Research Article

A Focused Bayesian Information Criterion

Georges Nguefack-Tsague¹ and Ingo Bulla²

¹ Biostatistics Unit, Department of Public Health, University of Yaoundé I, P.O. Box 1364, Yaoundé, Cameroon

² Theoretical Biology and Biophysics, Group T-6, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Correspondence should be addressed to Georges Nguefack-Tsague; nguefacktsague@yahoo.fr

Received 31 May 2014; Revised 18 September 2014; Accepted 25 September 2014; Published 14 October 2014

Academic Editor: Vito Mr Muggeo

Copyright © 2014 G. Nguefack-Tsague and I. Bulla. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Myriads of model selection criteria (Bayesian and frequentist) have been proposed in the literature aiming at selecting a single model regardless of its intended use. An honorable exception in the frequentist perspective is the “focused information criterion” (FIC) aiming at selecting a model based on the parameter of interest (focus). This paper takes the same view in the Bayesian context; that is, a model may be good for one estimand but bad for another. The proposed method exploits the Bayesian model averaging (BMA) machinery to obtain a new criterion, the *focused Bayesian model averaging* (FoBMA), for which the best model is the one whose estimate is closest to the BMA estimate. In particular, for two models, this criterion reduces to the classical Bayesian model selection scheme of choosing the model with the highest posterior probability. The new method is applied in linear regression, logistic regression, and survival analysis. This criterion is specially important in epidemiological studies in which the objective is often to determine a risk factor (focus) for a disease, adjusting for potential confounding factors.

1. Introduction

A variety of model selection criteria (Bayesian or frequentist) have been proposed in the literature; most of them aim at selecting a single model for any purposes. For an overview of model selection criteria, see the studies by Leeb and Poetscher [1] and Zucchini [2]; for inference after model selection, see the studies by Nguefack-Tsague [3], Nguefack-Tsague and Zucchini [4], Zucchini et al. [5], Behl et al. [6], and Nguefack-Tsague [7–9]. Allen [10], within the context of Mallows’ C_p [11], developed a criterion that depends on a given prediction. In a frequentist approach, Claeskens and Hjort [12] developed a focused information criterion (FIC) for model selection which, unlike common model selection criteria that lead to a single model for all purposes, selects different models for different purposes. Thus Allen’s criterion can be considered as an early precursor of FIC. So far the FIC is gaining in popularity as evidenced by its applications in various fields and specific models. Some of these applications include missing response (Sun et al. [13]), energy substitution (Behl et al. [6]), economic applications (Behl et al. [14]), Tobit model (Zhang et al. [15]), additive partial models (Zhang and Liang [16]), volatility forecasting (Brownlees and Gallo [17]),

and Cox proportional hazard regression models (Hjort and Claeskens [18]). Focused information criterion and model averaging can be found in the studies by Sueishi [19] and Sun et al. [13]. A more recent development of FIC on quantile regression can be found in the studies by Du et al. [20], Behl et al. [21], and Xu et al. [22]. The motivation for the new method is based on the fact that this concept appears, up to our knowledge, to be virtually unknown and overlooked in Bayesian model selection; thus, there is a need to develop a Bayesian counterpart. The present paper is organized as follows. Section 2 presents the concept of Bayesian model averaging and model selection while Section 3 introduces the new criterion. Section 4 provides practical examples, while Section 5 provides discussions. The paper ends with concluding remarks.

2. Bayesian Model Selection and Model Averaging

2.1. Framework. Consider a situation in which some quantity of interest, μ , is to be estimated from a sample of observations that can be regarded as realizations from some unknown

probability distribution and that, in order to do so, it is necessary to specify a model for the distribution. There are usually many alternative plausible models available and, in general, they each lead to different estimates of μ .

Consider a sample of data, x , and a set of K models $\mathcal{M} = (M_1, \dots, M_K)$, which we will assume to contain the true model M_t . Each M_k consists of a family of distributions $P(x | \eta_k, M_k)$, where η_k represents a parameter (or vector of parameters).

The prior probability that M_k is the true model is denoted by $P(M_k)$ and the prior distribution of the parameters of M_k (given that M_k is true) by $P(\eta_k | M_k)$. Conditioning on the data x and integrating out the parameter η_k , one obtains the following posterior model probabilities:

$$P(M_k | x) = \frac{P(x | M_k) P(M_k)}{\sum_{j=1}^K P(x | M_j) P(M_j)}, \quad (1)$$

where

$$P(x | M_j) = \int P(x | \eta_k, M_k) P(\eta_k | M_k) d\eta_k \quad (2)$$

is the integrated likelihood under M_k . If $P(\eta_k | M_k)$ is a discrete distribution, the integral in (2) is replaced by a sum.

2.2. Bayesian Model Selection. Within this framework, classical Bayesian model selection involves selecting the model with the highest posterior. Sometimes Bayes factors are used. From the Bayes factor framework, the Bayes factor (Kass and Raftery [23]) for M_i versus model M_j is defined by

$$B_{ij} = \frac{P(x | M_i)}{P(x | M_j)}. \quad (3)$$

Model M_i is chosen if $B_{ij} > 1$. Under certain assumptions and approximations (in particular the Laplace approximation) and taking all candidate models as a priori equally likely to be true, this leads to the Bayesian information criterion (BIC), also known as the Schwarz criterion [24]. More information on Bayesian model selection and applications can be found in the studies by Guan and Stephens [25], Nguefack-Tsague [26], Carvalho and Scott [27], Fridley [28], Robert [29], Liang et al. [30], and Bernardo and Smith [31].

2.3. Focused Information Criterion (FIC). As one can notice, in classical Bayesian model selection, a single (selected) model is used to explain all aspects of data, that is, regardless of the purpose of the selection, irrespective of the inference to follow (Claeskens and Hjort [12]). Allen [10] first developed this idea of focusing on a parameter of interest in a prediction problem in which the prediction at a particular value (target) of the regressor vector different from the values in the sample is of interest. Geisser [32] focused on prediction as a major inferential goal, rather than estimation, under many circumstances. In that method, the steps in the derivation of Mallows' C_p are repeated for this target; thus, this criterion depends on that particular value and is therefore an early precursor to the FIC. In the FIC framework, a parameter of

interest, say, μ , must have a definition making it meaningful across competing models. FIC methodology uses general parametric models and maximum likelihood as the estimation method in a general large sample theory. The FIC is derived as the result of establishing an unbiased estimation of the limiting risk of any submodel-based estimator of the parameter μ . FIC is based on the (crucial) assumption that the true data generating mechanism is contained in the largest parametric model considered. The candidate model with the smallest value of FIC is chosen. As one can see, FIC actually is based on frequentist approach; so far, up to our knowledge, Bayesian equivalent has not yet been considered.

2.4. Bayesian Model Averaging. Bayesian model averaging (BMA) is used to deal with the problem of model uncertainty. A discussion on the issue of model uncertainty is given in the study by Clyde and George [33]. Let μ be a quantity of interest depending on x , for example, a future observation from the same process that generated x . The idea is to use a weighted average of the estimates of μ obtained using each of the alternative models, rather than the estimate obtained using any single model. More precisely, the posterior distribution of μ is given by

$$P(\mu | x) = \sum_{k=1}^K P(\mu | x, M_k) P(M_k | x). \quad (4)$$

Note that $P(\mu | x)$ is a weighted average of the posterior distributions $P(\mu | M_k, x)$, $k = 1, \dots, K$, where the k th weight, $P(M_k | x)$, is the posterior probability that M_k is the true model.

The posterior distribution of μ , conditioned on M_k being true, is given by

$$P(\mu | x, M_k) = \int P(\mu | \eta_k, M_k) P(\eta_k | x, M_k) d\eta_k. \quad (5)$$

The posterior mean and posterior variance are given by

$$\begin{aligned} \hat{\mu}_{\text{bma}} &= E_{\text{bma}}(\mu | x) = \sum_{k=1}^K E(\mu | x, M_k) P(M_k | x), \\ \text{Var}_{\text{bma}}(\mu | x) &= \sum_{k=1}^K \left[\text{Var}(\mu | x, M_k) \right. \\ &\quad \left. + (E(\mu | x, M_k) - \hat{\mu}_{\text{bma}})^2 \right] P(M_k | x). \end{aligned} \quad (6)$$

Raftery et al. [34] call this averaging scheme *Bayesian model averaging*. Leamer [35] and Draper [36] advocate the same idea. Madigan and Raftery [37] note that BMA provides better predictive performance than any single model if the measure of performance is Good's [38] logarithm score rule, under the posterior distribution of μ given x .

Hoeting et al. [39] give an extensive framework of BMA methodology and applications for different statistical models. Various real data and simulation studies have investigated the predictive performance of BMA (Clyde [40]; Clyde and George [33]). Nguefack-Tsague [41] uses BMA in the context of estimating a multivariate mean.

2.5. Challenges. Implementing BMA is demanding, especially the computation of the integrated likelihood. Software for BMA implementation, as well as some BMA papers, can be found at “<http://www.research.att.com/~volinsky/bma.html>”. An R [42] package for BMA is now available for computational purposes; this package provides ways for carrying out BMA for linear regression, generalized linear models, and survival analysis using Cox proportional hazard models. For computations, Monte Carlo methods, or approximating methods, are used; thus, many BMA applications are based on the BIC, an asymptotic approximation of the log posterior odds when the prior odds are all equal.

Another problem is the selection of priors for both models and parameters. In most cases, a uniform prior is used for each model; that is, $P(M_k) = 1/K$, $k = 1, 2, \dots, K$. When the number of models is large, model search strategies are sometimes used to reduce the set of models, by eliminating those that seem comparatively less compatible with the data.

3. Using BMA for Model Selection

The purpose of this section is to define the focused Bayesian model averaging (FoBMA).

Definition 1. For a set of K models $\mathcal{M} = (M_1, \dots, M_K)$, M_k with focus parameter μ_k , under BMA framework, the selection criterion FoBMA consists of choosing $M_j \in \mathcal{M}$ for which its estimate $\hat{\mu}_j$ is closest (in terms of squared error) to BMA estimate.

Proposition 2. Under the square error loss and the weighted posterior probability of (4), FoBMA is an optimal model choice.

Proof. Conditioning on all models, that is, under (4), the optimal choice $M_j \in \mathcal{M}$ is the one for which its estimate $\hat{\mu}_j$ minimizes

$$\int (\hat{\mu}_k - \mu)^2 P(\mu | x) d\mu. \quad (7)$$

This is equivalent to minimizing

$$\begin{aligned} & \int (\hat{\mu}_k - \mu)^2 \left\{ \sum_{l=1}^K P(\mu | x, M_l) P(M_l | x) \right\} d\mu \\ &= \sum_{l=1}^K P(M_l | x) \left\{ \int (\hat{\mu}_k - \hat{\mu}_l + \hat{\mu}_l - \mu)^2 P(\mu | x, M_l) d\mu \right\} \\ &= \underbrace{\sum_{l=1}^K P(M_l | x) (\hat{\mu}_k - \hat{\mu}_l)^2}_{(1)} + \underbrace{\sum_{l=1}^K P(M_l | x) \text{Var}(\mu | x, M_l)}_{(2)}. \end{aligned} \quad (8)$$

The term (2) does not depend on $\hat{\mu}_k$ and, denoting

$$\hat{\mu}_{\text{bma}} = \sum_{l=1}^K \hat{\mu}_l P(M_l | x), \quad (9)$$

TABLE 1: Description of variables used for linear regression (Ehrlich [44]).

Label	Description
M	Percentage of males aged 14–24
So	Indicator variable for a Southern state (South = 1 and others = 0)
Ed	Mean years of schooling
$Po1$	Police expenditure in 1960
$Po2$	Police expenditure in 1959
LF	Labour force participation rate
$M.F$	Number of males per 1000 females
Pop	State population
NW	Number of nonwhites per 1000 people
$U1$	Unemployment rate of urban males 14–24
$U2$	Unemployment rate of urban males 35–39
GDP	Gross domestic product per head
$Ineq$	Income inequality
$Prob$	Probability of apprehension and imprisonment
$Time$	Average time served in state prisons (years)
Y	Rate of crimes in a particular category per head of population

the term (1) can be rearranged as (see also the study by Bernardo and Smith [31])

$$\begin{aligned} & \sum_{l=1}^K (\hat{\mu}_k - \hat{\mu}_l)^2 P(M_l | x) \\ &= \sum_{l=1}^K ((\hat{\mu}_{\text{bma}} - \hat{\mu}_l) + (\hat{\mu}_k - \hat{\mu}_{\text{bma}}))^2 P(M_l | x) \\ &= \underbrace{\sum_{l=1}^K (\hat{\mu}_{\text{bma}} - \hat{\mu}_l)^2 P(M_l | x)}_{(1')} + \underbrace{\sum_{l=1}^K (\hat{\mu}_k - \hat{\mu}_{\text{bma}})^2 P(M_l | x)}_{(2')} \\ &\quad + 2 \underbrace{\sum_{l=1}^K (\hat{\mu}_{\text{bma}} - \hat{\mu}_l) (\hat{\mu}_k - \hat{\mu}_{\text{bma}}) P(M_l | x)}_{(3')}. \end{aligned} \quad (10)$$

Since

$$(2') = (\hat{\mu}_k - \hat{\mu}_{\text{bma}})^2 \sum_{l=1}^K P(M_l | x) = (\hat{\mu}_k - \hat{\mu}_{\text{bma}})^2, \quad (11)$$

as well as

$$\begin{aligned} (3') &= (\hat{\mu}_k - \hat{\mu}_{\text{bma}}) \sum_{l=1}^K (\hat{\mu}_{\text{bma}} - \hat{\mu}_l) P(M_l | x) \\ &= (\hat{\mu}_k - \hat{\mu}_{\text{bma}}) \left(\hat{\mu}_{\text{bma}} - \sum_{l=1}^K \hat{\mu}_l P(M_l | x) \right) = 0, \end{aligned} \quad (12)$$

TABLE 2: Linear regression example: selected (best) model according to the focus parameter (FoBMA) compared to its BIC rank and value, the probability that the focus is not null ($100P(\beta \neq 0)$), its BMA mean ($\hat{\mu}_{\text{bma}} = E_{\text{bma}}(\mu | x)$), and standard deviation ($SD_{\text{bma}}(\mu | x)$).

Focus	Best model (FoBMA)	-BIC	Rank	$\beta \neq 0$	Mean	SD
<i>M</i>	<i>M, Ed, Po1, Pop, NW, 1, U2, GDP, neq, Prob</i>	51.6	34	97.5	1.4	0.53
<i>So</i>	<i>M, So, Ed, Po1, NW, U2, Ineq, Prob</i>	52.1	30	6.3	0.007	0.04
<i>Ed</i>	<i>M, Ed, Po1, LF, NW, U2, Ineq, Prob, Time</i>	52.2	29	100	2.13	0.51
<i>Po1</i>	<i>M, Ed, Po1, NW, U1, U2, GDP, Ineq, Prob, Time</i>	52.5	23	75.4	0.67	0.42
<i>Po2</i>	<i>M, Ed, Po2, NW, U1, U2, GDP, Ineq, Prob, Time</i>	51.2	45	24.6	0.22	0.40
<i>LF</i>	<i>M, Ed, Po1, LF, NW, U2, GDP, Ineq, Prob, Time</i>	50.9	46	2.1	0.003	0.08
<i>M.F</i>	<i>M, Ed, Po1, M.F, NW, U2, Ineq, Prob, Time</i>	52.3	27	5.2	-0.07	0.45
<i>Pop</i>	<i>M, Ed, Po1, Pop, NW, U2, Ineq, Prob, Time</i>	52.9	18	28.9	-0.02	0.04
<i>NW</i>	<i>M, Ed, Po1, Pop, NW, U1, U2, Ineq, Prob</i>	52.5	22	89.6	0.09	0.05
<i>U1</i>	<i>M, Ed, Po1, NW, U1, U2, Ineq, Prob</i>	52.3	28	11.9	-0.04	0.14
<i>U2</i>	<i>M, Ed, Po1, NW, U2, GDP, Ineq, Prob, Time</i>	54.7	3	83.9	0.27	0.19
<i>GDP</i>	<i>M, Ed, Po1, U2, GDP, Ineq, Prob</i>	51.4	37	33.8	0.20	0.35
<i>Ineq</i>	<i>M, Ed, Po2, Pop, NW, U2, Ineq, Prob</i>	53.1	14	100.0	1.38	0.33
<i>Prob</i>	<i>Ed, Po1, Pop, NW, Ineq, Prob</i>	52.4	25	98.8	-0.25	0.10
<i>Time</i>	<i>M, Ed, Po1, Pop, NW, U2, Ineq, Prob, Time</i>	52.9	18	44.5	-0.13	0.18

and since $(1')$ does not depend on $\hat{\mu}_k$, the only term that depends on $\hat{\mu}_k$ is $(2')$. Hence, the preferred M_j is the one whose estimate $\hat{\mu}_j$ is closest to the BMA estimate $\hat{\mu}_{\text{bma}}$. $\square$

In particular, for two models, Corollary 3 shows that FoBMA reduces to the classical Bayesian model selection scheme of choosing the model with the highest posterior probability.

Corollary 3. *For two models, the selected model is the one with the highest posterior probability.*

Proof. From Proposition 2, let us find the distance between each model and BMA model. Consider

$$\begin{aligned}
 (\hat{\mu}_1 - \hat{\mu}_{\text{bma}})^2 &= (\hat{\mu}_1 - P(M_1 | x) \hat{\mu}_1 - P(M_2 | x) \hat{\mu}_2)^2 \\
 &= ((1 - P(M_1 | x)) \hat{\mu}_1 - P(M_2 | x) \hat{\mu}_2)^2 \\
 &= (P(M_2 | x) \hat{\mu}_1 - P(M_2 | x) \hat{\mu}_2)^2 \\
 &= P^2(M_2 | x) (\hat{\mu}_1 - \hat{\mu}_2)^2.
 \end{aligned} \tag{13}$$

Similarly,

$$(\hat{\mu}_2 - \hat{\mu}_{\text{bma}})^2 = P^2(M_1 | x) (\hat{\mu}_2 - \hat{\mu}_1)^2. \tag{14}$$

Therefore, M_1 is selected if

$$P^2(M_2 | x) (\hat{\mu}_1 - \hat{\mu}_2)^2 < P^2(M_1 | x) (\hat{\mu}_2 - \hat{\mu}_1)^2, \tag{15}$$

that is, if $P(M_1 | x) > P(M_2 | x)$. Hence, FoBMA is equivalent to selecting the model with the highest posterior probability. $\square$

4. Applications

In this section, we apply the methodology to three models: linear regression, logistic regression, and survival analysis.

TABLE 3: Description of variables used for logistic regression (Hosmer and Lemeshow [46]).

Label	Description
<i>LBWT</i>	Indicator of birth weight (BWT) less than 2.5 kg
<i>AGE</i>	Mother's age in years
<i>LWT</i>	Mother's weight in pounds at last menstrual period
<i>RACE</i>	Mother's race (1 = white, 2 = black, and 3 = other)
<i>SMOKE</i>	Smoking status during pregnancy
<i>PTL</i>	Number of previous premature labours
<i>HT</i>	History of hypertension
<i>UI</i>	Presence of uterine irritability
<i>FTV</i>	Number of physician visits during the first trimester
<i>BWT</i>	Birth weight in grams

The three following examples have been widely used in Bayesian model averaging (Raftery et al. [43]) and are available in the R packages *MASS* and *survival*. They are also used as tutorial in the R package *BMA*. Since these are parametric models, the focus parameter, μ , is in every case the regression coefficient (β). FoBMA is compared to the classical well-known Bayesian information criterion (BIC). All computations were performed with R [42].

4.1. Linear Regression. In this subsection, FoBMA is applied to data of the effect of punishment regimes on crime rates (Ehrlich [44]), used in the study by Raftery [45]. It can be downloaded in the R package *MASS*.

4.1.1. Data Description. Criminologists are interested in the effect of punishment regimes on crime rates. This is a dataset

TABLE 4: Logistic regression example: selected (best) model according to the focus parameter (FoBMA) compared to its BIC rank and value, the probability that the focus is not null ($100P(\beta \neq 0)$), its BMA mean ($\hat{\mu}_{\text{bma}} = E_{\text{bma}}(\mu | x)$), and standard deviation ($SD_{\text{bma}}(\mu | x)$).

Focus	Best model (FoBMA)	-BIC	Rank	$\beta \neq 0$	Mean	SD
AGE	AGE, LWT, HT	749.8	20	12.6	0.008	0.024
LWT	LWT, SMOKE, UI	748.6	38	68.7	-0.012	0.010
RACE2	RACE, SMOKE, HT	748.1	46	9.6	0.115	0.390
RACE3	LWT, RACE, HT	748.2	44	9.6	0.092	0.320
SMOKE	LWT, SMOKE, PTL, HT, UI	748.5	40	33.2	0.256	0.43
PTL1	LWT, SMOKE, PTL, HT, UI	748.5	40	35.1	0.62	0.089
PTL2	LWT, SMOKE, PTL, HT, UI	748.5	40	35.1	0.170	0.61
PTL3	AGE, PTL, HT	749.3	29	35.1	-4.91	520
HT	RACE, SMOKE, HT	748.1	46	65.4	1.167	1
UI	LWT, PTL, UI	748.5	39	33.8	0.310	0.51
FTV	LWT, HT, FTV	748.7	33	1.5	-0.001	0.024

on per capita crime rates in 47 U.S. states in 1960 originally published in the study by Ehrlich [44]. The variables have been rescaled to convenient numbers.

Table 1 describes all variables used for linear regression. The dependent variable is the rate of crimes in a particular category per head, and 15 potential independent variables perceived to be associated with crime rates (Ehrlich [44]).

4.1.2. Results. Table 2 shows that the classical BIC (no focus) selects the model with variables *M, Ed, Po1, LF, NW, U2, Ineq, Prob, Time* (-BIC = 55.91). Since the number of explanatory variables in this case is 15, the model space is small enough, which allows for full enumeration of model space. If the focus was the *probability of imprisonment* (the number of offenders imprisoned per offense known), the selected model is *Ed, Po1, Pop, NW, Ineq, Prob* which is ranked 25 with BIC. If the focus was the *average time spent in state prisons*, the selected model is *M, Ed, Po1, Pop, NW, U2, Ineq, Prob, Time* which is ranked 18 with BIC. The other focuses parameters show that there is a great discrepancy between the FoBMA and the classical BIC.

4.2. Logistic Regression. In this subsection, FoBMA is applied to data of risk factors associated with low infants birth weights (Hosmer and Lemeshow [46]). These data were also used in the study by Raftery [45] and can be downloaded in the R package MASS. The aim was to study the risk factors associated with low infants birth weights.

4.2.1. Data Description. The “birthwt” data frame has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Massachusetts, during 1986.

4.2.2. Results. Table 3 describes all variables used for logistic regression. The outcome variable is low (indicator of birth weight (BWT) less than 2.5 kg), and 7 potential independent variables perceived to be associated with low birth weight (Hosmer and Lemeshow [46]).

TABLE 5: Description of variables used for logistic regression (Hosmer and Lemeshow [46]).

Label	Description
<i>Age</i>	Age of patients in years
<i>Albumin</i>	Serum albumin (mg/dL)
<i>Alk.phos</i>	Alkaline phosphatase (U/liter)
<i>Ascites</i>	Presence of ascites
<i>Ast</i>	Aspartate aminotransferase, once called SGOT (U/mL)
<i>Bili</i>	Serum bilirubin (mg/dL)
<i>Chol</i>	Serum cholesterol (mg/dL)
<i>Copper</i>	Urine copper (ug/day)
<i>Edema</i>	0 no edema, 0.5 untreated or successfully treated edema, and 1 edema despite diuretic therapy
<i>Hepato</i>	Presence of hepatomegaly or enlarged liver
<i>Platelet</i>	Platelet count
<i>Protime</i>	Standardized blood clotting time
<i>Sex</i>	Male/female
<i>Spiders</i>	Blood vessel malformations in the skin
<i>Stage</i>	Histologic stage of disease (which needs biopsy)
<i>Status</i>	Status at endpoint, 0/1/2 for censored, transplant, or dead
<i>Time</i>	Number of days between the date of registration and the earlier date of death or transplantation
<i>Trt</i>	1/2/NA for D-penicillamine, placebo, not randomized
<i>Trig</i>	Triglycerides (mg/dL)

Table 4 shows that BIC selects the one with variables *LWT, HT* (-BIC = 753). If the focus was the *LWT*, the selected model is *LWT, SMOKE, UI* which is ranked 38 with BIC. If the focus was the *UI*, the selected model is *LWT, PTL, UI* which is ranked 39 with BIC. The other focuses parameters show that there is a great discrepancy between the FoBMA and BIC. If hypertension (HT) is the focus, the selected model is the one with variables *RACE, SMOKE, HT*; but this model is ranked 20 if the focus is *SMOKE*. The model

TABLE 6: Survival regression example: selected (best) model according to the focus parameter (FoBMA) compared to its BIC rank and value, the probability that the focus is not null ($100P(\beta \neq 0)$), its BMA mean ($\hat{\mu}_{\text{bma}} = E_{\text{bma}}(\mu | x)$), and standard deviation ($SD_{\text{bma}}(\mu | x)$).

Focus	Best model (FoBMA)	-BIC	Rank	$\beta \neq 0$	Mean	SD
<i>Age</i>	<i>Age, Bili, Edem, Prot, Copp, Albu</i>	174.8	1	100	0.031	0.009
<i>AlkP</i>	<i>Age, AlkP, Bili, Edem, Stag, Copp, Albu</i>	168.9	34	7.8	0.000	0.000
<i>Asci</i>	<i>Age, Asci, Bili, Edem, Prot, Copp, Albu</i>	168.9	24	1.7	0.003	0.045
<i>Bili</i>	<i>Age, Bili, Prot, Stag, Copp, Albu</i>	170.8	14	100	0.78	0.13
<i>Edem</i>	<i>Age, Bili, Edem, Plat, Prot, Copp, Albu</i>	170.8	15	84.7	0.74	0.43
<i>Hepa</i>	<i>Age, Bili, Edem, Hepa, Prot, Copp, Albu</i>	170.7	16	2.8	0.006	0.05
<i>Plat</i>	<i>Age, Bili, Edem, Plat, Prot, Stag, Copp, Albu</i>	168.8	39	5.4	0.000	0.000
<i>Prot</i>	<i>Age, Bili, Edem, Prot, Stag, Albu</i>	171.0	13	78	2.5	1.7
<i>Sex</i>	<i>Age, Bili, Edem, Prot, Sex, Copp, Albu</i>	170.4	18	5.6	-0.018	0.1
<i>Ast</i>	<i>Age, Bili, Prot, Ast, Copp, Albu</i>	169.6	28	20.9	0.097	0.23
<i>Spid</i>	<i>Age, Bili, Edem, Prot, Spid, Copp, Albu</i>	170.0	27	1.4	0.000	0.026
<i>Stag</i>	<i>Age, AlkP, Bili, Edem, Prot, Stag, Copp, Albu</i>	169.5	29	36.6	0.1	0.16
<i>Trt</i>	<i>Age, Bili, Edem, Prot, Trt, Copp, Albu</i>	170.2	23	1.6	0.002	0.027
<i>Copp</i>	<i>Age, Bili, Edem, Prot, Ast, Stag, Copp, Albu</i>	171.3	11	73.8	0.26	0.2
<i>Albu</i>	<i>Age, Bili, Edem, Prot, Ast, Copp, Albu</i>	172.0	7	100	-2.8	8

AGE, LWT, PTL, HT, UI is ranked 4 if the focus is *AGE*, 11 if *LWT*, 14 if *PTL*, 23 if *HT*, and 5 if *UI*. It is not best (number 1) to estimate any of these focuses. Other applications of FIC for logistic regression in dental restoration can be found in the study by Candolo [47].

4.3. Survival Analysis. In this subsection, FoBMA is applied to the Mayo Clinic Primary Biliary Cirrhosis Data (Therneau and Grambsch [48]). It was used in the study by Volinsky et al. [49] and can be downloaded in the R package *survival*.

4.3.1. Data Description. The data are from the Mayo Clinic trial on primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. The first 312 cases in the dataset participated in the randomized trial and contain largely complete data. The additional 112 cases did not participate in the clinical trial but consented to have basic measurements recorded and to be followed for survival. Six of those cases were lost to follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants.

4.3.2. Results. Table 5 describes all variables used for survival analysis. The dependent variable was the survival time while the 18 others were the independent variables.

Table 6 shows that BIC selects the model with variables *Age, Bili, Edem, Prot, Copp, Albu*. If the focus was the *Bili*, the selected model is *Age, Bili, Prot, Stag, Copp, Albu* which is ranked 14 with BIC. If the focus was the *Albu*, the selected model is *Age, Bili, Edem, Prot, Ast, Copp, Albu* which is ranked 7 with BIC. The other focuses parameters show that there is a great discrepancy between the FoBMA

and BIC. If the focus was on age, BIC and FoBMA select the same model.

5. Discussion

5.1. Comparison to Other Bayesian Approaches. As with any other Bayesian approach, parameter priors and model priors are a source of uncertainty in FoBMA. FoBMA shares the same drawbacks of BMA, namely, the computation of integrated likelihoods of various models and the choice of model space. As model selection criterion, compared to Bayesian likes, it is simpler after BMA estimates have been obtained. It was shown in Corollary 3 that FoBMA is equivalent to the classical approach of selecting the model with the highest posterior model probability if there are only two competing models.

5.2. Interpretation of Focus. It was shown in Paragraph 4 that, in general, FoBMA is found to select a model different from the one selected without a focus. Hereby, the difference in terms of ranking is sometimes large. For example, in studying the risk factors associated with low infants birth weight, the classical Bayesian model selection approach selects the one with variables *LWT, HT*. However, if the focus was the *LWT* (mother's weight in pounds at last menstrual period), that is, if the main objective was to find out whether *LWT* was a risk factor, adjusted for possible confounding factors, the selected model is *LWT, SMOKE, UI* which is ranked 38 with BIC. Thus, a poor model for estimating a focus μ_1 may be the best for μ_2 .

5.3. Heuristic Explanation of Focus. Consider a set of data $x_1, \dots, x_n$, each x_i with PDF $f(x_i | \theta)$. Mean, median, and mode measure central tendencies. Mean is obtained by maximizing the square error loss, the median by maximizing

the absolute risk, and the mode by maximizing 0-1 loss. In this context of model selection, the square error loss was used; then it is reasonable to find out the “mean,” which can be identified as BMA estimate. Since BMA estimate is not any of the competing models, it seems appropriate to select the closest one to it.

6. Concluding Remarks

The present paper has derived a new model selection criterion (FoBMA), which, in contrast to other Bayesian model selection criteria, focuses on the parameter singled out for interest. The methodology was applied to concrete examples. The method needs to be applied in a variety of conditions; in particular, more works need to be done to find the asymptotic properties of FoBMA. It is expected that this paper brings motivations for more researches on focus-related criteria in Bayesian model selection.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank three anonymous referees for helpful comments that led to improvement on earlier versions of this paper.

References

- [1] H. Leeb and B. M. Poetscher, “Model selection,” in *Handbook of Financial Time Series*, T. G. Anderson, R. A. Davis, J.-P. Kreiss, and T. Mikosch, Eds., pp. 785–821, Springer, New York, NY, USA, 2008.
- [2] W. Zucchini, “An introduction to model selection,” *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 41–61, 2000.
- [3] G. Nguefack-Tsague, “An alternative derivation of some common distributions functions: a post-model selection approach,” *International Journal of Applied Mathematics and Statistics*, vol. 42, no. 12, pp. 138–147, 2013.
- [4] G. Nguefack-Tsague and W. Zucchini, “Post-model selection inference and model averaging,” *Pakistan Journal of Statistics and Operation Research*, vol. 7, no. 2, pp. 347–361, 2011.
- [5] W. Zucchini, G. Claeskens, and G. Nguefack-Tsague, “Model selection,” in *International Encyclopedia of Statistical Sciences*, M. Lovric, Ed., chapter part 13, pp. 830–833, Springer, Berlin, Germany, 2011.
- [6] P. Behl, H. Dette, M. Frondel, and H. Tauchmann, “Energy substitution: when model selection depends on the focus,” *Energy Economics*, vol. 39, pp. 233–238, 2013.
- [7] G. Nguefack-Tsague, “On bootstrap and post-model selection inference,” *International Journal of Mathematics and Computation*, vol. 21, no. 4, pp. 51–64, 2013.
- [8] G. Nguefack-Tsague, “Estimation of a multivariate mean under model selection uncertainty,” *Pakistan Journal of Statistics and Operation Research*, vol. 10, no. 2, pp. 131–145, 2014.
- [9] G. Nguefack-Tsague, “On optimal weighting scheme in model averaging,” *American Journal of Applied Mathematics and Statistics*, vol. 2, no. 3, pp. 150–156, 2014.
- [10] D. M. Allen, “Mean square error of prediction as a criterion for selecting variables,” *Technometrics*, vol. 13, no. 3, pp. 469–475, 1971.
- [11] C. L. Mallows, “Some comments on C_p ,” *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.
- [12] G. Claeskens and N. L. Hjort, “The focused information criterion,” *Journal of the American Statistical Association*, vol. 98, no. 464, pp. 900–916, 2003.
- [13] Z. Sun, Z. Su, and J. Ma, “Focused vector information criterion model selection and model averaging regression with missing response,” *Metrika*, vol. 77, no. 3, pp. 415–432, 2014.
- [14] P. Behl, H. Dette, M. Frondel, and H. Tauchmann, “Choice is suffering: a focused information criterion for model selection,” *Economic Modelling*, vol. 29, no. 3, pp. 817–822, 2012.
- [15] X. Zhang, A. T. Wan, and S. Z. Zhou, “Focused information criteria, model selection, and model averaging in a Tobit model with a nonzero threshold,” *Journal of Business and Economic Statistics*, vol. 30, no. 1, pp. 132–142, 2012.
- [16] X. Zhang and H. Liang, “Focused information criterion and model averaging for generalized additive partial linear models,” *Annals of Statistics*, vol. 39, no. 1, pp. 174–200, 2011.
- [17] C. T. Brownlees and G. M. Gallo, “On variable selection for volatility forecasting: the role of focused selection criteria,” *Journal of Financial Econometrics*, vol. 6, no. 4, pp. 513–539, 2008.
- [18] N. L. Hjort and G. Claeskens, “Focused information criteria and model averaging for the Cox’s hazard regression model,” *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1449–1464, 2006.
- [19] N. Sueishi, “Generalized empirical likelihood-based focused information criterion and model averaging,” *Econometrics*, vol. 1, no. 2, pp. 141–156, 2013.
- [20] J. Du, Z. Zhang, and T. Xie, “Focused information criterion and model averaging in quantile regression,” *Communications in Statistics: Theory and Methods*, vol. 42, no. 20, pp. 3716–3734, 2013.
- [21] P. Behl, G. Claeskens, and H. Dette, “Focused model selection in quantile regression,” *Statistica Sinica*, vol. 24, pp. 601–624, 2014.
- [22] G. Xu, S. Wang, and J. Z. Huang, “Focused information criterion and model averaging based on weighted composite quantile regression,” *Scandinavian Journal of Statistics*, vol. 41, no. 2, pp. 365–381, 2014.
- [23] R. E. Kass and A. E. Raftery, “Bayes factors,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.
- [24] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [25] Y. Guan and M. Stephens, “Bayesian variable selection regression for genome-wide association studies and other large-scale problems,” *The Annals of Applied Statistics*, vol. 5, no. 3, pp. 1780–1815, 2011.
- [26] G. Nguefack-Tsague, “Using bayesian networks to model hierarchical relationships in epidemiological studies,” *Epidemiology and Health*, vol. 33, no. 1, Article ID e2011006, 2011.
- [27] C. M. Carvalho and J. G. Scott, “Objective Bayesian model selection in Gaussian graphical models,” *Biometrika*, vol. 96, no. 3, pp. 497–512, 2009.

- [28] B. L. Fridley, "Bayesian variable and model selection methods for genetic association studies," *Genetic Epidemiology Series B*, vol. 33, no. 1, pp. 27–37, 2009.
- [29] C. P. Robert, *The Bayesian choice*, Springer, New York, NY, USA, 2001.
- [30] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *Journal of the American Statistical Association*, vol. 103, no. 481, pp. 410–423, 2008.
- [31] J.-M. Bernardo and A. F. M. Smith, *Bayesian Theory*, John Wiley & Sons, 1994.
- [32] S. Geisser, *Predictive Inference: An Introduction*, Prentice Hall, New York, NY, USA, 1993.
- [33] M. A. Clyde and E. I. George, "Model uncertainty," *Statistical Science*, vol. 19, no. 1, pp. 81–94, 2004.
- [34] A. E. Raftery, D. Madigan, and J. A. Hoeting, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 179–191, 1997.
- [35] E. E. Leamer, *Specification Searches: Ad Hoc Inference with Experimental Data*, John Wiley & Sons, New York, NY, USA, 1978.
- [36] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society Series B*, vol. 57, no. 1, pp. 45–97, 1995.
- [37] D. Madigan and A. E. Raftery, "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, vol. 89, pp. 1535–1546, 1994.
- [38] I. J. Good, "Rational decisions," *Journal of the Royal Statistical Society Series B: Methodological*, vol. 14, pp. 107–114, 1952.
- [39] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–417, 1999.
- [40] M. A. Clyde, "Bayesian model averaging and model search strategies," in *Bayesian Statistics VI*, pp. 157–185, Oxford University Press, Oxford, UK, 1999.
- [41] G. Nguefack-Tsague, "Bayesian estimation of a multivariate mean under model uncertainty," *International Journal of Mathematics and Statistics*, vol. 13, no. 1, pp. 83–92, 2013.
- [42] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [43] A. E. Raftery, I. S. Painter, and C. T. Volinsky, "BMA: an R package for Bayesian model averaging," *Rnews*, vol. 5, no. 2, pp. 2–8, 2005.
- [44] I. Ehrlich, "Participation in illegitimate activities: a theoretical and empirical investigation," *Journal of Political Economy B*, vol. 81, pp. 521–565, 1973.
- [45] A. E. Raftery, "Approximate Bayes factors and accounting for model uncertainty in generalised linear models," *Biometrika*, vol. 83, no. 2, pp. 251–266, 1996.
- [46] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley, New York, NY, USA, 1989.
- [47] C. Candolo, "The focused information criterion in logistic regression to predict repair of dental restorations," *Revista Brasileira de Biometria*, vol. 31, no. 4, pp. 547–557, 2014.
- [48] T. Therneau and P. Grambsch, *Modeling Survival Data: Extending the Cox Model*, University Press, Cambridge, Mass, USA, 2000.
- [49] C. T. Volinsky, D. Madigan, A. E. Raftery, and R. A. Kronmal, "Bayesian model averaging in proportional hazard models: assessing the risk of a stroke," *Journal of the Royal Statistical Society C: Applied Statistics*, vol. 46, no. 4, pp. 433–448, 1997.