

A Simple Test for Regression Specification with Non-Nested Alternatives

Andreas Hagemann*

University of Illinois at Urbana-Champaign

March 1, 2011

In this paper, I introduce a simple test for the presence of the data-generating process among several non-nested alternatives. The test is an extension of the classical J test for non-nested regression models. I also provide a bootstrap version of the test that avoids possible size distortions inherited from the J test.

JEL classification: C12, C15, C52

Keywords: non-nested testing, multiple testing, bootstrap, finite-sample size

1. Introduction

Non-nested testing problems typically do not have a natural null hypothesis. For example, it is a priori not clear what should be the null hypothesis when testing whether a specific covariate enters the regression equation in level or in log form. For the Davidson and MacKinnon (1981) J test and the Cox (1961, 1962) test, the literature therefore usually suggests a sequence of tests where each possible null hypothesis is considered; see, among others, Fisher and McAleer (1979), Dastoor (1981), and Pesaran and Weeks (2003). In this paper, I introduce a simple test for the presence of the correct model among several non-nested specifications that avoids sequential testing. The test, which I refer to as the MJ (*minimum J*) test, is an extension of the J test and bases its decision on the model with the least significant J statistic.

Non-nested hypothesis tests such as the J or the Cox tests rely heavily on the assumption that one of the models under consideration is correct, and therefore all other non-nested specifications must be wrong. However, it may well happen that a non-nested hypothesis test does not reject a model in the presence of an alternative model, but also does not reject

*Department of Economics, University of Illinois at Urbana-Champaign, 214 David Kinley Hall, 1407 W. Gregory, Urbana, Illinois 61801, MC-707, USA. Tel.: +1-217-333-0120. Fax: +1-217-244-6571. E-mail address: ahagama2@illinois.edu.

the alternative in the presence of the original model when the hypotheses are reversed. This leaves the researcher in the unfortunate situation of having to conclude that both specifications “explain the data equally well” even though at most one of them can be correct. Similar problems arise when all models are rejected. A further issue is that the sequential testing is typically conducted without regard to overall size, and thus two researchers working with the same data can arrive at different specifications simply because they used different levels of significance. Non-nested testing procedures have been subject to substantial criticism because of these features; see, e.g., Granger, King, and White (1995) and Shi (2010).

I show that the MJ test is robust to each of these concerns: It does not require the correct model to be among the considered specifications and avoids ambiguous test outcomes. The MJ test determines with asymptotically correct size if the correct model is among the specifications under consideration. If the correct model is present, it is chosen with probability approaching one as the sample size becomes large. I also provide a bootstrap version of the MJ test that possesses all of these properties.

As the Monte Carlo study in this paper shows, the bootstrap is crucial here because the MJ test can over-reject in empirically relevant cases, but a careful application of the bootstrap transforms it into an almost exact test even in quite small samples. This feature of the MJ test is inherited from the traditional J test, which is known to over-reject severely (Godfrey and Pesaran, 1983) unless the bootstrap is used; see Fan and Li (1995), Godfrey (1998), and Davidson and MacKinnon (2002). My simulation study also suggests that the bootstrap test has good power except when the researcher makes a particularly bad guess about the correct model, i.e., when there is little correlation between the designs of the true model and the models under consideration.

The paper is organized as follows: Section 2 establishes the large sample properties of the MJ test. Section 3 discusses bootstrap methods. Section 4 contains the simulation study. Section 5 concludes. The Appendix presents auxiliary results and proofs.

I will use the following notation throughout the paper: For an index set of increasing integers $\mathcal{I} = \{1, \dots, I\}$, $(a_i)_{i \in \mathcal{I}}$ denotes the column vector $(a_1, a_2, \dots, a_I)^\top \in \mathbb{R}^I$, and an $I \times I$ matrix with generic element $a_{i,i'}$ in its i^{th} row and i'^{th} column is denoted by $(a_{i,i'})_{i,i' \in \mathcal{I}}$. If the integers in \mathcal{I} are not consecutive, the notation is meant to indicate that the a_i and $a_{i,i'}$ enter $(a_i)_{i \in \mathcal{I}}$ and $(a_{i,i'})_{i,i' \in \mathcal{I}}$ sequentially from smallest to largest index. Convergence in distribution as $n \rightarrow \infty$ is denoted by \rightsquigarrow ; $|\cdot|$ is the Euclidean norm.

2. The MJ Test for Non-Nested Regression Models

This section extends the J test for non-nested linear regression models to handle heteroscedastic errors. I then introduce the MJ test (Procedure 2.6 below).

Suppose we observe covariates $\{(x_{i,1}^\top, \dots, x_{i,M}^\top)^\top \in \mathbb{R}^{d_1 + \dots + d_M} : i = 1, \dots, n\}$ that give rise to $M \geq 2$ different possible linear regression models for $y := (y_1, \dots, y_n)^\top \in \mathbb{R}^n$, i.e.,

$$y = X_m \beta_m + u_m, \quad m \in \mathcal{M} := \{1, \dots, M\}, \quad (2.1)$$

where $X_m := (x_{1,m}, \dots, x_{n,m})^\top \in \mathbb{R}^{n \times d_m}$ is the design matrix of model m . The matrices X_1, \dots, X_M are assumed to be *non-nested*, i.e., for any two matrices with index $m \neq l$ in \mathcal{M} , no matrix can be obtained from another by a linear transformation. This does not rule out the possibility that some of the columns of X_m and X_l are identical or that they may be nonlinear transformations of one another. In addition, there is an observed or unobserved design matrix $X_{m^*} := (x_{1,m^*}, \dots, x_{n,m^*})^\top$ associated with the correct model m^* . Suppose that $\mathcal{F} := \{X_m : m \in \mathcal{M} \cup \{m^*\}\}$ has the following properties:

Assumption 2.1. $\{(y_i, (x_{i,m})_{m \in \mathcal{M} \cup \{m^*\}})^\top : i \geq 1\}$ is a sequence of iid random vectors. We have $\mathbb{E}|x_{i,m}|^4 < \infty$ for all $m \in \mathcal{M} \cup \{m^*\}$, where the number of elements of \mathcal{M} does not depend on n . For all $m \in \mathcal{M}$, the matrices $\mathbb{E} x_{i,m} x_{i,m}^\top$ are positive definite.

Tests constructed for non-nested environments such as the Davidson and MacKinnon (1981) J test typically assume that there is an $m^* \in \mathcal{M}$ such that the conditional mean of y can be written as $\mathbb{E}(y | \mathcal{F}) = X_{m^*} \beta_{m^*}$. In the following, I depart from this condition and only assume that a correct specification exists.

Assumption 2.2. Model m^* satisfies $\mathbb{E}(y | \mathcal{F}) = X_{m^*} \beta_{m^*}$. Let $u_{i,m^*} := y_i - x_{i,m^*}^\top \beta_{m^*}$ for all $i \geq 1$ and $\mathbb{E} u_{i,m^*}^4 < \infty$.

Remarks. 1. Although I only consider linear regression models with independent data, the results of the paper can be extended to nonlinear (parametric) models with weakly dependent data; see also the discussion below Theorem 2.7.

2. Davidson and MacKinnon (1981) point out that the assumption $m^* \in \mathcal{M}$ is not crucial since the J test is capable of rejecting all models in \mathcal{M} .

The setup of the J test presumes that for some predetermined $m \in \mathcal{M}$, the researcher wants to test the null hypothesis $m = m^*$ against $m \neq m^*$ in the presence of the non-nested alternatives $l \in \mathcal{M} \setminus \{m\}$. This can be done by artificially nesting the models in (2.1) via an additional parameter vector $\alpha_m := (\alpha_{l,m})_{l \in \mathcal{M} \setminus \{m\}} \in \mathbb{R}^{M-1}$ such that

$$y = X_m b_m + \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m} X_l \beta_l + u, \quad \text{where } b_m := \left(1 - \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m}\right) \beta_m.$$

Since the vectors $(\alpha_{l,m}, \beta_l)_{l \in \mathcal{M} \setminus \{m\}}$ of the nesting model may not be identified, Davidson and MacKinnon replace the β_l in the preceding display by the least squares (LS) estimates $\hat{\beta}_l = (X_l^\top X_l)^{-1} X_l^\top y$. After redefining the error term appropriately, this yields

$$y = X_m b_m + \sum_{l \in \mathcal{M} \setminus \{m\}} \alpha_{l,m} X_l \hat{\beta}_l + u, \tag{2.2}$$

which can be estimated by LS. A Wald test for $\alpha_m = 0$ is a J test for the validity of model m in the presence of the alternatives $\mathcal{M} \setminus \{m\}$. To construct the test statistic, let

$$\lambda_{n,m} := n^{-1/2} \left(y^\top P_l M_m y \right)_{l \in \mathcal{M} \setminus \{m\}} \quad \text{and}$$

$$\hat{\Sigma}_{n,m} := n^{-1} \left(y^\top P_l M_m \hat{\Omega}_{n,m} M_m P_{l'} y \right)_{l,l' \in \mathcal{M} \setminus \{m\}},$$

where $P_m := X_m (X_m^\top X_m)^{-1} X_m^\top$ and $M_m := I_n - P_m$ are the usual projection matrices and $\hat{\Omega}_{n,m}$ is an “estimate” of $\mathbb{E}(u_{m^*} u_{m^*}^\top | \mathcal{F})$. The J test statistic for model m is then

$$J_{n,m} := \lambda_{n,m}^\top \hat{\Sigma}_{n,m}^{-1} \lambda_{n,m}, \quad (2.3)$$

and the hypothesis that model m is the true model is rejected for large values of $J_{n,m}$. In practice, the J statistic is easily computed by running regression (2.2) and performing a Wald test for $\alpha_m = 0$ with weighting matrix $\hat{\Omega}_{n,m}$.

Given the independence of the observations, $\hat{\Omega}_{n,m}$ should be a diagonal matrix of squared residuals. For heteroscedasticity-robust testing, the literature frequently recommends using the residuals of the model under the null hypothesis; see, e.g., Davidson and MacKinnon (1985). Hence, I use the residuals $\hat{u}_{i,m} := y_i - x_{i,m}^\top \hat{\beta}_m$ and define

$$\hat{\Omega}_{n,m} := \text{diag}(\hat{u}_{1,m}^2, \dots, \hat{u}_{n,m}^2),$$

although other approaches are clearly possible. If Assumptions 2.1 and 2.2 hold, I show in Appendix A that we can then write $\hat{\Sigma}_m = \Sigma_m + \Pi_m + o_{\mathbb{P}}(1)$, where Σ_m is $\hat{\Sigma}_{n,m}$ with products of the form $n^{-1} X_m^\top X_l$ replaced by $\mathbb{E} x_{i,m} x_{i,l}^\top$ and products of the form $n^{-1} X_l^\top \hat{\Omega}_{n,m} X_{l'}$ replaced by $\mathbb{E}(u_{i,m^*}^2 x_{i,l} x_{i,l'}^\top)$, and Π_m is a positive semi-definite matrix; in particular, Π_{m^*} is equal to zero. To ensure that the J statistics are well behaved in large samples, I make the following assumption:

Assumption 2.3. For all $m \in \mathcal{M}$, Σ_m is positive definite.

Remarks. Assumption 2.3 is needed to identify the artificial parameters α_m , $m \in \mathcal{M}$. This condition fails when the covariates of m^* are completely uncorrelated with the covariates of one of the models in \mathcal{M} or when the design matrices of any two models in \mathcal{M} are nested, i.e., one matrix is a linear transformation of the other. Michelis (1999) discusses the consequences of near-singularity of Σ_m for the case $M = 2$.

For the J test to have power against the alternative $m^* \notin \mathcal{M}$, I also assume that the covariates of m^* have enough idiosyncratic variation so that the following condition holds:

Assumption 2.4. For all $m \in \mathcal{M} \setminus \{m^*\}$, there exists $l \in \mathcal{M}$ such that

$$\begin{aligned} & |\beta_{m^*}^\top \mathbb{E} x_{i,m^*} x_{i,l}^\top (\mathbb{E} x_{i,l} x_{i,l}^\top)^{-1} [\mathbb{E} x_{i,l} x_{i,m^*}^\top \\ & \quad - \mathbb{E} x_{i,l} x_{i,m}^\top (\mathbb{E} x_{i,m} x_{i,m}^\top)^{-1} \mathbb{E} x_{i,m} x_{i,m^*}^\top] \beta_{m^*}| > 0. \end{aligned} \quad (2.4)$$

Remarks. In addition to the restrictions imposed via Assumption 2.3, this property rules out that the covariates of m^* are fully correlated with (but not identical to) the covariates of any model m in \mathcal{M} , e.g., if $X_m = X_{m^*} + Z$, where Z is an independent measurement error with mean zero. In such a case, the J test would mistake X_m for X_{m^*} .

The following Lemma summarizes the large sample properties of the $J_{n,m}$ statistics under Assumptions 2.1–2.4.

Lemma 2.5. *Suppose Assumptions 2.1–2.4 are satisfied.*

- (i) *If $m^* \in \mathcal{M}$, then $J_{n,m^*} \rightsquigarrow \chi_{M-1}^2$.*
- (ii) *For every $m \in \mathcal{M} \setminus \{m^*\}$ and every $B \in \mathbb{R}$, we have $\lim_{n \rightarrow \infty} \mathbb{P}(J_{n,m} > B) = 1$.*

The J test does not have a natural null hypothesis, and therefore the researcher is expected to have a preferred model m to test the null hypothesis $m = m^*$. Because there is usually little guidance in applied work about what m could be, this makes a seemingly simple task such as testing whether to include a specific covariate in level or in log form a surprisingly difficult problem. However, a closer look at Lemma 2.5 reveals an easy way out of this dilemma: If one of the models under consideration is the correct model, then its J statistic has an asymptotic distribution and the statistics of the other models diverge; if, instead, the correct model is not among the M models, then *all* statistics will diverge. Thus, only the model with the smallest J statistic can possibly be the correct model and we can reject the hypothesis $m^* \in \mathcal{M}$ when the smallest J statistic is large. This motivates the following alternative to the traditional sequential testing procedures:

Procedure 2.6 (*MJ Test*). 1. For each $m \in \mathcal{M}$, run regression (2.2) and compute $J_{n,m}$ as in (2.3). Let $\mathcal{J}_n := \{J_{n,m} : m \in \mathcal{M}\}$ and define $MJ_n := \min \mathcal{J}_n$.
 2. Reject the hypothesis $H_0: m^* \in \mathcal{M}$ in favor of $H_1: m^* \notin \mathcal{M}$ if $MJ_n > c_{1-\alpha}$, where $c_{1-\alpha}$ is the $1 - \alpha$ quantile of the χ_{M-1}^2 distribution.

As a referee points out, the *MJ* test is an instance of an intersection–union test (Berger, 1982): The null hypothesis is the union of different nulls, and the rejection region is thus the intersection of the rejection regions of the respective nulls. Indeed, the null is that one of the models in \mathcal{M} is correctly specified, but it can be any of them, so that the null is the union of the sets of data-generating processes, each of which constitutes one of the models in \mathcal{M} , and, for the null to be false, each of those hypotheses must be false.

The following Theorem shows that the *MJ* test is indeed a valid test for $m^* \in \mathcal{M}$ and, in particular, the asymptotic distributions of J_{n,m^*} and MJ_n coincide. The reason for this result is that the number of models in \mathcal{M} does not depend on n and the M different J statistics are asymptotically independent since $M - 1$ of them diverge—two features that are typically not available for nested testing problems, but can be exploited in non-nested environments to construct simple tests such as Procedure 2.6.

Theorem 2.7. *Suppose Assumptions 2.1–2.4 are true.*

- (i) *If $m^* \in \mathcal{M}$, then $MJ_n \rightsquigarrow \chi_{M-1}^2$ and*
- (ii) *if $m^* \notin \mathcal{M}$, then $\lim_{n \rightarrow \infty} \mathbb{P}(MJ_n > B) = 1$ for all $B \in \mathbb{R}$.*

Remarks. 1. The J test can be generalized to nonlinear regressions (Davidson and MacKinnon, 1981), and to models with dependent errors via standard heteroscedasticity and autocorrelation consistent covariance matrix estimators or fixed- b asymptotics as in Choi and Kiefer (2008). See also MacKinnon, White, and Davidson (1983) for an extension to models with weakly dependent and endogenous covariates. Since the *MJ* test is nothing

but a J test for a specific model, these generalizations are also available for the minimum J approach as long as an identification condition analogous to (2.4) holds.

2. Related tests for non-nested models such as the JA test of Fisher and McAleer (1981) can be extended in a similar way.

The MJ test chooses a model $\hat{m} = \arg \min \mathcal{J}_n$ for the test, and if the test provides evidence that $m^* \in \mathcal{M}$, then \hat{m} is the natural candidate for m^* . This can be interpreted as the non-nested analogue of general-to-specific testing: Selecting \hat{m} amounts to choosing a model for which there is the most evidence that it is not outperformed by the other models under consideration in terms of explanatory power. Indeed, as the following Lemma shows, the MJ test consistently finds the true model among the specifications, both unconditionally and conditional on having accepted the null hypothesis:

Lemma 2.8. *If Assumptions 2.1–2.4 hold and $m^* \in \mathcal{M}$, then $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{m} = m^*) = 1$ and $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{m} = m^* \mid MJ_n \leq c_{1-\alpha}) = 1$.*

The Monte Carlo results in Section 4 show that \hat{m} reliably chooses the correct model, but nonetheless one should be careful about how to proceed from there; see Leeb and Pötscher (2009) and the references therein.

The MJ test might be expected to inherit the well-known size distortion present in the J test (see Davidson and MacKinnon, 2002, and the references therein). However, I will show in Section 4 that these effects can be ameliorated as long as the bootstrap is used. The next section provides the necessary modifications.

3. Bootstrapping the MJ Test Statistic

This section presents a bootstrap version of the MJ test (Procedure 3.1). The idea is to replace $J_{n,\hat{m}}$ in \mathcal{J}_n by realizations of a bootstrap J statistic such that repeatedly computing the minimum over this new set of J statistics mimics the behavior of MJ_n . To account for the possible heterogeneity in the errors, I use the wild bootstrap of Liu (1988) and Mammen (1992); it perturbs the residuals with iid copies η_1, η_2, \dots of a random variable η with $\mathbb{E}(\eta \mid y, \mathcal{F}) = 0$, $\mathbb{E}(\eta^2 \mid y, \mathcal{F}) = 1$ and $\mathbb{E}(|\eta|^{2+\delta} \mid y, \mathcal{F}) < \infty$ for some $\delta > 0$. Let also $H := \text{diag}(\eta_1, \dots, \eta_n)$.

Procedure 3.1 (Bootstrap MJ Test). Let $\hat{m} = \arg \min \mathcal{J}_n$. (If $\arg \min \mathcal{J}_n$ is not unique, choose \hat{m} to be any element of $\arg \min \mathcal{J}_n$.)

1. (a) Use LS to obtain the residuals $\hat{u} = y - X_{\hat{m}}\hat{\beta}_{\hat{m}}$ and compute $u^* := H\hat{u}$.
- (b) Generate bootstrap data $y^* := X_{\hat{m}}\hat{\beta}_{\hat{m}} + u^*$ and calculate

$$\begin{aligned} \lambda_{n,\hat{m}}^* &:= n^{-1/2} \left(y^{*\top} P_l M_{\hat{m}} y^* \right)_{l \in \mathcal{M} \setminus \{\hat{m}\}} \quad \text{and} \\ \hat{\Sigma}_{n,\hat{m}}^* &:= n^{-1} \left(y^{*\top} P_l M_{\hat{m}} \hat{\Omega}_{n,\hat{m}} M_{\hat{m}} P_{l'} y^* \right)_{l,l' \in \mathcal{M} \setminus \{\hat{m}\}}. \end{aligned} \tag{3.1}$$

- (c) Compute the bootstrap J statistic $J_{n,\hat{m}}^* := \lambda_{n,\hat{m}}^{*\top} \hat{\Sigma}_{n,\hat{m}}^{*-1} \lambda_{n,\hat{m}}^*$.

2. Let $\mathcal{J}_n^* := \{J_{n,\hat{m}}^*\} \cup \mathcal{J}_n \setminus \{J_{n,\hat{m}}\}$ and calculate $MJ_n^* := \min \mathcal{J}_n^*$.
3. Repeat steps 1–2 R times, each with a new realization of H . Reject the hypothesis $m^* \in \mathcal{M}$ if MJ_n is larger than $c_{n,1-\alpha}^*$, the $1 - \alpha$ empirical quantile of the MJ_n^* .

Remarks. 1. The recommended choice for η in practice is a Rademacher variable that takes on the value 1 with probability 1/2 and the value -1 with probability 1/2. Distributions other than the Rademacher distribution could be used for η , in particular if the error distribution is skewed, but there is no evidence that they would lead to better inference; see Davidson, Monticini, and Peel (2007) for a discussion.

2. Step 1 is similar to the standard residual bootstrap used in Fan and Li (1995), Godfrey (1998), and Davidson and MacKinnon (2002), who deal with homoscedastic errors; Choi and Kiefer (2008) use the block bootstrap. None of these authors establish the validity of their bootstrap method.

3. There is no need to approximate $\hat{\Omega}_{n,\hat{m}}$ in $\hat{\Sigma}_{n,\hat{m}}^*$ by a bootstrap version because the variance of the bootstrap errors u^* is $\text{var}(u^*u^{*\top} \mid y, \mathcal{F}) = \hat{\Omega}_{n,\hat{m}}$; see also Flachaire (2002).

4. Step 1 only approximates the distribution of $J_{n,\hat{m}}$ conditional on \hat{m} , and computing the minimum over the J statistics after replacing $J_{n,\hat{m}}$ with $J_{n,\hat{m}}^*$ in step 2 mimics the additional randomness from choosing \hat{m} out of the M models. If $m^* \in \mathcal{M}$, \hat{m} and the minimizer of \mathcal{J}_n^* coincide in large samples, and hence this step is not crucial for the validity of the bootstrap. However, the simulation study in Section 4 suggests that, if anything, the Bootstrap MJ test can behave conservatively in small samples. Step 2 can therefore improve both the size and power of the bootstrap test in finite samples because it restricts the magnitude of large realizations of $J_{n,\hat{m}}^*$.

By choosing the number of bootstrap repetitions R in Procedure 3.1 large enough, we can approximate the quantiles of the distribution of MJ_n^* with arbitrary precision. I therefore let $R \rightarrow \infty$ in the following and define the quantiles of MJ_n^* directly as $c_{n,1-\alpha}^* := \inf\{x \in \mathbb{R} : \mathbb{P}^*(MJ_n^* \leq x) \geq 1 - \alpha\}$, where $\mathbb{P}^*(\cdot)$ abbreviates $\mathbb{P}(\cdot \mid y, \mathcal{F})$.

To ensure that the bootstrap approximation is well-behaved, a further condition similar to Assumption 2.3 is needed: Since $\mathbb{E}(X_l^\top u^* \mid y, \mathcal{F}) = 0$ and $\text{var}(X_l^\top u^*/n \mid y, \mathcal{F}) = X_l^\top \hat{\Omega} X_l/n^2 = o_{\mathbb{P}}(1)$ provided Assumptions 2.1 and 2.2 hold, we can with \mathbb{P} -probability approaching one as $n \rightarrow \infty$ write

$$\hat{\Sigma}_{n,m}^* = n^{-1} \left(\hat{\beta}_m^\top X_m^\top P_l M_m \hat{\Omega}_{n,m} M_m P_l X_m \hat{\beta}_m \right)_{l,l' \in \mathcal{M} \setminus \{m\}} + o_{\mathbb{P}^*}(1).$$

Define $\hat{\Xi}_{n,m}$ as the first term on the right-hand side. In the same way as $\hat{\Sigma}_{n,m} = \Sigma_m + \Pi_m + o_{\mathbb{P}}(1)$, it can be seen that $\hat{\Xi}_{n,m} = \Xi_m + \Psi_m + o_{\mathbb{P}}(1)$, where Ξ_m is $\hat{\Xi}_{n,m}$ with expressions of the form $n^{-1} X_m^\top X_l$ replaced by $\mathbb{E} x_{i,m} x_{i,l}^\top$ and expressions of the form $n^{-1} X_l^\top \hat{\Omega}_{n,m} X_{l'}$ replaced by $\mathbb{E}(u_{i,m}^2 x_{i,l} x_{i,l'}^\top)$; Ψ_m is a positive semi-definite matrix with Ψ_m^* equal to zero. Details are provided in Appendix A. I impose the following condition:

Assumption 3.2. For all $m \in \mathcal{M}$, Ξ_m is positive definite.

Remarks. In addition to the restrictions imposed by Assumptions 2.3 and 2.4, this condition

rules out that the covariates of any two models in \mathcal{M} are orthogonal. In practice, it should not limit the applicability of the Bootstrap MJ test because all models under consideration are supposed to explain the same variable y . Therefore, it is not much of a restriction to assume that they have some common features.

The following result establishes the consistency of the Bootstrap MJ test.

Theorem 3.3. *Suppose Assumptions 2.1–2.4 and 3.2 hold with $\mathbb{E}|u_{i,m^*}|^{4+\delta} < \infty$ and $\mathbb{E}|x_{i,m}|^{4+\delta} < \infty$ for all $m \in \mathcal{M} \cup \{m^*\}$ and some $\delta > 0$. Let $\alpha \in (0, 1)$. Procedure 3.1 has the following properties:*

- (i) *If $m^* \in \mathcal{M}$, then $\lim_{n \rightarrow \infty} \mathbb{P}(MJ_n > c_{n,1-\alpha}^*) = \alpha$, and*
- (ii) *if $m^* \notin \mathcal{M}$, then $\lim_{n \rightarrow \infty} \mathbb{P}(MJ_n > c_{n,1-\alpha}^*) = 1$.*

The consistency of \hat{m} conditional on accepting the null hypothesis of the bootstrap test is then an immediate consequence.

Corollary 3.4. *Suppose the conditions of Theorem 3.3 are satisfied. If $m^* \in \mathcal{M}$, then $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{m} = m^* \mid MJ_n \leq c_{n,1-\alpha}^*) = 1$.*

The next section illustrates this consistency property and the finite sample behavior of the MJ and Bootstrap MJ tests in a small simulation study.

4. Simulation Study

The J test is known to severely over-reject when (1) the sample size is small, (2) the error variance is large, (3) the number of regressors differs among the models, or (4) the correlation between the models under consideration is small; see Davidson and MacKinnon (2002) for a thorough analysis of why this is the case. Focusing on the correlation structure, this section investigates the impact of these properties on the performance of the MJ test and the Bootstrap MJ test.

The true model m^* for the following experiments is

$$y_i = x_{i,m^*}^\top \beta_{m^*} + u_i, \quad u_i = (v_i - 1)[|x_{i,1,m^*} x_{i,2,m^*}| (|x_{i,3,m^*}| + 2.5)]^{1/2},$$

where β_{m^*} is a d_{m^*} -vector of ones, the first element of x_{i,m^*} is one and the other components, denoted as $(x_{i,2,m^*}, \dots, x_{i,d_{m^*},m^*})^\top$, are uncorrelated standard normal variables; the v_i are independently distributed as χ_1^2 . I experimented with the form of the heteroscedasticity in u_i , but found that it had little effect on the performance of the bootstrap test. I therefore settled for a worst-case scenario and chose an error structure that is known to cause size distortions when heteroscedasticity-robust estimators are employed (see Long and Ervin, 2000). The covariates $x_{i,m} = (x_{i,1,m}, \dots, x_{i,d_m,m})^\top$ of any other model $m \in \mathcal{M} \setminus \{m^*\}$ are given by $x_{i,1,m} = 1$,

$$x_{i,j,m} = \frac{\rho}{\sqrt{1-\rho^2}} x_{i,j,m^*} + z_{i,j,m}, \quad j = 2, \dots, \min\{d_{m^*}, d_m\},$$

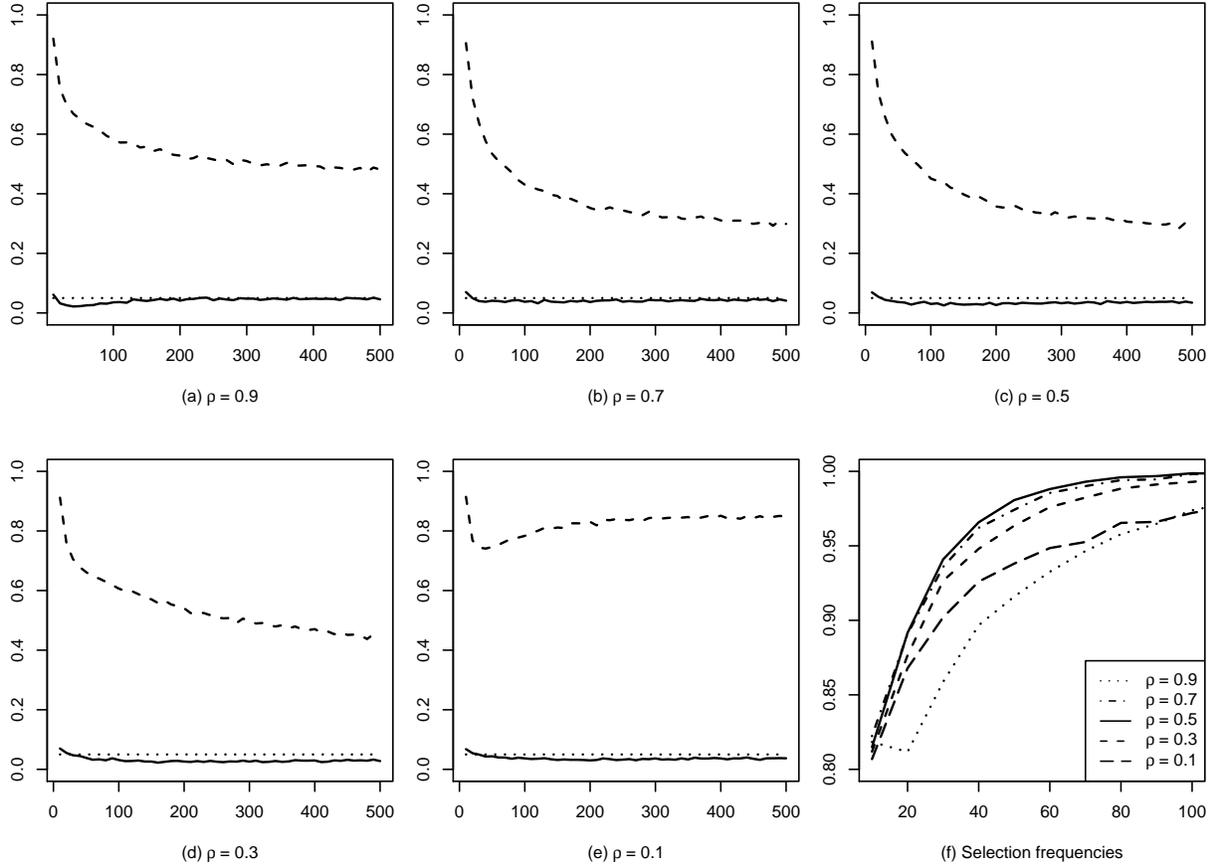


Figure 1: Panels (a)–(e) show rejection frequencies of the MJ test (dashed line) and Bootstrap MJ test (solid) under H_0 with $M = 2$ at the 5% level (dotted) for different values of ρ as a function of sample size n . Panel (f) plots the relative frequency of the event $\hat{m} = m^*$ for the experiments in panels (a)–(e) as a function of n .

and $x_{i,j,m} = z_{i,j,m}$ for $d_m > d_{m^*}$, where the $z_{i,j,m}$ are independent standard normals. This ensures that the correlation between the random components of x_{i,m^*} and the corresponding components of $x_{i,m}$ is exactly ρ .

I used 10,000 replications for each sample size $n \in \{10, 20, \dots, 500\}$ to investigate the behavior of the MJ test for $M \in \{2, 3\}$ and $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ under both the null and the alternative hypothesis. For each of these cases, I recorded the test decision of the MJ test at a significance level of 5% and the model \hat{m} with the smallest J statistic. I used the warp-speed method of Giacomini, Politis, and White (2007) to estimate size and power of the Bootstrap MJ test; this method considerably sped up the simulations because only one bootstrap replication per Monte Carlo replication was needed. For notational simplicity, let also $m^* = 1$ when $m^* \in \mathcal{M}$.

Panels (a)–(e) of Figure 1 show the rejection frequencies of the MJ test and the bootstrap version for the case $M = 2$ when $H_0: m^* \in \mathcal{M}$ is true as a function of n and ρ . Here and in the following, I used the same sequence of random numbers for each correlation structure in order to make the results comparable. The number of regressors in the true model was $d_{m^*} = 5$ and the other specification had $d_2 = 7$ covariates. The MJ test over-rejected severely in all cases at all sample sizes, although the size of the test improved considerably in larger samples for $\rho \in \{0.3, 0.5, 0.7, 0.9\}$. The case $\rho = 0.9$ was the slowest to adjust because the two models under consideration were so similar. The MJ test broke down completely for $\rho = 0.1$ because the distribution of the J test statistic is not well approximated by the χ_1^2 distribution when the correlation among the models is small; see Davidson and MacKinnon (2002). In contrast, the Bootstrap MJ test was almost exact for $\rho \in \{0.5, 0.7, 0.9\}$ at all sample sizes, and behaved mildly conservatively for the low correlation structures. Experiments with larger values of d_2 and larger variation in v_i increased the size distortion of the MJ test even further, whereas the Bootstrap MJ test remained unaffected.

The last panel in Figure 1 shows the relative number of times \hat{m} was indeed the true model in the experiments presented in panels (a)–(e). Selecting the model with the smallest J statistic worked well in small samples even when the models were highly correlated. In samples larger than 100 (not reported), the selection frequency of the true model was essentially one.

Table 1: Conditional and Unconditional Selection Frequencies

		Relative Frequency of $\hat{m} = m^*$		
		Unconditional	$\cdot \mid MJ_n \leq c_{1-\alpha}$	$\cdot \mid MJ_n \leq c_{n,1-\alpha}^*$
$n = 30$	$\rho = 0.9$	0.855	0.921	0.858
	$\rho = 0.5$	0.940	0.996	0.948
	$\rho = 0.1$	0.902	0.998	0.908
$n = 60$	$\rho = 0.9$	0.933	0.979	0.936
	$\rho = 0.5$	0.986	1.000	0.989
	$\rho = 0.1$	0.950	1.000	0.958
$n = 90$	$\rho = 0.9$	0.967	0.997	0.970
	$\rho = 0.5$	0.996	1.000	0.997
	$\rho = 0.1$	0.967	1.000	0.976

Figure 1 does not show the selection frequencies conditional on accepting the null hypothesis since the warp-speed method cannot estimate this quantity for the Bootstrap MJ test. I therefore calculated the actual rejection and selection frequencies for $n \in \{30, 60, 90\}$; the bootstrap test was based on $R = 399$ bootstrap replications. As Table 1 shows, the conditional selection frequencies for the Bootstrap MJ test and the relative frequencies of the event $\hat{m} = m^*$ were almost identical. The discrepancy was larger for the conditional

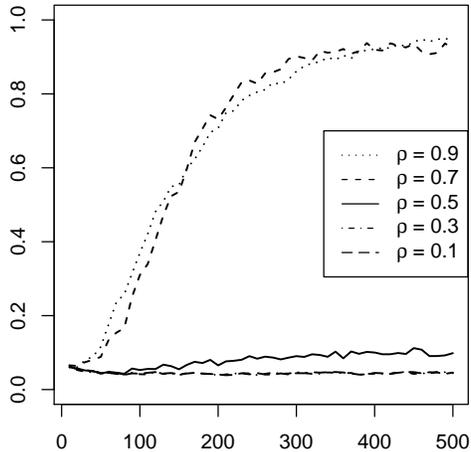


Figure 2: Rejection frequencies of the Bootstrap MJ test under H_1 with $M = 2$ at the 5% level for different values of ρ as a function of sample size n .

selection frequencies based on the MJ test, which was mostly driven by its inability to control the nominal size.

Figure 2 plots the rejection frequency of the Bootstrap MJ test for $M = 2$ under $H_1: m^* \notin \mathcal{M}$ as a function of n and ρ . I do not report power estimates for the MJ test because its size distortion invalidates the power estimates. As before, the true model had $d_{m^*} = 5$ covariates, and the other specifications had $d_1 = 5$ and $d_2 = 7$ regressors. The bootstrap test had good power in moderately large samples as long as the correlation between the models in \mathcal{M} and the true models was high. However, the power dropped considerably for $\rho = 0.5$ and was essentially zero for smaller correlations. Hence, if the researcher uses models that are very different from the correct model, the Bootstrap MJ test will most likely not be able to detect this problem. Experiments with larger error variances gave qualitatively similar results, but larger sample sizes were required to reach the same level of power. I also experimented with the number of covariates of the models in \mathcal{M} , but found that the impact was relatively small.

Figure 3 repeats the experiments shown in Figure 1 for $M = 3$. The models under consideration now had $d_{m^*} = 5$, $d_2 = 5$, and $d_3 = 7$ regressors. The size distortion of the MJ test was even more extreme than before, whereas the Bootstrap MJ test was less conservative and even more precise than in the case $M = 2$. The selection frequency of the true model was slightly worse and larger samples were needed to reliably detect the true model.

Finally, Figure 4 shows the power experiment from Figure 2 for $M = 3$. The true model had $d_{m^*} = 5$ regressors, and the models in \mathcal{M} now had $d_1 = 5$, $d_2 = 6$, and $d_3 = 7$ covariates, but the results remained almost unchanged.

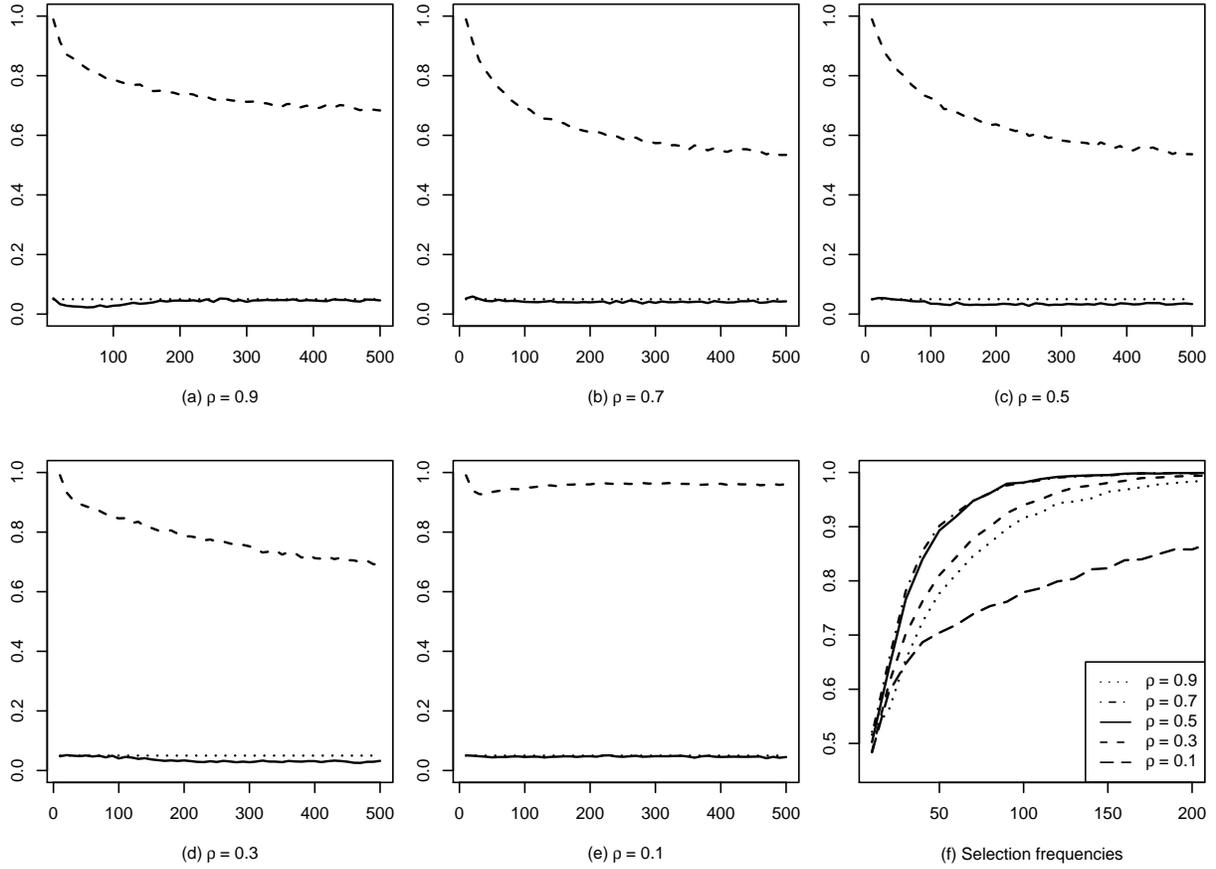


Figure 3: Panels (a)–(e) show rejection frequencies of the MJ test (dashed line) and Bootstrap MJ test (solid) under H_0 with $M = 3$ at the 5% level (dotted) for different values of ρ as a function n . Panel (f) plots the relative frequency of the event $\hat{m} = m^*$ for the experiments in panels (a)–(e) as a function of n .

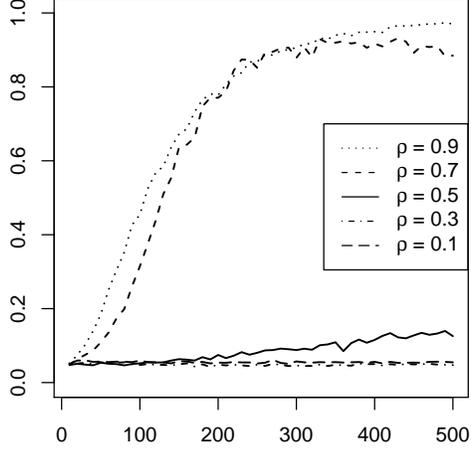


Figure 4: Rejection frequencies of the Bootstrap MJ test under H_1 with $M = 3$ at the 5% level for different values of ρ as a function of sample size n .

5. Conclusion

In this paper, I introduced a simple test for the presence of the data-generating process among several non-nested regression models. In contrast to classical approaches to non-nested testing, the MJ test does not require the correct model to be among the considered specifications and avoids ambiguous test outcomes.

Appendix

A. Auxiliary Results and Definitions

This section states two results that are needed for the proofs below and defines $\hat{\Sigma}_{n,m} = \Sigma_m + \Pi_m + o_{\mathbb{P}}(1)$ and $\hat{\Xi}_{n,m} = \Xi_m + \Psi_m + o_{\mathbb{P}}(1)$ explicitly. To simplify notation, for a random variable \mathcal{Z} and random sequences $\mathcal{Z}_n, \mathcal{Z}'_n$, the expression “ $\mathcal{Z}_n \rightsquigarrow \mathcal{Z}$ in probability” abbreviates “ $\mathbb{P}(\sup_z |\mathbb{P}^*(\mathcal{Z}_n \leq z) - \mathbb{P}(\mathcal{Z} \leq z)| > \varepsilon) = o(1)$ for every $\varepsilon > 0$,” and “ $\mathcal{Z}_n = \mathcal{Z}'_n + o_{\mathbb{P}^*}(1)$ in probability” abbreviates “ $\mathbb{P}(\mathbb{P}^*(|\mathcal{Z}_n - \mathcal{Z}'_n| > \varepsilon) > \varepsilon) = o(1)$ for every $\varepsilon > 0$.” The Frobenius norm $\sqrt{\text{trace}(A^\top A)}$ of a matrix A is denoted by $\|A\|$.

Given any $m, l, l' \in \mathcal{M}$, repeated application of the Law of Large Numbers yields $n^{-1}X_l^\top \hat{\Omega}_{m,n} X_{l'} = \mathbb{E}(u_{i,m}^2 x_{i,l} x_{i,l'}^\top) + \mathbb{E}[(x_{i,m}^\top \beta_{m^*} - x_{i,m}^\top \beta_m)^2 x_{i,l} x_{i,l'}^\top] + o_{\mathbb{P}}(1)$, provided that Assumptions 2.1 and 2.2 hold, where $\beta_m := (\mathbb{E} x_{i,m} x_{i,m}^\top)^{-1} \mathbb{E} x_{i,m} x_{i,m^*}^\top \beta_{m^*}$. Define $\Gamma_{m,l} := (\mathbb{E} x_{i,m} x_{i,m}^\top)^{-1} \mathbb{E} x_{i,m} x_{i,l}^\top$ and $v_{i,m}^2 := (x_{i,m^*}^\top \beta_{m^*} - x_{i,m}^\top \beta_m)^2$; then

$$n^{-1} y^\top P_l M_m \hat{\Omega}_{n,m} M_m P_l y$$

$$\begin{aligned}
&= \left[\beta_{m^*}^\top \Gamma_{l,m^*}^\top \left(\mathbb{E}(u_{i,m^*}^2 x_{i,l} x_{i,l'}^\top) - \Gamma_{m,l}^\top \mathbb{E}(u_{i,m^*}^2 x_{i,m} x_{i,l'}^\top) \right. \right. \\
&\quad \left. \left. + \Gamma_{m,l}^\top \mathbb{E}(u_{i,m^*}^2 x_{i,m} x_{i,m}^\top) \Gamma_{m,l'} - \mathbb{E}(u_{i,m^*}^2 x_{i,l} x_{i,m}^\top) \Gamma_{m,l'} \right) \Gamma_{l',m^*} \beta_{m^*} \right] \\
&\quad + \left[\beta_{m^*}^\top \Gamma_{l,m^*}^\top \left(\mathbb{E}(v_{i,m}^2 x_{i,l} x_{i,l'}^\top) - \Gamma_{m,l}^\top \mathbb{E}(v_{i,m}^2 x_{i,m} x_{i,l'}^\top) \right. \right. \\
&\quad \left. \left. + \Gamma_{m,l}^\top \mathbb{E}(v_{i,m}^2 x_{i,m} x_{i,m}^\top) \Gamma_{m,l'} - \mathbb{E}(v_{i,m}^2 x_{i,l} x_{i,m}^\top) \Gamma_{m,l'} \right) \Gamma_{l',m^*} \beta_{m^*} \right] + o_{\mathbb{P}}(1).
\end{aligned}$$

Denote the first term in square brackets by $\sigma_{m,l,l'}$ and the second term in square brackets by $\pi_{m,l,l'}$. Define $\Sigma_m := (\sigma_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}}$ and $\Pi_m := (\pi_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}}$. Similarly, replace y by $P_m y$, $\Gamma_{l,m^*} \beta_{m^*}$ by $\Gamma_{l,m} \Gamma_{m,m^*} \beta_{m^*}$, and $\Gamma_{l',m^*} \beta_{m^*}$ by $\Gamma_{l',m} \times \Gamma_{m,m^*} \beta_{m^*}$ in the preceding display to define $\xi_{m,l,l'}$ as the resulting first term in square brackets and $\psi_{m,l,l'}$ as the new second term in square brackets. Let $\Xi_m := (\xi_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}}$ and $\Psi_m := (\psi_{m,l,l'})_{l,l' \in \mathcal{M} \setminus \{m\}}$. By construction, for all $m \in \mathcal{M}$ we then have $\hat{\Sigma}_{n,m} = \Sigma_m + \Pi_m + o_{\mathbb{P}}(1)$ and $\hat{\Xi}_{n,m} = \Xi_m + \Psi_m + o_{\mathbb{P}}(1)$, and Σ_m , Π_m , Ξ_m , and Ψ_m are positive semi-definite with $\Pi_{m^*} = \Psi_{m^*} = 0$.

Finally, I state two results that are needed below. In particular, Lemma A.1 establishes the asymptotic distribution of $J_{n,\hat{m}}^*$ defined in step 1 of Procedure 3.1 with the random index \hat{m} equal to a fixed index $m \in \mathcal{M}$; this statistic is denoted by $J_{n,m}^*$.

Lemma A.1. *Suppose we are in the situation of Theorem 3.3. For all $m \in \mathcal{M}$, we have $J_{n,m}^* \rightsquigarrow \mathcal{X}$ in probability, where $\mathcal{X} \sim \chi_{M-1}^2$.*

Lemma A.2. *Suppose the conditions of Theorem 3.3 hold. If $m^* \in \mathcal{M}$, then*

$$\text{plim}_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| \mathbb{P}^*(MJ_n^* \leq x) - \mathbb{P}(MJ_n \leq x) \right| = 0. \tag{A.1}$$

B. Proofs

Proof of Lemma 2.5. Both parts can be established by routine arguments for Wald statistics. The details are therefore omitted. \square

Proof of Theorem 2.7. (i) In view of Lemma 2.5, this can be shown by arguments similar to those used to prove Theorem 1 of Berger (1982). (ii) Use Lemma 2.5(ii) and the continuity of the minimum function. \square

Proof of Lemma 2.8. By Lemma 2.5, for every $\varepsilon > 0$ we can find $B > 0$ such that

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \mathbb{P}(\hat{m} \neq m^*) &= \limsup_{n \rightarrow \infty} \mathbb{P}(\exists m \in \mathcal{M} \setminus \{m^*\} : J_{n,m} < J_{n,m^*}) \\
&\leq \sum_{m \in \mathcal{M} \setminus \{m^*\}} \lim_{n \rightarrow \infty} \mathbb{P}(J_{n,m} \leq B) + \sup_{n \in \mathbb{N}} \mathbb{P}(J_{n,m^*} > B) < \varepsilon,
\end{aligned}$$

which can be made arbitrarily small by choosing B large enough. To see the second part, note that $\mathbb{P}(\hat{m} = m^* \mid MJ_n \leq c_{1-\alpha}) = \mathbb{P}(J_{n,m^*} \leq c_{1-\alpha}) / \mathbb{P}(MJ_n \leq c_{1-\alpha})$, which converges to one by Lemma 2.5(i) and Theorem 2.7(i). \square

Proof of Lemma A.1. I first show $\hat{\Sigma}_{n,m}^{*-1/2} \lambda_{n,m}^* \rightsquigarrow N_{M-1}(0, I_{M-1})$ in probability. Since

$$\hat{\Sigma}_{n,m}^{*-1/2} \lambda_{n,m}^* = \hat{\Xi}_{n,m}^{-1/2} \left(n^{-1/2} \hat{\beta}_m^\top X_m^\top P_l M_m H \hat{u} \right)_{l \in \mathcal{M} \setminus \{m\}} + o_{\mathbb{P}^*}(1) \text{ in probability,}$$

it suffices to argue that the first term on the right-hand side satisfies a Liapounov condition in probability.

Let $\mathbb{E}^*(\cdot) := \mathbb{E}(\cdot \mid y, \mathcal{F})$, $\hat{\Gamma}_{m,l} := (X_m^\top X_m)^{-1} X_m^\top X_l$, and note that

$$\begin{aligned} & n^{-1-\delta} \sum_{i=1}^n \mathbb{E}^* \left| \left(\hat{\beta}_m^\top \hat{\Gamma}_{l,m}^\top (x_{i,l} - \hat{\Gamma}_{m,l}^\top x_{i,m}) \eta_i \hat{u}_{i,m} \right)_{l \in \mathcal{M} \setminus \{m\}} \right|^{2+\delta} \\ & \leq \left(\frac{2(M-1)}{n} \right)^{1+\delta} \sum_{i=1}^n \sum_{l \in \mathcal{M} \setminus \{m\}} \mathbb{E}^* \left(\left| \hat{\beta}_m^\top \hat{\Gamma}_{l,m}^\top x_{i,l} \eta_i \hat{u}_{i,m} \right|^{2+\delta} \right. \\ & \quad \left. + \left| \hat{\beta}_m^\top \hat{\Gamma}_{l,m}^\top \hat{\Gamma}_{m,l}^\top x_{i,m} \eta_i \hat{u}_{i,m} \right|^{2+\delta} \right) \\ & \leq \left(\frac{2(M-1)}{n} \right)^{1+\delta} \sum_{l \in \mathcal{M} \setminus \{m\}} \left(\left(\|\hat{\beta}_m\| \|\hat{\Gamma}_{l,m}\| \right)^{2+\delta} \mathbb{E}^* |\eta|^{2+\delta} \sum_{i=1}^n |x_{i,l} \hat{u}_{i,m}|^{2+\delta} \right. \\ & \quad \left. + \left(\|\hat{\beta}_m\| \|\hat{\Gamma}_{l,m}\| \|\hat{\Gamma}_{m,l}\| \right)^{2+\delta} \mathbb{E}^* |\eta|^{2+\delta} \sum_{i=1}^n |x_{i,m} \hat{u}_{i,m}|^{2+\delta} \right) = O_{\mathbb{P}}(n^{-\delta}), \end{aligned}$$

where the first inequality uses the c_r inequality and the second inequality applies the fact that η is iid and submultiplicativity of the Frobenius norm. The equality follows from another application of the c_r inequality to the sums involving n such that

$$\begin{aligned} \sum_{i=1}^n |x_{i,l} \hat{u}_{i,m}|^{2+\delta} & \leq 3^{1+\delta} \sum_{i=1}^n \left(|x_{i,l} u_{i,m^*}|^{2+\delta} + \|x_{i,l} x_{i,m^*}^\top\|^{2+\delta} |\beta_{m^*}|^{2+\delta} \right. \\ & \quad \left. + \|x_{i,l} x_{i,m}^\top\|^{2+\delta} |\hat{\beta}_m|^{2+\delta} \right) \end{aligned}$$

which is $O_{\mathbb{P}}(n)$, and $|\hat{\beta}_m|$, $\|\hat{\Gamma}_{l,m}\|$, and $\|\hat{\Gamma}_{m,l}\|$ are $O_{\mathbb{P}}(1)$. Pólya's Theorem (Theorem 11.2.9 of Lehmann and Romano, 2005, p. 429) then implies $\hat{\Sigma}_{n,m}^{*-1/2} \lambda_{n,m}^* \rightsquigarrow N_{M-1}(0, I_{M-1})$ in probability and therefore $J_{n,m}^* \rightsquigarrow \chi_{M-1}^2$ in probability by continuity. \square

Proof of Lemma A.2. Let $\hat{m}^* := \arg \min \mathcal{J}_n^*$. I first show that if $m^* \in \mathcal{M}$, then \hat{m}^* approximates m^* . Notice that for all $m \in \mathcal{M} \setminus \{\hat{m}^*\}$, we have $\mathbb{P}^*(J_{n,m} \leq B) = 1(J_{n,m} \leq B)$ since $J_{n,m}$ is constant with respect to \mathbb{P}^* and therefore

$$\mathbb{P}^*(\hat{m}^* \neq \hat{m}) \leq 1(\exists m \in \mathcal{M} \setminus \{\hat{m}^*\} : J_{n,m} \leq B) + \mathbb{P}^*(J_{n,\hat{m}^*}^* > B). \quad (\text{B.1})$$

Without loss of generality, fix any $0 < \varepsilon < 1$; then $\mathbb{P}[1(\exists m \in \mathcal{M} \setminus \{\hat{m}^*\} : J_{n,m} \leq B) > \varepsilon] \leq \mathbb{P}(\exists m \in \mathcal{M} \setminus \{m^*\} : J_{n,m} \leq B) + \mathbb{P}(\hat{m}^* \neq m^*)$, which converges to zero as $n \rightarrow \infty$ by

Lemmas 2.5(ii) and 2.8. Further, $\mathbb{P}[\mathbb{P}^*(J_{n,\hat{m}}^* > B) > \varepsilon] \leq \mathbb{P}(\hat{m} \neq m^*) + \sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(J_{n,m^*}^* > B) > \varepsilon]$. Since the second term on the right-hand side can be shown to be small for large B by arguments similar to those used to prove Prohorov's Theorem (Theorem 2.4 of van der Vaart, 1998, p. 8), conclude that (B.1) and therefore also $\mathbb{P}^*(\hat{m}^* \neq \hat{m})$ is $o_{\mathbb{P}}(1)$; but then $\mathbb{P}[\mathbb{P}^*(\hat{m}^* \neq m^*) > \varepsilon] \leq \mathbb{P}[\mathbb{P}^*(\hat{m}^* \neq \hat{m}) > \varepsilon] + \mathbb{P}(\hat{m} \neq m^*) = o(1)$ for all $\varepsilon > 0$. Since $\mathbb{P}^*(J_{n,m^*}^* \leq x) = \mathbb{P}^*(J_{n,\hat{m}^*}^* \leq x) + \mathbb{P}^*(\{J_{n,m^*}^* \leq x\} \cap \{\hat{m}^* \neq m^*\})$ for all $x \in \mathbb{R}$, and $\mathbb{P}^*(\{J_{n,m^*}^* \leq x\} \cap \{\hat{m}^* \neq m^*\}) \leq \mathbb{P}^*(\hat{m}^* \neq m^*) = o_{\mathbb{P}}(1)$, we have $\text{plim}_{n \rightarrow \infty} \mathbb{P}^*(J_{n,\hat{m}^*}^* \leq x) = K(x)$, where K is the distribution function of a χ_{M-1}^2 variable. Pólya's Theorem then implies $MJ_n^* = J_{n,\hat{m}^*}^* \rightsquigarrow \chi_{M-1}^2$ in probability. Finally, because the supremum in (A.1) is bounded above by $\sup_x |\mathbb{P}^*(MJ_n^* \leq x) - K(x)| + \sup_x |\mathbb{P}(MJ_n \leq x) - K(x)|$, the result follows from Theorem 2.7(i). \square

Proof of Theorem 3.3. (i) Lemma A.2 combined with Lehmann and Romano's (2005, p. 430) Lemma 11.2.1 implies $\text{plim}_{n \rightarrow \infty} c_{n,1-\alpha}^* = K^{-1}(1 - \alpha)$, and therefore $MJ_n - c_{n,1-\alpha}^* \rightsquigarrow \mathcal{X} - K^{-1}(1 - \alpha)$ by the Slutsky Lemma, where $\mathcal{X} \sim \chi_{M-1}^2$; but $\mathbb{P}[\mathcal{X} - K^{-1}(1 - \alpha) \leq x] = K[K^{-1}(1 - \alpha) + x]$ is continuous in x and thus

$$|\mathbb{P}(MJ_n > c_{n,1-\alpha}^*) - \alpha| \leq \sup_{x \in \mathbb{R}} \left| \mathbb{P}(MJ_n - c_{n,1-\alpha}^* \leq x) - K[K^{-1}(1 - \alpha) + x] \right|$$

converges to zero by Pólya's Theorem.

(ii) Suppose for now that the random index \hat{m} that selects the model that is bootstrapped is equal to some fixed $l \in \mathcal{M}$. Define $\mathcal{J}_n^*(l) := \{J_{n,l}^*\} \cup \mathcal{J}_n \setminus \{J_{n,l}\}$ and notice that $\mathcal{J}_n^*(\hat{m}) = \mathcal{J}_n^*$. Lemma A.1 implies that $J_{n,l}^* \rightsquigarrow \chi_{M-1}^2$ in probability. To deal with the other elements of $\mathcal{J}_n^*(l)$, pick any $m \in \mathcal{M} \setminus \{l\}$ and note that for any $0 < \varepsilon < 1$, we can find a $B > 0$ such that $K(B) > 1 - \varepsilon$. Thus, $\mathbb{P}^*(J_{n,m} \leq B) \geq \mathbb{P}^*[|1 - K(J_{n,m})| > \varepsilon]$. Because $\mathbb{P}(J_{n,m} \leq B) = o_{\mathbb{P}^*}(1)$ in probability by Lemma 2.5(ii), the asymptotic distribution of $K(J_{n,m})$, $m \in \mathcal{M} \setminus \{l\}$, therefore converges to \mathbb{P}^* -point mass at 1 in probability.

By an ‘‘in probability’’ version of Theorem 2.7(v) of van der Vaart (1998, p. 10), we can now strengthen the marginal convergence of each element of $\mathcal{J}_n^*(l)$ to the joint convergence of the vector $\mathcal{K}_l := (K(J_{n,1}), \dots, K(J_{n,l}^*), \dots, K(J_{n,M}))$ such that

$$\text{plim}_{n \rightarrow \infty} \mathbb{P}^*(\mathcal{K}_l \leq x) = \mathbb{P}[(1, \dots, K(\mathcal{X}_l), \dots, 1) \leq x] \quad (\text{B.2})$$

for each $x \in \mathbb{R}^M$ at which the right-hand side is continuous, where $\mathcal{X}_l \sim \chi_{M-1}^2$. The Continuous Mapping Theorem and Pólya's Theorem in probability then yield $\min \mathcal{K}_l \rightsquigarrow \min(K(\mathcal{X}_l), 1) \sim \mathcal{U}_l \sim \text{Uniform}(0, 1)$ in probability.

There are M different ways of choosing the index l and therefore there are M different bootstrap procedures that can be carried out. The bootstrapped variables of each of the procedures, i.e., $(J_{n,1}^*, \dots, J_{n,M}^*)$, are independent conditional on y and \mathcal{F} , and therefore the components of the vector $(\min \mathcal{K}_1, \dots, \min \mathcal{K}_M)$ are also conditionally independent. By Lévy's Continuity Theorem (Theorem 11.2.2 of Lehmann and Romano, 2005, p. 426)

and Pólya's Theorem in probability, the marginal convergence then implies the joint convergence $(\min \mathcal{K}_1, \dots, \min \mathcal{K}_M) \rightsquigarrow (\mathcal{U}_1, \dots, \mathcal{U}_M)$ in probability, where the \mathcal{U}_l , $l \in \mathcal{M}$, are independent. Since K^{-1} is continuous and K is increasing, conclude that

$$\overline{MJ}_n^* := \max_{l \in \mathcal{M}} \min \mathcal{J}_n^*(l) = K^{-1} \left(\max_{l \in \mathcal{M}} \min \mathcal{K}_l \right) \rightsquigarrow K^{-1} \left(\max_{l \in \mathcal{M}} \mathcal{U}_l \right) \quad \text{in probability.}$$

Notice that $MJ_n^* = \min \mathcal{J}_n^*(\hat{m})$ and thus $\mathbb{P}^*(MJ_n^* > B) \leq \mathbb{P}^*(\overline{MJ}_n^* > B)$. By an ‘‘in probability’’ version of Prohorov's Theorem, for any $0 < \varepsilon < 1$ and $\delta > 0$, we can then find a B such that $\sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(MJ_n^* > B) > \varepsilon] \leq \sup_{n \in \mathbb{N}} \mathbb{P}[\mathbb{P}^*(\overline{MJ}_n^* > B) > \varepsilon] < \delta$. Conclude from Lemma 21.1(i) of van der Vaart (1998, p. 304) that $\mathbb{P}(c_{n,1-\alpha}^* > B) = \mathbb{P}[\mathbb{P}^*(MJ_n^* > B) > \alpha] < \delta$ uniformly in n for any fixed $\alpha \in (0, 1)$. This result and Lemma 2.5(ii) imply that for large enough n the right-hand side of the inequality $\mathbb{P}(MJ_n \leq c_{n,1-\alpha}^*) \leq \mathbb{P}(MJ_n \leq B) + \mathbb{P}(c_{n,1-\alpha}^* > B)$ can be made as small as desired, which completes the proof. \square

Proof of Corollary 3.4. Identical to the proof of Lemma 2.8, mutatis mutandis. \square

Acknowledgements

I would like to thank Anil Bera, Dan Bernhardt, Roger Koenker, Rolf Tschernig, three referees, and an associate editor for helpful discussions and comments. I gratefully acknowledge financial support from a Paul W. Boltz Fellowship. All errors are my own.

References

- Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* 24, 295–300.
- Choi, H.-s. and N. M. Kiefer (2008). Robust nonnested testing and the demand for money. *Journal of Business and Economic Statistics* 26, 9–17.
- Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 105–123.
- Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B* 24, 406–424.
- Dastoor, N. K. (1981). A note on the interpretation of the Cox procedure for non-nested hypotheses. *Economics Letters* 8, 113–119.
- Davidson, J., A. Monticini, and D. Peel (2007). Implementing the wild bootstrap using a two-point distribution. *Economics Letters* 96, 309–315.
- Davidson, R. and J. G. MacKinnon (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49, 781–793.
- Davidson, R. and J. G. MacKinnon (1985). Heteroskedasticity-robust tests in regression directions. *Annales de l'insee* 59-60, 183–218.
- Davidson, R. and J. G. MacKinnon (2002). Bootstrap J tests of nonnested linear regression models. *Journal of Econometrics* 109, 167–193.

- Fan, Y. and Q. Li (1995). Bootstrapping J -type tests for non-nested regression models. *Economics Letters* 48, 107–112.
- Fisher, G. and M. McAleer (1979). On the interpretation of the Cox test in econometrics. *Economics Letters* 4, 145–150.
- Fisher, G. R. and M. McAleer (1981). Alternative procedures and associated tests of significance for non-nested hypotheses. *Journal of Econometrics* 16, 103–119.
- Flachaire, E. (2002). Bootstrapping heteroskedasticity consistent covariance matrix estimator. *Computational Statistics* 17, 501–506.
- Giacomini, R., D. N. Politis, and H. White (2007). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Unpublished manuscript, Department of Economics, University of California, Los Angeles.
- Godfrey, L. G. (1998). Tests of non-nested regression models: Some small sample behaviour and the bootstrap. *Journal of Econometrics* 84, 59–74.
- Godfrey, L. G. and M. H. Pesaran (1983). Tests on non-nested regression models: Small sample adjustments and Monte Carlo evidence. *Journal of Econometrics* 21, 133–154.
- Granger, C. W. J., M. L. King, and H. White (1995). Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics* 67, 173–188.
- Leeb, H. and B. M. Pötscher (2009). Model selection. In T. G. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 889–926. Springer.
- Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer, New York.
- Liu, R. Y. (1988). Bootstrap procedures under some non-I.I.D. models. *Annals of Statistics* 16, 1696–1708.
- Long, J. S. and L. H. Ervin (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician* 54, 217–224.
- MacKinnon, J. G., H. White, and R. Davidson (1983). Several tests for model specification in the presence of alternative hypotheses: Some further results. *Journal of Econometrics* 21, 53–70.
- Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*, Volume 77 of *Lecture Notes in Statistics*. Springer, New York.
- Michelis, L. (1999). The distributions of the J and Cox non-nested tests in regression models with weakly correlated regressors. *Journal of Econometrics* 93, 369–401.
- Pesaran, M. H. and M. Weeks (2003). Nonnested hypothesis testing: An overview. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Chapter 13, pp. 279–309. Blackwell Publishing.
- Shi, X. (2010). Model selection tests for nonnested moment inequality models. Unpublished manuscript, Department of Economics, Yale University.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.