)$. For a particular realization of the original t-statistic $\tau(\beta_0)$, Figure 1 (upper graph) shows the asymptotic distributions of the t-statistics $\tau^{*''}(\beta_0)$ and $\tau^{*''}(\hat{\beta}_i)$. From this figure, we see that the original test statistic $\tau(\beta_0)$ always lies between the bootstrap quantiles $\tau^{*''}_{r,\alpha}$ and $\tau^{*''}_{r,1-\alpha}$ for $\alpha < 0.5$. So, if $\tau(\beta_0)$ is used as test statistic and the critical values are based on the bootstrap distribution of $\tau^{*''}(\beta_0)$, then one will never reject the null hypothesis irrespective of the true value of β_i and the tested value β_0. Hence, this implementation of bootstrap hypothesis testing can be classified as fundamentally unsound.

Insert Figure 1 about here.

Although Proposition 3 seems to suggest that the asymptotic rejection probability of $\tau^{*'}(\hat{\beta}_i)$ is also zero, this is not the case. It can be shown that the asymptotic size of inference based on the distribution of $\tau^{*'}(\hat{\beta}_i)$ equals $\Phi(\frac{1}{2}z_\alpha)$, which is greater than the nominal level for $\alpha < 0.5$, *i.e.*

$$\begin{aligned}
\alpha_2(\tau^{*'}_u) &= \mathbb{P}[\tau(\beta_0) \leq \tau^{*'}_{u,\alpha} \,|\, \beta_i = \beta_0] \\
&= \mathbb{P}[\tau(\beta_0) \leq \{-\tau(\beta_0) + \tau^{*'}(\hat{\beta}_0)\}_\alpha + o_p(1) \,|\, \beta_i = \beta_0] \\
&= \mathbb{P}[\tau(\beta_0) \leq \tfrac{1}{2}z_\alpha \,|\, \beta_i = \beta_0] + o_p(1) \to \Phi(\tfrac{1}{2}z_\alpha).
\end{aligned} \tag{24}$$

Here z_α denotes the α-quantile of the standard normal distribution. A similar result holds for inference based on the critical value $\tau^{*'}_{u,1-\alpha}$ for the other-sided alternative. Since $\Phi(\frac{1}{2}z_{0.05}) = 0.205$, we see that the difference between the nominal level and the asymptotic rejection probability of this bootstrap test can be very substantial, and hence the test procedure based on the distribution of $\tau^{*'}(\hat{\beta}_i)$ is found to be unsound too. The over-rejection of inference based on the distribution of $\tau^{*'}(\hat{\beta}_i)$ can be explained as follows. The test statistic $\tau^{*'}(\hat{\beta}_i)$ can be rewritten as

$$\tau^{*'}(\hat{\beta}_i) = -\frac{s}{s^*}\tau(\beta_0) + \tau^{*'}(\hat{\beta}_0). \tag{25}$$

Contrary to $\tau^{*^u}(\beta_0)$, which is centred around the observed t-statistic $\tau(\beta_0)$ under the null, the asymptotic bootstrap distribution of $\tau^{*^r}(\hat{\beta}_i)$ is centred around $-\tau(\beta_0)$, *i.e.* $\mathbb{E}_*[\tau^{*^r}(\hat{\beta}_i)] \to -\tau(\beta_0)$. Hence, the quantiles based on the distribution of $\tau^{*^r}(\hat{\beta}_i)$ are approximately shifted by the amount of $-\tau(\beta_0)$ with respect to the quantiles based on the distribution of $\tau^{*^r}(\hat{\beta}_0)$, *i.e.* $\tau^{*^r}_{u,\alpha} \to \tau^{*^r}_{r,\alpha} - \tau(\beta_0)$. This shift in the bootstrap distribution is shown in the lower graph of Figure 1, where the same realization of the t-statistic $\tau(\beta_0)$ is used as in the upper graph of Figure 1. Note that we reject the null hypothesis $H_0 : \beta_i = \beta_0$ against the alternative $H_2 : \beta_i < \beta_0$ if the t-statistic $\tau(\beta_0)$ is significantly negative. However, when $\tau(\beta_0)$ is negative, the critical value $\tau^{*^r}_{u,\alpha}$ tends to be greater than $\tau^{*^r}_{r,\alpha}$ since the quantity $-\tau(\beta_0)$ is positive. Hence, the inequality $\tau(\beta_0) \leq \tau^{*^r}_{u,\alpha}$ has a greater probability to occur than the event $\tau(\beta_0) \leq \tau^{*^r}_{r,\alpha}$, resulting in an asymptotic size that is greater than the nominal level.

Before turning to dynamic regression models, we return as promised earlier, to the issue of using restricted or unrestricted residuals in the bootstrap resampling procedures. When the null hypothesis is true, restricted estimators are of course more efficient than unrestricted estimators. Hence, at first sight it seems advantageous to resample from residuals based on restricted estimators. However, if the null hypothesis is false, the empirical distribution function of the restricted residuals will in general represent the population distribution not very well. To illustrate this point, we consider again the simple regression model as given in (1). Suppose we wish to test the hypothesis $H_0 : \delta = \delta_0$ for $\delta_0 = 0$, whereas the true value δ is different from zero, *i.e.* $\delta \neq 0$. Under the restriction $\delta = 0$, the restricted intercept equals the sample mean of y, *i.e.* $\hat{\mu}_r = \bar{y}$ where $\bar{y} = n^{-1}\sum_1^n y_i$. Since $\bar{y} = \mu + \delta\bar{x} + \bar{\varepsilon}$, we obtain the following expression for the restricted residuals

$$\hat{\varepsilon}_r = \delta(x - \bar{x}) + (\varepsilon - \bar{\varepsilon}). \tag{26}$$

If ε_i^* is drawn randomly with replacement from the restricted residuals $\hat{\varepsilon}_r$, then the variance of the bootstrap disturbances does not converge to σ^2, since

$$\mathrm{Var}_*(\varepsilon_i^*) \to \delta^2\sigma_x^2 + \sigma^2, \tag{27}$$

where $\sigma_x^2 = \lim_{n\to\infty} n^{-1}\sum_1^n (x_i - \bar{x})^2$. Hence, the asymptotic variance of the bootstrap disturbances based on restricted residuals will in general be larger than the asymptotic variance based on unrestricted residuals when the null is false. Moreover, if $n^{-1}\sum_1^n (x_i - \bar{x})^3$ is substantially different from zero, the distribution of ε_i^* will be skewed as well, even if the population distribution does not exhibit any skewness. Hence, a p-value obtained by the bootstrap is only correct when δ_0 equals the true value δ. This phenomenon is especially undesirable when a confidence interval is constructed by inverting a test statistic. Such a confidence interval requires bootstrapping a p-value for a whole range of test values δ_0. Since there is only one value of δ_0 which equals the true but unknown value δ, the bootstrapped p-values for $\delta_0 \neq \delta$ will be distorted if restricted residuals are used in the resampling scheme. Therefore, if the aim of the analysis is to approximate the null distribution of the test statistic even if the null hypothesis is false, we recommend to use the unrestricted residuals, which have the additional computational advantage of remaining the same irrespective of the value being tested.

4 Bootstrap Hypothesis Testing in Dynamic Models

By the theorems given in Freedman (1984), the asymptotic results of Propositions 2 and 3 immediately carry over to the stable autoregressive model of order one with fixed exogenous variables

$$y = \rho y_{-1} + X\beta + \varepsilon, \qquad |\rho| < 1, \tag{28}$$

where $y_{-1} = (y_0, ..., y_{n-1})$ and $\varepsilon_i \sim F_\varepsilon$. However, the one-to-one relation between the test statistic and confidence region approach, as established in Proposition 1 for fixed-regressors models, no longer holds here. In dynamic models with a lagged-dependent variable, we have to resort to a recursive resampling scheme. One of the consequences is that now the t-statistics $\tau^{*^u}(\hat{\beta}_i)$ and $\tau^{*^r}(\beta_0)$ are not identical in general and neither are the F-statistics $\lambda^{*^u}(\hat{\beta})$ and $\lambda^{*^r}(\hat{\beta}_r)$. We shall first illustrate this in the Gaussian AR(1) model with intercept and subsequently through a small-scale Monte Carlo study in the autoregressive moving-average of order (1,1) –ARMA(1,1)– model. Since the procedures based on the hybrid implementations are afflicted with serious size and power problems also in dynamic models, they are not considered in this section.

Consider the stationary Gaussian AR(1) model

$$y_t = \rho y_{t-1} + \mu + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \qquad t = 1, ..., n, \tag{29}$$

where the AR(1) parameter ρ lies in the stationarity zone, *i.e.* $|\rho| < 1$, and the starting value y_0 is drawn from the stationary distribution, *i.e.* $y_0 \sim \mathcal{N}(\mu/(1-\rho), \sigma^2/(1-\rho^2))$. The AR(1) model can also be written as $y = X\beta + \varepsilon$ with $X_t' = (y_{t-1}, 1)$ and $\beta = (\rho, \mu)'$, although now the matrix X is stochastic due to the presence of a lagged-dependent variable. We focus on testing $H_0\colon \rho = \rho_0$ using the t-statistic

$$\tau(\rho_0) = (\hat{\rho} - \rho_0)/s_{\hat{\rho}} \tag{30}$$

based on the maximum likelihood (ML) estimator conditional on y_0, which is equivalent to the least-squares estimator. Since y_0 is drawn from the stationary distribution, the distribution of the t-statistic $\tau(\rho_0)$ does not depend on the nuisance parameters μ and σ^2; see for instance Nankervis and Savin (1988). The recursive dynamic (d) analogues of resampling schemes I and II based on the parametric bootstrap are:

$$\text{Scheme I}^d\ (*^u): \qquad y_t^{*^u} = \hat{\rho}\, y_{t-1}^{*^u} + \hat{\mu} + \varepsilon_t^*,$$
$$\text{Scheme II}^d\ (*^r): \qquad y_t^{*^r} = \rho_0\, y_{t-1}^{*^r} + \hat{\mu}_r + \varepsilon_t^*,$$

where y_0^* is defined in such a way that y_t^* is stationary, $\varepsilon_t^* \sim \mathcal{N}(0, \hat{\sigma}^2)$ and $t = 1, ..., n$. Since the bootstrap distributions of $\tau^{*''}(\hat{\rho})$ and $\tau^{*'}(\rho_0)$ are invariant to the estimator $\hat{\sigma}^2$, it makes no difference whether $\hat{\sigma}^2$ is based on the unrestricted or restricted residuals. More information about bootstrapping general stationary ARMA(p, q)-models can be found in Kreiss and Franke (1992).

Since $\tau(\rho_0)$ is invariant with respect to μ and σ^2, the distribution of $\tau(\rho_0)$ only depends on F_ε and ρ_0. The latter is assumed to be known under the null hypothesis. In the *parametric* 'bootstrap world', the distribution of $\tau^{*''}(\hat{\rho})$ depends on the estimator $\hat{\rho}$ only while the distribution of $\tau^{*'}(\rho_0)$ depends on the parameter ρ_0 only. Because of these invariance properties, the bootstrap distribution of $\tau^{*'}(\rho_0)$ mimics the finite-sample distribution of $\tau(\rho_0)$ exactly, which results in an exact inference procedure. In effect, the bootstrap test reduces to a Monte Carlo test, which has a size which does not exceed the nominal level. So, even if the critical value $\tau_{r,\alpha}^{*'}$ is approximated by a finite number of bootstrap replications, bootstrap inference based on the test statistic approach yields exact inference when the number of bootstrap replications B is chosen such that $\alpha(B+1)$ is an integer; see *inter alia* Hall (1986). Hence, we conclude that when F_ε is known

$$\alpha_2(\tau_r^{*'}) = \mathbb{P}[\tau(\rho_0) \leq \tau_{r,\alpha}^{*'}] = \alpha. \tag{31}$$

Next, we consider the finite sample properties of the confidence region approach. Since $\hat{\rho}$ is stochastic and in general different from ρ_0, inference based on the critical value $\tau_{u,\alpha}^{*''}$ will not be exact. To illustrate this point, Figure 2 shows the critical values based on the test statistic approach, *i.e.* $\tau_{r,\alpha}^{*'}$, and the confidence region approach, *i.e.* $\tau_{u,\alpha}^{*''}$, as the estimator $\hat{\rho}$ varies for the case $\rho_0 = 0.8$, $n = 25$, $\alpha \in \{0.05, 0.95\}$ and F_ε is normal. Of course, $\tau_{r,\alpha}^{*'}$ remains constant for fixed values of ρ_0 and n, just like the α-quantile based on the exact finite-sample null distribution. Note that due to the parametric bootstrap, which makes the distribution of $\tau^{*''}(\hat{\rho})$ invariant with respect to the estimated intercept and variance, the quantiles of the bootstrap distribution of $\tau^{*''}(\hat{\rho})$ are constant *conditional* on a realization of the estimator $\hat{\rho}$. However, $\hat{\rho}$ is stochastic in finite sample so that $\tau_{u,\alpha}^{*''}$ will be different for different realizations of $\hat{\rho}$. Figure 2, which is based on 100,000 simulation replications for a grid of 190 $\hat{\rho}$ values, shows that the quantiles based on the confidence region approach, *i.e.* $\tau_{u,\alpha}^{*''}$, vary substantially with realizations of the estimator $\hat{\rho}$. It appears that the parametric bootstrap distribution of $\tau^{*''}(\hat{\rho})$ is quite close to the non-central t-distribution[2] with a non-centrality parameter equal to $\mathbb{E}_*[\tau^{*''}(\hat{\rho})]$. By the results in Tanaka (1983, formula (3.3)), this expectation for a given value of $\hat{\rho}$ can be approximated

[2]Based on extensive simulations of the distribution of $\tau(\rho)$ for $T \geq 25$ and $|\rho| \leq 0.95$, Nankervis and Savin (1988, p. 127) observe that "the skewness is essentially zero and that the kurtosis is about the same as for Student's t". This observation also applies to the parametric bootstrap distribution of $\tau^{*''}(\hat{\rho})$, since this distribution is equal to the distribution of $\tau(\rho)$ for $\rho = \hat{\rho}$.

as

$$\mathbb{E}_*[\tau^{*''}(\hat{\rho})] = \mathbb{B}_{\tau^*}(\hat{\rho}) + O(T^{-1}), \qquad \text{with } \mathbb{B}_{\tau^*}(\hat{\rho}) = -\frac{1}{\sqrt{T}}\frac{2\hat{\rho}+1}{\sqrt{1-\hat{\rho}^2}}. \tag{32}$$

Note that the approximation of the expectation of $\tau^{*'}(\rho_0)$ up to first order is given by $\mathbb{B}_{\tau^*}(\rho_0)$, so that $\tau^{*''}_{u,\alpha} \approx \tau^{*'}_{r,\alpha} - [\mathbb{B}_{\tau^*}(\rho_0) - \mathbb{B}_{\tau^*}(\hat{\rho})]$. Because the function $\mathbb{B}_{\tau^*}(\hat{\rho})$ is decreasing in $\hat{\rho}$ for $\hat{\rho} \in (-1, 1)$, we find the inequality $\tau^{*''}_{u,\alpha} > \tau^{*'}_{r,\alpha}$ for $-1 < \hat{\rho} < \rho_0$ and $\rho_0 > 0$, which is evident in Figure 2 for $\rho_0 = 0.8$. The inequality $\tau^{*''}_{u,\alpha} > \tau^{*'}_{r,\alpha}$ leads to a rejection probability against the one-sided alternative $H_2\colon \rho < \rho_0$ which is larger than α in finite samples, *i.e.* $\alpha_2(\tau^{*''}_u) > \alpha_2(\tau^{*'}_r) = \alpha$. We can substantiate this as follows. Let $f_n(x)$ denote the finite-sample density function of the OLS estimator $\hat{\rho}$. If α is small enough such that $\tau^{*''}_{u,\alpha}$ and $\tau^{*'}_{r,\alpha}$ are negative ($\alpha < 0.5$), then the rejection probability of the confidence region approach can be bounded from below by

$$\begin{aligned}
\alpha_2(\tau^{*''}_u) &= \mathbb{P}[\tau(\rho_0) \le \tau^{*''}_{u,\alpha}] = \textstyle\int_{-\infty}^{+\infty} \mathbb{P}[\tau(\rho_0) \le \tau^{*''}_{u,\alpha} \mid \hat{\rho} = x]\; f_n(x)\,\mathrm{d}x \\
&= \textstyle\int_{-\infty}^{\rho_0} \mathbb{P}[\tau(\rho_0) \le \tau^{*''}_{u,\alpha} \mid \hat{\rho} = x]\; f_n(x)\,\mathrm{d}x \\
&> \textstyle\int_{-\infty}^{\rho_0} \mathbb{P}[\tau(\rho_0) \le \tau^{*'}_{r,\alpha} \mid \hat{\rho} = x]\; f_n(x)\,\mathrm{d}x \\
&= \textstyle\int_{-\infty}^{+\infty} \mathbb{P}[\tau(\rho_0) \le \tau^{*'}_{r,\alpha} \mid \hat{\rho} = x]\; f_n(x)\,\mathrm{d}x = \alpha_2(\tau^{*'}_r) = \alpha.
\end{aligned} \tag{33}$$

The second equality in the first line is a consequence of the law of iterated expectations by writing $\mathbb{P}[\tau(\rho_0) \le \tau^{*''}_{u,\alpha}] = \mathbb{E}[I\{\tau(\rho_0) \le \tau^{*''}_{u,\alpha}\}] = \mathbb{E}_{\hat{\rho}}\left[\mathbb{E}[I\{\tau(\rho_0) \le \tau^{*''}_{u,\alpha}\}|\hat{\rho}]\right]$, where $I\{\cdot\}$ denotes the indicator function. The equalities in the second and fourth lines result from the fact that the hypothesis $H_0 : \rho = \rho_0$ is rejected against the alternative $H_2 : \rho < \rho_0$ if the t-statistic $\tau(\rho_0) = (\hat{\rho} - \rho_0)/s_{\hat{\rho}}$ is sufficiently negative because $\tau^{*''}_{u,\alpha} < 0$ and $\tau^{*'}_{r,\alpha} < 0$ by assumption. Over the one-sided interval $\hat{\rho} > \rho_0$, $\tau(\rho_0)$ is positive so the probabilities of the events $\{\tau(\rho_0) < \tau^{*''}_{u,\alpha}\}$ and $\{\tau(\rho_0) < \tau^{*'}_{r,\alpha}\}$ for $\hat{\rho} > \rho_0$ are zero. The inequality in the third line is due to the fact that $\mathbb{P}[\tau(\rho_0) \le \upsilon] > \mathbb{P}[\tau(\rho_0) \le w]$ for $\upsilon > w$. A similar argument shows that the rejection probability for the other one-sided alternative, *i.e.* $\alpha_3(\tau^{*''}_u)$, is smaller than the nominal level α (for small α such that $\tau^{*''}_{u,1-\alpha} > 0$ and $\tau^{*'}_{r,1-\alpha} > 0$) since $\tau^{*'}(\rho_0)$ is less biased than $\tau^{*''}(\hat{\rho})$ when $\hat{\rho} > \rho_0$. Overall, we conclude that in the stationary Gaussian AR(1) model with an intercept only the test statistic approach leads to exact parametric bootstrap inference in finite samples.

Insert Figure 2 about here.

If, instead of the parametric bootstrap, the non-parametric bootstrap is used a similar picture emerges although now the situation is less unambiguous. This can be attributed to the fact that, in the non-parametric case, the α-quantile $\tau^{*'}_r(\alpha)$ will be stochastic even conditional on a realization of the estimator $\hat{\rho}$ since now the bootstrap distribution depends among other things on the realization of the empirical distribution function of the residuals. For the non-parametric bootstrap, Bose (1988) has shown that the bootstrap distribution based on the confidence region approach is capable of making a first-order Edgeworth correction in AR(1) models, *i.e.* the bootstrap is second-order accurate. Since simulation results reported in Giersbergen and Kiviet (1994) and Davidson (1999) show that the test statistic approach outperforms the confidence region approach, we conjecture that the bootstrap distribution based on the test statistic approach is at least second-order accurate. However, if the test statistic approach is precisely second-order accurate, just like the confidence region approach, the theory of Edgeworth expansions cannot be used to explain the finite-sample difference between the two approaches.

Next, we consider the ARMA(1,1) model

$$y_t = \rho y_{t-1} + \mu + \varepsilon_t + \theta \varepsilon_{t-1}, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \qquad t = 1, ..., n, \tag{34}$$

where ρ and θ are the parameters of interest. The finite-sample performance of the two bootstrap test procedures is investigated by a small Monte Carlo experiment. All simulations were carried out using the matrix programming language Gauss 2.2. We generated 1000 ($= N$) samples according to model (34) for the parameter values $(\rho, \theta, \mu) = (0.53, 0.55, 0.82)$, $\sigma^2 = 0.15$ and $n = 99$. These parameter values were inspired by the ARMA(1,1) estimates obtained from annual unemployment data which are analysed in the next section. The unrestricted residuals were multiplied by $\sqrt{n/(n-3)}$ to correct for the loss of degrees of freedom; see Bickel and Freedman (1983). For each sample $\{y_t\}$, the bootstrap critical values were calculated from 999 ($= B$) bootstrap realizations of the appropriate test statistic. The test statistics were based on the unconditional ML estimators of the parameters ρ and θ, where the starting values were generated according to the procedure proposed by McLeod and Hipel (1978, Wasim 2). The Wasim procedure is based on the Cholesky decomposition of the covariance matrix of $(y_1, y_0, \varepsilon_1, \varepsilon_0)$. We restrict ourselves to the non-parametric bootstrap since this implementation is most used in practice. The starting values in the bootstrap procedure were generated by starting up the ARMA process and disregarding the first 30 observations. Table 1 shows the rejection frequencies of the null hypotheses $H_0\colon \rho = 0.53$ and $H_0\colon \theta = 0.55$ against their one-sided alternatives at 5% significance level in the Monte Carlo experiment. Note that the two hypotheses are satisfied by the data generation process, so that the rejection frequencies yield an indication of the actual size of the test procedures. Since the true parameter values are far from the instability boundary, the events $|\hat{\rho}| > 1$ or $|\hat{\theta}| > 1$ did not occur in the simulations.

Insert Table 1 about here.

The Monte Carlo results can be summarised as follows. Inference on the AR(1) parameter based on the asymptotic approximation appears to be reasonably accurate, although the rejection frequencies are significantly different from the nominal level α. Asymptotic inference on the MA(1) parameter is less reliable, since the rejection frequencies are approximately two times larger/smaller than the nominal level α. The bootstrap approximation yields a uniform improvement over the asymptotic approximation. Note that inference based on the distribution of $\tau^{*''}(\hat{\theta})$, *i.e.* the confidence region approach, seems to over-correct, since the rejection frequency for the alternative $\theta > 0.55$ is significantly too low whereas the rejection frequency based on the asymptotic approximation is significantly too high. Inference based on the test statistic approach mimics the behaviour of an exact inference procedure since none of the rejection frequencies are significantly different from the nominal level. Note that the test statistic approach employs a resampling scheme which avoids the stochastic variability due to the parameter under test. For instance, when testing the hypothesis $H_0 : \rho = \rho_0$ the resampling scheme incorporates ρ_0 instead of $\hat{\rho}$. Of course, the test statistic approach cannot avoid the stochastic variability due to the nuisance parameters. However, it is expected that the finite-sample distribution of a bootstrap test statistic is more affected by the value of the parameter under test than by the values of the estimated nuisance parameters. For example, in the AR(1) model (29) using resampling scheme $y_t^* = \rho_0\, y_{t-1}^* + \hat{\mu}_r + \varepsilon_t^*$, the distribution of $\tau^{*'}(\rho_0)$ does not depend on the value of $\hat{\mu}_r$. Since the t-statistics $\tau^{*''}(\hat{\rho})$ and $\tau^{*''}(\hat{\theta})$ in the ARMA(1,1) model are not invariant with respect to $\hat{\rho}$ and $\hat{\theta}$ respectively, the use of the restricted estimators, which incorporate additional non-sample information, in the resampling scheme seems helpful in improving the bootstrap approximation, especially for the MA(1) parameter. Overall, bootstrap inference based on the test statistic approach performs best, followed by bootstrap inference based on the confidence region approach. Inference based on the first-order asymptotic approximation is least accurate. Although our Monte Carlo design is limited, the results are useful for explaining the empirical findings that are obtained in the next section.

5 Empirical Illustration

To compare the different (bootstrap) approximations in practice, we consider annual data on unemployment for the U.S. from 1890 through 1988 as used by Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991). Note that the unemployment rate is the only time series of the fourteen considered by Nelson and Plosser which appears to be stationary. We focus on the construction of confidence intervals. Instead of fitting a AR(p) model, we chose to model the data as an ARMA(1,1) model. ML yields the following estimates (t-values between parentheses):

$$y_t = \underset{(5.11)}{0.527} y_{t-1} + \underset{(4.96)}{0.822} + \hat{\varepsilon}_t + \underset{(5.37)}{0.554} \hat{\varepsilon}_{t-1}. \tag{35}$$

There appears to be no significant autocorrelation in the residuals; a p-value of 0.18 is obtained for a likelihood ratio test against the ARMA(2,1) specification and a p-value of 0.52 is obtained for the

Box-Ljung test statistic based on the first 10 residual autocorrelations. The non-parametric bootstrap will be employed since the hypothesis of normality is rejected; a p-value of 0.001 is obtained for the Jarque-Bera normality test. The non-normality appears to be caused mainly by two 'large' residuals in the years 1893 and 1918.

The 95% equal-tail bootstrap confidence intervals for ρ and θ based on the confidence region approach are obtained as follows. Bootstrap observations $\{y_t^*\}$ are generated by the 'unrestricted' resampling scheme

$$y_t^* = 0.527 y_{t-1}^* + 0.822 + \varepsilon_t^* + 0.554 \varepsilon_{t-1}^*, \qquad t = 1, ..., 99, \tag{36}$$

where ε_t^* is drawn randomly with replacement from the unrestricted but rescaled residuals $\hat{\varepsilon}_t$. Let $\tau_{\alpha(B+1)}^{\rho}$ denote the $\alpha(B+1)$-th order statistic from the B bootstrap realizations $\{\tau^{*^u}(\hat{\rho})_b\}_{b=1}^{B}$. A 95% confidence interval for ρ is given by

$$(\hat{\rho} - s_{\hat{\rho}}\, \tau_{0.975(B+1)}^{\rho}, \hat{\rho} - s_{\hat{\rho}}\, \tau_{0.025(B+1)}^{\rho}). \tag{37}$$

A similar interval for θ is defined with respect to the quantities $\hat{\theta}$, $s_{\hat{\theta}}$ and $\tau_{\alpha(B+1)}^{\theta}$. The bootstrap confidence intervals for ρ and θ based on the test statistic approach are obtained as follows. For simplicity, we focus on the construction of the interval for ρ. Since the confidence interval is obtained by inverting a test statistic, a whole range of parameter values ρ_0 has to be tested. Hence, the test statistic approach is more computer-intensive than the confidence region approach[3], especially when constructing a confidence region for a high dimensional parameter vector. Bootstrap observations $\{y_t^*\}$ under $H_0 : \rho = \rho_0$ are generated by the 'restricted' resampling scheme

$$y_t^* = \rho_0 y_{t-1}^* + \hat{\mu}_r + \varepsilon_t^* + \hat{\theta}_r \varepsilon_{t-1}^*, \qquad t = 1, ..., 99, \tag{38}$$

where $\hat{\mu}_r$ and $\hat{\theta}_r$ denote restricted estimators, and ε_t^* is drawn randomly with replacement from the unrestricted but rescaled residuals $\hat{\varepsilon}_t$. The bootstrapped p-value, based on B bootstrap realizations $\{\tau^{*^r}(\rho_0)_b\}_{b=1}^{B}$ for $H_0 : \rho = \rho_0$ against $H_1 : \rho < \rho_0$ using the observed t-statistic $\tau(\rho_0)$ is given by

$$p^*\text{-value} = (B+1)^{-1} \sum_{b=1}^{B} I\{\tau^{*^r}(\rho_0)_b \leq \tau(\rho_0)\}. \tag{39}$$

The upper confidence limit is equal to the value of ρ_0 for which the p^*-value based on the lower-sided alternative $H_1 : \rho < \rho_0$ equals 2.5%, whereas the lower confidence limit is equal to the value ρ_0 for which the p^*-value based on the upper-sided alternative $H_2 : \rho > \rho_0$ equals 2.5%. The confidence

[3]In a recent paper, Hansen (1999) proposes to use a limited grid of parameter values and estimate the dependence of the bootstrap quantiles upon the parameter value by kernel regression to reduce the computational burden.

interval for θ is obtained in a similar way. Figure 3 shows the p^*-values for the one-sided alternatives when testing the parameter sets $\rho_0 \in (0.2, 0.8)$ and $\theta_0 \in (0.15, 0.8)$. Graphically the confidence limits are found as the intersection of the solid lines, which denote the p^*-values, with the dashed lines, which denote the 2.5% nominal level.

Insert Figure 3 about here.

Table 2 shows the confidence intervals that were obtained by the different approximations. The entries in the column below R/L are a measure of the asymmetry of the various confidence intervals. Contrary to the symmetric confidence intervals based on the first-order asymptotic approximation, the bootstrap confidence intervals appear to be asymmetric, especially for the MA(1) parameter, even though the sample size is quite large. The Monte Carlo results of the previous section suggest that the confidence intervals based on the test statistic approach are most accurate. The intervals based on the confidence region approach seem to be too asymmetric, as was also found in the Monte Carlo study.

Insert Table 2 about here.

6 Concluding Remarks

We have contrasted two approaches for testing a hypothesis using the bootstrap. The test statistic approach is based on *restricted* estimators in both the test statistic and resampling scheme, while the confidence region approach employs *unrestricted* estimators in the test statistic and resampling scheme. Two hybrid approaches emerge if the test statistic approach is combined with the confidence region approach. The hybrid approaches are shown to be cursed with serious size (and power) problems since the bootstrap distribution revolves around the realization of the test statistic (possibly with opposite sign) obtained from the original sample. Hence, these hybrid implementations are found to be completely inadequate for inference purposes.

In static regression models, the test statistic and confidence region approaches are shown to be identical, even when various restrictions are tested jointly. Hence, also in the 'bootstrap world', there is a one-to-one correspondence between testing hypotheses and constructing confidence regions in static models. However, in dynamic models, this equivalence breaks down and the finite-sample performance of the two approaches can be quite different. In the stationary Gaussian AR(1) model, the test statistic approach leads to exact parametric bootstrap inference. The inaccuracy of the confidence region approach stems from the fact that the critical bootstrap values depend on the unrestrictedly estimated AR(1) parameter. We conjecture that in general dynamic models, the test statistic approach has to be preferred to the confidence region approach in finite samples since the unrestricted resampling scheme (utilised in the confidence region approach) is inflicted by stochastic variation which may be reduced or even completely annihilated by the restricted resampling scheme (utilised in the test statistic approach). This observation implies that confidence intervals based on inverting a test statistic will in general have a smaller coverage error than confidence intervals based on the confidence region approach. Among others, Carpenter (1999), Davidson (1999) and Hansen (1999) share this conclusion.

Note that in the class of models we examined the test statistic and confidence region approaches are both consistent, *i.e.* their bootstrap distributions converge to the correct asymptotic distribution. So, the conclusion to favour the former approach in stable dynamic models is based on its finite-sample performance. Note, however, that in AR models containing a unit root, Basawa *et al.* (1991) have demonstrated the inconsistency of the confidence region approach, whereas the test statistic approach does lead to correct bootstrap inference; see for instance Ferretti and Romo (1996) or Hansen (1999). Hence, the superiority of the test statistic approach in unit root models is not simply a finite-sample phenomenon, but follows from asymptotic theory.

A Proof of Lemma 1

Lemma 1 *In the linear fixed-regressors model (8),* $\mathbb{P}[\lambda^{*''}(\hat{\beta}) \pm \gamma^* \leq 0] \leq 0.5$ *as* $n \to \infty$.

Proof. The quantity γ^* involves the term $\nabla = (R\beta - r_0)$, which is non-zero when the null hypothesis $H_0 : R\beta = r_0$ is false. To avoid exploding quantities in the analysis, we only consider local alternatives. Define $\upsilon_m \sim \mathcal{N}(\nabla, \Sigma)$ and $\upsilon_m^* \sim \mathcal{N}(0, \Sigma)$, where $\Sigma = RQ^{-1}R'$. From Freedman (1981), it follows that

$$(\lambda^{*''}(\hat{\beta}) \pm \gamma^*) \xrightarrow{d} \tfrac{1}{m}(\upsilon_m^{*\prime}\Sigma^{-1}\upsilon_m^* \pm 2\upsilon_m'\Sigma^{-1}\upsilon_m^*). \tag{A.1}$$

Since Σ is positive definite, there exists a non-singular $m \times m$ matrix C such that $C'C = \Sigma^{-1}$. Put $u_m = C\upsilon_m$ and $u_m^* = C\upsilon_m^*$, so that $u_m \sim N(C\nabla, I_m)$ and $u_m^* \sim N(0, I_m)$. Then we have

$$\begin{aligned}\mathbb{P}[\lambda^{*''}(\hat{\beta}) + \gamma^* \leq 0] &\to \mathbb{P}[(C\upsilon_m^*)'(C\upsilon_m^*) + 2(C\upsilon_m)'(C\upsilon_m^*) \leq 0] \\ &= \mathbb{P}[u_m^{*\prime}u_m^* + 2u_m'u_m^* \leq 0].\end{aligned} \tag{A.2}$$

Let $J = \left(\begin{smallmatrix}1 & 1\\ 1 & 0\end{smallmatrix}\right)$ and $J_{2m} = J \otimes I_m$. Defining $u_{2m} = (u_m^{*\prime}\ u_m')'$, we can write $u_m^{*\prime}u_m^* + 2u_m'u_m^*$ as $u_{2m}'J_{2m}u_{2m}$. Since J is a real symmetric matrix, it can be diagonalized by a real orthogonal matrix K, *i.e.*

$$K'JK = \Lambda, \tag{A.3}$$

where Λ =diag(λ_1, λ_2) with $\lambda_1 = \frac{1}{2}(1-\sqrt{5}) \approx -0.618$ and $\lambda_2 = \frac{1}{2}(1+\sqrt{5}) \approx 1.618$. In fact, K is given by

$$K = \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}-1}{\sqrt{10-2\sqrt{5}}} & \frac{-\sqrt{2}}{\sqrt{5-\sqrt{5}}} \\ \frac{-\sqrt{2}}{\sqrt{5-\sqrt{5}}} & \frac{1-\sqrt{5}}{\sqrt{10-2\sqrt{5}}} \end{bmatrix} \approx \begin{bmatrix} 0.526 & -0.851 \\ -0.851 & -0.526 \end{bmatrix}.$$

If $K_{2m} = K \otimes I_m$, then $K_{2m}'J_{2m}K_{2m} = \Lambda \otimes I_m$. Since $K_{2m}'K_{2m} = I_{2m}$, we obtain $w_{2m} = (w_m^{*\prime}\, w_m')' = K_{2m}'u_{2m} \sim \mathcal{N}(\mu_{2m}, I_{2m})$, where $\mu_{2m} = (\mu_m^{*\prime}\, \mu_m')'$ with $\mu_m^* = \kappa_{21}C\nabla$ and $\mu_m = \kappa_{22}C\nabla$. Now

$$\begin{aligned}\mathbb{P}[u_m^{*\prime}u_m^* + 2u_m'u_m^* \leq 0] &= \mathbb{P}[(K_{2m}'u_{2m})'K_{2m}'J_{2m}K_{2m}(K_{2m}'u_{2m}) \leq 0] \\ &= \mathbb{P}[w_{2m}'(\Lambda \otimes I_m)w_{2m} \leq 0] \\ &= \mathbb{P}[\textstyle\sum_{i=1}^m w_i^2 / \sum_{i=1}^m w_i^{*2} \leq -\lambda_1/\lambda_2] \\ &= \mathbb{P}[F_{m,m}(\delta_1, \delta_2) \leq \lambda],\end{aligned} \tag{A.4}$$

where $\lambda \equiv -\lambda_1/\lambda_2 = \frac{\sqrt{5}-1}{\sqrt{5}+1} \approx 0.382$ and $F_{m,m}(\delta_1, \delta_2)$ is a random variable having a non-central F-distribution with (m, m) degrees of freedom and non-centrality parameters $\delta_1 = \mu_m'\mu_m$ and $\delta_2 = \mu_m^{*\prime}\mu_m^*$. Next, we consider two case: (i) the null is true; and (ii) the null is false. When the null is true, $\delta_1 = \delta_2 = 0$ and the lemma follows from the fact that $\mathbb{P}[F_{m,m}(0,0) \leq 0.382] < \mathbb{P}[F_{m,m}(0,0) \leq 1] = 0.5$. When the null is false, the lemma is proved as follows. First, note that $\kappa_{22}^2/\kappa_{21}^2 = \lambda$, so that $\delta_1 = \lambda\delta_2$. Hence, $\delta_1 < \delta_2$ and $\mathbb{P}[F_{m,m}(\delta_1, \delta_2) \leq \lambda]$ is an increasing function of $||\delta_2||$, where $||\cdot||$ denotes some well-defined matrix norm. However, we shall show that the probability is bounded from above by 0.5 as the non-centrality parameters become large. Define the two random vectors $z_m \sim \mathcal{N}(0, I_m)$ and $z_m^* \sim \mathcal{N}(0, I_m)$. Now, the probability given in (A.4) can be written as

$$\mathbb{P}[(z_m + \mu_m)'(z_m + \mu_m) \leq \lambda(z_m^* + \mu_m^*)'(z_m^* + \mu_m^*)], \tag{A.5}$$

where $\mu_m'\mu_m = \lambda\mu_m^{*\prime}\mu_m^*$. Furthermore, if $||\nabla||$ becomes large, then $z_m'z_m$ and $\lambda z_m^{*\prime}z_m^*$ become negligibly small in comparison to $2\mu_m'z_m$ and $2\lambda\mu_m^{*\prime}z_m^*$ respectively. Hence, as $||\nabla|| \to \infty$, the probability given in (A.5) can be accurately approximated by

$$\mathbb{P}[\mu_m'z_m \leq \lambda\mu_m^{*\prime}z_m^*] = 0.5, \tag{A.6}$$

since $\mathbb{P}[x \leq y] = 0.5$ for $x \sim \mathcal{N}(0, \sigma_x^2)$ and $y \sim \mathcal{N}(0, \sigma_y^2)$. Hence, we find that $\mathbb{P}[\lambda^{*''}(\hat{\beta}) + \gamma^* \leq 0] \leq 0.5$ asymptotically irrespective whether the null is true or false. Hence, Lemma 1 is proven in case of addition, *i.e.* $\lambda^{*''}(\hat{\beta}) + \gamma^*$. In case of subtraction, *i.e.* $\lambda^{*''}(\hat{\beta}) - \gamma^*$, we define the matrix $J^- = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$. Since the eigenvalues of J^- are the same as the eigenvalues of J, the proof also holds for $\lambda^{*''}(\hat{\beta}) - \gamma^*$ except that a slightly different matrix K^- is used to diagonalize the matrix J^-, *viz.* κ_{11} and κ_{22} are interchanged. □

References

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., Taylor, R. L., 1991. Bootstrapping unstable first order autoregressive processes, The Annals of Statistics 19, 1089–1101.

Beran, R., 1986. Simulated power functions, The Annals of Statistics 14, 151–173.

Beran, R., 1987. Prepivoting to reduce level error of confidence sets, Biometrika 74, 151–173.

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series, Econometric Reviews 19, 1–48.

Bickel, P. J., Freedman, D. A., 1983. Bootstrapping regression models with many parameters, in: Bickel, P. J., Doksum, K., Hodges, J. C. eds., A Festschrift for Erich Lehmann (California: Wadsworth), 28–48.

Bose, A., 1988. Edgeworth correction by bootstrap in autoregressions, The Annals of Statistics 16, 1709–1722.

Carpenter, J., 1999. Test inversion bootstrap confidence intervals, Journal of the Royal Statistical Society (Series B) 61, 159–172.

Davidson, R., 1999. Comment on 'Recent developments in bootstrapping time series', Econometric Reviews 19, 49–54.

Davidson, R., MacKinnon, J. G., 1999. The size distortion of bootstrap tests. Econometric Theory 15, 361–376.

Davison, A. C., Hinkley, D. V., 1997. Bootstrap Methods and their Applications (Cambridge University Press, Cambridge).

DiCiccio, T. J., Romano, J. P., 1990. Nonparametric confidence limits by resampling methods and least favorable families, International Statistical Review 58, 59–76.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap (Chapman & Hall, London).

Ferretti, N., Romo, J., 1996. Unit root bootstrap tests for AR(1) models, Biometrika 83, 849–860.

Freedman, D. A., 1981. Bootstrapping regression models, The Annals of Statistics 9, 1218–1228.

Freedman, D. A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models, The Annals of Statistics 12, 827–842.

Giersbergen, N. P. A. v., Kiviet, J. F., 1994. How to Implement Bootstrap Hypothesis Testing in Static and Dynamic Regression Models. Discussion paper TI 94-130 (Tinbergen Institute, Amsterdam). Paper presented at ESEM 1994 (Maastricht, The Netherlands) and EC^2 1993 (Oxford, U.K.).

Hall, P., 1986. On the number of bootstrap simulations required to construct a confidence interval, The Annals of Statistics 14, 1453–1462.

Hall, P., 1989. Unusual properties of bootstrap confidence intervals in regression problems, Probablity Theory and Related Fields 81, 247–273.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion (Springer-Verlag).

Hall, P., Wilson, S. R., 1991. Two guidelines for bootstrap hypothesis testing, Biometrics 47, 757–762.

Hansen, B., 1999. The grid bootstrap and the autoregressive model, The Review of Economics and Statistics 81, 594–607.

Horowitz, J. L., 1994. Bootstrap-based critical values for the Information-matrix test, Journal of Econometrics 61, 395–411.

Horowitz, J. L., 1997. Bootstrap Methods in Econometrics: Theory and Numerical Performance, in: Kreps, D., and Wallis, K., eds., Advances in Economics and Econometrics: Theory and Applications (Cambridge University Press, Cambridge), Vol. 3, 188–222.

Kreiss, J.-P., Franke, J., 1992. Bootstrapping stationary autoregressive moving-average Models, Journal of Time Series Analysis 13, 297–317.

Li, H., Maddala, G. S., 1996. Bootstrapping time series models (with discussion), Econometric Reviews 15, 115–158.

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models, The Annals of Statistics 21, 255–285.

McLeod, A. I., Hipel, K. W., 1978. Simulation procedures for Box-Jenkins models, Water Resources Research 14, 969–975.

Nankervis, J. C., Savin, N. E., 1988. The Student's t approximation in a stationary first order autoregressive model, Econometrica 56, 119–145.

Navidi, W., 1989. Edgeworth expansions for bootstrapping regression models, The Annals of Statistics 17, 1472–1478.

Nelson, C. R., Plosser, C. I., 1982. Trends and random walks in macroeconomic time series, Journal of Monetary Economics 10, 139–162.

Schotman, P. C., Dijk, H. K. v., 1991. On bayesian routes to unit roots, Journal of Applied Econometrics 6, 387–401.

Tanaka, K., 1983. Asymptotic expansions associated with the AR(1) model with unknown mean, Econometrica 51, 1221–1231.

Tibshirani, R., 1992. Comment on 'Two guidelines for bootstrap hypothesis testing' by P. Hall and S.R. Wilson, Biometrics 48, 969–970.

Wu, C. F. J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis (with discussion), The Annals of Statistics 14, 1261–1295.

Table 1: Rejection frequencies (in %) for $H_0 : \rho = 0.53$ and $H_0 : \theta = 0.55$ against their one-sided alternatives in the model $y_t = \rho y_{t-1} + \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ at nominal significance level $\alpha = 5\%$ (N=1000, B=999, n=99).

	alternative		alternative	
	$\rho < 0.53$	$\rho > 0.53$	$\theta < 0.55$	$\theta > 0.55$
$\mathcal{N}(0,1)$	6.7	3.2	2.4	11.0
$\tau^{*^u}(\hat{\beta}_i), \quad \hat{\beta}_i \in \{\hat{\rho}, \hat{\theta}\}$	4.8	5.2	4.5	3.2
$\tau^{*^r}(\beta_0), \quad \beta_0 \in \{\rho, \theta\}$	5.0	4.4	4.7	4.5

Table 2: Various empirical 95% equal-tail confidence intervals for the ARMA(1,1) parameters (B=4999).

	$(\hat{\rho}_L, \hat{\rho}_U)$	R/L	$(\hat{\theta}_L, \hat{\theta}_U)$	R/L
$\mathcal{N}(0,1)$	(0.325,0.729)	1.00	(0.352,0.757)	1.00
$\tau^{*^u}(\hat{\beta}_i), \quad \hat{\beta}_i \in \{\hat{\rho}, \hat{\theta}\}$	(0.343,0.748)	1.20	(0.245,0.720)	0.54
$\tau^{*^r}(\beta_0), \quad \beta_0 \in \{\rho_0, \theta_0\}$	(0.331,0.751)	1.14	(0.283,0.716)	0.60

Note: $R/L = (\hat{\beta}_U - \hat{\beta})/(\hat{\beta} - \hat{\beta}_L)$.

Figure 1: Graphical comparison of the bootstrap distributions based on $\tau^{*^{u}}(\beta_0)$ (straight line, upper figure), $\tau^{*^{u}}(\hat{\beta}_i)$ (dashed line, upper figure), $\tau^{*^{r}}(\hat{\beta}_i)$ (straight line, lower figure) and $\tau^{*^{r}}(\beta_0)$ (dashed line, lower figure).

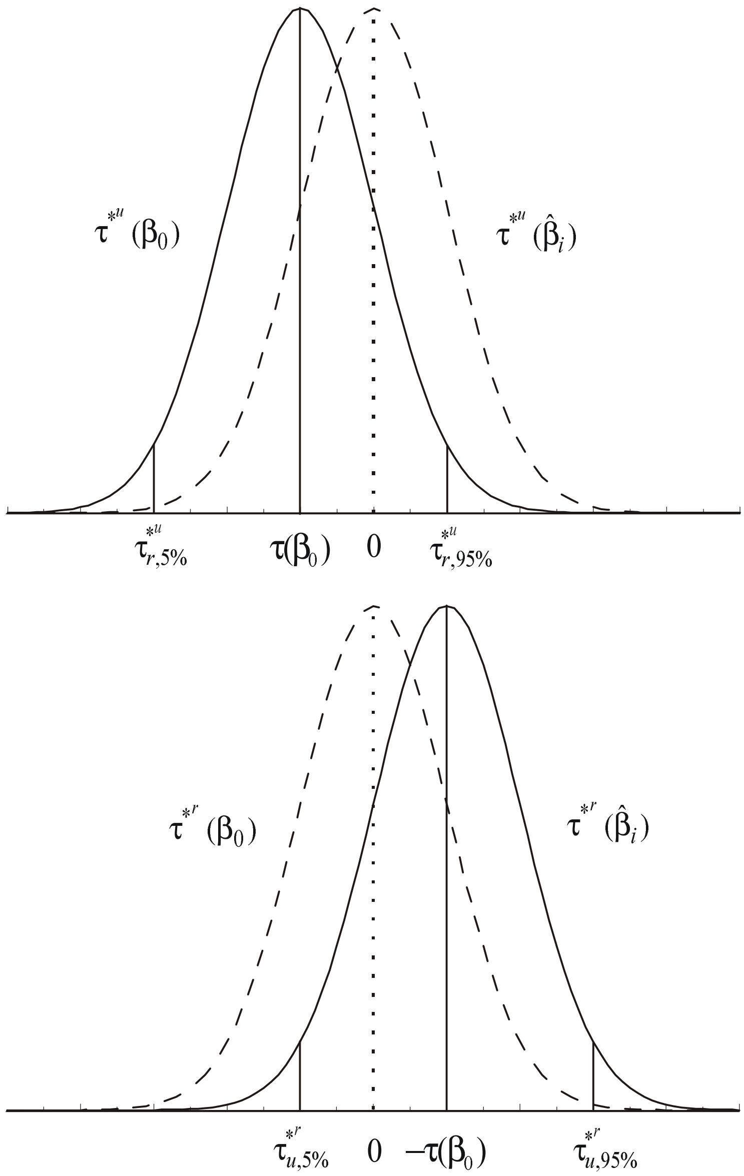

Figure 2: Assessment of the dependence on $\hat{\rho}$ of the parametric bootstrap quantiles $\tau^{*^u}_{u,\alpha}$, $\tau^{*^r}_{r,\alpha}$ and the actual expectation $\mathbb{E}_*[\tau^{*^u}(\hat{\rho})]$ in the Gaussian AR(1) model (n=25, ρ_0=0.8, $\alpha \in \{0.05, 0.95\}$).

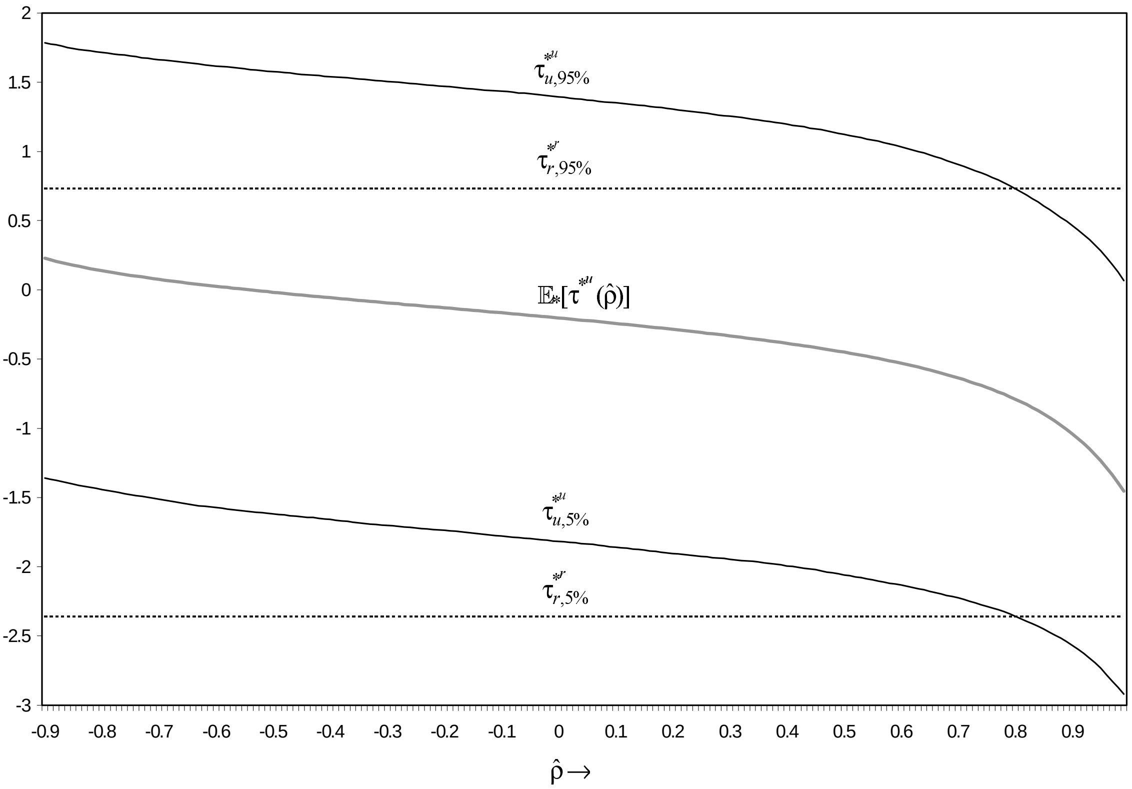

Figure 3: Bootstrapped p-values based on the test statistic approach for the null hypotheses $H_0: \rho = \rho_0$ and $H_0: \theta = \theta_0$ against their one-sided alternatives ($B = 4999$).