The Annals of Statistics
2014, Vol. 42, No. 6, 2441–2468
DOI: 10.1214/14-AOS1258

ASYMPTOTIC THEORY OF GENERALIZED INFORMATION CRITERION FOR GEOSTATISTICAL REGRESSION MODEL SELECTION

BY CHIH-HAO CHANG, HSIN-CHENG HUANG[1] AND CHING-KANG ING[2]

National University of Kaohsiung, Academia Sinica and Academia Sinica

Information criteria, such as Akaike's information criterion and Bayesian information criterion are often applied in model selection. However, their asymptotic behaviors for selecting geostatistical regression models have not been well studied, particularly under the fixed domain asymptotic framework with more and more data observed in a bounded fixed region. In this article, we study the generalized information criterion (GIC) for selecting geostatistical regression models under a more general mixed domain asymptotic framework. Via uniform convergence developments of some statistics, we establish the selection consistency and the asymptotic loss efficiency of GIC under some regularity conditions, regardless of whether the covariance model is correctly or wrongly specified. We further provide specific examples with different types of explanatory variables that satisfy the conditions. For example, in some situations, GIC is selection consistent, even when some spatial covariance parameters cannot be estimated consistently. On the other hand, GIC fails to select the true polynomial order consistently under the fixed domain asymptotic framework. Moreover, the growth rate of the domain and the degree of smoothness of candidate regressors in space are shown to play key roles for model selection.

1. Introduction. With the advent of data collection technologies, more and more data, such as remote sensing data or environmental monitoring data, are collected in space and managed by geographical information systems. In many applications, a response of interest is observed on a set of sites in space, and it is of interest to apply a geostatistical regression model to predict the response at unsampled sites with the aid of auxiliary/explanatory variables. For example, in precision agriculture, it is of interest to predict crop yield based on some explanatory variables involving, for example, climatic conditions, soil types, fertilizers, cropping practices, weeds and topographic features. Not only do we aim to identify the important explanatory variables, but the precision of yield also depends on how well

Received March 2014; revised July 2014.

[1]Supported by National Science Council of Taiwan under Grant 100-2628-M-001-004-MY3.

[2]Supported by National Science Council of Taiwan under Grant 99-2118-M-001-008-MY2 and the Academia Sinica Investigator Award.

MSC2010 subject classifications. Primary 63M30; secondary 62F07, 62F12.

Key words and phrases. Akaike's information criterion, Bayesian information criterion, fixed domain asymptotic, selection consistency, increasing domain asymptotic, variable selection.

the explanatory variables are chosen, which if not chosen properly, may result in poor performance, particularly when the number of explanatory variables is large. Clearly, model selection is essential in geostatistics.

There are two different asymptotic frameworks in geostatistics. One is called the increasing domain asymptotic framework, where the observation region grows with the sample size. The other is called the fixed domain asymptotic (or infill asymptotic) framework, where the observation region is bounded and fixed with more and more data taken more densely in the region. It is known that the two frameworks lead to possibly different asymptotic behaviors in covariance parameter estimation. However, little is known about their effects on model selection. In general, asymptotic behaviors of the estimated parameters under the increasing domain framework are more standard. For example, the maximum likelihood estimates of covariance parameters are typically consistent and asymptotically normal when fitted by a correct model [Mardia and Marshall (1984)]. In contrast, not all covariance parameters can be estimated consistently under the fixed domain asymptotic framework, even for the simple exponential covariance model in one dimension with no consideration of explanatory variables [Ying (1991); Chen, Simpson and Ying (2000)]. The readers are refereed to Stein (1999) for more details regarding fixed domain asymptotics. Some discussion concerning which asymptotic framework is more appropriate can also be found in Zhang and Zimmerman (2005).

Many model selection methods have been applied in geostatistical regression, such as Akaike's information criterion [AIC, Akaike (1973)], Bayesian information criterion [BIC, Schwartz (1978)], the generalized information criterion [GIC, Nishii (1984)] and the cross validation method [Stone (1974)]. Note that GIC contains a range of criteria, including both AIC and BIC, governed by a tuning parameter. Although theoretical properties of these selection methods have been thoroughly established in linear regression and time series model selection [e.g., Shao (1997), McQuarrie and Tsai (1998), Ing and Wei (2005), Ing (2007)], only limited results are available for selecting geostatistical regression models. For example, Hoeting et al. (2006) provided some heuristic arguments for AIC in geostatistical model selection when the spatial process of interest is observed with no measurement error. They show via a simulation study that spatial dependence has to be considered, which if ignored, may result in unsatisfactory results. Huang and Chen (2007) developed a technique of estimating the mean squared prediction error for a general spatial prediction procedure using the generalized degrees of freedom and derived some asymptotic efficiency results. For linear mixed models, Jiang and Rao (2003) developed some consistent procedures similar to GIC. Pu and Niu (2006) derived conditions under which GIC is selection consistent. Jiang et al. (2008) introduced a fence method for mixed model selection and showed its consistency under some regularity conditions. Jones (2011) proposed a modified BIC, which replaces the sample size in the penalty of the original BIC by an effective sample size to account for correlations in linear mixed models. Vaida and Blanchard (2005) proposed the conditional Akaike's information criterion (CAIC) and argued that it is more appropriate than AIC when the focus is on subjects/clusters requiring prediction of random effects. In addition, selection among semiparametric regression models and penalized smoothing spline models [e.g., Chapter 4, Ruppert, Wand and Carroll (2003)] can also be formulated in terms of random-effect selection in linear mixed models. The asymptotic theory of AIC for this type of model was given by Shi and Tsai (1999), and that for BIC was given by Bunea (2004). A recent review of linear and generalized linear mixed model selection can also be found in Müller, Scealy and Welsh (2013).

Although the geostatistical regression model can be regarded as a linear mixed model with one random effect, its asymptotic behavior is surprisingly subtler than a usual linear mixed model for the following three reasons. First, variables in a geostatistical regression model are sampled from a spatial process, resulting in small "effective sample size" unless the spatial domain is allowed to grow quickly. Second, unlike some random-effect models with independent random components, spatial dependence forces all variables to depend in a complex way, making it very difficult to handle asymptotically. Third, under the fixed domain asymptotic framework, classical regularity conditions are generally not satisfied, and traditional approaches for establishing asymptotic results are typically not applicable. To the best of our knowledge, asymptotic properties of GIC for geostatistical regression model selection have yet to be developed, particularly under the fixed domain asymptotic framework, where nonstandard behaviors are often expected. In this article, we focus on GIC for geostatistical regression model selection regardless of whether the covariance model is correctly or wrongly specified. Although a conditional-type criterion, such as CAIC may be more appropriate when spatial prediction is of main interest, it is beyond the scope of this paper. Major accomplishments are listed in the following:

(1) We establish a general theory of GIC for the selection consistency and the asymptotic loss efficiency under mild regularity conditions in a general mixed domain asymptotic framework, which includes both the fixed and increasing domain asymptotics. In particular, we allow the possibilities that some covariance parameters may converge to a nondegenerate distribution and the covariance model may be mis-specified.

(2) We provide some examples that satisfy the aforementioned regularity conditions under exponential covariance models in one and two dimensions, and demonstrate how selection consistency is affected by candidate regressors.

We shall show that the asymptotic behaviors of GIC are related to how fast the domain grows with the sample size. In addition, some nonstandard properties of GIC under the fixed domain asymptotic framework will be highlighted. For example, under fixed domain asymptotics, GIC fails to identify the correct order of polynomial consistently regardless of the tuning parameter value, even when the underlying covariance model is correctly specified. On the other hand, for a properly chosen tuning parameter value, GIC is selection consistent when candidate explanatory variables are generated from some spatial dependent processes.

This article is organized as follows. Section 2 gives a brief introduction of geostatistical regression models and GIC. Our main results for the consistency and the asymptotic loss efficiency of GIC are presented in Sections 3 and 4. Specifically, in Section 3, we assume that the covariance model is specified correctly. While in Section 4, we consider the covariance model to be mis-specified. In Section 5, we provide some examples that satisfy the regularity conditions. Finally, a brief discussion is provided in Section 6.

2. Models and criteria.

2.1. *Geostatistical regression models.* Consider a spatial process, $\{S(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$. Suppose that we observe data $\{Z(\mathbf{s}_{n1}), \dots, Z(\mathbf{s}_{nn})\}$ according to the following measurement equation:

$$\begin{aligned} Z(\mathbf{s}_{ni}) &= S(\mathbf{s}_{ni}) + \epsilon(\mathbf{s}_{ni}) \\ &= \mu_0(\mathbf{s}_{ni}) + \eta(\mathbf{s}_{ni}) + \epsilon(\mathbf{s}_{ni}); \qquad i = 1, \dots, n, \end{aligned} \tag{2.1}$$

where $\mu_0(\cdot)$ is the mean function, $\eta(\cdot)$ is a zero-mean Gaussian spatial dependent process with $\sup_{\mathbf{s} \in D} \mathrm{E}(\eta^2(\mathbf{s})) < \infty$ and $\{\epsilon(\mathbf{s}_{ni}) : i = 1, \dots, n\}$ are Gaussian white-noise variables with variance v^2, independent of $S(\cdot) = \mu_0(\cdot) + \eta(\cdot)$, corresponding to measurement errors.

In addition to $Z(\mathbf{s}_{ni})$'s, we observe $\mathbf{x}(\mathbf{s}_{ni}) = (1, x_1(\mathbf{s}_{ni}), \dots, x_{p_n}(\mathbf{s}_{ni}))'$, a $(p_n + 1)$-vector of explanatory variables, for $i = 1, \dots, n$. We consider the geostatistical regression model

$$Z(\mathbf{s}_{ni}) = \mathbf{x}(\mathbf{s}_{ni})' \boldsymbol{\beta}_n + \eta(\mathbf{s}_{ni}) + \epsilon(\mathbf{s}_{ni}); \qquad \mathbf{s}_{ni} \in D, i = 1, \dots, n,$$

where $\boldsymbol{\beta}_n = (\beta_0, \beta_1, \dots, \beta_{p_n})'$. This model reduces to the usual linear regression model when $\eta(\cdot)$ is absent. Similarly to linear regression, a large model that contains many insignificant variables may produce a large variance, resulting in low predictive power. On the other hand, a model that ignores some important variables may suffer from a large bias. To strike a good balance between (squared) bias and variance, it is essential to include only significant variables in the model. Clearly, variable selection is essential not only in regression but also in geostatistical regression.

We use $\alpha \subseteq \{1, \dots, p_n\}$ to denote a model, which consists of the indices of the corresponding explanatory variables. Let $\mathcal{A}_n \subseteq 2^{\{1, \dots, p_n\}}$ be the set of all candidate models with $\varnothing$ being the intercept-only model. Let $\mathbf{X}_n$ be the $n \times (p_n + 1)$ matrix with the ith row, $\mathbf{x}(\mathbf{s}_{ni})'$; $1 \le i \le n$. Also let $\mathbf{X}_n(\alpha)$ be an $n \times (p(\alpha) + 1)$ sub-matrix of $\mathbf{X}_n$ containing a column $\mathbf{1}$ (corresponding to the intercept) and the columns corresponding to $\alpha \in \mathcal{A}_n$, and $\boldsymbol{\beta}_n(\alpha)$ be the sub-vector of $\boldsymbol{\beta}_n$ corresponding to

$\mathbf{X}_n(\alpha)$. A model α is said to be correct if $\mu_0(\mathbf{s})$ can be written as $\beta_0+\sum_{j\in\alpha}\beta_j x_j(\mathbf{s})$ for all $\mathbf{s}\in D$. If there exists a correct model, we denote the correct model having the smallest number of variables by $\alpha_n^0=\arg\min_{\alpha\in\mathcal{A}_n^0} p(\alpha)$, where $\mathcal{A}_n^0$ is the set of all correct models.

The geostatistical regression model α can be written in a matrix form as

$$\mathbf{Z}_n=\big(Z(\mathbf{s}_{n1}),\dots,Z(\mathbf{s}_{nn})\big)'=\mathbf{X}_n(\alpha)\boldsymbol{\beta}_n(\alpha)+\boldsymbol{\eta}_n+\boldsymbol{\epsilon}_n;\qquad \alpha\in\mathcal{A}_n, \tag{2.2}$$

where $\boldsymbol{\eta}_n=(\eta(\mathbf{s}_{n1}),\dots,\eta(\mathbf{s}_{nn}))'\sim N(0,\boldsymbol{\Sigma}_{n\eta})$ and $\boldsymbol{\epsilon}_n=(\epsilon(\mathbf{s}_{n1}),\dots,\epsilon(\mathbf{s}_{nn}))'\sim N(0,v^2\mathbf{I}_n)$ with $\boldsymbol{\Sigma}_{n\eta}=\mathrm{E}(\boldsymbol{\eta}_n\boldsymbol{\eta}_n')$ and $\mathbf{I}_n$ denoting the $n\times n$ identity matrix. Hence the mean and the variance of $\mathbf{Z}_n$ conditional on $\mathbf{X}_n$ based on model $\alpha\in\mathcal{A}_n$ are $\mathbf{X}_n(\alpha)\boldsymbol{\beta}_n(\alpha)$ and

$$\boldsymbol{\Sigma}_n(\boldsymbol{\theta})=\boldsymbol{\Sigma}_{n\eta}+v^2\mathbf{I}_n, \tag{2.3}$$

where $\boldsymbol{\theta}$ is a covariance parameter vector belonging to some parameter space Θ. Throughout the paper, we assume that $\boldsymbol{\Sigma}_n(\boldsymbol{\theta})$ is continuous on $\boldsymbol{\theta}\in\Theta$. We denote the true covariance matrix by $\boldsymbol{\Sigma}_{n0}$ and the true mean of $\mathbf{Z}_n$ conditional on $\mathbf{X}_n$ by $\boldsymbol{\mu}_{n0}$. In other words, given $\mathbf{X}_n$, the data $\mathbf{Z}_n$ are generated from $N(\boldsymbol{\mu}_{n0},\boldsymbol{\Sigma}_{n0})$.

In order to facilitate mathematical exposition, the asymptotic results established in Sections 3 and 4 focus only on the case where $\mathbf{X}_n$ is nonrandom. These results are also valid in the almost sure sense when $\mathbf{X}_n$ is random, provided that the required conditions involving $\mathbf{X}_n$ hold for almost all sequences $\mathbf{X}_n$; $n\in\{1,2,\dots\}$. We further illustrate these results in Section 5 using either random or nonrandom $\mathbf{X}_n$.

2.2. *Generalized information criterion.* For notational simplicity, we suppress the dependence of $\mathbf{X}_n,\mathbf{X}_n(\alpha),\boldsymbol{\beta}_n,\boldsymbol{\beta}_n(\alpha),\mathbf{Z}_n$, $\boldsymbol{\eta}_n$, $\boldsymbol{\epsilon}_n$, $\boldsymbol{\Sigma}_{n\eta}$, $\mathbf{I}_n$, $\boldsymbol{\Sigma}_n(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_{n0}$, $\boldsymbol{\mu}_{n0}$ and $\mathbf{s}_{ni}$ on n in the rest of this paper. To estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we consider maximum likelihood (ML). We assume that $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$ and $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X})^{-1}$ exist for $\boldsymbol{\theta}\in\Theta$. The ML estimate of $\boldsymbol{\theta}$ based on $\alpha\in\mathcal{A}_n$, denoted by $\hat{\boldsymbol{\theta}}(\alpha)$, is obtained by maximizing the following profile log-likelihood function:

$$\begin{aligned}\ell(\alpha;\boldsymbol{\theta})=&-\tfrac{1}{2}n\log(2\pi)-\tfrac{1}{2}\log\det\big(\boldsymbol{\Sigma}(\boldsymbol{\theta})\big)\\&-\tfrac{1}{2}\big(\mathbf{Z}-\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})\big)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big(\mathbf{Z}-\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})\big),\end{aligned}$$

where $\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})=\mathbf{X}(\alpha)\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta})$, and

$$\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta})=\big(\mathbf{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}(\alpha)\big)^{-1}\mathbf{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{Z}.$$

Specifically, $\ell(\alpha;\hat{\boldsymbol{\theta}}(\alpha))=\sup_{\boldsymbol{\theta}\in\Theta}\ell(\alpha;\boldsymbol{\theta})$, and $\hat{\boldsymbol{\beta}}(\alpha;\hat{\boldsymbol{\theta}}(\alpha))$ is the ML estimate of $\boldsymbol{\beta}(\alpha)$. For $\alpha\in\mathcal{A}_n$ and $\boldsymbol{\theta}\in\Theta$, let

$$\mathbf{M}(\alpha;\boldsymbol{\theta})=\mathbf{X}(\alpha)\big(\mathbf{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}(\alpha)\big)^{-1}\mathbf{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}), \tag{2.4}$$

$$\mathbf{A}(\alpha;\boldsymbol{\theta})=\mathbf{I}-\mathbf{M}(\alpha;\boldsymbol{\theta}). \tag{2.5}$$

Then $\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) = \mathbf{M}(\alpha;\boldsymbol{\theta})\mathbf{Z}$ and $\mathbf{Z} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) = \mathbf{A}(\alpha;\boldsymbol{\theta})\mathbf{Z}$. Note that $\mathbf{M}^2(\alpha;\boldsymbol{\theta}) = \mathbf{M}(\alpha;\boldsymbol{\theta})$, $\mathbf{M}(\alpha;\boldsymbol{\theta})\mathbf{X}(\alpha) = \mathbf{X}(\alpha)$, and

$$\mathbf{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{M}(\alpha;\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{M}(\alpha;\boldsymbol{\theta}),$$
$$\mathbf{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{A}(\alpha;\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{A}(\alpha;\boldsymbol{\theta}).$$

Therefore, by (2.4) and (2.5), the profile log-likelihood function can also be written as

$$\begin{aligned}\ell(\alpha;\boldsymbol{\theta}) = {}& -\tfrac{1}{2}n\log(2\pi) - \tfrac{1}{2}\log\det\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta})\bigr) - \tfrac{1}{2}\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu}_0 \\ & - \boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{A}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \tfrac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ & + \tfrac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}); \qquad \alpha \in \mathcal{A}_n, \boldsymbol{\theta} \in \Theta.\end{aligned} \tag{2.6}$$

To identify the smallest correct model α_n^0, one may adopt the GIC of Nishii (1984),

$$\Gamma_{\tau_n}(\alpha) = -2\ell\bigl(\alpha;\hat{\boldsymbol{\theta}}(\alpha)\bigr) + \tau_n p(\alpha); \qquad \alpha \in \mathcal{A}_n, \tag{2.7}$$

where τ_n is a tuning parameter controlling the trade-off between goodness-of-fit and the model parsimoniousness. The criterion includes AIC (when $\tau_n = 2$) and BIC [when $\tau_n = \log(n)$] as special cases, and has been widely used in many statistical areas. The model selected by GIC based on τ_n is denoted by $\hat{\alpha}_{\tau_n} = \arg\min_{\alpha\in\mathcal{A}_n} \Gamma_{\tau_n}(\alpha)$. In the next section, we shall first investigate GIC for variable selection when the covariance model is correctly specified.

3. Variable selection under a correct covariance model. The asymptotic properties of GIC will be derived in terms of the Kullback–Leibler (KL) loss, which for $\alpha \in \mathcal{A}_n$ and $\boldsymbol{\theta} \in \Theta$ is given by

$$\begin{aligned}L(\alpha;\boldsymbol{\theta}) &= \int_{\mathbf{Y}\in\mathbb{R}^n} f(\mathbf{Y};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)\log\frac{f(\mathbf{Y};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)}{f(\mathbf{Y};\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}),\boldsymbol{\Sigma}(\boldsymbol{\theta}))}\,d\mathbf{Y} \\ &= \frac{1}{2}\log\det\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta})\bigr) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}_0) + \frac{1}{2}\operatorname{tr}\bigl(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\bigr) \\ &\quad - \frac{n}{2} + \frac{1}{2}\bigl(\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) - \boldsymbol{\mu}_0\bigr)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\bigl(\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) - \boldsymbol{\mu}_0\bigr),\end{aligned}$$

where $\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) = \mathbf{X}(\alpha)\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta})$ and $f(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the Gaussian density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Note that $L(\alpha;\boldsymbol{\theta}) \geq 0$, for any $\alpha \in \mathcal{A}_n$ and $\boldsymbol{\theta} \in \Theta$. When $\boldsymbol{\mu}_0$ is known, the KL loss for $\boldsymbol{\theta} \in \Theta$ is given by

$$L_0(\boldsymbol{\theta}) = \tfrac{1}{2}\bigl\{\log\det\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta})\bigr) - \log\det(\boldsymbol{\Sigma}_0) + \operatorname{tr}\bigl(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\bigr) - n\bigr\}.$$

Then the optimal vector of $\boldsymbol{\theta} \in \Theta$, which minimizes the KL loss, is given by

$$\boldsymbol{\theta}_0 = \arg\inf_{\boldsymbol{\theta}\in\Theta} L_0(\boldsymbol{\theta}).$$

Clearly, $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and $L_0(\boldsymbol{\theta}_0) = 0$, if the covariance model class contains the correct model. In this case, $\boldsymbol{\theta}_0$ is the true covariance parameter vector of $\boldsymbol{\theta}$. Let $R(\alpha; \boldsymbol{\theta}) = \mathrm{E}(L(\alpha; \boldsymbol{\theta}))$. By (2.4) and (2.5), we have

$$\begin{aligned} L(\alpha; \boldsymbol{\theta}) = {} & L_0(\boldsymbol{\theta}) + \tfrac{1}{2}\boldsymbol{\mu}_0' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu}_0 \\ & + \tfrac{1}{2}(\boldsymbol{\eta} + \boldsymbol{\epsilon})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{M}(\alpha; \boldsymbol{\theta})(\boldsymbol{\eta} + \boldsymbol{\epsilon}), \end{aligned} \tag{3.1}$$

$$\begin{aligned} R(\alpha; \boldsymbol{\theta}) = {} & L_0(\boldsymbol{\theta}) + \tfrac{1}{2}\boldsymbol{\mu}_0' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu}_0 \\ & + \tfrac{1}{2}\operatorname{tr}\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{M}(\alpha; \boldsymbol{\theta}) \boldsymbol{\Sigma}_0\big), \end{aligned} \tag{3.2}$$

for $\alpha \in \mathcal{A}_n$ and $\boldsymbol{\theta} \in \Theta$, where $\boldsymbol{\mu}_0' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu}_0 = \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}) \mathbf{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu}_0\|^2$, which results from using a wrong regression model, and is equal to 0 when $\alpha \in \mathcal{A}_n^0$. In particular, for $\alpha \in \mathcal{A}_n^0$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$,

$$L(\alpha; \boldsymbol{\theta}_0) = \tfrac{1}{2}(\boldsymbol{\eta} + \boldsymbol{\epsilon})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\alpha; \boldsymbol{\theta}_0)(\boldsymbol{\eta} + \boldsymbol{\epsilon}), \tag{3.3}$$

$$R(\alpha; \boldsymbol{\theta}_0) = \tfrac{1}{2} p(\alpha). \tag{3.4}$$

Consider a model selection procedure $\hat{\alpha}$ that maps data to $\alpha \in \mathcal{A}_n$. We say that $\hat{\alpha}$ is consistent if $\lim_{n \to \infty} P\{\hat{\alpha} = \alpha_n^0\} = 1$, and $\hat{\alpha}$ is asymptotically loss efficient if

$$\frac{L(\hat{\alpha}; \hat{\boldsymbol{\theta}}(\hat{\alpha}))}{\min_{\alpha \in \mathcal{A}_n} L(\alpha; \hat{\boldsymbol{\theta}}(\alpha))} \xrightarrow{P} 1, \tag{3.5}$$

as $n \to \infty$. When $\eta(\cdot)$ is absent, geostatistical regression reduces to the usual linear regression with a property that $\lim_{n \to \infty} P\{L(\alpha_n^0) = \inf_{\alpha \in \mathcal{A}_n} L(\alpha)\} = 1$; see Shao (1997) for more details. In this case, pursuing consistency is equivalent to finding the model with the smallest KL loss. However, α_n^0 may not always lead to the smallest KL loss when $\boldsymbol{\Sigma}_\eta$ has to be estimated, making asymptotic loss efficiency more difficult to derive. In addition, the possible inconsistency of $\hat{\boldsymbol{\theta}}(\alpha)$ for $\alpha \in \mathcal{A}_n$ under the fixed domain asymptotic framework further complicates the development of asymptotic theory for GIC.

Let $\lambda_{\min}(\mathbf{Q})$ and $\lambda_{\max}(\mathbf{Q})$ be the smallest and the largest eigenvalue of a square matrix $\mathbf{Q}$. We impose the following regularity conditions for model selection:

(C1) $\lambda_{\min}(\boldsymbol{\Sigma}(\boldsymbol{\theta})) > 0$ for all n and $\boldsymbol{\theta} \in \Theta$, and

$$\limsup_{n \to \infty} \sup_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}\big(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\big) < \infty.$$

(C2) For $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$, there exists $\boldsymbol{\theta}_\alpha \in \Theta$, not depending on n, such that

$$\sup_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \left| \frac{\ell(\alpha; \hat{\boldsymbol{\theta}}(\alpha)) - \ell(\alpha; \boldsymbol{\theta}_\alpha)}{R(\alpha; \boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)} \right| = o_p(1),$$

$$\sup_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \left| \frac{L(\alpha; \hat{\boldsymbol{\theta}}(\alpha)) - L(\alpha; \boldsymbol{\theta}_\alpha)}{R(\alpha; \boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)} \right| = o_p(1).$$

Moreover,

$$\sup_{\alpha\in\mathcal{A}_n^0}\left|\ell(\alpha;\hat{\boldsymbol{\theta}}(\alpha))-\ell(\alpha;\boldsymbol{\theta}_0)\right|=O_p(1),$$

$$\sup_{\alpha\in\mathcal{A}_n^0}\left|L(\alpha;\hat{\boldsymbol{\theta}}(\alpha))-L(\alpha;\boldsymbol{\theta}_0)\right|=O_p(1).$$

(C3) For $\boldsymbol{\theta}_\alpha$ defined in (C2),

$$\lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\frac{1}{(R(\alpha;\boldsymbol{\theta}_\alpha)-L_0(\boldsymbol{\theta}_0))^q}=0,$$

for some $q>0$.

(C4) For $\boldsymbol{\theta}_\alpha$ defined in (C2),

$$\lim_{n\to\infty}\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{\operatorname{tr}(\boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha))\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha))}{R(\alpha;\boldsymbol{\theta}_\alpha)-L_0(\boldsymbol{\theta}_0)}\right|=0.$$

(C5) For $\boldsymbol{\theta}_\alpha$ defined in (C2),

$$\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{\operatorname{tr}(((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})'-\boldsymbol{\Sigma}_0)(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)))}{R(\alpha;\boldsymbol{\theta}_\alpha)-L_0(\boldsymbol{\theta}_0)}\right|=o_p(1).$$

While $L_0(\boldsymbol{\theta}_0)=0$ for a correct spatial covariance model, we still keep $L_0(\boldsymbol{\theta}_0)$ in (C2)–(C5) because $L_0(\boldsymbol{\theta}_0)\neq 0$ under covariance mis-specification, which will be discussed in Section 4. In the rest of this section, we shall assume $\boldsymbol{\Sigma}_0=\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, yielding $L_0(\boldsymbol{\theta}_0)=0$. Condition (C1), imposing some constraints on the family of covariance matrices parameterized by $\boldsymbol{\theta}\in\Theta$, is usually satisfied when Θ is compact. Condition (C2) generally holds when $\hat{\boldsymbol{\theta}}(\alpha)$ converges in probability to some $\boldsymbol{\theta}_\alpha\in\Theta$, not necessarily equal to $\boldsymbol{\theta}_0$. Surprisingly, it can hold even if $\hat{\boldsymbol{\theta}}(\alpha)$ does not converge in probability; see Section 5 for some examples in which the domain D is fixed with n. Condition (C3) is easily met when $|\mathcal{A}_n\setminus\mathcal{A}_n^0|$ (i.e., the number of models in $\mathcal{A}_n\setminus\mathcal{A}_n^0$) is bounded and

$$\min_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left\|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}_0\right\|^2\to\infty,$$

as $n\to\infty$. Moreover, (C5) can be verified using some moment bounds for quadratic forms in $\boldsymbol{\eta}+\boldsymbol{\epsilon}$, and (C4) is ensured by (C3) when p_n is bounded.

Conditions (C1)–(C5) appear to be natural generalizations of the conditions used to establish the asymptotic loss efficiency in usual linear regression models. To see this, note that if $\boldsymbol{\Sigma}_0=\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is known (or, equivalently, $\Theta=\{\boldsymbol{\theta}_0\}$), then (C1), (C2), (C4) and (C5) become redundant, and only (C3) is needed, which corresponds to (A.3) of Li (1987) or (2.6) of Shao (1997). This is the only assumption needed to derive the asymptotic loss efficiency of AIC under model (2.2) with

$\eta(\cdot)=0$, v^2 known, $\mathbf{s}_i=i$; $i=1,\dots,n$, and $|\mathcal{A}_n^0|\le 1$. For more details, see Theorem 1 of Shao (1997). On the other hand, when $\boldsymbol{\theta}_0$ is unknown, (C1), (C2), (C4) and (C5) seem indispensable for dealing with the inherent difficulties in model selection under (2.2). That is, the ML estimate of $\boldsymbol{\theta}$ may not only vary across candidate models, but may also converge to wrong parameter vectors or have no probability limits. In the following theorem, these four conditions will be used in conjunction with (C3) to establish the consistency and the asymptotic loss efficiency of AIC, extending Theorem 1 of Shao (1997) to the geostatistical model described in (2.2) and (2.3).

THEOREM 3.1. *Consider the data generated from* (2.1) *and the model given by* (2.2) *and* (2.3) *with* $\boldsymbol{\theta}_0$ *being the true covariance parameter vector* [*i.e.*, $\mathrm{var}(\mathbf{Z})=\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$]. *Suppose that conditions* (C1)–(C5) *are satisfied*:

(i) *If* $|\mathcal{A}_n^0|\le 1$, *then* $\hat{\alpha}_2$ *is asymptotically loss efficient. If, in addition,* $|\mathcal{A}_n^0|=1$ *and* $\limsup_{n\to\infty} p(\alpha_n^0)<\infty$, *then* $\hat{\alpha}_2$ *is consistent.*

(ii) *If* $|\mathcal{A}_n^0|\ge 2$ *for sufficiently large n and either of the following is satisfied for some* $m>0$,

$$\lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}_n^0}\frac{1}{p^m(\alpha)}=0, \tag{3.6}$$

$$\lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}_n^0\setminus\{\alpha_n^0\}}\frac{1}{(p(\alpha)-p(\alpha_n^0))^m}=0. \tag{3.7}$$

Then $\hat{\alpha}_2$ *is asymptotically loss efficient. If, in addition,* (3.7) *holds and* $\limsup_{n\to\infty} p(\alpha_n^0)<\infty$, *then* $\hat{\alpha}_2$ *is consistent.*

PROOF. We begin by showing that

$$\Gamma_2(\alpha)=\nu+2L(\alpha;\boldsymbol{\theta}_\alpha)+o_p\big(L(\alpha;\boldsymbol{\theta}_\alpha)\big), \tag{3.8}$$

uniformly for $\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0$, where $\nu=n\log(2\pi)+\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))+(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})$ is independent of α. By (2.7) and (C2), we have

$$\begin{aligned}
\Gamma_2(\alpha)&=-2\ell(\alpha;\boldsymbol{\theta}_\alpha)+2p(\alpha)+o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha)\big)\\
&=n\log(2\pi)+\log\det\big(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)\big)+\mathbf{Z}'\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\mathbf{Z}\\
&\quad+2p(\alpha)+o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha)\big)\\
&=n\log(2\pi)+\log\det\big(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)\big)+\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}_0\\
&\quad+2\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\quad+(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})+2p(\alpha)+o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha)\big),
\end{aligned}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. It follows from (3.1) that

$$\begin{aligned}
\Gamma_2(\alpha) &= n\log(2\pi) + \log\det\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\bigr) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad - \operatorname{tr}\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\bigr) + n + 2L(\alpha;\boldsymbol{\theta}_\alpha) \\
&\quad - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad + 2\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2p(\alpha) + o_p\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr) \\
&= n\log(2\pi) + \log\det\bigl(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\bigr) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad + \operatorname{tr}\bigl(\bigl((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\bigr)\bigl(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\bigr)\bigr) \\
&\quad + 2L(\alpha;\boldsymbol{\theta}_\alpha) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2p(\alpha) \\
&\quad + 2\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr),
\end{aligned} \tag{3.9}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. Therefore, by (C5), for (3.8) to hold, it suffices to show that

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha) = o_p\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr), \tag{3.10}$$

$$\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = o_p\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr), \tag{3.11}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$, and

$$\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{L(\alpha;\boldsymbol{\theta}_\alpha)}{R(\alpha;\boldsymbol{\theta}_\alpha)} - 1\right| = o_p(1). \tag{3.12}$$

First, we prove (3.10). By (C4), we have

$$\mathrm{E}\bigl\{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\bigr\} - p(\alpha) = o\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr),$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. Let $c(\alpha) = \operatorname{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha))/p(\alpha)$. Then by (2.4) and (C1), $\limsup_{n\to\infty}\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} c(\alpha) < \infty$. Thus for (3.10) to hold, it suffices to show that

$$\begin{aligned}
&(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha) \\
&\qquad = o_p\bigl(R(\alpha;\boldsymbol{\theta}_\alpha)\bigr),
\end{aligned}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. Applying Chebyshev's inequality, we have for any $\varepsilon > 0$,

$$\begin{aligned}
&P\left\{\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha)}{R(\alpha;\boldsymbol{\theta}_\alpha)}\right| > \varepsilon\right\} \\
&\qquad \le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\frac{\mathrm{E}|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha)|^{2q}}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}
\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_1\{\operatorname{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha))\}^q}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_2 p^q(\alpha)}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_3}{\varepsilon^{2q}R^{q}(\alpha;\boldsymbol{\theta}_\alpha)},
\end{aligned}
$$

for some constants $c_1, c_2, c_3 > 0$, where the second inequality follows from Theorem 2 of Whittle (1960) that $\mathrm{E}(|\mathbf{y}'\mathbf{A}\mathbf{y} - \mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y})|)^{2q} \le c_1(\operatorname{tr}(\mathbf{A}^2))^q$ for $\mathbf{y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \sim N(\mathbf{0},\mathbf{I})$ and $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\theta}_0)$, the third inequality follows from (C1), and the last inequality follows from (C4). Therefore by (C3), we obtain (3.10).

Next, we prove (3.11). Similar to the proof of (3.10), we have

$$
\begin{aligned}
P&\left\{\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})}{R(\alpha;\boldsymbol{\theta}_\alpha)}\right| > \varepsilon\right\}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{\mathrm{E}|\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|^{2q}}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_4(\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}_0)^q}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_5(\boldsymbol{\mu}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}_0)^q}{\varepsilon^{2q}R^{2q}(\alpha;\boldsymbol{\theta}_\alpha)}\\
&\le \sum_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{c_6}{\varepsilon^{2q}R^{q}(\alpha;\boldsymbol{\theta}_\alpha)},
\end{aligned}
$$

for some constant $c_4, c_5, c_6 > 0$, where the second inequality follows from Theorem 2 of Whittle (1960) that $\mathrm{E}(|\mathbf{a}'\mathbf{y}|)^{2q} \le c_4(\mathbf{a}'\mathbf{a})^q$ for $\mathbf{y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \sim N(\mathbf{0},\mathbf{I})$ and $\mathbf{a} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\theta}_0)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}_0$, the third inequality follows from (C1), and the last inequality follows from (3.2). Therefore by (C3), we obtain (3.11).

It remains to prove (3.12). By (3.1) and (3.2), for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$,

$$
\begin{aligned}
L(\alpha;\boldsymbol{\theta}_\alpha) - R(\alpha;\boldsymbol{\theta}_\alpha) = {}& (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
& - \operatorname{tr}\bigl(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\bigr).
\end{aligned}
$$

It follows from (C1), (C3) and an argument similar to one used to prove (3.10) that

$$\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{L(\alpha;\boldsymbol{\theta}_\alpha)}{R(\alpha;\boldsymbol{\theta}_\alpha)}-1\right| = \sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})}{R(\alpha;\boldsymbol{\theta}_\alpha)} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))}{R(\alpha;\boldsymbol{\theta}_\alpha)}\right| = o_p(1).$$

This gives (3.12). Thus (3.8) is established.

(i) If $|\mathcal{A}_n^0|=0$, it follows from (3.8), (3.12) and (C2) that $\hat{\alpha}_2$ is asymptotically loss efficient. If $|\mathcal{A}_n^0|=1$ and $\lim_{n\to\infty} p(\alpha_n^0)=\infty$, by (3.8), to show the asymptotic loss efficiency of $\hat{\alpha}_2$, it suffices to show that

$$\Gamma_2(\alpha)=\nu+2L(\alpha;\boldsymbol{\theta}_0)+o_p\big(L(\alpha;\boldsymbol{\theta}_0)\big);\qquad \alpha\in\mathcal{A}_n^0. \tag{3.13}$$

By (2.6), (3.3) and (C2),

$$\begin{aligned}\Gamma_2(\alpha)&=-2\ell(\alpha;\boldsymbol{\theta}_0)+2p(\alpha)+O_p(1)\\ &=\nu-2\big\{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})-p(\alpha)\big\}\\ &\quad+2L(\alpha;\boldsymbol{\theta}_0)+O_p(1);\qquad \alpha\in\mathcal{A}_n^0.\end{aligned} \tag{3.14}$$

Therefore, by (3.3), (3.4) and an argument similar to that used to prove (3.8), we have

$$\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha_n^0;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})-p(\alpha_n^0)}{p(\alpha_n^0)}\right|=o_p(1), \tag{3.15}$$

$$\left|\frac{L(\alpha_n^0;\boldsymbol{\theta}_0)}{R(\alpha_n^0;\boldsymbol{\theta}_0)}-1\right|=o_p(1). \tag{3.16}$$

These together with (3.14) give (3.13). If $|\mathcal{A}_n^0|=1$ and $\limsup_{n\to\infty} p(\alpha_n^0)<\infty$, then the consistency and the asymptotical loss efficiency are ensured by

$$L\big(\alpha;\hat{\boldsymbol{\theta}}(\alpha)\big)-L\big(\alpha_n^0;\hat{\boldsymbol{\theta}}(\alpha_n^0)\big)\xrightarrow{P}\infty, \tag{3.17}$$

$$\Gamma_2(\alpha)-\Gamma_2\big(\alpha_n^0\big)\xrightarrow{P}\infty, \tag{3.18}$$

uniformly for $\alpha\in\mathcal{A}_n\setminus\{\alpha_n^0\}$, as $n\to\infty$. First, (3.17) follows from

$$\begin{aligned}L\big(\alpha_n^0;\hat{\boldsymbol{\theta}}(\alpha_n^0)\big)&=L\big(\alpha_n^0;\boldsymbol{\theta}_0\big)+O_p(1)\\ &=\tfrac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}\big(\alpha_n^0;\boldsymbol{\theta}_0\big)(\boldsymbol{\eta}+\boldsymbol{\epsilon})+O_p(1)\\ &=o_p\big(L\big(\alpha;\hat{\boldsymbol{\theta}}(\alpha)\big)\big),\end{aligned} \tag{3.19}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \{\alpha_n^0\}$, where the first equality follows from (C2), the second equality follows from (3.1) and the last equality follows from (3.12), (C2), (C3) and $\limsup_{n\to\infty} p(\alpha_n^0) < \infty$. It remains to prove (3.18). By (3.14), we have

$$\begin{aligned}\Gamma_2(\alpha_n^0) &= \nu - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha_n^0;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2p(\alpha_n^0) + O_p(1)\\ &= \nu + o_p(L(\alpha;\hat{\boldsymbol{\theta}}(\alpha))),\end{aligned} \tag{3.20}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \{\alpha_n^0\}$, where the last equality follows from an argument similar to that used to prove (3.19). This together with (3.8) implies (3.18). This completes the proof of (i).

(ii) First, suppose that (3.6) is satisfied. In view of (3.8), it suffices to show that (3.13) holds uniformly for $\alpha \in \mathcal{A}_n^0$. Similarly to the proofs of (3.15) and (3.16), we only need to show that

$$\sup_{\alpha\in\mathcal{A}_n^0}\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha)}{p(\alpha)}\right| = o_p(1), \tag{3.21}$$

$$\sup_{\alpha\in\mathcal{A}_n^0}\left|\frac{L(\alpha;\boldsymbol{\theta}_0)}{R(\alpha;\boldsymbol{\theta}_0)} - 1\right| = o_p(1). \tag{3.22}$$

By an argument similar to that used to prove (3.10), we have

$$\begin{aligned}&P\left\{\sup_{\alpha\in\mathcal{A}_n^0}\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha)}{p(\alpha)}\right| > \varepsilon\right\}\\ &\quad\le \sum_{\alpha\in\mathcal{A}_n^0}\frac{\mathrm{E}|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha)|^{2m}}{\varepsilon^{2m}p^{2m}(\alpha)} \le \sum_{\alpha\in\mathcal{A}_n^0}\frac{c_7}{\varepsilon^{2m}p^m(\alpha)},\end{aligned}$$

for some constant $c_7 > 0$, as $n \to \infty$. This together with (3.3), (3.4) and (3.6) gives (3.21) and (3.22). Therefore, (3.13) holds uniformly for $\alpha \in \mathcal{A}_n^0$.

Finally, suppose that (3.7) is satisfied. If $\lim_{n\to\infty} p(\alpha_n^0) = \infty$, it implies (3.6) and hence $\hat{\alpha}_2$ is asymptotically loss efficient. If $\limsup_{n\to\infty} p(\alpha_n^0) < \infty$, by (3.17) and (3.18), it remains to show that

$$L(\alpha;\hat{\boldsymbol{\theta}}(\alpha)) - L(\alpha_n^0;\hat{\boldsymbol{\theta}}(\alpha_n^0)) \xrightarrow{P} \infty, \tag{3.23}$$

$$\Gamma_2(\alpha) - \Gamma_2(\alpha_n^0) \xrightarrow{P} \infty, \tag{3.24}$$

uniformly for $\alpha \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}$, as $n \to \infty$. First, we prove (3.23). By (3.4) and (C2),

$$\begin{aligned}&L(\alpha;\hat{\boldsymbol{\theta}}(\alpha)) - L(\alpha_n^0;\hat{\boldsymbol{\theta}}(\alpha_n^0))\\ &\quad= L(\alpha;\boldsymbol{\theta}_0) - L(\alpha_n^0;\boldsymbol{\theta}_0) + O_p(1)\\ &\quad= \tfrac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\{\mathbf{M}(\alpha;\boldsymbol{\theta}_0) - \mathbf{M}(\alpha_n^0;\boldsymbol{\theta}_0)\}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + O_p(1)\\ &\quad= \tfrac{1}{2}(p(\alpha) - p(\alpha_n^0)) + o_p(p(\alpha) - p(\alpha_n^0)),\end{aligned} \tag{3.25}$$

uniformly for $\alpha \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}$, where the last equality follows from

$$\sup_{\alpha \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}} \left| \frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha)}{p(\alpha)-p(\alpha_n^0)} \right.$$
$$\left. - \frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha_n^0;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha_n^0)}{p(\alpha)-p(\alpha_n^0)} \right| = o_p(1),$$

which can be obtained in a way similar to the proof of (3.15). This together with (3.7) gives (3.23). Next, we prove (3.24). By (3.14) and (3.25), we have

$$\begin{aligned}\Gamma_2(\alpha) - \Gamma_2(\alpha_n^0) &= 2L(\alpha;\hat{\boldsymbol{\theta}}(\alpha)) - 2L(\alpha_n^0;\hat{\boldsymbol{\theta}}(\alpha_n^0)) + o_p(p(\alpha)-p(\alpha_n^0)) \\ &= p(\alpha) - p(\alpha_n^0) + o_p(p(\alpha)-p(\alpha_n^0)),\end{aligned}$$

uniformly for $\alpha \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}$. This together with (3.7) gives (3.24). This completes the proof of (ii). □

REMARK 3.1. When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is known, Theorem 3.1 reduces to the standard asymptotic theory of AIC in linear regression; see Theorem 1 of Shao (1997). In this case, (C1), (C2), (C4) and (C5) are not needed.

REMARK 3.2. Although Theorem 3.1 only obtains the consistency of $\hat{\alpha}_2$ under $\limsup_{n\to\infty} p(\alpha_n^0) < \infty$, the consistency result can be extended to $\lim_{n\to\infty} p(\alpha_n^0) = \infty$ if $p(\alpha_n^0) = o(\inf_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} R(\alpha;\boldsymbol{\theta}_0))$.

REMARK 3.3. When $|\mathcal{A}_n^0| \geq 2$, AIC is generally not able to identify α_n^0 almost surely. A heavier penalty τ_n of GIC (e.g., BIC) is needed for consistency.

THEOREM 3.2. *Consider the data generated from* (2.1) *and the model given by* (2.2) *and* (2.3) *with* $\boldsymbol{\theta}_0$ *being the true covariance parameter vector* [*i.e.*, $\mathrm{var}(\mathbf{Z}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$]. *Suppose that* (C1)–(C5) *are satisfied. In addition, suppose that* $\lim_{n\to\infty}\tau_n = \infty$, *and for* $\boldsymbol{\theta}_\alpha$ *defined in* (C2),

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{\tau_n p_n}{R(\alpha;\boldsymbol{\theta}_\alpha)} = 0. \tag{3.26}$$

(i) *If* $|\mathcal{A}_n^0| = 0$, *then* $\hat{\alpha}_{\tau_n}$ *is asymptotically loss efficient.*
(ii) *If* $|\mathcal{A}_n^0| \geq 1$ *and*

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}_n^0} \frac{1}{p^m(\alpha)} < \infty, \tag{3.27}$$

for some $m > 0$, *then* $\hat{\alpha}_{\tau_n}$ *is consistent.*

PROOF. (i) By (3.8) and (3.26), we have

$$\Gamma_{\tau_n}(\alpha) = \nu + 2L(\alpha; \boldsymbol{\theta}_\alpha) + o_p\big(L(\alpha; \boldsymbol{\theta}_\alpha)\big), \tag{3.28}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. Thus by (3.12) and (C2), $\hat{\alpha}_{\tau_n}$ is asymptotically loss efficient.

(ii) By (2.6) and (C2), we have for $\alpha \in \mathcal{A}_n^0$,

$$\begin{aligned} \Gamma_{\tau_n}(\alpha) &= -2\ell(\alpha; \boldsymbol{\theta}_0) + \tau_n p(\alpha) + O_p(1) \\ &= \nu - (\boldsymbol{\eta} + \boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha; \boldsymbol{\theta}_0)(\boldsymbol{\eta} + \boldsymbol{\epsilon}) + \tau_n p(\alpha) + O_p(1), \end{aligned} \tag{3.29}$$

where ν is defined in (3.8). By (3.27) and an argument similar to that used to prove (3.21), we have

$$\sup_{\alpha \in \mathcal{A}_n^0} \left| \frac{(\boldsymbol{\eta} + \boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha; \boldsymbol{\theta}_0)(\boldsymbol{\eta} + \boldsymbol{\epsilon}) - p(\alpha)}{\tau_n p(\alpha)} \right| = o_p(1). \tag{3.30}$$

This and (3.29) give

$$\Gamma_{\tau_n}(\alpha) = \nu + (\tau_n - 1)p(\alpha) + o_p\big(\tau_n p(\alpha)\big), \tag{3.31}$$

uniformly for $\alpha \in \mathcal{A}_n^0$. Thus

$$\lim_{n \to \infty} P\{\hat{\alpha}_{\tau_n} \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}\} = 0. \tag{3.32}$$

By (3.26), (3.28) and (3.31), we have

$$\min_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \Gamma_{\tau_n}(\alpha) - \Gamma_{\tau_n}\big(\alpha_n^0\big) \xrightarrow{P} \infty,$$

as $n \to \infty$. This together with (3.32) implies that $\hat{\alpha}_{\tau_n}$ is consistent. This completes the proof. □

Unlike the KL loss function in usual linear regression models, $L(\alpha, \hat{\boldsymbol{\theta}}(\alpha))$ does not necessarily have the minimum at $\alpha = \alpha_n^0$, and hence selection consistency may not lead to asymptotic loss efficiency in geostatistical regression models. Nevertheless, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is known, Theorem 3.2 reduces to the standard asymptotic theory of GIC in linear regression [see Theorem 2 of Shao (1997)], in which selection consistency is known to imply asymptotic loss efficiency. This property continues to hold if $\hat{\boldsymbol{\theta}}(\alpha)$ in (2.7), and (3.5) is replaced by a common estimate $\hat{\boldsymbol{\theta}}$, independent of α. Then for $\alpha \in \mathcal{A}_n^0 \setminus \{\alpha_n^0\}$,

$$L(\alpha; \hat{\boldsymbol{\theta}}) - L\big(\alpha_n^0; \hat{\boldsymbol{\theta}}\big) = (\boldsymbol{\eta} + \boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})\big(\mathbf{M}(\alpha; \hat{\boldsymbol{\theta}}) - \mathbf{M}\big(\alpha_n^0; \hat{\boldsymbol{\theta}}\big)\big)(\boldsymbol{\eta} + \boldsymbol{\epsilon}) \geq 0,$$

almost surely.

COROLLARY 3.1. *Consider the data generated from* (2.1) *and the model defined in* (2.2) *and* (2.3) *with* $\boldsymbol{\theta}_0$ *being the true covariance parameter vector* [*i.e.*, $\mathrm{var}(\mathbf{Z}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$]. *Suppose that* (C1)–(C5) *are satisfied with* $\hat{\boldsymbol{\theta}}(\alpha)$ *and* $\boldsymbol{\theta}_\alpha$ *in* (C2)–(C5) *being replaced by* $\hat{\boldsymbol{\theta}}$ *and a constant vector* $\boldsymbol{\theta}_c \in \Theta$, *independent of* α. *Let* $\hat{\alpha}_{\tau_n}$ *be the model selected by a modified GIC criterion with* $\hat{\boldsymbol{\theta}}(\alpha)$ *in* (2.7) *being replaced by* $\hat{\boldsymbol{\theta}}$. *In addition, suppose that* $\lim_{n\to\infty} \tau_n = \infty$, *and* $\lim_{n\to\infty} \sup_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \frac{\tau_n p_n}{R(\alpha;\boldsymbol{\theta}_c)} = 0$.

(i) *If* $|\mathcal{A}_n^0| = 0$, *then* $\hat{\alpha}_{\tau_n}$ *is asymptotically loss efficient in the sense that* $L(\hat{\alpha}_{\tau_n};\hat{\boldsymbol{\theta}})/\inf_{\alpha\in\mathcal{A}_n} L(\alpha;\hat{\boldsymbol{\theta}}) \xrightarrow{P} 1$, *as* $n\to\infty$.

(ii) *If* $|\mathcal{A}_n^0| \geq 1$ *and* (3.27) *holds, then* $\hat{\alpha}_{\tau_n}$ *is consistent and asymptotically loss efficient in the sense that* $L(\hat{\alpha}_{\tau_n};\hat{\boldsymbol{\theta}})/\inf_{\alpha\in\mathcal{A}_n} L(\alpha;\hat{\boldsymbol{\theta}}) \xrightarrow{P} 1$, *as* $n\to\infty$.

4. Variable selection under an incorrect covariance model. In this section, we establish the asymptotic theory of GIC for variable selection, when the covariance model is mis-specified with $\boldsymbol{\Sigma}_0 \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, yielding $L_0(\boldsymbol{\theta}_0) \neq 0$. To ensure that the asymptotic optimality of GIC for $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ carries over to this case, we need a stronger condition in place of (C4):

(C4′) For $\boldsymbol{\theta}_\alpha$ defined in (C2),

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0} \frac{p_n}{R(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)} = 0.$$

THEOREM 4.1. *Consider the data generated from* (2.1) *and the model given by* (2.2) *and* (2.3). *Suppose that the conditions* (C1)–(C3), (C4′) *and* (C5) *are satisfied*:

(i) *If* $|\mathcal{A}_n^0| \leq 1$, *then* $\hat{\alpha}_2$ *is asymptotically loss efficient. If* $|\mathcal{A}_n^0| = 1$, *then* $\hat{\alpha}_2$ *is consistent*.

(ii) *If* $|\mathcal{A}_n^0| \geq 2$ *for sufficient large* n, $|\mathcal{A}_n^0|^q = o(L_0(\boldsymbol{\theta}_0))$ *for some* $q > 0$, *and*

$$\lim_{n\to\infty} \frac{p_n}{L_0(\boldsymbol{\theta}_0)} = 0, \tag{4.1}$$

then $\hat{\alpha}_2$ *is asymptotically loss efficient*.

PROOF. Let $L^*(\alpha;\boldsymbol{\theta}_\alpha) = L(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)$; $\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0$. We begin by showing that

$$\Gamma_2(\alpha) = \nu + 2L^*(\alpha;\boldsymbol{\theta}_\alpha) + o_p\big(L^*(\alpha;\boldsymbol{\theta}_\alpha)\big), \tag{4.2}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$, where ν is defined in (3.8) and is independent of α. By an argument similar to that used to prove (3.9), we have

$$\begin{aligned}\Gamma_2(\alpha) &= n\log(2\pi) + \log\det(\mathbf{\Sigma}_0) + n - \operatorname{tr}\big(\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\big)\\ &\quad + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &\quad + \operatorname{tr}\big(\big((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \mathbf{\Sigma}_0\big)\big(\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)\big)\big)\\ &\quad + 2L(\alpha;\boldsymbol{\theta}_\alpha) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2p(\alpha)\\ &\quad + 2\boldsymbol{\mu}_0'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha)\big)\\ &= \nu + \operatorname{tr}\big(\big((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \mathbf{\Sigma}_0\big)\big(\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)\big)\big)\\ &\quad + 2L^*(\alpha;\boldsymbol{\theta}_\alpha) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2p(\alpha)\\ &\quad + 2\boldsymbol{\mu}_0'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha)\big),\end{aligned}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. Hence by (C5) and an argument similar to that used to prove (3.8), for (4.2) to hold, it suffices to show that

$$\begin{aligned}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{M}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha) &= o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)\big),\\ \boldsymbol{\mu}_0'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\mathbf{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= o_p\big(R(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)\big),\end{aligned}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$, and

$$\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\left|\frac{L^*(\alpha;\boldsymbol{\theta}_\alpha)}{R(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)} - 1\right| = o_p(1). \tag{4.3}$$

The above three equations follow from arguments similar to those used to prove (3.10)–(3.12).

(i) Clearly, (4.2) implies (3.8). Therefore, if $|\mathcal{A}_n^0| = 0$, it follows from (4.3) and (C2) that $\hat{\alpha}_2$ is asymptotically loss efficient. On the other hand, if $|\mathcal{A}_n^0| = 1$, it suffices to show (3.17) and (3.18). First, we prove (3.17). By (C3), (C4′) and an argument similar to that used to prove (3.19), we have $L^*(\alpha_n^0; \hat{\boldsymbol{\theta}}(\alpha_n^0)) = o_p(L^*(\alpha; \hat{\boldsymbol{\theta}}(\alpha)))$, uniformly for $\alpha \in \mathcal{A}_n \setminus \{\alpha_n^0\}$. Next, we prove (3.18). By (C3), (C4′) and an argument similar to that used to prove (3.20), we have $\Gamma_2(\alpha_n^0) = \nu^* + o_p(L^*(\alpha; \hat{\boldsymbol{\theta}}(\alpha)))$, uniformly for $\alpha \in \mathcal{A}_n \setminus \{\alpha_n^0\}$. This together with (4.2) implies (3.18), and hence the proof of (i) is complete.

(ii) In view of (3.8), it suffices to show that

$$\Gamma_2(\alpha) = \nu^* + 2L(\alpha;\boldsymbol{\theta}_0) + o_p\big(L(\alpha;\boldsymbol{\theta}_0)\big), \tag{4.4}$$

uniformly for $\alpha \in \mathcal{A}_n^0$, where $\nu^* = \nu - 2L_0(\boldsymbol{\theta}_0)$ with ν being defined in (3.8). By an argument similar to that used to prove (3.14), we have

$$\begin{aligned}\Gamma_2(\alpha) &= \nu^* - 2\big\{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha)\big\}\\ &\quad + 2L(\alpha;\boldsymbol{\theta}_0) + O_p(1); \qquad \alpha \in \mathcal{A}_n^0.\end{aligned} \tag{4.5}$$

Therefore, by an argument similar to that used to prove (3.13), we only need to show that

$$(4.6)\qquad (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - p(\alpha) = o_p\big(L(\alpha;\boldsymbol{\theta}_0)\big),$$

uniformly for $\alpha \in \mathcal{A}_n^0$ and

$$(4.7)\qquad \sup_{\alpha\in\mathcal{A}_n^0}\left|\frac{L(\alpha;\boldsymbol{\theta}_0)}{R(\alpha;\boldsymbol{\theta}_0)} - 1\right| = o_p(1).$$

First, we prove (4.6). Clearly, by (2.4) and (C1), we have

$$(4.8)\qquad \mathrm{E}\big((\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big) = c(\alpha)p(\alpha),$$

where $\limsup_{n\to\infty}\sup_{\alpha\in\mathcal{A}_n} c(\alpha) < \infty$. Hence by (3.2) and (4.1), $c(\alpha)p(\alpha) - p(\alpha) = o(R(\alpha;\boldsymbol{\theta}_0))$ uniformly for $\alpha \in \mathcal{A}_n^0$. It remains to show that

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha) = o_p\big(R(\alpha;\boldsymbol{\theta}_0)\big),$$

uniformly for $\alpha \in \mathcal{A}_n^0$. Applying Chebyshev's inequality, we have for any $\varepsilon > 0$,

$$\begin{aligned}
&P\left\{\sup_{\alpha\in\mathcal{A}_n^0}\left|\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha)}{R(\alpha;\boldsymbol{\theta}_0)}\right| > \varepsilon\right\}\\
&\quad\le \sum_{\alpha\in\mathcal{A}_n^0}\frac{\mathrm{E}|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha)|^{2m}}{\varepsilon^{2m}R^{2m}(\alpha;\boldsymbol{\theta}_0)}\\
&\quad\le \sum_{\alpha\in\mathcal{A}_n^0}\frac{c_1\{\mathrm{tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0))\}^m}{\varepsilon^{2m}R^{2m}(\alpha;\boldsymbol{\theta}_0)}\\
&\quad\le \sum_{\alpha\in\mathcal{A}_n^0}\frac{c_2 p^m(\alpha)}{\varepsilon^{2m}L_0^{2m}(\boldsymbol{\theta}_0)} \le \sum_{\alpha\in\mathcal{A}_n^0}\frac{c_3}{\varepsilon^{2m}L_0^{m}(\boldsymbol{\theta}_0)},
\end{aligned}$$

where the second-to-last equality follows from (C1) and $R(\alpha;\boldsymbol{\theta}_0) \ge L_0(\boldsymbol{\theta}_0)$, for $\alpha \in \mathcal{A}_n$, and the last equality follows from (4.1). Taking $m = 1/q$, we obtain (4.6). Next, we prove (4.7). By (3.1), (3.2) and (4.8), we have for $\alpha \in \mathcal{A}_n^0$,

$$L(\alpha;\boldsymbol{\theta}_0) - R(\alpha;\boldsymbol{\theta}_0) = \tfrac{1}{2}\big\{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha;\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - c(\alpha)p(\alpha)\big\},$$

where $\limsup_{n\to\infty}\sup_{\alpha\in\mathcal{A}_n^0} c(\alpha) < \infty$. Thus (4.7) follows from an argument similar to that used to prove (4.6). Thus we obtain (4.4). This completes the proof. □

THEOREM 4.2. *Under the setup of Theorem* 4.1, *suppose that* $\lim_{n\to\infty}\tau_n = \infty$, *and*

$$(4.9)\qquad \lim_{n\to\infty}\sup_{\alpha\in\mathcal{A}_n\setminus\mathcal{A}_n^0}\frac{\tau_n p_n}{R(\alpha;\boldsymbol{\theta}_\alpha) - L_0(\boldsymbol{\theta}_0)} = 0.$$

(i) *If* $|\mathcal{A}_n^0| = 0$, *then* $\hat{\alpha}_{\tau_n}$ *is asymptotically loss efficient.*

(ii) *If* $|\mathcal{A}_n^0| \geq 1$, $|\mathcal{A}_n^0|^q = o(L_0(\boldsymbol{\theta}_0))$ *for some* $q > 0$, *and* (3.27) *is satisfied, then* $\hat{\alpha}_{\tau_n}$ *is consistent and asymptotically loss efficient.*

PROOF. (i) By (4.2) and (4.9), we have $\Gamma_{\tau_n}(\alpha) = \nu + 2L^*(\alpha; \boldsymbol{\theta}_\alpha) + o_p(L^*(\alpha; \boldsymbol{\theta}_\alpha))$, uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$, and hence

$$\Gamma_{\tau_n}(\alpha) = \nu^* + 2L(\alpha; \boldsymbol{\theta}_\alpha) + o_p\big(L(\alpha; \boldsymbol{\theta}_\alpha)\big), \tag{4.10}$$

uniformly for $\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0$. In addition, (4.3) gives

$$\sup_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \left| \frac{L(\alpha; \boldsymbol{\theta}_\alpha)}{R(\alpha; \boldsymbol{\theta}_\alpha)} - 1 \right| = o_p(1). \tag{4.11}$$

These together with (C2) imply that $\hat{\alpha}_{\tau_n}$ is asymptotically loss efficient.

(ii) First, we prove the asymptotic loss efficiency of $\hat{\alpha}_{\tau_n}$. By (4.7) and (4.11), we have

$$\sup_{\alpha \in \mathcal{A}_n} \left| \frac{L(\alpha; \boldsymbol{\theta}_0)}{R(\alpha; \boldsymbol{\theta}_0)} - 1 \right| = o_p(1). \tag{4.12}$$

By (4.9) and an argument similar to that used to prove (4.4), we have

$$\Gamma_{\tau_n}(\alpha) = \nu^* + 2L(\alpha; \boldsymbol{\theta}_0) + o_p\big(L(\alpha; \boldsymbol{\theta}_0)\big),$$

uniformly for $\alpha \in \mathcal{A}_n^0$. This together with (4.10), (4.12) and (C2) implies that $\hat{\alpha}_{\tau_n}$ is asymptotically loss efficient.

Next, we prove the consistency of $\hat{\alpha}_{\tau_n}$. By (4.5) and (4.8), we have for $\alpha \in \mathcal{A}_n^0$,

$$\begin{aligned} \Gamma_{\tau_n}(\alpha) = \nu - \big\{&(\boldsymbol{\eta} + \boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha; \boldsymbol{\theta}_0)(\boldsymbol{\eta} + \boldsymbol{\epsilon}) - c(\alpha)p(\alpha)\big\} \\ &+ \big(\tau_n - c(\alpha)\big)p(\alpha) + o_p\big(\tau_n p(\alpha)\big). \end{aligned} \tag{4.13}$$

By (3.27) and an argument similar to that used to prove (3.30), we have

$$\sup_{\alpha \in \mathcal{A}_n^0} \left| \frac{(\boldsymbol{\eta} + \boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{M}(\alpha; \boldsymbol{\theta}_0)(\boldsymbol{\eta} + \boldsymbol{\epsilon}) - c(\alpha)p(\alpha)}{\tau_n p(\alpha)} \right| = o_p(1).$$

Hence by (4.13),

$$\Gamma_{\tau_n}(\alpha) = \nu + \big(\tau_n - c(\alpha)\big)p(\alpha) + o_p\big(\tau_n p(\alpha)\big), \tag{4.14}$$

uniformly for $\alpha \in \mathcal{A}_n^0$. Thus we obtain (3.32). In addition, by (4.9), (4.10) and (4.14),

$$\min_{\alpha \in \mathcal{A}_n \setminus \mathcal{A}_n^0} \Gamma_{\tau_n}(\alpha) - \Gamma_{\tau_n}(\alpha_n^0) \xrightarrow{P} \infty,$$

as $n \to \infty$. This together with (3.32) implies that $\hat{\alpha}_{\tau_n}$ is consistent. This completes the proof. □

REMARK 4.1. Recall that in (ii) of Theorem 3.2, asymptotic loss efficiency of GIC is generally not satisfied, unless $\hat{\boldsymbol{\theta}}(\alpha)$'s are replaced by a common estimate. In contrast, in (ii) of Theorem 4.2, we have, from (3.1) and an argument similar to that used to prove (4.6) that $L(\alpha; \boldsymbol{\theta}_0) = L_0(\boldsymbol{\theta}_0) + o_p(L_0(\boldsymbol{\theta}_0))$, uniformly for $\alpha \in \mathcal{A}_n^0$, which leads to

$$\frac{L(\alpha; \hat{\boldsymbol{\theta}}(\alpha))}{\min_{\alpha' \in \mathcal{A}_n} L(\alpha'; \hat{\boldsymbol{\theta}}(\alpha'))} \xrightarrow{P} 1,$$

for any $\alpha \in \mathcal{A}_n^0$, indicating that the asymptotic loss efficiency can be achieved for any correct model.

5. Examples. In this section, we provide some specific examples for GIC that satisfy regularity conditions (C1)–(C5). Throughout this section, we assume that $p_n = p$, $\mathcal{A}_n = \mathcal{A}$, $\mathcal{A}_n^0 = \mathcal{A}^0$ and $\alpha_n^0 = \alpha^0$ are fixed, and give proofs of the theoretical results in the supplemental material [Chang, Huang and Ing (2014)].

5.1. *One-dimensional examples.* First, we consider spatial models in the one-dimensional space with $D = [0, n^\delta] \subseteq \mathbb{R}$; $\delta \in [0, 1)$. We assume the exponential covariance model for $\eta(\cdot)$,

$$\operatorname{cov}\big(\eta(s), \eta(s^*)\big) = \sigma^2 \exp\big(-\kappa |s - s^*|\big); \qquad s, s^* \in D, \tag{5.1}$$

where $\sigma^2 > 0$ is the variance parameter, and $\kappa > 0$ is a spatial dependence parameter. We also assume that the data are uniformly sampled at $s_i = in^{-(1-\delta)}$; $i = 1, \dots, n$, $s_i \in D$. Clearly, $\delta = 0$ corresponds to the fixed domain asymptotic framework with $D = [0, 1]$, and a larger δ corresponds to a faster growth rate of the domain. Note that $\sigma^2\kappa$ is often referred to as a microergodic parameter under fixed domain asymptotics [Stein (1999)].

The following proposition allows us to replace (C1)–(C5) in Theorems 3.1 and 3.2 by simpler conditions.

PROPOSITION 5.1. *Consider* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *in* (2.3), *where* $\boldsymbol{\Sigma}_\eta$ *is given by* (5.1) *and* $s_i = in^{-(1-\delta)}$; $i = 1, \dots, n$, *for some* $\delta \in [0, 1)$. *Let* $\boldsymbol{\theta} = (v^2, \sigma^2, \kappa)'$. *Then for any compact set* $\Theta \subseteq (0, \infty)^3$ *and any* $\boldsymbol{\theta}_0 = (v_0^2, \sigma_0^2, \kappa_0)' \in \Theta$,

$$\begin{aligned} 0 &< \liminf_{n \to \infty} \inf_{\boldsymbol{\theta} \in \Theta} \lambda_{\min}\big(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\big) \\ &\le \limsup_{n \to \infty} \sup_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}\big(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\big) < \infty. \end{aligned} \tag{5.2}$$

PROOF. The proof follows directly from Proposition 2.1 of Chang, Huang and Ing (2013). □

THEOREM 5.1. *Consider the data generated from* (2.1) *and the model given by* (2.2) *and* (2.3) *with* $\boldsymbol{\theta}_0$ *being the true covariance parameter vector* [*i.e.*, $\mathrm{var}(\mathbf{Z}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$]. *Assume the setup of Proposition* 5.1 *with* $\delta \in (0, 1)$. *Suppose that* $\hat{\boldsymbol{\theta}}(\alpha) \xrightarrow{P} \boldsymbol{\theta}_\alpha$ *for some* $\boldsymbol{\theta}_\alpha \in \Theta$; $\alpha \in \mathcal{A}$, *and*

$$\min_{\alpha \in \mathcal{A}\setminus\mathcal{A}^0} R(\alpha; \boldsymbol{\theta}_\alpha) \to \infty, \tag{5.3}$$

as $n \to \infty$. *Then* $\hat{\alpha}_2$ *is asymptotically loss efficient if* $|\mathcal{A}^0| \le 1$. *In addition, suppose that* $\lim_{n\to\infty} \tau_n = \infty$ *and* $\tau_n = o(\min_{\alpha \in \mathcal{A}\setminus\mathcal{A}^0} R(\alpha; \boldsymbol{\theta}_\alpha))$.

(i) *If* $|\mathcal{A}^0| = 0$, *then* $\hat{\alpha}_{\tau_n}$ *is asymptotically loss efficient.*
(ii) *If* $|\mathcal{A}^0| \ge 1$, *then* $\hat{\alpha}_{\tau_n}$ *is consistent.*

REMARK 5.1. The assumption, $\hat{\boldsymbol{\theta}}(\alpha) \xrightarrow{P} \boldsymbol{\theta}_\alpha$; $\alpha \in \mathcal{A}$, is generally satisfied under the increasing domain asymptotic framework, and is guaranteed to hold when $R(\alpha; \boldsymbol{\theta}_0) = o(n^\delta)$, for all $\alpha \in \mathcal{A} \setminus \mathcal{A}^0$; see Theorem 2.3 of Chang, Huang and Ing (2013). In fact, as given by Theorems 5.2–5.4, the assumption continues to hold even if $R(\alpha; \boldsymbol{\theta}_0) > cn^\delta$ for $\alpha \in \mathcal{A} \setminus \mathcal{A}^0$ and some constant $c > 0$.

Although the theorem is established under the increasing domain asymptotic framework, the theorem remains valid in some situations even when $\hat{\boldsymbol{\theta}}(\alpha)$ fails to converge for some $\alpha \in \mathcal{A}$ under the fixed domain asymptotic framework with $\delta = 0$. As mentioned at the end of Section 2.1, our asymptotic results of GIC are still valid for random $\mathbf{X}$. In what follows, we provide three examples based on different classes of regressors that are either random or fixed. We derive the consistency of GIC not only for $\delta \in (0, 1)$ but also for $\delta = 0$ without requiring the regularity conditions. The three examples below can be seen to have increasing degrees of smoothness in space, leading to different conditions to ensure the consistency of GIC.

EXAMPLE 5.1 (White-noise processes). Consider p regressors, $x_j(\cdot)$; $j = 1, \dots, p$, generated from independent white-noise processes with

$$x_j(s) \sim N(0, v_j^2); \qquad s \in [0, n^\delta], j = 1, \dots, p,$$

for some $\delta \in [0, 1)$, where $v_j^2 > 0$; $j = 1, \dots, p$.

EXAMPLE 5.2 (Spatially dependent processes). Consider p regressors, $x_j(\cdot)$; $j = 1, \dots, p$, generated from independent zero-mean Gaussian spatial processes with covariance functions

$$\mathrm{cov}(x_j(s), x_j(s')) = \sigma_j^2 \exp(-\kappa_j |s - s'|); \qquad s, s' \in [0, n^\delta],$$

for some $\delta \in [0, 1)$, where $\sigma_j^2, \kappa_j > 0$; $j = 1, \dots, p$.

EXAMPLE 5.3 (Monomials). Consider p regressors, $x_j(\cdot)$; $j = 1, \ldots, p$,

$$x_j(s) = n^{-\delta j} s^j; \qquad s \in [0, n^{\delta}],$$

for some $\delta \in [0, 1)$. Note that a scaling factor $n^{-\delta j}$ is introduced to standardize $x_j(\cdot)$ so that $\frac{1}{n^{\delta}} \int_0^{n^{\delta}} (x_j(s) - \bar{x}_j)^2 \, ds$ does not depend on n, where $\bar{x}_j = \frac{1}{n^{\delta}} \int_0^{n^{\delta}} x_j(s) \, ds$.

THEOREM 5.2. *Consider the model defined in* (2.2) *with the white-noise regressors given by Example* 5.1. *Suppose that* $\mathbf{Z} \sim N(\mathbf{X}\boldsymbol{\beta}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$ *conditional on* $\mathbf{X}$, *where* $\boldsymbol{\beta}_0 = (\beta_{0,0}, \ldots, \beta_{0,p})' \in \mathbb{R}^{p+1}$ *and* $\boldsymbol{\theta}_0 = (v_0^2, \sigma_0^2, \kappa_0)' \in \Theta \subseteq (0, \infty)^3$ *are constant vectors, and* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ *is given by Proposition* 5.1 *for some* $\delta \in [0, 1)$. *Assume that* Θ *is compact and*

$$\boldsymbol{\theta}_0 + \biggl(\sum_{j \in \alpha^0 \setminus \alpha} \beta_{0,j}^2 v_j^2, 0, 0 \biggr)' \in \Theta; \qquad \alpha \in \mathcal{A}.$$

If $\lim_{n \to \infty} \tau_n = \infty$ *and* $\tau_n = o(n)$, *then* $\lim_{n \to \infty} P\{\hat{\alpha}_{\tau_n} = \alpha^0\} = 1$.

REMARK 5.2. Theorem 5.2 assumes $\mathcal{A}^0 \neq \varnothing$. Suppose that $\mu_0(\cdot)$ has an additional unobserved term $\zeta(\cdot)$, which is also a white-noise process,

$$\mu_0(s) = \beta_{0,0} + \sum_{j=1}^{p} \beta_{0,j} x_j(s) + \zeta(s); \qquad s \in D, \tag{5.4}$$

and hence $|\mathcal{A}^0| = 0$. Then by Theorem 5.1 and an argument similar to that in proof of Theorem 5.2 for $\delta = 0$, GIC is also asymptotically loss efficient for $\delta \in [0, 1)$, provided that $\lim_{n \to \infty} \tau_n = \infty$ and $\tau_n = o(n)$.

THEOREM 5.3. *Consider the model defined in* (2.2) *with the spatially dependent regressors given by Example* 5.2. *Suppose that* $\mathbf{Z} \sim N(\mathbf{X}\boldsymbol{\beta}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$ *conditional on* $\mathbf{X}$, *where* $\boldsymbol{\beta}_0 = (\beta_{0,0}, \ldots, \beta_{0,p})' \in \mathbb{R}^{p+1}$ *and* $\boldsymbol{\theta}_0 = (v_0^2, \sigma_0^2, \kappa_0)' \in \Theta \subseteq (0, \infty)^3$ *are constant vectors, and* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ *is given by Proposition* 5.1 *with* $\delta \in [0, 1)$. *Assume that* Θ *is compact and* $\boldsymbol{\theta}_0 + (0, \sum_{j \in \alpha^0 \setminus \alpha} \beta_{0,j}^2 \sigma_j^2, \kappa_\alpha^*)' \in \Theta$ *for any* $\alpha \in \mathcal{A}$, *where* $\kappa_\alpha^* = (\sigma_0^2 + \sum_{j \in \alpha^0 \setminus \alpha} \beta_{0,j}^2 \sigma_j^2)^{-1} (\sum_{j \in \alpha^0 \setminus \alpha} \beta_{0,j}^2 \sigma_j^2 (\kappa_j - \kappa_0))$. *If* $\lim_{n \to \infty} \tau_n = \infty$ *and* $\tau_n = o(n^{(1+\delta)/2})$, *then* $\lim_{n \to \infty} P\{\hat{\alpha}_{\tau_n} = \alpha^0\} = 1$.

REMARK 5.3. Theorem 5.3 assumes $\mathcal{A}^0 \neq \varnothing$. Suppose that $\mu_0(\cdot)$ is given by (5.4), where $\zeta(\cdot)$ is an unobserved spatial dependent process given in Example 5.2. Then by Theorem 5.1 and an argument similar to that in proof of Theorem 5.3 for $\delta = 0$, GIC is also asymptotically loss efficient for $\delta \in [0, 1)$, provided that $\lim_{n \to \infty} \tau_n = \infty$ and $\tau_n = o(n^{(1+\delta)/2})$.

THEOREM 5.4. *Consider the model defined in* (2.2) *with the monomial regressors given by Example* 5.3. *Suppose that* $\mathbf{Z} \sim N(\mathbf{X}\boldsymbol{\beta}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$, *where* $\boldsymbol{\beta}_0 = (\beta_{0,0}, \dots, \beta_{0,p})' \in \mathbb{R}^{p+1}$ *and* $\boldsymbol{\theta}_0 = (v_0^2, \sigma_0^2, \kappa_0)' \in \Theta \subseteq (0, \infty)^3$ *are constant vectors, and* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ *is given by Proposition* 5.1 *with* $\delta \in (0, 1)$. *Assume that* $\mathcal{A} = \{\varnothing, \{1\}, \{1, 2\}, \dots, \{1, \dots, p\}\}$, Θ *is compact, and* $\boldsymbol{\theta}_0 + (0, \gamma(k), -(\sigma_0^2 + \gamma(k))^{-1}\gamma(k)\kappa_0)' \in \Theta$; $k = 0, 1, \dots, p$, *where* $\gamma(k) = \boldsymbol{\beta}_0'\mathbf{V}_{p,p}\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0'\mathbf{V}_{p,k}\mathbf{V}_{k,k}^{-1}\mathbf{V}_{k,p}\boldsymbol{\beta}_0$ *and* $\mathbf{V}_{k,p} = (\frac{1}{i+j-1})_{(k+1)\times(p+1)}$. *If* $\lim_{n\to\infty} \tau_n = \infty$ *and* $\tau_n = o(n^\delta)$, *then* $\lim_{n\to\infty} P\{\hat{\alpha}_{\tau_n} = \alpha^0\} = 1$.

REMARK 5.4. Theorem 5.4 assumes $\mathcal{A}^0 \neq \varnothing$. Suppose that $\mu_0(\cdot)$ is given by (5.4), where $\zeta(s) = n^{-\delta k}s^k$; $s \in D$, is an unobserved function with $k > p$. Then by Theorem 5.1, GIC can be shown to be asymptotically loss efficient for $\delta \in (0, 1)$, provided that $\lim_{n\to\infty} \tau_n = \infty$ and $\tau_n = o(n^\delta)$.

The results of Theorems 5.2–5.4 show that the consistency of GIC depends on not only the smoothness of regressors in space but also the growth rate of the domain. Evidently, GIC is more difficult to identify the true model when the candidate regressors are smoother in space. Although there exists τ_n such that GIC is consistent for either white-noise regressors or spatially dependent regressors under the fixed domain asymptotic framework, interestingly, as shown in the next theorem, consistent polynomial order selection turns out not possible when the true model has at least one nonzero regression coefficient and $|\mathcal{A}^0| \geq 2$ under the fixed domain asymptotic framework.

THEOREM 5.5 (Inconsistency). *Consider the same setup as in Theorem* 5.4, *except that* $\delta = 0$:

(i) *If* $\lim_{n\to\infty} \tau_n = \infty$, *then* $\lim_{n\to\infty} P\{\hat{\alpha}_{\tau_n} = \{\varnothing\}\} = 1$.
(ii) *If* $\alpha^0 \neq \{\varnothing\}$ *and* $\liminf_{n\to\infty} \tau_n > 0$, *then* $\lim_{n\to\infty} P\{\hat{\alpha}_{\tau_n} = \alpha^0\} < 1$.

5.2. *A two-dimensional exponential model.* Consider the multiplicative exponential covariance model

$$\operatorname{cov}(\eta(\mathbf{s}), \eta(\mathbf{s}^*)) = \sigma^2 \exp(-\kappa\{|s_1 - s_1^*| + |s_2 - s_2^*|\}), \tag{5.5}$$

parameterized by $\sigma^2 > 0$ and $\kappa > 0$, where $\mathbf{s} = (s_1, s_2)$ and $\mathbf{s}^* = (s_1^*, s_2^*) \in D = [0, n^{\delta/2}]^2 \subseteq \mathbb{R}^2$; $\delta \in [0, 1)$. Clearly, $\delta = 0$ corresponds to the fixed domain asymptotic framework with $D = [0, 1]^2$, and a larger δ corresponds to a faster growth rate of the domain.

Similarly to the one-dimensional case, we first prove (5.2), which is the key to show (C1)–(C5).

PROPOSITION 5.2. *Consider* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *in* (2.3) *with* $\boldsymbol{\Sigma}_\eta$ *given by* (5.5), $v^2 = 0$, *and* $\mathbf{s}_k = (im^{-(1-\delta)}, jm^{-(1-\delta)})$; $k = i + (j-1)m$; $i, j = 1, \dots, m$, *for some integer* $m = n^{1/2}$, *where* $\delta \in [0, 1)$. *Let* $\boldsymbol{\theta} = (\sigma^2, \kappa)'$. *Then* (5.2) *holds for any compact set* $\Theta \subseteq (0, \infty)^2$ *and any* $\boldsymbol{\theta}_0 = (\sigma_0^2, \kappa_0)' \in \Theta$.

PROOF. Write

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \mathbf{B}(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta}), \tag{5.6}$$

where $\mathbf{B}(\boldsymbol{\theta}) = (\rho^{|i-j|})_{m\times m}$ and $\rho = \exp(-\kappa m^{-(1-\delta)})$. By (5.6),

$$\begin{aligned}
&\lambda_{\max}\big(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta})\big) \\
&\quad \le \frac{\sigma_0^2}{\sigma^2}\lambda_{\max}\big(\big(\mathbf{B}(\boldsymbol{\theta}_0) \otimes \mathbf{B}(\boldsymbol{\theta}_0)\big)\big(\mathbf{B}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{B}^{-1}(\boldsymbol{\theta})\big)\big) \\
&\quad = \frac{\sigma_0^2}{\sigma^2}\lambda_{\max}\big(\big(\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{B}^{-1}(\boldsymbol{\theta})\big) \otimes \big(\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{B}^{-1}(\boldsymbol{\theta})\big)\big) \\
&\quad = \frac{\sigma_0^2}{\sigma^2}\lambda_{\max}^2\big(\big(\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{B}^{-1}(\boldsymbol{\theta})\big)\big) < \infty,
\end{aligned}$$

where the last inequality follows from Proposition 2.1 of Chang, Huang and Ing (2013). This gives the last inequality of (5.2). The proof for the first inequality of (5.2) is analogous and omitted. This completes the proof. □

THEOREM 5.6. *Consider the data generated from* (2.1), *the model given by* (2.2) *and* (2.3) *and the setup of Proposition* 5.2 *with* $\delta \in [0, 1)$. *Suppose that* $\hat{\boldsymbol{\theta}}(\alpha) \xrightarrow{P} \boldsymbol{\theta}_\alpha$ *for some* $\boldsymbol{\theta}_\alpha \in \Theta$; $\alpha \in \mathcal{A}$, *and* (5.3) *holds. Then* $\hat{\alpha}_2$ *is asymptotically loss efficient if* $|\mathcal{A}^0| \le 1$. *In addition, suppose that* $\lim_{n\to\infty} \tau_n = \infty$ *and* $\tau_n = o(\min_{\alpha\in\mathcal{A}\setminus\mathcal{A}^0} R(\alpha; \boldsymbol{\theta}_\alpha))$.

(i) *If* $|\mathcal{A}^0| = 0$, *then* $\hat{\alpha}_{\tau_n}$ *is asymptotically loss efficient.*
(ii) *If* $|\mathcal{A}^0| \ge 1$, *then* $\hat{\alpha}_{\tau_n}$ *is consistent.*

REMARK 5.5. As in the one-dimensional case, the assumption, $\hat{\boldsymbol{\theta}}(\alpha) \xrightarrow{P} \boldsymbol{\theta}_\alpha$; $\alpha \in \mathcal{A} \setminus \mathcal{A}^0$, is generally satisfied. In fact, the assumption is guaranteed to hold when $R(\alpha; \boldsymbol{\theta}_0) = o(n^{(1+\delta)/2})$, for any $\alpha \in \mathcal{A}$; see Lemma A.5 of Chang, Huang and Ing (2014).

Here we consider only a multiplicative exponential model because of two difficulties. First, for the two-dimensional exponential covariance model, the asymptotic distribution of the ML estimate of $(\sigma^2\kappa, \kappa)'$ is needed but has yet to be derived unless κ is assumed known [Du, Zhang and Mandrekar (2009), Wang and Loh (2011)]. Second, our proof relies on a decomposition of the log-likelihood into different layers having different orders of magnitude. Such a decomposition requires an innovative treatment of the log-likelihood for the two-dimensional exponential model. Further research is needed to characterize the asymptotic behavior of GIC under the two-dimensional exponential covariance model or the more general Matérn covariance model [Matérn (1986)], but is beyond the scope of this paper.

6. Summary and discussion. In this article, we study the asymptotic properties of GIC for geostatistical model selection regardless of whether the covariance model is correct or wrong, and establish conditions under which GIC is consistent and asymptotically loss efficient. Some specific examples that satisfy the regularity conditions are also provided. To the best of our knowledge, this research is the first to provide such results for GIC in geostatistical regression model selection.

The method we developed also sheds some light for solving linear mixed model selection problems involving parameters that cannot be estimated consistently. For example, consider a simple Laird–Ware model [Laird and Ware (1982)],

$$Z_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \eta_i + \epsilon_{ij}; \qquad i = 1, \dots, m, \ j = 1, \dots, n_i, \tag{6.1}$$

where $\mathbf{x}_{ij}$'s are p-vector of fixed effects, and $\eta_i \sim N(0, \sigma^2)$ is the random effect for subject i, independent of $\epsilon_{ij} \sim N(0, v^2)$. Here $\boldsymbol{\beta} \in \mathbb{R}^p$ is the regression-coefficient vector, and $\boldsymbol{\theta} = (\sigma^2, v^2)'$ consists of random-effect parameters. Clearly, σ^2 in (6.1) cannot be estimated consistently when m is fixed [Longford (2000)]. Nevertheless, as shown below, it is still possible to derive a condition analogous to (C2). For simplicity, we consider a simple case of (6.1) with mean zero and no fixed effect. Let $\hat{\boldsymbol{\theta}}$ be the ML estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0 = (v_0^2, \sigma_0^2)'$ be the true parameter value. Applying an argument similar to that used to prove (2.10) of Chang, Huang and Ing (2013), twice the negative log-likelihood of (6.1) can be written as

$$-2\ell(\boldsymbol{\theta}) = n\log(2\pi) + \sum_{j=1}^{m} \log n_j + n \log v^2 + n\frac{v_0^2}{v^2} + h(\boldsymbol{\theta}) + O_p(1), \tag{6.2}$$

where $n = \sum_{i=1}^m n_i$, $h(\boldsymbol{\theta}) = \sum_{i=1}^m \{\boldsymbol{\epsilon}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i - \mathrm{E}(\boldsymbol{\epsilon}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i)\}$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{i,n_i})'$ and $\boldsymbol{\Sigma}_j = \sigma^2\mathbf{1}_{n_j}\mathbf{1}_{n_j}' + v^2\mathbf{I}_{n_j}$. We shall show that $\ell(\hat{\boldsymbol{\theta}}) = \ell(\boldsymbol{\theta}_0) + O_p(1)$. Applying an argument similar to that used to prove Theorem 2.2 in Chang, Huang and Ing (2013),

$$\hat{\boldsymbol{\theta}} = (v_0^2, \sigma_0^2)' + \big(O_p(n^{-1/2}), O_p(1)\big)'. \tag{6.3}$$

Let $\Theta_n = \{\boldsymbol{\theta} \in \Theta : |\sigma^2 - \sigma_0^2| < M, |v^2 - v_0^2| \le Mn^{-1/2}\}$ for any constant $M > 0$. By Lemma B.1 of Chan and Ing (2011) and an argument similar that used to prove (2.12) in Chang, Huang and Ing (2013), we have

$$\begin{aligned}
&\mathrm{E}\Big(\sup_{\boldsymbol{\theta}\in\Theta_n} |h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}_0)|^2\Big) \\
&\quad \le \sup_{\boldsymbol{\theta}\in\Theta_n} \left\{(v^2 - v_0^2)^2 \operatorname{var}\Big(\frac{\partial}{\partial v^2} h(\boldsymbol{\theta})\Big) + (\sigma^2 - \sigma_0^2)^2 \operatorname{var}\Big(\frac{\partial}{\partial \sigma^2} h(\boldsymbol{\theta})\Big)\right\} \\
&\quad = O(1),
\end{aligned}$$

which implies $h(\hat{\boldsymbol{\theta}}) - h(\boldsymbol{\theta}_0) = O_p(1)$. This together with (6.2) and (6.3) gives $\ell(\hat{\boldsymbol{\theta}}) = \ell(\boldsymbol{\theta}_0) + O_p(1)$, indicating some possibility to establish the asymptotic theory of GIC for the Laird–Ware model, even when some random-effect parameter cannot be consistently estimated.

TABLE 1
Frequencies of models selected by BIC based on 100 *simulation replicates, where* $\varnothing$ *denotes the intercept only model and* α^0 *denotes the correct model*

	$\mu(s) = 1 + s^2 \sin(\pi/s)$		$\mu(s) = 1 + s \sin(\pi/s)$	
n	$\varnothing$	α^0	$\varnothing$	α^0
100	67	33	38	62
500	66	34	23	77
1000	76	24	8	92

In this article, we focus only on variable selection under a certain covariance model. Clearly, simultaneous selection of both variables and covariance models is an interesting problem that deserves further investigation. Although we believe that the framework we developed in this article can be generalized to this problem, it will require introducing more complex notation.

In addition, some more efforts are needed to completely characterize GIC, even for the exponential covariance model in one dimension. We note that both the candidate regressors in Examples 5.1 and 5.2 are not of bounded variation (BV), whereas the polynomial regressors given by Example 5.3 are BV functions. It is of interest to know if BV plays an important role. We conducted a small test simulation experiment under the setup of (2.2) with only one regressor $x(\cdot)$ and $v^2 = 0.5$, where $\mu(s) = 1 + x(s)$, $\eta(\cdot)$ is given by (5.1) with $\sigma^2 = 0.5$ and $\kappa = 1$, and data are sampled at $\{1/n, 2/n, \dots, 1\}$. We consider two functions for $x(\cdot)$, which are $f_1(s) = s^2 \sin(\pi/s)$ and $f_2(s) = s \sin(\pi/s)$, in combination with three different sample sizes ($n = 100, 500, 1000$). Note that $f_1(\cdot)$ is of bounded variation on [0, 1], and $f_2(\cdot)$ is not. The results based on 100 simulation replicates with known σ^2, κ and v^2 are shown in Table 1. Clearly, GIC has better ability in identifying the correct model when $f_2(\cdot)$, rather than $f_1(\cdot)$, is used as the regressor, which partially supports that BV may be an important factor.

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotic theory of generalized information criterion for geostatistical regression model selection" (DOI: 10.1214/14-AOS1258SUPP; .pdf). The supplement materials contain the proofs of all theorems in Section 5.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) (B. N. Petrov and F. Csáki, eds.) 267–281. Akadémiai Kiadó, Budapest. MR0483125

BUNEA, F. (2004). Consistent covariate selection and post model selection inference in semiparametric regression. *Ann. Statist.* **32** 898–927. MR2065193

CHAN, N. H. and ING, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39** 1526–1550. MR2850211

CHANG, C.-H., HUANG, H.-C. and ING, C.-K. (2013). Asymptotic properties of maximum likelihood estimators in geostatistical regression models. Technical report 2013-02, Institute of Statistical Science, Academia Sinica, Taiwan. Available at http://www3.stat.sinica.edu.tw/library/c_tec_rep/2013-02_fulltext-1030630.pdf.

CHANG, C., HUANG, H. and ING, C. (2014). Supplement to "Asymptotic theory of generalized information criterion for geostatistical regression model selection." DOI:10.1214/14-AOS1258SUPP.

CHEN, H.-S., SIMPSON, D. G. and YING, Z. (2000). Infill asymptotics for a stochastic process model with measurement error. *Statist. Sinica* **10** 141–156. MR1742105

DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. MR2549562

HOETING, J. A., DAVIS, R. A., MERTON, A. A. and THOMPSON, S. E. (2006). Model selection for geostatistical models. *Ecological Applications* **16** 87–98.

HUANG, H.-C. and CHEN, C.-S. (2007). Optimal geostatistical model selection. *J. Amer. Statist. Assoc.* **102** 1009–1024. MR2411661

ING, C.-K. (2007). Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series. *Ann. Statist.* **35** 1238–1277. MR2341705

ING, C.-K. and WEI, C.-Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474. MR2211091

JIANG, J. and RAO, J. S. (2003). Consistent procedures for mixed linear model selection. *Sankhyā* **65** 23–42. MR2016775

JIANG, J., RAO, J. S., GU, Z. and NGUYEN, T. (2008). Fence methods for mixed model selection. *Ann. Statist.* **36** 1669–1692. MR2435452

JONES, R. H. (2011). Bayesian information criterion for longitudinal and clustered data. *Stat. Med.* **30** 3050–3056. MR2851399

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.

LI, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15** 958–975. MR0902239

LONGFORD, N. T. (2000). On estimating standard errors in multilevel analysis. *The Statistician* **49** 389–398.

MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. MR0738334

MATÉRN, B. (1986). *Spatial Variation*, 2nd ed. *Lecture Notes in Statistics* **36**. Springer, Berlin. MR0867886

MCQUARRIE, A. D. R. and TSAI, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific, River Edge, NJ. MR1641582

MÜLLER, S., SCEALY, J. L. and WELSH, A. H. (2013). Model selection in linear mixed models. *Statist. Sci.* **28** 135–167. MR3112403

NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12** 758–765. MR0740928

PU, W. and NIU, X.-F. (2006). Selecting mixed-effects models based on a generalized information criterion. *J. Multivariate Anal.* **97** 733–758. MR2236499

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SHAO, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. MR1466682

SHI, P. and TSAI, C.-L. (1999). Semiparametric regression model selections. *J. Statist. Plann. Inference* **77** 119–139. MR1677812

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York. MR1697409

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **36** 111–147. MR0356377

VAIDA, F. and BLANCHARD, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92** 351–370. MR2201364

WANG, D. and LOH, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Stat.* **5** 238–269. MR2792553

WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305. MR0133849

YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36** 280–296. MR1096671

ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

C.-H. CHANG
INSTITUTE OF STATISTICS
NATIONAL UNIVERSITY OF KAOHSIUNG
700 KAOHSIUNG UNIVERSITY ROAD
KAOHSIUNG 811
TAIWAN
E-MAIL: jhow@stat.sinica.edu.tw

H.-C. HUANG
C.-K. ING
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
128 ACADEMIA ROAD SECTION 2
TAIPEI 115
TAIWAN
E-MAIL: hchuang@stat.sinica.edu.tw
cking@stat.sinica.edu.tw
URL: http://www.stat.sinica.edu.tw/