



# Model Confidence Set Based on Kullback-Leibler Divergence Distance

Ghobad Barmalzan<sup>†,\*</sup> and Amir T. Payandeh Najafabadi<sup>‡</sup>

<sup>†</sup> Zabol University

<sup>‡</sup> Shahid Beheshti University

**Abstract.** Consider the problem of estimating true density,  $h(\cdot)$  based upon a random sample  $X_1, \dots, X_n$ . In general,  $h(\cdot)$  is approximated using an appropriate (in some sense, see below) model  $f_\theta(x)$ . This article using Vuong's (1989) test along with a collection of  $k(> 2)$  non-nested models constructs a set of appropriate models, say model confidence set, for unknown model  $h(\cdot)$ . Application of such confidence set has been confirmed through a simulation study.

**Keywords.** Kullback-Leibler divergence distance; confidence set; model selection; non-nested models; Vuong's test.

MSC 2010: 62E30, 62F03.

## 1 Introduction

Model selection play an important role in statistical data analysis. Once a model has been identified, various forms of inferences such as prediction, information extraction, risk evaluation and decision making can be done using a given framework of deductive argument. Based upon data, model selection has two approaches: hypothesis testing and model selection criteria. Most of the time one chooses a model which is at least not falsified. Clearly, this approach is different from classical hypothesis testing. Two given models may be either nested or non-nested. The nested models are frequently studied in both theoretical and applied statistics. Surprisingly, the non-nested

---

\* Corresponding author

models do not receive enough attention from authors. Cox (1961, 1962), Atkinson (1970), Pesaran (1974), Pesaran and Deaton (1978), Davidson and MaKinnson (1981), Fishwer and McAleer (1981), Dastoor (1983) and Vuong (1989) were authors who worked on non-nested hypothesis tests. Among such authors Cox (1961, 1962) and Vuong (1989) developed two statistical approaches to inference for non-nested models. Both Cox's and Vuong's tests are a generalization of the likelihood ratio test (LRT) for non-nested hypotheses tests and non-nested model selection, respectively. In the Vuong's test viewpoint, an appropriate model is a model that maximizes the relevant part of the Kullback–Leibler divergence distance, say  $\mathcal{KL}$ . The Vuong's test built up based upon the expectation of likelihood ratio of the two competing models for unknown model  $h(\cdot)$ . Therefore, under the null hypothesis that two competing models are equivalent, the Vuong's test is zero. One may notice that decision-making procedure by the Vuong's test does not depend on unknown model  $h(\cdot)$ .

The idea of model selection starts with a collection of competing models which are appropriate (in some senses such visual investigation, nature of problem in hand, etc) for unknown model  $h(\cdot)$ . The decision-making process based upon such collection can be done using the Cox's test, the Vuong's test, or the AIC criterion. The AIC criterion (Hirotugu Akaike, 1973) is widely used to choose an appropriate model among a collection of competing parametric models. Shimodaira (1998) was pioneer author who developed a producer to construct sub-collection of models among a collection of competing parametric models. He obtained a model confidence set based on  $\mathcal{H}_0 : E(AIC_i) \leq \min_{i \neq j} E(AIC_j)$  against  $\mathcal{H}_1 : E(AIC_i) > \min_{i \neq j} E(AIC_j)$  by bootstrap simulation in the regression models.

Instead of choosing a single model that maximizes the log-likelihood, this article proposes a version of multiple comparisons method to construct a confidence set for unknown model  $h(\cdot)$  based upon a collection of non-nested models. The rest of this article is organized as follows. Some mathematical background for the problem has been collected in Section 2. Based on the relevant part of Kullback-Leibler divergence distance, Section 3 develops a new statistical test. Section 4 constructs confidence set for unknown model  $h(\cdot)$ . Practical application of our results have been study through a simulations study in Section 5.

## 2 Definitions and Notation

Kullback-Leibler divergence distance, say  $\mathcal{KL}$  is widely used statistical criteria which employs to select an appropriate model among a collection given competing models. The  $\mathcal{KL}$  of  $f_\theta(X)$  with respect to unknown model  $h(\cdot)$  and observation  $X$  is defined by expectation of  $\log h(X)/f_\theta(X)$ . In other words,

$$\begin{aligned}\mathcal{KL}\{h(X); f_\theta(X)\} &= E_h \left\{ \log \left( \frac{h(X)}{f_\theta(X)} \right) \right\} \\ &= E_h \{\log h(X)\} - E_h \{\log f_\theta(X)\}.\end{aligned}$$

The above two expressions  $E_h \{\log h(X)\}$  and  $E_h \{\log f_\theta(X)\}$  can be viewed as an irrelevant and relevant part of the  $\mathcal{KL}$ , respectively. Since model selection procedure does not depend on whole of the  $\mathcal{KL}$ , we just focus on its relevant part. It should be worthwhile to mention that, the  $\mathcal{KL}$  (i) takes its value from an interval  $[0, \infty)$ ; (ii) cannot be a metric but it has additive property over marginals of product measures; and (iii)  $\mathcal{KL}\{h(X); f_\theta(X)\} = 0$  implies that  $h(\cdot) = f_{\theta_0}(\cdot)$ .

Suppose  $\mathbf{X} = (X_1, \dots, X_n)$  is a random sample from unknown model  $h(\cdot)$  and

$$\begin{aligned}\mathcal{F}_\Theta &= \{f_\theta(\cdot) : \theta \in \Theta \subseteq \mathbb{R}^p\}, \\ \mathcal{G}_\Gamma &= \{g_\gamma(\cdot) : \gamma \in \Gamma \subseteq \mathbb{R}^q\},\end{aligned}$$

are two family of models which are appropriate for unknown model  $h(\cdot)$ . Two models  $\mathcal{F}_\Theta$  and  $\mathcal{G}_\Gamma$  are non-nested models if and only if  $\mathcal{F}_\Theta \cap \mathcal{G}_\Gamma = \emptyset$  and  $\mathcal{G}_\Gamma$  is nested in  $\mathcal{F}_\Theta$  if and only if  $\mathcal{G}_\Gamma \subset \mathcal{F}_\Theta$ . Therefore, two competing models are non-nested if one models cannot be reduced to another by imposing some restrictions on certain parameter.

**Definition 1.** The model set  $\mathcal{F}_\Theta$  with respect to  $h(\cdot)$  is called well-specified if and only if there exists  $\theta_0 \in \Theta$  such that  $h(\cdot) \equiv f_{\theta_0}(\cdot)$ . Otherwise,  $\mathcal{F}_\Theta$  is called misspecified model with respect to  $h(\cdot)$ .

For misspecified model set  $\mathcal{F}_\Theta$ , one may find a  $\theta_\star \in \Theta$  which minimizes the discrepancy between  $h(\cdot)$  and  $f_\theta(\cdot)$ . The quasi maximum likelihood estimator (QMLE), and the pseudo  $\mathcal{KL}$  estimator can be appropriate choices for such  $\theta_\star$ . The following provides definition of the QML and the pseudo  $\mathcal{KL}$  estimators.

**Definition 2.** Suppose  $\mathbf{X} = (X_1, \dots, X_n)$  stands for a random sample from the unknown model  $h(\cdot)$ . Moreover, suppose that  $f_\theta(x)$  is an appropriate (in some sense) candidate for  $h(\cdot)$ . Then, the quasi maximum likelihood and pseudo  $\mathcal{KL}$  estimators for  $\theta$ , respectively, are given by

$$\begin{aligned}\hat{\theta}_{qMLE} &= \arg \sup_{\theta \in \Theta} \mathcal{L}_n^f(\theta); \\ \theta_{\mathcal{KL}} &= \arg \max_{\theta \in \Theta} E_h \left\{ \frac{1}{n} \mathcal{L}_n^f(\theta) \right\},\end{aligned}$$

where the quasi log likelihood  $\mathcal{L}_n^f(\theta)$  is given by  $\mathcal{L}_n^f(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$ .

The QMLE and the classical maximum likelihood estimator (MLE) coincide whenever  $\mathcal{F}_\Theta$  with respect to  $h(\cdot)$  is a well-specified model set. Wald (1949) showed that under some regularity conditions, the QMLE coincides on  $\theta_0$  which returns minimum  $\mathcal{KL}\{h(\cdot); f_\theta(\cdot)\}$ . Moreover, Huber (1976) showed that for misspecified model sets,  $\hat{\theta}_{qMLE}$  is a consistent estimator of  $\theta_*$ . As mentioned above, the  $\mathcal{KL}$  study just though  $E_h\{\log f_\theta(X)\}$  depends on unknown model  $h(\cdot)$ . The expectation  $E_h\{\log f_\theta(X)\}$  can be estimated in two following steps:

**Step 1 (ML estimator for  $h(\cdot)$ )** Based upon random sample  $\mathbf{X} = (X_1, \dots, X_n)$ , suppose that Bernoulli random variable  $Y_t(x)$  is defined by  $Y_t(x) := 1(X_t \leq x)$ . The ML estimator for distribution function  $H(x) := P(X_1 \leq x)$  is given by

$$\hat{H}(x) = \frac{1}{n} \sum_{t=1}^n Y_t = \frac{1}{n} \sum_{t=1}^n 1(X_t \leq x).$$

Moreover, the empirical estimator for density  $h(\cdot)$  is given by  $\hat{h}(x_i) = 1/n$ , for  $i = 1, \dots, n$ . Substituting such empirical estimation leads to

$$\begin{aligned}E_{\hat{h}}\{\log f_\theta(X)\} &= \int \log f_\theta(x) d\hat{H}(x) \\ &= \frac{1}{n} \sum_{i=1}^n \log f_\theta(x_i)\end{aligned}$$

which is still unknown because  $\theta$  is so.

**Step 2 (ML estimator for  $\theta$ )** Now by replacing  $\hat{\theta}_{qmle}$  for  $\theta$ , expectation  $E_h\{\log f_{\theta}(X)\}$  can be estimated by

$$\frac{1}{n} \sum_{i=1}^n \log f_{\hat{\theta}_{qmle}}(x_i).$$

The likelihood ratio statistic, say  $\mathcal{LR}$ , for the model  $\mathcal{F}_{\Theta}$  against the model  $\mathcal{G}_{\Gamma}$ ,  $\mathcal{LR}_n(\hat{\theta}_{qmle}, \hat{\gamma}_{qmle})$ , is given by

$$\begin{aligned} \mathcal{LR}_n(\hat{\theta}_{qmle}, \hat{\gamma}_{qmle}) &= \mathcal{L}_n^f(\hat{\theta}_{qmle}) - \mathcal{L}_n^g(\hat{\gamma}_{qmle}) \\ &= \sum_{i=1}^n \log \frac{f_{\hat{\theta}_{qmle}}(X_i)}{g_{\hat{\gamma}_{qmle}}(X_i)}. \end{aligned}$$

White (1982) showed that under some regularity conditions

$$\frac{1}{n} \mathcal{LR}_n(\hat{\theta}_{qmle}, \hat{\gamma}_{qmle}) \xrightarrow{a.s.} E_h \left\{ \log \frac{f_{\theta_*}(X)}{g_{\gamma_*}(X)} \right\}.$$

Moreover, Vuong (1989) showed that  $f_{\theta_*}(\cdot) \neq g_{\gamma_*}(\cdot)$  whenever

$$n^{-\frac{1}{2}} \mathcal{LR}_n(\hat{\theta}_{qmle}, \hat{\gamma}_{qmle}) - n^{\frac{1}{2}} E_h \left\{ \log \frac{f_{\theta_*}(X)}{g_{\gamma_*}(X)} \right\} \xrightarrow{D} N(0, w_{\star}^2),$$

where  $w_{\star}^2$  denotes the variance of random variable  $\log(f_{\theta_*}(X)/g_{\gamma_*}(X))$  with respect to  $h(\cdot)$ . A natural estimator for  $w_{\star}^2$  could be

$$\hat{w}_n^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \log \frac{f_{\hat{\theta}_{qmle}}(X_i)}{g_{\hat{\gamma}_{qmle}}(X_i)} \right\}^2 - \left\{ \frac{1}{n} \sum_{i=1}^n \log \frac{f_{\hat{\theta}_{qmle}}(X_i)}{g_{\hat{\gamma}_{qmle}}(X_i)} \right\}^2.$$

## Main Results

### 3 A Generalization for the Vuong's Test

The classical hypotheses testing cannot be employed whenever hypotheses do not belong to the same parametric family. A pioneering contributions was Cox (1961, 1962) who modified the classical procedure for two non-nested models. The model selection criteria does not talk about truth or falsity of a hypothesis. It talks about closeness of a candidate model to the true

model. Hypothesis testing asks whether there is any statistically significant evidence of departure from the null hypothesis in the direction of one or more alternative hypotheses. On the other hand, in the case of hypothesis testing for model selection (nested or non-nested hypothesis testing) it is possible for all models under consideration to be rejected. The null hypothesis for classical and non-nested hypothesis testing play different roles. When the models are nested the classical hypothesis testing approach select the most parsimonious model as the null hypothesis. But, when the model are non-nested the null hypothesis is selected on a priori grounds. The Vuong's test implies the equivalence of two competing models from the aspect of being close to true model  $h(\cdot)$ . This section generalizes this problem to  $k(\geq 2)$  non-nested competing models. Multiple comparison techniques are methods of performing the tests so that if all the hypotheses are true, then the probability of rejecting any of the hypotheses is no greater than some specified value. Suppose that we have a list of  $k$  competing non-nested models which could be used to describe random sample  $\mathbf{X}$ . Moreover suppose that  $\mathcal{U}$  denotes a collection of such non-nested competing models. Each model  $\mathcal{F}_{\Theta(i)} \in \mathcal{U}$  can be restated as

$$\mathcal{F}_{\Theta(i)} = \{f_{\theta(i)}(\cdot) : \theta(i) \in \Theta(i) \subseteq \mathbb{R}^{p_i}\}, \quad i \in \mathcal{M} := \{1, \dots, k\}.$$

For each model  $\mathcal{F}_{\Theta(i)}$ , there exists a member  $f_{\theta_*(i)}(X)$  such that is closest member to true density  $h(\cdot)$ . Then, under the  $\mathcal{KL}$  model  $f_{\theta_*(i)}(\cdot)$  is an appropriate approximation for  $h(\cdot)$ . Now, for each  $i \in \mathcal{M}$  consider the following hypothesis

$$\begin{cases} \mathcal{H}_{0i} : E_h\{\log f_{\theta_*(i)}(X)\} \geq \max_{j \in \mathcal{M} \setminus \{i\}} E_h\{\log f_{\theta_*(j)}(X)\}, \\ \mathcal{H}_{1i} : E_h\{\log f_{\theta_*(i)}(X)\} < \max_{j \in \mathcal{M} \setminus \{i\}} E_h\{\log f_{\theta_*(j)}(X)\}. \end{cases}$$

Testing the null hypothesis  $\mathcal{H}_{0i}$  is reasonable because the null hypothesis  $\mathcal{H}_{0i}$  shows that  $\mathcal{F}_{\Theta(i)}$  has the maximum expected log-likelihood and consequently has the minimum  $\mathcal{KL}$  between  $(k - 1)$  non-nested models. Therefore, it is natural to select  $\mathcal{F}_{\Theta(i)}$  from aspect of being close to the unknown true density  $h(\cdot)$ .

**Theorem 1.** *Suppose  $X_1, \dots, X_n$  be a random sample from unknown density  $h(\cdot)$ . Moreover, suppose that  $\mathcal{U}$  denotes a class of these non-nested competing*

models for approximation  $h(\cdot)$ . For testing

$$\begin{cases} \mathcal{H}_{0i} : E_h \{ \log f_{\theta_*(i)}(X) \} \geq \max_{j \in \mathcal{M} \setminus \{i\}} E_h \{ \log f_{\theta_*(j)}(X) \}, \\ \mathcal{H}_{1i} : E_h \{ \log f_{\theta_*(i)}(X) \} < \max_{j \in \mathcal{M} \setminus \{i\}} E_h \{ \log f_{\theta_*(j)}(X) \}, \end{cases}$$

the hypothesis  $\mathcal{H}_{0i}$  will be rejected at significant level  $\alpha$  if and only if

$$\left\{ \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} < -c \right\} = \left\{ \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} < z_{\frac{\alpha}{k-1}} \right\},$$

where

$$T_{ij} = \frac{n^{-\frac{1}{2}} \left\{ \mathcal{L}_n^{f_i}(\widehat{\theta}_*(i)) - \mathcal{L}_n^{f_j}(\widehat{\theta}_*(j)) - (p_i - p_j) \right\}}{\widehat{w}_{ij}}.$$

**Proof.** The null hypothesis  $\mathcal{H}_{0i}$  contains the following  $(k-1)$  individual hypothesis.

$$\mathcal{H}'_{0j} : E_h \{ \log f_{\theta_*(i)}(X) \} \geq E_h \{ \log f_{\theta_*(j)}(X) \} \quad \text{for all } j \in \mathcal{M} \setminus \{i\},$$

against

$$\mathcal{H}'_{1j} : E_h \{ \log f_{\theta_*(i)}(X) \} < E_h \{ \log f_{\theta_*(j)}(X) \},$$

each individual hypothesis testing  $\mathcal{H}'_{0j}$  against  $\mathcal{H}'_{1j}$  is at the significant level of  $\alpha/(k-1)$ . Now, consider

$$\mathcal{H}''_{0j} : E_h \{ \log f_{\theta_*(i)}(X) \} = E_h \{ \log f_{\theta_*(j)}(X) \},$$

against

$$\mathcal{H}'_{1j} : E_h \{ \log f_{\theta_*(i)}(X) \} < E_h \{ \log f_{\theta_*(j)}(X) \}.$$

Since the models  $\mathcal{F}_{\Theta(i)}$  and  $\mathcal{F}_{\Theta(j)}$  are non-nested, then results of White (1982) and Vuong (1989) imply that  $f_{\theta_*(i)}(\cdot) \neq f_{\theta_*(j)}(\cdot)$ . By using the Vuong's approach, test statistic for individual hypothesis  $\mathcal{H}''_{0j}$  against  $\mathcal{H}'_{1j}$  is

$$T_{ij} = \frac{n^{-\frac{1}{2}} \left\{ \mathcal{L}_n^{f_i}(\widehat{\theta}_*(i)) - \mathcal{L}_n^{f_j}(\widehat{\theta}_*(j)) - (p_i - p_j) \right\}}{\widehat{w}_{ij}}.$$

Which asymptotically has standard normal distribution under  $\mathcal{H}_{0j}''$ . The individual hypothesis  $\mathcal{H}_{0j}''$  is rejected if  $T_{ij} < -c$  and  $c$  satisfies  $\Phi(-c) = \frac{\alpha}{k-1}$ , where  $\Phi(z)$  is the cumulative distribution function of standard normal. Now, set

$$\Delta_{ij} = E_h\{\log f_{\theta_{*}(i)}(X)\} - E_h\{\log f_{\theta_{*}(j)}(X)\}.$$

Under the  $\mathcal{H}'_{0j}$ , the values of  $\Delta_{ij}$  is positive. So

$$\begin{aligned} \frac{\alpha}{k-1} &= P(T_{ij} < -c) \\ &\geq P\left(T_{ij} < -c - \frac{\Delta_{ij}}{\widehat{w}_{ij}}\right). \end{aligned}$$

The latter inequality indicates that the class of tests  $\mathcal{H}''_{0j}$  contains the class of test  $\mathcal{H}'_{0j}$ . On the other hand, the hypothesis  $\mathcal{H}''_{0j}$  means the equality in the equations of  $\mathcal{H}'_{0j}$ . It follows that the class of test  $\mathcal{H}''_{0j}$  is larger than class of  $\mathcal{H}'_{0j}$ . Since the hypothesis  $\mathcal{H}'_{0j}$  is subclass of  $\mathcal{H}''_{0j}$ . Therefore,  $T_{ij} < -c$  is also critical region for  $\mathcal{H}'_{0j}$ .

For  $i, j \in \mathcal{M}$ , each of individual tests has an acceptance region of the form  $\{T_{ij} > -c\}$ , where  $c$  dose not depend on  $i$  and  $j$ . Therefore, the hypothesis  $\mathcal{H}_{0i}$  will be accepted if and only if

$$T_{ij} > -c \quad \text{for all } j \in \mathcal{M} \setminus \{i\},$$

this is equivalent to

$$\left\{ \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} > -c \right\}.$$

Consequently,  $\mathcal{H}_{0i}$  will be rejected if and only if

$$\left\{ \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} < -c \right\} = \left\{ \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} < z_{\frac{\alpha}{k-1}} \right\}.$$

□

The calculation of above test statistics is not difficult. The R software is an appropriate statistical software which is used. If the hypotheses to be tested are chosen after looking at data, the individual significance level of  $\alpha/(k-1)$  are invalid, so experiment error rate has not been controlled. The Bonferroni method can be taken into account to handle for a finite number of hypotheses. This paper doesn't assume that  $h \in \mathcal{F}_{\Theta(i)}$  for some  $i \in \mathcal{M}$ .

## 4 A $100(1 - \alpha)\%$ Model Confidence Set

A model confidence set, MCS, is a set of density functions that it will contain the true model  $h(\cdot)$  with a given level of confidence. This section investigates the following main question which arise in the model selection:

*How is it possible to build a MCS from a collection of  $k(\geq 2)$  appropriate non-nested models?*

By having such collection of  $k(\geq 2)$  appropriate non-nested models, say  $\mathcal{U}$ , our investigation is restricted within  $\mathcal{U}$ . Then by considering a test based on relevant part of the  $\mathcal{KL}$  one has to choose a set of appropriate models rather than a single model. This set is called a MCS. The MCS as dual of hypothesis testing is constructed to assess which models are more appropriate than theirs in terms of the  $\mathcal{KL}$  divergence.

To illustrate such idea, consider the most simplest case where one has just two non-nested models, say  $\mathcal{F}_{\Theta(1)}$  and  $\mathcal{F}_{\Theta(2)}$ . In the first step, consider the null hypothesis test

$$E_h\{\log f_{\theta_{\star(1)}}(X)\} \geq E_h\{\log f_{\theta_{\star(2)}}(X)\},$$

against

$$E_h\{\log f_{\theta_{\star(1)}}(X)\} < E_h\{\log f_{\theta_{\star(2)}}(X)\},$$

and include  $\mathcal{F}_{\Theta(1)}$  in MCS unless the null hypothesis is rejected. In the next step, repeat the same hypothesis testing, but the roles of  $\mathcal{F}_{\Theta(1)}$  and  $\mathcal{F}_{\Theta(2)}$  must be interchanged. The MCS will be one of  $\{\mathcal{F}_{\Theta(1)}\}$ ,  $\{\mathcal{F}_{\Theta(2)}\}$  or  $\{\mathcal{F}_{\Theta(1)}, \mathcal{F}_{\Theta(2)}\}$ .

The above idea readily can be extend to the situation that there exists a collection of  $k(\geq 2)$  non-nested appropriate models, say  $\{\mathcal{F}_{\Theta(i)} \mid i \in \mathcal{M} = \{1, \dots, k\}\}$ . For each  $i \in \mathcal{M}$ , one may consider

$$\begin{cases} \mathcal{H}_{0i} : E_h\{\log f_{\theta_{\star(i)}}(X)\} \geq \max_{j \in \mathcal{M} \setminus \{i\}} E_h\{\log f_{\theta_{\star(j)}}(X)\}, \\ \mathcal{H}_{1i} : E_h\{\log f_{\theta_{\star(i)}}(X)\} < \max_{j \in \mathcal{M} \setminus \{i\}} E_h\{\log f_{\theta_{\star(j)}}(X)\}, \end{cases}$$

and include  $\mathcal{F}_{\Theta(i)}$  in MCS unless is rejected at a prescribed significance level. The MCS cannot be empty because at least it contains the maximum  $E_h\{\log f_{\theta_{\star(i)}}(X)\}$  model at an error smaller than specified significant level.  $\tau$  is a MCS whenever

$$P(\mathcal{F}_{\Theta(m)} \in \tau) = P(\mathcal{H}_{0m} \text{ is not rejected}) \geq 1 - \alpha.$$

On the other hand, the MCS at a given significance level is obtained from

$$\tau = \left\{ \mathcal{F}_{\Theta(i)} \in \mathcal{U} \mid \min_{j \in \mathcal{M} \setminus \{i\}} T_{ij} > z_{\frac{\alpha}{k-1}}; \quad i \in \mathcal{M} \right\}.$$

The MCS in model selection approach is different from confidence set in classical approach. There is not exists the parameter related to MCS of models. Confidence set of models makes sense because

$$\begin{aligned} P(\mathcal{F}_{\Theta(m)} \in \tau) &= P(\mathcal{H}_{0m} \text{ is not rejected}) \\ &= P\left(\min_{j \in \mathcal{M} \setminus \{m\}} T_{mj} > -c\right) \\ &= P(T_{m1} > -c, \dots, T_{m,m-1} > -c, T_{m,m+1} \\ &\quad > -c, \dots, T_{mk} > -c) \\ &\geq \left\{ \sum_{j=1, j \neq m}^k P(T_{mj} > -c) \right\} - (k-2) \\ &\hspace{15em} \text{(Bonferroni Inequality)} \\ &= (k-1) \left(1 - \frac{\alpha}{k-1}\right) - (k-2) \\ &= 1 - \alpha. \end{aligned}$$

Therefore, this procedure leads to a quantitative measure of reliability. Result of the above method is different from those found out by Shimodaira (1998). He employed the bootstrap-based multiple comparisons procedure while in the above method the Bonferroni correction of pairwise tests of log-likelihood differences has been employed. These two procedures are essentially the same but there is some difference in commotional manner and critical constants of significance tests.

## 5 Simulation Study

This section through a simulation study shows that how one may employ the above findings in practical applications. For this simulation study, we

consider the following three non-nested models.

$$\begin{aligned}\mathcal{F}_{\Theta(1)} &= \left\{ f_{\theta(1)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}; \quad \theta(1) = (\mu, \sigma^2) \right\}; \\ \mathcal{F}_{\Theta(2)} &= \left\{ f_{\theta(2)}(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\frac{x}{\lambda}\right\}; \quad \theta(2) = (\alpha, \lambda) \right\}; \text{ and} \\ \mathcal{F}_{\Theta(3)} &= \left\{ f_{\theta(3)}(x) = \frac{\beta}{\gamma} \left(\frac{x}{\gamma}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\gamma}\right)^\beta\right\}; \quad \theta(3) = (\beta, \gamma) \right\}.\end{aligned}$$

as a possible competing for unknown model  $h(\cdot)$ . Based upon a random sample  $X_1, \dots, X_n$  the QML estimator for parameters of such three non-nested models are:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{t=1}^n \log X_t; \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (\log X_t - \hat{\mu})^2; \\ \hat{\lambda} &= \frac{1}{n} \sum_{i=1}^n X_i; \quad \log \hat{\alpha} - \psi(\hat{\alpha}) = \log \hat{\lambda} - \frac{1}{n} \sum_{i=1}^n \log X_i; \\ \hat{\gamma} &= \frac{1}{n} \sum_{i=1}^n X_i^{\hat{\beta}}; \quad \hat{\beta} = \left\{ \frac{\sum_{i=1}^n X_i^{\hat{\beta}} \log X_i}{\sum_{i=1}^n X_i^{\hat{\beta}}} - \frac{\sum_{i=1}^n \log X_i}{n} \right\}^{-1},\end{aligned}$$

where  $\psi(x) = \frac{\partial}{\partial x} \ln \Gamma(x)$ .

Using the above three candidate models along with the following three hypothesis tests, one may build up a  $(1 - \alpha)\%$  MCS for true model  $h(\cdot)$ .

$$\mathcal{H}_{01} : E_h\{\log f_{\hat{\theta}(1)}(X)\} \geq \max_{j \in \{2,3\}} E_h\{\log f_{\hat{\theta}(j)}(X)\} \text{ v.s. } \mathcal{H}_{11} : \text{Reject } \mathcal{H}_{01};$$

$$\mathcal{H}_{02} : E_h\{\log f_{\hat{\theta}(2)}(X)\} \geq \max_{j \in \{1,3\}} E_h\{\log f_{\hat{\theta}(j)}(X)\} \text{ v.s. } \mathcal{H}_{12} : \text{Reject } \mathcal{H}_{02};$$

$$\mathcal{H}_{03} : E_h\{\log f_{\hat{\theta}(3)}(X)\} \geq \max_{j \in \{1,2\}} E_h\{\log f_{\hat{\theta}(j)}(X)\} \text{ v.s. } \mathcal{H}_{13} : \text{Reject } \mathcal{H}_{03}.$$

(1)

Using the Monte-Carlo method with  $10^3$  irritations, the following examples generate data (with sample size  $n = 50, 100, 200$ ) from the above non-nested models. Then, using such generated data each example builds up an 95% MCS for true model  $h(\cdot)$ .

**Example 1.** Suppose random sample  $x_1, \dots, x_n$  ( $n = 50, 100, 200$ ) have been generated from a Lognormal model with parameters  $\theta = (\mu = 2, \sigma^2 = 0.5)$ .

To build up a 95% MCS for true model  $h(\cdot)$ , one has to conduct Hypothesis tests (1) at significant level  $\alpha = 0.05$ .

Table 1 shows decision of the above hypothesis test at significant level  $\alpha = 0.05$  for sample size  $n = 50, 100, 200$ .

**Table 1.** Results of the above hypotheses test (1) at level  $\alpha = 0.05$  based upon  $10^3$  Monte-Carlo irritations and sample size  $n = 50, 100$  and  $200$ .

Sample Size	Hypothesis Test					
	$H_{01}$ v.s. $H_{11}$		$H_{02}$ v.s. $H_{12}$		$H_{03}$ v.s. $H_{13}$	
	Test Statistic	Conclusion	Test Statistic	Conclusion	Test Statistic	Conclusion
$n = 50$	0.9902	$\mathcal{H}_{01}$ is accepted	-1.0547	$\mathcal{H}_{02}$ is accepted	-4.5339	$\mathcal{H}_{03}$ is rejected
$n = 100$	1.3504	$\mathcal{H}_{01}$ is accepted	-1.3624	$\mathcal{H}_{02}$ is accepted	-6.7850	$\mathcal{H}_{03}$ is rejected
$n = 200$	1.8656	$\mathcal{H}_{01}$ is accepted	-1.8656	$\mathcal{H}_{02}$ is accepted	-9.7171	$\mathcal{H}_{03}$ is rejected

Using results of Table 1, one may build up the following 95% MCS for the true density function Lognormal(2, 0.5) is given by

$$\tau = \{\mathcal{F}_{\Theta(1)}; \mathcal{F}_{\Theta(2)}\} = \{\text{Lognormal}; \text{Gamma}\}.$$

The MCS shows that the method works properly and the true model (Lognormal) falls in MCS. For the interpretation of equivalence of two competing models, see Sayyareh et. al (2011). □

**Example 2.** Suppose random sample  $x_1, \dots, x_n$  ( $n = 50, 100, 200$ ) have been generated from a Gamma density function with parameters  $\theta = (\alpha = 2, \lambda = 3)$ . To build up a 95% MCS for true model  $h(\cdot)$ , one has to conduct Hypothesis tests (1) at significant level  $\alpha = 0.05$ .

Table 2 shows decision of the above hypothesis test at significant level  $\alpha = 0.05$  for sample size  $n = 50, 100, 200$ .

**Table 2.** Results of the above hypotheses test (1) at level  $\alpha = 0.05$  based upon  $10^3$  Monte-Carlo irritations and sample size  $n = 50, 100$  and  $200$ .

Sample Size	Hypothesis Test					
	$H_{01}$ v.s. $H_{11}$		$H_{02}$ v.s. $H_{12}$		$H_{03}$ v.s. $H_{13}$	
	Test Statistic	Conclusion	Test Statistic	Conclusion	Test Statistic	Conclusion
$n = 50$	-1.1444	$\mathcal{H}_{01}$ is accepted	-0.2142	$\mathcal{H}_{02}$ is accepted	0.1841	$\mathcal{H}_{03}$ is accepted
$n = 100$	-1.4681	$\mathcal{H}_{01}$ is accepted	0.1120	$\mathcal{H}_{02}$ is accepted	0.5929	$\mathcal{H}_{03}$ is accepted
$n = 200$	-1.5058	$\mathcal{H}_{01}$ is accepted	0.5352	$\mathcal{H}_{02}$ is accepted	1.1168	$\mathcal{H}_{03}$ is accepted

Using results of Table 2, one may build up the following 95% MCS for the true density function Gamma(2, 3) is given by

$$\tau = \{\mathcal{F}_{\Theta(1)}; \mathcal{F}_{\Theta(2)}; \mathcal{F}_{\Theta(3)}\} = \{\text{Lognormal}; \text{Gamma}; \text{Weibull}\}.$$

□

**Example 3.** Suppose random sample  $x_1, \dots, x_n$  ( $n = 50, 100, 200$ ) have been generated from a Weibull density function with parameters  $\theta = (\beta = 2, \gamma = 5)$ . To build up a 95% MCS for true model  $h(\cdot)$ , one has to conduct Hypothesis tests (1) at significant level  $\alpha = 0.05$ .

Table 3 shows decision of the above hypothesis test at significant level  $\alpha = 0.05$  for sample size  $n = 50, 100, 200$ .

**Table 3.** Results of the above hypotheses test (1) at level  $\alpha = 0.05$  based upon  $10^3$  Monte-Carlo irritations and sample size  $n = 50, 100$  and  $200$ .

Sample Size	Hypothesis Test					
	$H_{01}$ v.s. $H_{11}$		$H_{02}$ v.s. $H_{12}$		$H_{03}$ v.s. $H_{13}$	
	Test Statistic	Conclusion	Test Statistic	Conclusion	Test Statistic	Conclusion
$n = 50$	-2.1264	$\mathcal{H}_{01}$ is rejected	-0.8228	$\mathcal{H}_{02}$ is accepted	1.4831	$\mathcal{H}_{03}$ is accepted
$n = 100$	-6.7829	$\mathcal{H}_{01}$ is rejected	-1.0639	$\mathcal{H}_{02}$ is accepted	2.1042	$\mathcal{H}_{03}$ is accepted
$n = 200$	-3.5932	$\mathcal{H}_{01}$ is rejected	-1.3516	$\mathcal{H}_{02}$ is accepted	2.7686	$\mathcal{H}_{03}$ is accepted

Using results of Table 3, one may build up the following 95% MCS for the true density function Weibull (2, 5) is given by

$$\tau = \{\mathcal{F}_{\Theta(2)}; \mathcal{F}_{\Theta(3)}\} = \{\text{Gamma}; \text{Weibull}\}.$$

□

Using value of statistic test for the above examples, one may conclude that increasing sample size leads to more accurate MCS.

## 6 Conclusion and Suggestion

This article considered the problem of estimating an unknown model  $h(\cdot)$  based upon a random sample  $X_1, \dots, X_n$  and collection of  $k$  ( $k \geq 2$ ) non-nested models. This article using  $k - 1$  hypothesis tests with significant level  $\alpha/(k - 1)$  develops an  $100(1 - \alpha)\%$  MCS for  $h(\cdot)$ . A simulation study has been conducted to show applicability of our findings. The findings of this

article cannot employ whenever number of non-nested models is large and a single step procedure has to develop.

## References

- Akaike, H. (1973). Information Theory and An Extension of Maximum Likelihood Principle. *In Second International Symposium on Information Theory*. Akademia Kiado, 267-281.
- Atkinson, A.C. (1970). A Method for Discriminating Between Models. *Journal of the Royal Statistical Society*, **32**, 323-344.
- Cox, D.R. (1961). Tests of Separate Families of Hypothesis. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 105-123.
- Cox, D.R. (1962). Further Results on Tests of Separate Families of Hypothesis. *Journal of the Royal Statistical Society*, **24**, 406-424.
- Dastoor, N.K. (1983). Some Aspects of Testing Non-nested Hypothesis. *Journal of Econometrics*, **21**, 213-228.
- Davidson, R. and MacKinnon, J.G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Journal of Econometrics*, **29**, 213-228.
- Fisher, G.R. and McAleer, M. (1981). Alternative Procedures and Associated Tests of Significance for Non-nested Hypotheses. *Journal of Econometrics*, **16**, 103-119.
- Huber, P.J. (1976). The Behavior of Maximum Likelihood Estimate under Nonstandard Conditions. *Proc. Fifth Berkeley Symposium in Mathematical Statistics and Probability*. Berkeley, University of California Press, 221-223.
- Pesaran, M.H. (1974). On the General Test of Model Selection. *Review of Economic Studies*, **41**, 153-171.
- Pesaran, M.H. and Deaton, A.S. (1978). Testing Non-nested Nonlinear Regression Models. *Econometrica*, **46**, 667-694.
- Shimodaira, H. (1988). An Application of Multiple Comparison Techniques to Model Selection. *Annals of Institute of Statistical Mathematics*, **50**, 1-13.
- Vuong, Q.H. (1989). Likelihood Ratio Test for Model Selection and Non-nested Hypotheses. *Econometrica*, **57**, 307-333.
- Wald, A. (1949). Note on Consistency of Maximum Likelihood Estimate. *Annals of Mathematical Statistics*, **60**, 595-603.
- White, H. (1982). Maximum Likelihood Estimation of Mis-Specified Models. *Econometrica*, **50**, 1-26.

**Ghobad Barmalzan**

Department of Sciences,  
Zabol University,  
Sistan and Baluchestan, Zabol, Iran.  
email: *ghbarmalzan@uoz.ac.ir*

**Amir T. Payandeh Najafabadi**

Department of Mathematical Sciences,  
Shahid Beheshti University,  
Tehran, Iran.  
email: *amirtpayandeh@sbu.ac.ir*