國立交通大學

統計學研究所

博 士 論 文

空間統計模型選取之大樣本理論

Asymptotic Theory for Geostatistical Model Selection

研 究 生：張志浩

指導教授：黃信誠 博士

共同指導：銀慶剛 博士

中 華 民 國 一百 年 六 月

空間統計模型選取之大樣本理論
Asymptotic Theory for Geostatistical Model Selection

研 究 生：張志浩　　　　Student：Chih-Hao Chang

指導教授：黃信誠　　　　Advisor：Hsin-Cheng Huang

共同指導：銀慶剛　　　　Co-Advisor：Ching-Kang Ing

國 立 交 通 大 學
統 計 學 研 究 所

博 士 論 文

A Dissertation Submitted to

National Chiao Tung University, Hsinchu, Taiwan

for the Degree of

Doctor of Philosophy in Institute of Statistics

June 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年六月

空間統計模型選取之大樣本理論

學生：張志浩　　　　　　　　　　　　　　　　　　指導教授：黃信誠

共同指導：銀慶剛

國立交通大學 統計學研究所 博士班

摘　　要

在傳統迴歸模型中，模型選取的大樣本理論已被廣泛建立。然而在空間統計迴歸模型中，使用傳統模型選取準則的選模結果並未被完善的討論及研究，尤其當假設資料觀測空間為一固定區域而不隨著樣本增加而放大時，其大樣本理論可以預期會與傳統的理論結果有所差異。論文中，我們在一些常規假設下，建立了傳統模型選取準則的大樣本理論。而後在一維空間的一些例子下，我們發現這些常規假設的成立與否不僅與樣本空間放大的速度有關，也與所選取變數在空間中的平滑程度有緊密關係。當空間互變異函數參數未知時，我們同樣發現，參數估計及傳統模型選取準則的大樣本理論，也與樣本空間放大的速度和所選取變數在空間中的平滑程度有關。最後我們執行有限樣本的模擬實驗，並得到與大樣本理論一致的結果。

Asymptotic Theory for Geostatistical Model Selection

student：Chih-Hao Chang

Advisors：Dr. Hsin-Cheng Huang
Co-Advisor：Dr. Ching-Kang Ing

Institute of Statistics
National Chiao Tung University

ABSTRACT

Information criteria, such as Akaike's information criterion (AIC), Bayesian information criterion (BIC), and conditional AIC (CAIC) are often applied in model selection. However, their asymptotic behaviors under 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 thesis, we investigate two classes of criteria for geostatistical model selection: generalized information criterion (GIC) and conditional GIC (CGIC), which include AIC, BIC, and CAIC as special cases, under both the increasing domain asymptotic and fixed domain asymptotic frameworks. We establish conditions under which GIC and CGIC are selection consistent and asymptotically efficient even without assuming spatial covariance structure to be known. These conditions are further examined for GIC and CGIC in selecting one-dimensional geostatistical regression models with the exponential covariance function class under various settings. For example, under the fixed domain asymptotic framework, where some covariance parameters are not consistently estimable, we show that selection consistency not only depends on the tuning parameter of GIC, but also depends on smoothness of the explanatory variables in space. In addition, under the increasing domain framework, we show that asymptotic properties of GIC depend on the growing rates for the size of the domain. Moreover, some numerical experiments are provided to demonstrate the finite sample behavior of various criteria.

誌　謝

回首交通大學多年的求學生涯，首先我要感謝的是黃信誠老師，在學術研究的漫漫長路上，給了我很多大方向小細節的指導與叮嚀，也感謝老師付出了這麼多的勞力與心力，一一糾正我散漫不經心的錯誤，為論文劃下了一個完善的句點，並諄諄教誨的讓我謹記不至於在未來的路途上犯下重複的錯誤，以期在研究的路上能走得更遠更踏實。同時要感謝銀慶剛老師，在研究的方向與證明的細節上，不時的給予了非常即時的指導與建議，完善且充實了論文中的內容，並讓我在研究過程中少走了許多彎路，得到了更多的啟發，在此特別感謝兩位老師所給予的指導。

另外感謝在最後的口試當天，對於論文提出了寶貴的意見與建議的口試委員： 徐南蓉教授、王維菁教授、洪慧念教授、林培生教授，進而修補了論文中的不足之處。同時感謝所上的老師，再學校的授課過程中，讓我累積了足以通過資格考並進而得到了博士學位的學識。

而在學校的日子中，與同學們的相處，有一起考過資格考的喜、一起沒考過資格考的哀，也有一起舉杯八卦的樂，感謝秋婷、江村剛志、穗碧、達叔、弘家、家群，還有最後給予我很大幫忙的健霖，或許我們離開學校的日子不一樣，但我會珍惜我們一起相伴過的日子。也感謝在中研院擔任助理時，春樹學長的提點，幫助我更快融入中研院的環境。最後特別感謝郭碧芬郭姐，如果沒有郭姐，我可能會忘記考入學考，可能會忘記註冊，真的很感謝郭姐不時的即時提醒。

最後，我要感謝我的家人，感謝爸爸、媽媽、大姐、二姐、小弟，還有兩位姐夫，對於我的支持與包容，我知道妳們都相信我可以，只是不好意思讓大家等這麼久，謹將此本論文與我的博士學位獻給我愛與愛我的爸媽，謝謝你們長久以來的一路栽培，也希望在未來的人生大道中，在你們的相伴與敦促下，繼續的分享我的酸甜苦辣。

張志浩 謹識於

國立交通大學統計學研究所

中華民國一百年六月

Contents

List of Tables

List of Figures

Chapter 1

Introduction

More and more spatial data are collected in this world. In many problems, several variables are measured at some locations over a region in space, and it is of interest to predict a variable at some locations, where measurements may or may not be taken, based on all data available in the region. We can formulate the problem as a geostatistical regression problem by treating the variable of interest as the response and regressing it with other (explanatory) variables while accounting for spatial dependencies. However, inference and prediction generally depend on how explanatory variables are chosen, which if not chosen properly, may lead to poor inference and prediction, particularly when the number of explanatory variables is large. Clearly, model selection problem is essential in geostatistics.

For example, suppose that we are interested in knowing ground-level ozone concentrations for a region consisting of Taoyuan, Hsinchu and Miaoli in Taiwan based on data measured at some monitoring stations in Taiwan Air Quality Monitoring Network (see Figure 1.1). At each monitoring station, we collect hourly ozone concentrations together with some explanatory variables, including ozone precursors (such as nitrogen oxides and hydrocarbons), meteorological variables (such as wind speed and wind direction, temperature, mixing height, humidity and rainfall), altitude, population, etc. It is of interest to identify influential explanatory variables for ozone concentrations, and to predict ozone concentrations for the whole region by applying a geostatistical regression model. Although lower altitudes, lower wind speeds, and higher temperatures are expected to associate with higher ozone concentrations, some other variables may or may not have effects on ozone concentrations. Removing unrelated variables, while retaining important variables, will allow one to reduce estimation variability, thereby increase prediction accuracy.

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. It is known that these two frameworks lead to possibly different asymptotic behaviors on covariance parameter estimation, and hence are also expected to produce different asymptotic behaviors on model selection. In general, asymptotic behaviors under the increasing domain asymptotic framework are more standard. For example, the maximum likelihood (ML) estimates of covariance parameters are typically consistent and asymptotically normal (Mardia and Marshall 1984). In contrast, not all covariance parameters can be consistently estimated under the fixed domain asymptotic framework even for a simple one-dimensional example with the stationary exponential covariance model (Ying 1991). The readers are refereed to Stein

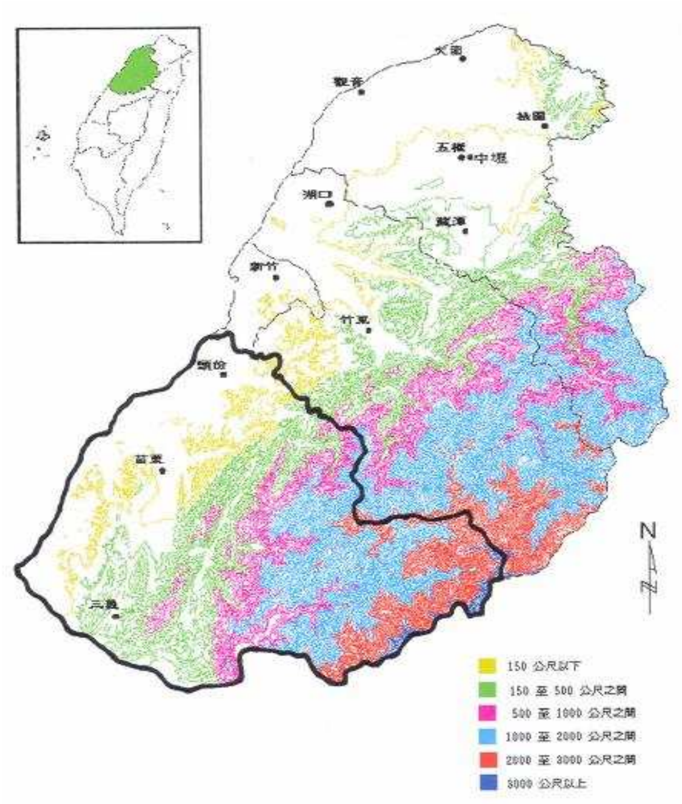


Figure 1.1: Locations of monitoring stations in Taoyuan, Hsinchu and Miaoli counties in Taiwan Air Quality Monitoring Network.

(1999) for more details regarding fixed domain asymptotics.

There are many model selection methods that have been applied in geostatistical model selection, such as Akaike's information criterion (AIC, Akaike 1973), Bayesian information criterion (BIC, Schwartz 1978), the generalized information criterion (GIC, Nishii 1984), and cross validation. Note that GIC contains a tuning parameter, which includes both AIC and BIC as special cases. Although asymptotic properties of these selection methods have been well studied in linear regression and time series model selection (e.g., Shao 1997; McQuarrie and Tsai 1989), they have not been well established for geostatistical model selection. In fact, there are only limited results available partly because asymptotic properties under the fixed domain asymptotic framework are generally nonstandard and difficult to handle. Hoeting *et al.* (2006) provided some heuristic arguments for AIC in geostatistical model selection under the assumption that the variable of interest is observed with no measurement error. They show via a simulation experiment that spatial dependence has to be considered, which if ignored, may lead to unsatisfactory results. For linear mixed-effect models, Pu and Niu (2006) provided conditions under which GIC is selection consistent. In addition, Vaida and Blanchard (2005) developed a criterion for linear mixed model selection, called the conditional Akaike's information criterion (CAIC). This criterion provides unbiased estimation of the mean squared prediction error, which appears to be more suitable than AIC for geostatistical model selection when spatial prediction is of main interest. Huang and Chen (2007) developed a general technique of estimating the mean squared prediction error for a general spatial prediction procedure, in which a concept called generalized degrees of freedom is used to provide an almost unbiased estimate. Their method is applicable to select among arbitrary spatial prediction methods, and is shown to achieve some asymptotic efficiency result.

In this thesis, we first study GIC for geostatistical model selection. Then we propose a new criterion, called conditional GIC (CGIC), which includes CAIC as a special case.

Major accomplishments are listed in the following:

1. Asymptotic properties of GIC under both the fixed domain asymptotic and the increasing domain asymptotic frameworks are established under some regularity conditions.

2. Asymptotic properties of CGIC under both the fixed domain asymptotic and the increasing domain asymptotic frameworks are established under some regularity conditions.

3. The above regularity conditions are explicitly checked for some examples in the one-dimensional space with various forms of explanatory variables under the exponential covariance model corresponding to the Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930).

We shall show that asymptotic behaviors of these criteria are related to how fast the domain increases with the sample size. In addition, some nonstandard behaviors of these criteria under the fixed domain asymptotic framework will be highlighted. For example, under the fixed domain asymptotic framework, GIC and CGIC fail to identify the correct set of polynomial variables consistently regardless of which tuning parameters are chosen. On the other hand, both BIC and CBIC are selection consistent when candidate variables are generated from either white-noise processes or some zero-mean spatial dependent processes.

We shall start by developing asymptotic results under known covariance parameters, and then allowing them to be unknown. However, under the fixed domain asymptotic framework, ML estimates may converge to nondegenerate distributions. In this situation, general asymptotic properties are very difficult to develop even for parameter estimation. Therefore, we shall focus only on some examples of geostatistical models defined over the one-dimensional space with the exponential covariance model.

The thesis is organized as follows. Chapter 2 gives a brief introduction of geostatistics, including various spatial covariance models, various spatial prediction and parameter estimation methods. In Chapter 3, we introduce the variable selection problem and consider two loss functions for comparing among different methods. In Chapter 4, some asymptotic properties of GIC are established under some regularity conditions. These conditions are further verified by some examples in the one-dimensional space with the exponential covariance model under either known or unknown covariance parameters in Chapter 5. Chapter 6 is devoted to CGIC and its asymptotic properties. Chapter 7 provides some simulation examples for comparing among various model selection criteria. Some conclusions and discussion regarding future research directions are provided in Chapter 8. Finally, the appendix contains proofs for all lemmas and propositions.

Chapter 2

Geostatistics

This chapter provides a brief introduction to geostatistics, including geostatistical models, spatial prediction, parameter estimation, and the two asymptotic frameworks.

2.1 Geostatistical Models

Consider a spatial process $\{S(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ of interest defined over a region $D \subset \mathbb{R}^d$ with $d \in \mathbb{N} \equiv \{1, 2, \dots\}$. Suppose that we observe data $\{(\boldsymbol{x}(\boldsymbol{s}_i), Z(\boldsymbol{s}_i)) : i = 1, \dots, n\}$ at locations $\boldsymbol{s}_i \in D$, where

$$\boldsymbol{x}(\boldsymbol{s}_i) = (1, x_1(\boldsymbol{s}_i), \dots, x_p(\boldsymbol{s}_i))', \tag{2.1}$$

is a p-vector of explanatory variables observed at $\boldsymbol{s}_i \in D$, and $Z(\boldsymbol{s}_i)$ is the corresponding response variable observed according to the following measurement equation:

$$Z(\boldsymbol{s}_i) = S(\boldsymbol{s}_i) + \epsilon(\boldsymbol{s}_i); \quad i = 1, \dots, n,$$

and $\{\epsilon(\boldsymbol{s}_i) : i = 1, \dots, n\}$ are white-noise variables corresponding to measurement errors with variance σ_ϵ^2. The spatial process $S(\cdot)$ is further decomposed into a linear combination of explanatory variables $\boldsymbol{x}(\cdot)'\boldsymbol{\beta}$ and a zero-mean spatial dependent process $\eta(\cdot)$:

$$Z(\boldsymbol{s}_i) = \boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta} + \eta(\boldsymbol{s}_i) + \epsilon(\boldsymbol{s}_i); \quad \boldsymbol{s}_i \in D, \ i = 1, \dots, n. \tag{2.2}$$

In general, $\eta(\cdot)$ is assumed to be L^2-continuous (i.e., $\mathrm{E}(\eta(\boldsymbol{s}) - \eta(\boldsymbol{s}'))^2 \to 0$ as $\|\boldsymbol{s} - \boldsymbol{s}'\| \to 0$) with its spatial dependence structure described by a variogram model or a covariance model (Section 2.2). The goal is either to make inference on $\boldsymbol{\beta}$ or more often to predict $\{S(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ based on data $\boldsymbol{Z} \equiv (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$. Commonly used loss functions for spatial prediction include $\int_{\boldsymbol{s} \in D} \left|\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s})\right|^2 d\boldsymbol{s}$ and $\sum_{i=1}^{n} \left(\hat{S}(\boldsymbol{s}_i) - S(\boldsymbol{s}_i)\right)^2$, where $\hat{S}(\boldsymbol{s})$ denotes a generic predictor of $S(\boldsymbol{s})$ at $\boldsymbol{s} \in D$.

2.2 Variograms and Covariance Functions

In geostatistical literature, spatial dependence is commonly described using a variogram, defined as

$$2\gamma^*(\boldsymbol{s}, \boldsymbol{s}') \equiv \mathrm{E}(Z(\boldsymbol{s}) - Z(\boldsymbol{s}'))^2; \quad \boldsymbol{s}, \boldsymbol{s}' \in D.$$

The function $\gamma^*(\cdot,\cdot)$ is usually called the semivariogram. Clearly, $\gamma^*(\boldsymbol{s},\boldsymbol{s}') \geq 0$ and $\gamma^*(\boldsymbol{s},\boldsymbol{s}) = 0$, for $\boldsymbol{s},\boldsymbol{s}' \in D$. A spatial process $S(\cdot)$ is said to be intrinsically stationary if it has a constant mean: $\mathrm{E}(S(\boldsymbol{s}+\boldsymbol{h}) - S(\boldsymbol{s})) = 0$, and its variogram can be written as

$$2\gamma(\boldsymbol{h}) \equiv 2\gamma^*(\boldsymbol{s}+\boldsymbol{h},\boldsymbol{s}) = \mathrm{E}(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2,$$

for any pairs $\boldsymbol{s}$ and $\boldsymbol{s}+\boldsymbol{h} \in D$. Note that the function $2\gamma(\boldsymbol{h})$ does not depend on $\boldsymbol{s}$.

Spatial dependence can also be described using a covariance function:

$$C(\boldsymbol{s},\boldsymbol{s}') \equiv \mathrm{cov}(S(\boldsymbol{s}), S(\boldsymbol{s}')); \quad \boldsymbol{s},\boldsymbol{s}' \in D.$$

Similar to an intrinsically stationary process, a spatial process $S(\cdot)$ is said to be second-order stationary if $\mathrm{E}(S(\boldsymbol{s}+\boldsymbol{h}) - S(\boldsymbol{s})) = 0$ and its covariance function can be written as:

$$K(\boldsymbol{h}) = C(\boldsymbol{s},\boldsymbol{s}+\boldsymbol{h}) = \mathrm{cov}(S(\boldsymbol{s}), S(\boldsymbol{s}+\boldsymbol{h})),$$

independent of $\boldsymbol{s}$, for any $\boldsymbol{s}$, $\boldsymbol{s}+\boldsymbol{h} \in D$. Clearly, $K(\boldsymbol{h}) = K(-\boldsymbol{h})$, and $|K(\boldsymbol{h})| \leq K(\boldsymbol{0}) = \mathrm{var}(S(\boldsymbol{s}))$, for $\boldsymbol{s} \in D$. Note that a second-order stationary process is an intrinsically stationary process with $\gamma(\boldsymbol{h}) = K(0) - K(\boldsymbol{h})$, but not necessary *vice versa*. For example, a Brownian motion is an intrinsically stationary process, but not a stationary process, because its variance is not a constant (see Cressie 1993).

In addition, a second-order stationary process is said to be isotropic if $K(\boldsymbol{h})$ depends on $\boldsymbol{h}$ only through $\|\boldsymbol{h}\|$, where $\|\cdot\|$ is the L^2 norm. In what follows, we introduce some isotropic stationary covariance function classes commonly used in the literature. The exponential covariance class is given by:

$$K(\boldsymbol{h}) = \sigma_\eta^2 \exp(-\kappa_\eta \|\boldsymbol{h}\|), \tag{2.3}$$

where $\sigma_\eta^2 > 0$ and $\kappa_\eta \geq 0$ is a range parameter. The Gaussian covariance class is given by:

$$K(\boldsymbol{h}) = \sigma_\eta^2 \exp(-\kappa_\eta \|\boldsymbol{h}\|^2), \tag{2.4}$$

where $\sigma_\eta^2 > 0$ and $\kappa_\eta \geq 0$ is a range parameter. The Matérn covariance class (Matérn 1986) is given by:

$$K(\boldsymbol{h}) = \frac{\sigma_\eta^2 (\kappa \|\boldsymbol{h}\|)^\nu}{\Gamma(\nu) 2^{\nu-1}} \mathcal{K}_\nu(\kappa_\eta \|\boldsymbol{h}\|), \tag{2.5}$$

where $\sigma_\eta^2 > 0$, $\kappa_\eta \geq 0$ is a range parameter, $\nu > 0$ is a smoothness parameter, and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind with order ν (Abramowitz and Stegun 1965). Note that the Matérn covariance class contains the exponential covariance class as a special case when $\nu = 0.5$. It also reduces to the Gaussian covariance class as $\nu \to \infty$.

2.3 Kriging and Spatial Prediction

Spatial prediction is commonly called kriging in geostatistics, which utilizes spatial dependence structure to interpolate or smooth the surface of a spatial stochastic process based on (noisy) data observed at some locations in space. It is named after a South African mining engineer, Daniel Gerhardus Krige (1951), who pioneered the field of geostatistics. In this section, we are going to introduce several kriging methods derived under different circumstances.

Simple Kriging

Suppose that we observe data $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$ according to (2.2), where σ_ϵ^2, $\mu(\boldsymbol{s}) = \mathrm{E}(S(\boldsymbol{s}))$; $\boldsymbol{s} \in D$, and $C(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{cov}(S(\boldsymbol{s}), S(\boldsymbol{s}'))$; $\boldsymbol{s}, \boldsymbol{s}' \in D$, are known. Then the predictor $\hat{S}(\boldsymbol{s})$ that minimizes the mean squared error, $\mathrm{E}(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s}))^2$ is $\hat{S}(\boldsymbol{s}) = \mathrm{E}(S(\boldsymbol{s})|\boldsymbol{Z})$, for $\boldsymbol{s} \in D$. This predictor is called the simple kriging predictor. If in addition, $S(\cdot)$ and $\{\epsilon(\boldsymbol{s}_1), \dots, \epsilon(\boldsymbol{s}_n)\}$ are both Gaussian. Then $\hat{S}(\boldsymbol{s})$ is a linear predictor and can be explicitly written as:

$$\hat{S}(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}), \tag{2.6}$$

where $\boldsymbol{\mu} \equiv (\mu(\boldsymbol{s}_1), \dots, \mu(\boldsymbol{s}_n))'$, $\boldsymbol{\sigma} \equiv (C(\boldsymbol{s}_1, \boldsymbol{s}), \dots, C(\boldsymbol{s}_n, \boldsymbol{s}))'$, $\boldsymbol{\Sigma} \equiv \left[C(\boldsymbol{s}_i, \boldsymbol{s}_j)\right]_{n \times n} + \sigma_\epsilon^2 \boldsymbol{I}_n$, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Under the Gaussian assumption, the kriging variance (or the mean squared prediction error) of $\hat{S}(\boldsymbol{s})$ at $\boldsymbol{s} \in D$ is

$$\mathrm{E}\big(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s})\big)^2 = C(\boldsymbol{s}, \boldsymbol{s}) - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}.$$

Ordinary Kriging

Suppose that we observe data $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$ according to (2.2), with a constant mean $\mu = \mathrm{E}(S(\boldsymbol{s}))$; $\boldsymbol{s} \in D$, which is unknown, whereas σ_ϵ^2 and $C(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{cov}(S(\boldsymbol{s}), S(\boldsymbol{s}'))$; $\boldsymbol{s}, \boldsymbol{s}' \in D$, are known. Then the best linear unbiased predictor of $S(\boldsymbol{s})$, which minimizes $\mathrm{E}(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s}))^2$ among all unbiased linear predictors is given by:

$$\hat{S}(\boldsymbol{s}) = \hat{\mu} + \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \hat{\mu}\mathbf{1}),$$

where $\hat{\mu} \equiv (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\Sigma}$ are defined in (2.6), and $\mathbf{1} \equiv (1, \dots, 1)'$. The predictor is usually called the ordinary kriging (OK) predictor. The ordinary kriging variance is

$$\mathrm{E}\big(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s})\big)^2 = C(\boldsymbol{s}, \boldsymbol{s}) - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + \frac{(1 - \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma})^2}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

Universal Kriging

Instead of having a constant mean as in the ordinary kriging model, suppose that we observe data $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$ according to (2.2) with mean,

$$\mu(\boldsymbol{s}) = \mathrm{E}(S(\boldsymbol{s})) = \sum_{k=0}^{p} \beta_k x_k(\boldsymbol{s}),$$

where $x_j(\cdot)$'s are known function corresponding to explanatory variables in (2.1) and β_j's are unknown regression coefficients. Here σ_ϵ^2 and $C(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{cov}(S(\boldsymbol{s}), S(\boldsymbol{s}'))$; $\boldsymbol{s}, \boldsymbol{s}' \in D$, are assumed known. Then the best linear unbiased predictor, usually called the universal kriging (UK) predictor, has the following linear form:

$$\hat{S}(\boldsymbol{s}) = \sum_{i=1}^{n} \omega_i Z(\boldsymbol{s}_i),$$

which minimizes $\mathrm{E}(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s}))^2$ subject to the following $p+1$ unbiasedness constraints:

$$\sum_{i=1}^{n} \omega_i x_j(\boldsymbol{s}_i) = x_j(\boldsymbol{s}); \quad j = 0, 1, \dots, p.$$

Let $\boldsymbol{X}$ be the $n \times (p+1)$ matrix with the ith row given by $\boldsymbol{x}(\boldsymbol{s}_i)$; $1 \leq i \leq n$, defined in (2.1). Then the UK predictor is given by:

$$\hat{S}(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\hat{\boldsymbol{\beta}} + \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}$ is the generalized least square estimate of $\boldsymbol{\beta}$, and $\boldsymbol{\sigma}$ and $\boldsymbol{\Sigma}$ are defined in (2.6). In particular, the UK predictor of $\boldsymbol{S} \equiv (S(\boldsymbol{s}_1), \ldots, S(\boldsymbol{s}_n))'$ is

$$\hat{\boldsymbol{S}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \mathrm{var}(\boldsymbol{S})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{2.7}$$

Note that OK is a special case of UK with $p = 0$. The UK kriging variance satisfies

$$\begin{aligned} \mathrm{E}\big(\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s})\big)^2 &= C(\boldsymbol{s}, \boldsymbol{s}) - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + \boldsymbol{\sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} \\ &\quad + \boldsymbol{x}(\boldsymbol{s})'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}(\boldsymbol{s}) - 2\boldsymbol{x}(\boldsymbol{s})'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}. \end{aligned}$$

Other Kriging Methods

The above kriging methods assume that the covariance parameters are known. When they are unknown, one may plug the estimated parameters into the expressions of the corresponding kriging predictors. Another solution is to apply a Bayesian approach by specifying a joint prior distribution for all the unknown parameters. Then a Bayesian kriging method is obtained by using either the posterior mean or the posterior median of $S(\cdot)$ as a predictor of $S(\cdot)$.

When either $S(\cdot)$ is not a Gaussian process or $\epsilon(\boldsymbol{s}_i)$'s are not Gaussian distributed, the optimal predictor, $\mathrm{E}(S(\boldsymbol{s})|\boldsymbol{Z})$ of $S(\boldsymbol{s})$ that minimizes $\mathrm{E}\|\hat{S}(\boldsymbol{s}) - S(\boldsymbol{s})\|^2$ is generally nonlinear and has a complex form. Under this situation, some nonlinear kriging methods, such as transGaussin kriging, disjunctive kriging, and indicator kriging have been developed. The readers are referred to Journel (1983), Cressie (1993), or Schabenberger and Gotway (2005) for more details.

2.4 Covariance Parameter Estimation

The kriging methods introduced previously basically require knowing σ_ϵ^2 and the variogram (or covariance function) of $S(\cdot)$. In practice, they are generally unknown and have to be estimated. To visualize the spatial dependence structure, it is common to plot the following empirical variogram at various spatial lags $h > 0$:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2,$$

where $N(h)$ denotes all the pairs of $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \approx h$, and $|N(h)|$ denotes the number of elements in $N(h)$. However, the empirical variogram cannot be computed at every lag distance due to limited amounts of data. It is common to estimate the variogram (or covariance function) of $S(\cdot)$ by specifying a parametric model after looking at the empirical variogram.

Hereafter, we consider a covariance model parameterized by $\boldsymbol{\theta}$. Denote $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to be the variance-covariance matrix of $\boldsymbol{Z}$ based on parameter $\boldsymbol{\theta}$. Let

$$f(\boldsymbol{Z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-1/2} \exp(-(\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})/2), \tag{2.8}$$

be the Gaussian density function with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Given $\boldsymbol{\theta}$, the ML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are given by

$$\hat{\beta}(\boldsymbol{\theta}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),$$

and $\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}) \equiv \boldsymbol{X}\hat{\beta}(\boldsymbol{\theta})$. Therefore, the ML estimate of $\boldsymbol{\theta}$ can be obtained by maximizing the profile log-likelihood function:

$$\begin{aligned} \ell(\boldsymbol{\theta}; \boldsymbol{Z}) &= \log f(\boldsymbol{Z}; \hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det \boldsymbol{\Sigma}(\boldsymbol{\theta})) - \frac{1}{2}(\boldsymbol{Z} - \boldsymbol{X}\hat{\beta}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{Z} - \boldsymbol{X}\hat{\beta}(\boldsymbol{\theta})). \end{aligned} \tag{2.9}$$

Alternatively, we can estimate $\boldsymbol{\theta}$ by restricted maximum likelihood (REML), obtained by maximizing the likelihood of some contrasts $\boldsymbol{Z}^{\dagger} = \boldsymbol{A}\boldsymbol{Z}$ such that $\boldsymbol{A}\boldsymbol{X} = 0$, where $\boldsymbol{A}$ is a $(n-p-1) \times n$ matrix with rank $n-p-1$, which can be chosen as $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$. Then the REML estimate of $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function of $\boldsymbol{Z}^{\dagger}$:

$$\begin{aligned} \log f\big(\boldsymbol{Z}^{\dagger}; \boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{A}'\big) &= -\frac{(n-p-1)}{2}\log(2\pi) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) \\ &\quad -\frac{1}{2}\log\det((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X})^{-1}) - \frac{1}{2}\boldsymbol{Z}^{\dagger\prime}(\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{A}')^{-1}\boldsymbol{Z}^{\dagger}. \end{aligned}$$

The covariance parameter vector $\boldsymbol{\theta}$ can also be estimated by some methods of moments, which have an advantage of not relying on the Gaussian assumption. For more details, the readers are referred to Cressie (1993) and Schabenberger and Gotway (2005).

2.5 Asymptotic Frameworks

There are two asymptotic frameworks in geostatistics having different assumptions on the domain D. One is called the fixed domain asymptotic framework, where data are sampled more and more densely in a bounded fixed region D. The other is called the increasing domain asymptotic framework with $|D| \to \infty$ as $n \to \infty$, which is often considered in time series analysis. The fixed domain asymptotic framework is somewhat unique in geostatistics, which tends to have some unusual asymptotic behavior due to limited information available in a bounded fixed region.

Asymptotic properties under the increasing domain asymptotic framework are more standard. Suppose that we observe data $\boldsymbol{Z}$ according to (2.2), where $\mu(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}$ is known, but $\mathrm{var}(\boldsymbol{Z})$ depends on some unknown parameter vector $\boldsymbol{\theta}$. Then Mardia and Marshall (1984) show under some regularity conditions that

$$\hat{\boldsymbol{\theta}}_{ML} \sim N(\boldsymbol{\theta}_0, I^{-1}(\boldsymbol{\theta}_0)), \tag{2.10}$$

where $\boldsymbol{\theta}_0$ is the true parameter vector and $I(\boldsymbol{\theta}_0)$ is the Fisher information. However, (2.10) is generally not satisfied under the fixed domain asymptotic framework, and in fact some parameters of $\boldsymbol{\theta}$ can not be consistently estimated. For example, suppose that $\eta(\cdot)$ is generated from a Matérn covariance function of (2.5) with ν known but σ_η^2 and κ_η unknown. Zhang (2004) shows that the ML estimates of σ_η^2 and κ_η are inconsistent under the fixed domain asymptotic framework. That is,

$$\lim_{n\to\infty} P\big(|\hat{\sigma}_\eta^2 - \sigma_{\eta,0}^2| > \varepsilon\} > 0,$$

and

$$\lim_{n\to\infty} P\big(|\hat{\kappa}_\eta - \kappa_{\eta,0}| > \varepsilon\big\} > 0.$$

for any $\varepsilon > 0$, where $\sigma^2_{\eta,0}$ and $\kappa_{\eta,0}$ are the corresponding true parameters. However, as shown in the following proposition, some function of σ^2_η and κ_η can be consistently estimated.

Proposition 1 *(Zhang, 2004) Consider an increasing sequence of finite subsets D_n of $\mathbb{R}^d$, for $d = 1, 2, 3$, such that $\cup_{n=1}^{\infty} D_n$ is bounded and infinite. Suppose that the data $\boldsymbol{Z}$ are observed on $D = D_n$ according to (2.2) with $\boldsymbol{\beta} = \mathbf{0}$ and $\sigma^2_\epsilon = 0$ known, where $\eta(\cdot)$ is a Gaussian process with a Matérn covariance function of (2.5) and $\nu > 0$ is known. Assume that $\sigma^2_{\eta,0} > 0$ and $\kappa_{\eta,0} > 0$ are the true parameters corresponding to σ^2 and κ_η. If κ_η is fixed at some constant $\kappa_1 > 0$, and $\hat{\sigma}^2_\eta$ is the ML estimate of σ^2. Then*

$$\hat{\sigma}^2_\eta \kappa_1^{2\nu} \xrightarrow{p} \sigma^2_{\eta,0}\kappa^{2\nu}_{\eta,0}, \quad \text{as } n \to \infty.$$

Also, Ying (1991) shows the similar results for exponential covariance function which is a special case of Matérn class for $\nu = 0.5$.

Proposition 2 *(Ying, 1991) Suppose that the data $\boldsymbol{Z}$ are observed on $D = [0, 1]$ according to (2.2) with $\boldsymbol{\beta} = \mathbf{0}$ and $\sigma^2_\epsilon = 0$ known, where $\eta(\cdot)$ is a zero-mean Gaussian process with an exponential covariance function of (2.3). Let Θ be the parameter space of $(\sigma^2_\eta, \kappa_\eta)'$. Assume that either $\Theta = [a, b] \times (0, \infty)$ or $\Theta = (0, \infty) \times [a, b]$, where $0 < a \leq b < \infty$, and the true parameter vector $(\sigma^2_{\eta,0}, \kappa_{\eta,0})' \in \Theta$.*

(i) Let $\hat{\sigma}^2_\eta$ and $\hat{\kappa}_\eta$ be the ML estimates of σ^2_η and κ_η. Then

$$\sqrt{n}(\hat{\sigma}^2_\eta \hat{\kappa}_\eta - \sigma^2_{\eta,0}\kappa_{\eta,0}) \xrightarrow{d} N(0, 2(\sigma^2_{\eta,0}\kappa_{\eta,0})^2), \quad \text{as } n \to \infty.$$

(ii) Suppose that κ_η is fixed at some constant $\kappa_1 > 0$ and $\hat{\sigma}^2_1$ is the corresponding ML estimate of σ^2_η. Then

$$\sqrt{n}\left(\hat{\sigma}^2_1 - \frac{\sigma^2_{\eta,0}\kappa_{\eta,0}}{\kappa_1}\right) \xrightarrow{d} N\left(0, 2\left(\frac{\sigma^2_{\eta,0}\kappa_{\eta,0}}{\kappa_1}\right)^2\right), \quad \text{as } n \to \infty.$$

(iii) Suppose that σ^2_η is fixed at some constant σ^2_1 and $\hat{\kappa}_1$ is the corresponding ML estimate of κ_η. Then

$$\sqrt{n}\left(\hat{\kappa}_1 - \frac{\sigma^2_{\eta,0}\kappa_{\eta,0}}{\sigma^2_1}\right) \xrightarrow{d} N\left(0, 2\left(\frac{\sigma^2_{\eta,0}\kappa_{\eta,0}}{\sigma^2_1}\right)^2\right), \quad \text{as } n \to \infty.$$

The parameters $\sigma^2_\eta \kappa^{2\nu}_\eta$ in Proposition 1 and $\sigma^2_\eta \kappa_\eta$ in Proposition 2 are called microergodic parameters (Matheron 1971, 1989; Stein 1999), which basically imply that both parameters can be recovered with probability 1 from observations in a bounded fixed region. These parameters have also been shown to play an important role in spatial prediction by Stein (1999). Specifically, consider the spectral density function of $K(\boldsymbol{h})$, $\boldsymbol{h} \in \mathbb{R}^d$:

$$f(\omega) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} exp(-i\omega'\boldsymbol{h})K(\boldsymbol{h})d\boldsymbol{h}; \quad \omega \in \mathbb{R}^d.$$

Stein shows that under the fixed domain asymptotic framework, $f(\omega)$ contributes to mean square prediction error mainly for large $|\omega|$, whose behavior is governed by some microergodic parameters. He also provides some specific examples for exponential and Matérn covariance functions.

For $\sigma_\epsilon^2 > 0$ in (2.2), Chen *et al.* (2000) provides the following results regarding the ML estimates of σ_η^2, κ_η and σ_ϵ^2.

Proposition 3 *(Chen et al. 2000) Suppose that the data $\boldsymbol{Z}$ are observed regularly on $D = [0,1]$ according to (2.2) with $\boldsymbol{\beta} = \mathbf{0}$ known, where $\eta(\cdot)$ is a zero-mean Gaussian process with an exponential covariance function of (2.3). Assume that $(\sigma_\eta, \kappa_\eta, \sigma_\epsilon^2)' \in \Theta$, where $\Theta \subset (0,\infty)^3$ is a compact set, and the true parameter vector $(\sigma_{\epsilon,0}^2, \sigma_{\eta,0}^2, \kappa_{\eta,0})' \in \Theta$.*

(i) Let $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_\eta^2$ and $\hat{\kappa}_\eta$ be the ML estimates of σ_ϵ^2, σ_η^2 and κ_η. Then, as $n \to \infty$,

$$\begin{pmatrix} n^{1/4}(\hat{\sigma}_\eta^2 \hat{\kappa}_\eta - \sigma_{\eta,0}^2 \kappa_{\eta,0}) \\ n^{1/2}(\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\sqrt{2}\sigma_{\epsilon,0}(\sigma_{\eta,0}^2 \kappa_{\eta,0})^{3/2} & 0 \\ 0 & 2\sigma_{\epsilon,0}^4 \end{pmatrix} \right).$$

(ii) Suppose that κ_η is known and $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\eta^2$ are the corresponding ML estimates of σ_ϵ^2 and σ_η^2. Then, as $n \to \infty$,

$$\begin{pmatrix} n^{1/4}(\hat{\sigma}_\eta^2 - \sigma_{\eta,0}^2) \\ n^{1/2}(\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\sqrt{2}\sigma_{\epsilon,0}\sigma_{\eta,0}^3 \kappa_{\eta,0}^{-1/2} & 0 \\ 0 & 2\sigma_{\epsilon,0}^4 \end{pmatrix} \right).$$

In Chapter 5, we shall provide the convergence rates for the ML estimates of σ_ϵ^2, σ_η^2 and κ_η under more general spatial domains with $D = [0, n^\delta]$ and $\delta \in [0,1)$. In addition, those convergence rates will be given under geostatistical regression models of (2.2) based on not only the true model, but also underfitted and overfitted models.

Chapter 3

Variable Selection

Consider the geostatistical regression model of (2.2). Suppose that we observe spatial data, $\{\boldsymbol{x}(\boldsymbol{s}_i), Z(\boldsymbol{s}_i)\}$; $\boldsymbol{s}_i \in D$ and $i = 1, \ldots, n$. This model reduces to a usual regression model when $\eta(\cdot) = 0$. Similar to linear regression, a large model with many insignificant variables tends to produce a large variance, resulting in low predictive power. On the other hand, a small model that ignores some important variable may produce large bias. To achieve good compromise between bias and variance, it is essential to identify significant variables. Clearly, variable selection is essential not only in regression but also in geostatistical regression.

We consider selecting a subset of $\{1, \ldots, p\}$ corresponding to p explanatory variables. Let $\mathcal{A} \subset 2^{\{1,\ldots,p\}}$ be the set of all candidate models, and let $\alpha \in \mathcal{A}$ denotes a candidate model. Note that intercept is always included in our models, and $\alpha = \emptyset$ corresponds to the intercept only model.

Let $\boldsymbol{X}(\alpha)$ be an $n \times p(\alpha)$ sub-matrix of $\boldsymbol{X}$ containing the columns corresponding to α, and let $\boldsymbol{\beta}(\alpha)$ be the sub-vector of $\boldsymbol{\beta}$ corresponding to $\boldsymbol{X}(\alpha)$. A model α is said to be correct if $\mu(\boldsymbol{s})$ can be written as $\sum_{j\in\alpha} \beta_j x_j(\boldsymbol{s})$, for $\boldsymbol{s} \in D$. Let $\mathcal{A}^c \subset \mathcal{A}$ be the set of all correct models and let $\alpha^c = \underset{\alpha\in\mathcal{A}}{\arg\min}\, |\alpha|$ be the correct model having the smallest number of variables. Then $\mathcal{A}^c = \{\alpha \in \mathcal{A} : \alpha^c \subset \alpha\}$.

The geostatistical regression model corresponding to $\alpha \in \mathcal{A}$ can be written in a matrix form as:

$$\boldsymbol{Z} = \boldsymbol{X}(\alpha)\boldsymbol{\beta}(\alpha) + \boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{3.1}$$

where $\boldsymbol{\eta} \equiv (\eta(\boldsymbol{s}_1), \ldots, \eta(\boldsymbol{s}_n))' \sim N(0, \boldsymbol{\Sigma}_\eta)$ and $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \boldsymbol{I})$. Hence the mean and the variance of $\boldsymbol{Z}$ under model $\alpha \in \mathcal{A}$ are $\boldsymbol{\mu}(\alpha) = \boldsymbol{X}(\alpha)\boldsymbol{\beta}(\alpha)$ and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_\eta + \sigma_\epsilon^2 \boldsymbol{I}, \tag{3.2}$$

where $\boldsymbol{\theta}$ is the covariance parameter vector associate with var$(\boldsymbol{Z})$.

3.1 Loss Functions

We consider two loss functions: the Kullback-Leibler (KL) loss function and the squared error loss function. First, for model α given in (3.1), the KL loss function is given by:

$$\begin{aligned} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \int_{\boldsymbol{Y}\in\mathbb{R}^n} f(\boldsymbol{Y};\boldsymbol{\mu},\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\log\frac{f(\boldsymbol{Y};\boldsymbol{\mu},\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))}{f(\boldsymbol{Y};\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}),\boldsymbol{\Sigma}(\boldsymbol{\theta}))}d\boldsymbol{Y} \\ &= \frac{1}{2}\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - \frac{n}{2} \\ &\quad +\frac{1}{2}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})), \end{aligned} \tag{3.3}$$

where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{Z})$ is the true mean vector and $\boldsymbol{\theta}_0$ is the true covariance parameter vector,

$$\begin{aligned} \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}) &= \boldsymbol{X}(\alpha)\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta}), \qquad (3.4)\\ \hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta}) &= (\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{Z}, \end{aligned}$$

and recall that $f(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the Gaussian density function defined in (2.8). Now, let

$$\boldsymbol{M}(\alpha;\boldsymbol{\theta}) = \boldsymbol{X}(\alpha)(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}), \tag{3.5}$$

$$\boldsymbol{A}(\alpha;\boldsymbol{\theta}) = \boldsymbol{I} - \boldsymbol{M}(\alpha;\boldsymbol{\theta}). \tag{3.6}$$

Note that when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})$ in (3.3) reduces to a simpler form:

$$L^{\mathrm{KL}}(\alpha) \equiv L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_0) = \frac{1}{2}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}(\alpha)), \tag{3.7}$$

where $\hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ are written as $\hat{\boldsymbol{\mu}}(\alpha)$ and $\boldsymbol{\Sigma}$ to simplify their notations. We can rewrite (3.7) as

$$L^{\mathrm{KL}}(\alpha) = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \frac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}), \tag{3.8}$$

where $\boldsymbol{A}(\alpha;\boldsymbol{\theta}_0)$ and $\boldsymbol{M}(\alpha;\boldsymbol{\theta}_0)$ are also simplified as $\boldsymbol{A}(\alpha)$ and $\boldsymbol{M}(\alpha)$. Clearly, the first term $\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu}$ on the righthand side of the equality in (3.8) vanishes when $\alpha \in \mathcal{A}^c$. Thus we have the following lemma.

Lemma 1 *Consider a class of models given by (3.1). Let $L^{KL}(\alpha)$ be the KL loss for model α defined in (3.7). Then*

$$E(L^{KL}(\alpha)) = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \frac{p(\alpha)}{2}; \quad \alpha \in \mathcal{A}, \tag{3.9}$$

where $\boldsymbol{A}(\alpha)$ is defined in (3.6). In particular, $E(L^{KL}(\alpha)) = p(\alpha)/2$, for $\alpha \in \mathcal{A}^c$.

Lemma 2 *Consider a class of models given by (3.1). Let $L^{KL}(\alpha)$ be the KL loss for model α defined in (3.7). Then*

$$\lim_{n\to\infty} P\big(\alpha^c = \arg\min_{\alpha\in\mathcal{A}^c} L^{KL}(\alpha)\big) = 1, \tag{3.10}$$

and

$$\alpha^c = \arg\min_{\alpha\in\mathcal{A}^c} E(L^{KL}(\alpha)). \tag{3.11}$$

In addition, if α^c is fixed, and

$$\lim_{n\to\infty} \inf_{\alpha\in\mathcal{A}\backslash\mathcal{A}^c} \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} = \infty, \tag{3.12}$$

where $\boldsymbol{A}(\alpha)$ is defined in (3.6), then

$$\lim_{n\to\infty} P\big(\alpha^c = \arg\min_{\alpha\in\mathcal{A}} L^{KL}(\alpha)\big) = 1. \tag{3.13}$$

In general, (3.12) is satisfied under the increasing domain asymptotic framework. However, under the fixed domain asymptotic framework, it may or may not be satisfied; see Theorem 9 in Section 5.2 and Theorem 12 in Section 5.3, for which (3.12) holds and Theorems 5 and 6 in Section 5.1 for which (3.12) fails. In fact, as shown in Theorems 5 and 6, the smallest true model α^c does not have the smallest KL loss under the fixed domain asymptotic framework. In other words, (3.13) is not always satisfied.

The other loss function we consider in this thesis is the squared error loss commonly used in geostatistics particularly for prediction purpose:

$$L(\alpha) = \|\hat{\boldsymbol{S}}(\alpha) - \boldsymbol{S}\|^2, \tag{3.14}$$

where $\hat{\boldsymbol{S}}(\alpha)$ is a generic predictor of $\boldsymbol{S}$ based on model $\alpha \in \mathcal{A}$. Throughout the thesis, we consider the universal kriging predictor of $\boldsymbol{S}$ in (2.7) unless indicated otherwise. For $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the universal kriging predictor based on model α can be written as:

$$\hat{\boldsymbol{S}}(\alpha) = \boldsymbol{H}(\alpha)\boldsymbol{Z}, \tag{3.15}$$

where

$$\boldsymbol{H}(\alpha) = \boldsymbol{M}(\alpha) + \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha), \tag{3.16}$$

with $\boldsymbol{M}(\alpha)$ and $\boldsymbol{A}(\alpha)$ defined in (3.5) and (3.6), respectively. Then the corresponding risk can be decomposed into the following:

$$\begin{aligned} \mathrm{E}(L(\alpha)) &= \mathrm{E}\|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z}) - \hat{\boldsymbol{S}}(\alpha) + \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 \\ &= \mathrm{E}\|\hat{\boldsymbol{S}}(\alpha) - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 - 2\mathrm{E}((\hat{\boldsymbol{S}}(\alpha) - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z}))'(\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z}))) + \mathrm{E}\|\boldsymbol{S} - (\boldsymbol{S}|\boldsymbol{Z})\|^2 \\ &= \mathrm{E}\|\hat{\boldsymbol{S}}(\alpha) - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 + \mathrm{E}\|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2, \end{aligned}$$

which is lower bounded by $\mathrm{E}\|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2$, independent of $\alpha \in \mathcal{A}$. The following lemma provides some more details regarding decomposition of $E(L(\alpha))$, which is useful in establishing some asymptotic properties concerning the squared error loss.

Lemma 3 *Consider a class of models given by (3.1). Let $\hat{\boldsymbol{S}}(\alpha)$ be the UK predictor of $\boldsymbol{S}$ given by (3.15) and $L(\alpha)$ be the corresponding squared error loss defined in (3.14). Then*

$$\begin{aligned} E(L(\alpha)) &= E\|\hat{\boldsymbol{S}}(\alpha) - E(\boldsymbol{S}|\boldsymbol{Z})\|^2 + E\|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2 \\ &= R_1(\alpha) + R_2(\alpha) + \sigma_\epsilon^2 tr(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}), \end{aligned} \tag{3.17}$$

where $E\|\hat{\boldsymbol{S}}(\alpha) - E(\boldsymbol{S}|\boldsymbol{Z})\|^2 = R_1(\alpha) + R_2(\alpha)$,

$$\begin{aligned} R_1(\alpha) &= \sigma_\epsilon^4 \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}, \\ R_2(\alpha) &= \sigma_\epsilon^4 tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)), \end{aligned} \tag{3.18}$$

where $\boldsymbol{M}(\alpha) = \boldsymbol{M}(\alpha;\boldsymbol{\theta}_0)$ is defined in (3.5).

Note that the term $R_1(\alpha)$ corresponds to the model misspecification error, which is smaller for a larger model α, and in particular, $R_1(\alpha) = 0$ for $\alpha \in \mathcal{A}^c$. The term $R_2(\alpha)$ corresponds to the estimation error, which generally increases with $p(\alpha)$ and is bounded by $\sigma_\epsilon^2 p(\alpha)$, since

$$\begin{aligned}
\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) &= \sigma_\epsilon^2 \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\big) \\
&= \mathrm{tr}\Big(\big(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)\big)^{-1}\boldsymbol{X}(\alpha)'\big(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-2}\big)\boldsymbol{X}(\alpha)\Big) \\
&\leq \mathrm{tr}\Big(\big(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)\big)^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)\Big) \\
&= \mathrm{tr}(\boldsymbol{I}_{p(\alpha)}) = p(\alpha).
\end{aligned}$$

In addition, the term $\mathrm{E}\|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 = \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})$ in (3.17) corresponds to the optimal mean squared prediction error, which provides a lower bound for $\mathrm{E}(L(\alpha))$.

In general, $\lim_{n\to\infty} \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})\big/R_2(\alpha) = \infty$, for $\alpha \in \mathcal{A}^c$. It follows from (3.17) that

$$\lim_{n\to\infty} \mathrm{E}(L(\alpha))\big/\mathrm{E}(L(\alpha^c)) = 1, \quad \text{for } \alpha \in \mathcal{A}^c.$$

In contrast, from (3.9),

$$\lim_{n\to\infty} \mathrm{E}(L^{\mathrm{KL}}(\alpha))\big/\mathrm{E}(L^{\mathrm{KL}}(\alpha^c)) > 1, \quad \text{for } \alpha \in \mathcal{A}^c \setminus \{\alpha^c\}.$$

Therefore, it would be preferable to select α^c among $\alpha \in \mathcal{A}^c$ under the KL loss.

3.2 Consistency and Asymptotic Loss Efficiency

Suppose that we are given a class of models (3.1) and a model selection procedure $\hat{\alpha}$. We consider two aspects in assessing asymptotic optimal properties of $\hat{\alpha}$ with respect to a given loss function $L^*(\cdot)$. First, a selection procedure $\hat{\alpha}$ is said to be asymptotic loss efficiency if it satisfies

$$\operatorname*{plim}_{n\to\infty} L^*(\hat{\alpha})\Big/\inf_{\alpha\in\mathcal{A}} L^*(\alpha) = 1, \tag{3.19}$$

where plim denotes convergence in probability. Second, a selection procedure $\hat{\alpha}$ is said to be consistent if it satisfies

$$\lim_{n\to\infty} P\big(\hat{\alpha} = \alpha^c\} = 1.$$

For the KL loss of (3.7), it is straightforward to show that

$$\alpha^c = \operatorname*{arg\,min}_{\alpha\in\mathcal{A}^c} L^{\mathrm{KL}}(\alpha).$$

In some situations,

$$\alpha^c = \operatorname*{arg\,min}_{\alpha\in\mathcal{A}} L^{\mathrm{KL}}(\alpha). \tag{3.20}$$

In this case, consistency automatically implies asymptotic loss efficiency. For example, when $\sigma_\eta^2 = 0$, (3.1) becomes a class of traditional linear regression models with (3.20) being satisfied under some mild conditions (Shao 1997). However, the results given in Shao (1997) can't be easily generalized here, because as to be established in Chapters 4-6, asymptotic behavior of geostatistical model selection depends not only on asymptotic frameworks but also on some smoothness property of the explanatory variables in space.

According to (3.17), $E(L(\alpha))$ is lower bounded by $E\|\boldsymbol{S}-\mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2$, which is sometimes a higher order term than both $R_1(\alpha)$ and $R_2(\alpha)$. Under this situation, asymptotic loss efficiency of (3.19) with $L^*(\cdot) = L(\cdot)$ can be achieved by an arbitrary model selection procedure. Therefore, it seems natural and preferable to consider the loss function, $L(\alpha) - \|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2$, leading to another version of asymptotic loss efficiency that is stronger than (3.19).

Definition 1 *Consider a class of models given by (3.1) and the squared error loss, $L(\alpha) = \|\hat{\boldsymbol{S}}(\alpha) - \boldsymbol{S}\|^2$, where $\hat{\boldsymbol{S}}(\alpha)$ is a predictor of $\boldsymbol{S}$ based on model α. A selection procedure $\hat{\alpha}$ is said to be strongly asymptotically loss efficient with respect to the squared error loss if*

$$\operatorname*{plim}_{n\to\infty} \frac{L(\hat{\alpha}) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2}{\inf_{\alpha\in\mathcal{A}} L(\alpha) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2} = 1. \tag{3.21}$$

Chapter 4

Generalized Information Criterion

Consider a class of models given by (3.1). The generalized information criterion (GIC) introduced by Nishii (1984) is given by:

$$\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) = -2\,(\text{maximum log-likelihood}) + \lambda\,(\text{number of parameters}), \tag{4.1}$$

where λ is a tuning parameter, providing control of compromise between goodness-of-fit corresponding to maximum log-likelihood and the model parsimoniousness corresponding to the number of parameters. The criterion includes some commonly used criteria, such as Akaike's information criterion (AIC) with $\lambda = 2$ and Bayesian information criterion (BIC) with $\lambda = \log(n)$, as special cases, and has been widely used in many statistical fields. The model selected by $\Gamma_{\mathrm{GIC}(\lambda)}$ is given by:

$$\hat{\alpha}_{\mathrm{GIC}(\lambda)} \equiv \arg\min_{\alpha\in\mathcal{A}} \Gamma_{\mathrm{GIC}(\lambda)}(\alpha). \tag{4.2}$$

4.1 Akaike's Information Criterion

We shall first consider AIC with $\boldsymbol{\theta}$ known, which corresponds to $\lambda = 2$ in (4.1) and can be written as:

$$\Gamma_{\mathrm{AIC}}(\alpha) = (\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha)) + 2\,p(\alpha), \tag{4.3}$$

where $\hat{\boldsymbol{\mu}}(\alpha)$ is given by (3.4), and the goodness-of-fit component becomes the generalized squared errors, $(\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha))$, which is smaller for a larger model α, and has a χ^2 distribution with $n - p(\alpha)$ degrees of freedom if $\alpha \in \mathcal{A}^c$. The model selected by AIC is given by:

$$\hat{\alpha}_{\mathrm{AIC}} \equiv \arg\min_{\alpha\in\mathcal{A}} \Gamma_{\mathrm{AIC}}(\alpha). \tag{4.4}$$

The following theorem provides some asymptotic properties of AIC when $\boldsymbol{\theta}$ is known.

Theorem 1 *Consider a class of models given by (3.1). Let $L^{KL}(\alpha)$ be the KL loss for model α and $\hat{\alpha}_{AIC}$ be defined in (4.4).*

(i) For $|\mathcal{A}^c| \leq 1$, if

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{1}{E(L^{KL}(\alpha))} = 0, \tag{4.5}$$

then

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{AIC}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

(ii) For $|\mathcal{A}^c| \geq 2$, if (4.5) holds, and either

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

or

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c\setminus\{\alpha^c\}} \frac{1}{p(\alpha) - p(\alpha^c)} = 0,$$

then

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{AIC}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

Proof. The proof is essentially the same as that for Theorem 1 of Shao (1997) after transforming $\boldsymbol{Z}$ into $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Z}$. We therefore omit the details. □

Equation (4.5) provides a condition for risks associated with underfitted models so that correct models can be distinguished from incorrect models. The following corollary provides an example for which (4.5) is replaced by a simple condition that can be easily checked.

Corollary 1 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from Gaussian white-noise processes of (5.7), where p is fixed and $\mathcal{A}^c \neq \emptyset$. If $\lim_{n\to\infty} tr(\boldsymbol{\Sigma}^{-1}) = \infty$, then*

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{AIC}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

Note that $\lim_{n\to\infty} \text{tr}(\boldsymbol{\Sigma}^{-1}) = \infty$ when $\sigma_\epsilon^2 > 0$. Applying the inequality, $\sum_{i=1}^n \omega_i^{-1}/n \geq n/\sum_{i=1}^n \omega_i$, where $\omega_i > 0$; $i = 1, \ldots, n$, we obtain a sufficient condition for $\lim_{n\to\infty} \text{tr}(\boldsymbol{\Sigma}^{-1}) = \infty$, given by $\lim_{n\to\infty} \text{tr}(\boldsymbol{\Sigma})/n^2 = 0$ (see an example in Theorem 12 of Section 5.3).

4.2 Generalized Information Criterion

When p is fixed and $|\mathcal{A}^c| \geq 2$, we have for $\alpha \in \mathcal{A}^c \setminus \{\alpha^c\}$,

$$\Gamma_{\text{AIC}}(\alpha^c) - \Gamma_{\text{AIC}}(\alpha) + 2(p(\alpha) - p(\alpha^c)) \sim \chi^2(p(\alpha) - p(\alpha^c)),$$

where $\chi^2(k)$ denotes the chi-square distribution with k degrees of freedom. This implies that $\lim_{n\to\infty} P(\hat{\alpha}_{\text{AIC}} = \alpha^c) < 1$. That is, AIC is not able to achieve selection consistency. Replacing the penalty 2 in (4.3) by a penalty parameter $\lambda > 0$ leads to the GIC of (4.1) given under $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

$$\Gamma_{\text{GIC}(\lambda)}(\alpha) = (\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \hat{\boldsymbol{\mu}}(\alpha)) + \lambda\, p(\alpha). \tag{4.6}$$

Choosing a tuning parameter such that $\lambda \to \infty$, we obtain for $\alpha \in \mathcal{A} \setminus \{\alpha^c\}$,

$$\lim_{n\to\infty} P\big(\big(\Gamma_{\text{GIC}(\lambda)}(\alpha^c) - \Gamma_{\text{GIC}(\lambda)}(\alpha)\big) < 0\big) = 1.$$

That is, GIC can identify α^c among models in $\mathcal{A}^c$ asymptotically. For linear regression models (i.e., $\sigma_\eta^2 = 0$ in (3.1)), Shao (1997) established asymptotic loss efficiency and consistency for GIC under some regularity conditions. For linear mixed models, Pu and Niu (2006) also developed some asymptotic optimal properties of GIC. Adapted from Pu and Niu (2006), we have the following theorem.

Theorem 2 *Consider a class of models given by (3.1). Let $L^{KL}(\alpha)$ be the KL loss for model α and $\hat{\alpha}_{GIC(\lambda)}$ be the model selected by GIC.*

(i) For $\mathcal{A}^c = \emptyset$, if

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\lambda p}{E(L^{KL}(\alpha))} = 0, \tag{4.7}$$

then

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{GIC(\lambda)}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

(ii) For $\mathcal{A}^c \neq \emptyset$, if $\lambda \to \infty$, (4.7) holds, and

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} \frac{1}{p(\alpha)} < \infty, \tag{4.8}$$

then $\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1$. In addition,

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{GIC(\lambda)}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

Proof. The proof is essentially the same as that for Theorem 2 of Shao (1997) and hence is omitted. □

Theorem 2 reduces to Theorem 2 of Shao (1997) if $\mathbf{\Sigma} = \sigma^2 \boldsymbol{I}$. Similar to (4.5), Equation (4.7) provides a condition for risks associated with underfitted models. Equation (4.8) is a weak technique condition that holds trivially when p is fixed. In fact, (4.7) is slightly weaker than the two conditions given in Theorem 2 of Shao (1997): $\lim_{n\to\infty} \inf_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \mathrm{E}(L^{\mathrm{KL}}(\alpha))/n > 0$ and $\lim_{n\to\infty} \lambda p/n = 0$. Similar to Corollary 1, we have the following corollary.

Corollary 2 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7), where p is fixed and $\mathcal{A}^c \neq \emptyset$. If $\lim_{n\to\infty} tr\big(\mathbf{\Sigma}^{-1}\big)/\lambda = \infty$ and $\lambda \to \infty$, then $\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1$. In addition,*

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\alpha}_{GIC(\lambda)}) \Big/ \inf_{\alpha\in\mathcal{A}} L^{KL}(\alpha) = 1.$$

Similar to the remark given right after Corollary 1, $\lim_{n\to\infty} \lambda \mathrm{tr}\big(\mathbf{\Sigma}\big)/n^2 = 0$ is sufficient for $\lim_{n\to\infty} \mathrm{tr}\big(\mathbf{\Sigma}^{-1}\big)/\lambda = \infty$. (see an example in Theorem 12 of Section 5.3).

4.3 Unknown Covariance Parameters

In practice, the covariance parameter vector $\boldsymbol{\theta}$ is usually unknown and needs to be estimated. Two approaches are commonly applied under this situation. The first one utilizes a two-step procedure by first estimating the covariance parameters using, for example, ML or REML, and then pretending the estimated parameters as known for subsequent inference or prediction. The other one applies a Bayesian method that requires specifying a joint prior distribution for all the unknown parameters. Here we consider only the former one with $\hat{\boldsymbol{\theta}}(\alpha)$ being the ML estimate of $\boldsymbol{\theta}$ for $\alpha \in \mathcal{A}$, obtained by maximizing the following profile log-likelihood function,

$$\begin{aligned}\ell(\boldsymbol{\theta};\alpha) &= -\frac{1}{2}n\log(2\pi) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) \\ &\quad -\frac{1}{2}(\boldsymbol{Z}-\boldsymbol{X}(\alpha)\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{Z}-\boldsymbol{X}(\alpha)\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta})), \end{aligned} \tag{4.9}$$

where $\hat{\boldsymbol{\beta}}(\alpha;\boldsymbol{\theta}) \equiv (\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{Z}$ and $\boldsymbol{\Sigma}$ is written as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to emphasis its dependence on $\boldsymbol{\theta}$. Let Θ be the parameter space for $\boldsymbol{\theta}$, and let $\boldsymbol{\theta}_0 \in \Theta$ be the true covariance parameter vector. We shall develop asymptotic properties of GIC,

$$\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) = -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + \lambda(p(\alpha)), \tag{4.10}$$

under both the fixed domain asymptotic and the increasing domain asymptotic frameworks. The main difficulty to overcome is that some components of $\hat{\boldsymbol{\theta}}(\alpha)$ may converge to nondegenerate distributions even for $\alpha \in \mathcal{A}^c$ under the fixed domain asymptotic framework.

We impose some regularity conditions for establishing asymptotic properties of GIC. Denote by $\lambda_{\min}(\boldsymbol{M})$ the smallest eigenvalue of a symmetric matrix $\boldsymbol{M}$. We consider some regularity conditions. Suppose that there exists $\tau_n \to \infty$ such that the following are satisfied;

(A.1) For $\boldsymbol{\theta} \in \Theta$, $\lim_{n\to\infty} \frac{1}{\tau_n} \inf_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} > 0$, where $\boldsymbol{A}(\alpha;\boldsymbol{\theta})$ is defined in (3.6).

(A.2) For $\boldsymbol{\theta} \in \Theta$, $\lim_{n\to\infty} \frac{1}{\tau_n}\lambda_{\min}\big(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\big) > 0$.

(A.3) For $\boldsymbol{\theta} \in \Theta$, $\lim_{n\to\infty} \frac{1}{\tau_n}\lambda_{\max}\big(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\big) < \infty$.

(A.4) For $\alpha \in \mathcal{A}$, there exists some $\boldsymbol{\theta}_\alpha \in \Theta$ such that $\operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n}\big(\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) - \ell(\boldsymbol{\theta}_\alpha;\alpha)\big) = 0$.

(A.5) For $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$ and $\boldsymbol{\theta}_\alpha$ given in (A.4), $\operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n}\big(L^{\mathrm{KL}}(\alpha;\hat{\boldsymbol{\theta}}(\alpha)) - L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha)\big) = 0$.

In most cases, τ_n can be chosen as $\inf_{j\in\alpha^c} \boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_j$ or $\lambda_{\min}\big(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\big)$, where $\boldsymbol{X}_j$ is the jth column of $\boldsymbol{X}$ (see Theorems 7, 10 and 13). Condition (A.1) provides the effect suffered from applying an incorrect model. Condition (A.2) ensures that the explanatory variables are not too much correlated. Obviously, (A.4) and (A.5) hold when $\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\theta}}(\alpha) = \boldsymbol{\theta}_\alpha$, for some $\boldsymbol{\theta}_\alpha \in \Theta$. In some situation, $\boldsymbol{\theta}_\alpha$ is different from $\boldsymbol{\theta}_0$. For example, when $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, $\hat{\boldsymbol{\theta}}(\alpha)$ generally does not converge in probability to $\boldsymbol{\theta}_0$. Surprisingly, (A.4) and (A.5) may hold even if $\hat{\boldsymbol{\theta}}(\alpha)$ converges to a nondegenerate distribution (see Theorems 7, 10 and 13).

Theorem 3 *Consider a class of models given by (3.1) with p fixed. Let Θ be a compact parameter space for $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_0 \in \Theta$ being the true parameter, and let $L^{KL}(\alpha)$ be the KL loss defined in (3.3). Suppose that for $\alpha \in \mathcal{A}$, $\ell(\boldsymbol{\theta};\alpha)$ defined in (4.9) is continuous in Θ, and (A.1)-(A.5) are satisfied for some $\tau_n \to \infty$.*

(i) For $\mathcal{A}^c = \emptyset$, if $\tau_n/\lambda \to \infty$, and the following two conditions hold for $\alpha \in \mathcal{A}$:

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{1}{\tau_n}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} \quad < \quad \infty, \tag{4.11}$$

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n} tr\big(((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\big)\big) \quad = \quad 0, \tag{4.12}$$

then GIC defined in (4.2) is asymptotically loss efficient:

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\boldsymbol{\theta}}(\hat{\alpha}_{GIC(\lambda)}); \hat{\alpha}_{GIC(\lambda)}) \Big/ \min_{\alpha\in\mathcal{A}} L^{KL}(\hat{\boldsymbol{\theta}}(\alpha);\alpha) = 1. \tag{4.13}$$

(ii) For $\mathcal{A}^c \neq \emptyset$, if $\lambda \to \infty$, $\tau_n/\lambda \to \infty$, (4.12) holds, and

$$\lim_{n\to\infty} \frac{1}{\tau_n}\big(\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) - \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + tr(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)) - n\big) \quad = \quad 0, \tag{4.14}$$

for $\alpha \in \mathcal{A}^c$, then $\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1$.

Proof. (i) We first prove that for $\alpha \in \mathcal{A}$,

$$\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) = n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) + o_p(\tau_n). \tag{4.15}$$

By (3.3) and (3.7), we can rewrite $2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha)$ as

$$\begin{aligned} 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) = {} & \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) - \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)) - n \\ & +\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\theta})\boldsymbol{\mu} \\ & +(\boldsymbol{\eta}+\boldsymbol{\epsilon})\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}), \end{aligned} \tag{4.16}$$

where $\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)$ and $\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)$ are defined in (3.5) and (3.6). By (4.10), we have for $\alpha \in \mathcal{A}$,

$$\begin{aligned} \Gamma_{\mathrm{GIC}(\lambda)}(\alpha) &= -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha;\alpha) - 2\ell(\boldsymbol{\theta}_\alpha;\alpha) + \lambda p(\alpha) \\ &= -2\ell(\boldsymbol{\theta}_\alpha;\alpha) + \lambda p(\alpha) + o_p(\tau_n) \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu} \\ &\quad -2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p(\tau_n) \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu} \\ &\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p(\tau_n) \\ &= 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}) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) \\ &\quad +\mathrm{tr}((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0))) + o_p(\tau_n) \\ &= 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}) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) \\ &\quad +o_p(\tau_n), \end{aligned}$$

where the second equality follows from (A.4), the third equality follows from (4.9), the fourth equality follows from the following two equations, which will be proved later:

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1); \quad \alpha\in\mathcal{A}, \tag{4.17}$$

$$\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = o_p(\tau_n); \quad \alpha\in\mathcal{A}, \tag{4.18}$$

the fifth equality follows from (4.16) and

$$\begin{aligned}&(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + n - \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha))\\ &= \mathrm{tr}((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0))),\end{aligned}$$

and the last equality follows from (4.12). It remains to show (4.17) and (4.18). For (4.17), we have

$$\begin{aligned}&(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &= \left(\frac{(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}(\alpha)}{\tau_n^{1/2}}\right)\left(\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}(\alpha)}{\tau_n}\right)^{-1}\\ &\quad\times\left(\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})}{\tau_n^{1/2}}\right).\end{aligned} \tag{4.19}$$

By (A.2),

$$\left(\frac{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}}{\tau_n}\right)^{-1} = O_p(1). \tag{4.20}$$

By (A.3),

$$\lim_{n\to\infty}\frac{1}{\tau_n}\mathrm{var}(\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})) = \lim_{n\to\infty}\frac{1}{\tau_n}\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}_j < \infty.$$

where $\boldsymbol{X}_j$ be the jth column of $\boldsymbol{X}$. This together with $\mathrm{E}(\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})) = 0$ imply that

$$\frac{1}{\tau_n^{1/2}}\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1). \tag{4.21}$$

Therefore, (4.17) follows from (4.19)-(4.21). Using

$$\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}),$$

(4.11) and the Markov's inequality, we have for any $\varepsilon > 0$,

$$\begin{aligned}&\lim_{n\to\infty}P\left(|\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})/\tau_n| > \varepsilon\right)\\ &\le \lim_{n\to\infty}P\left(|\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})/\tau_n|^2 \ge \varepsilon^2\right)\\ &\le \lim_{n\to\infty}\frac{1}{\varepsilon^2\tau_n^2}\mathrm{E}\big(\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu}\big) = 0.\end{aligned}$$

This gives (4.18). Thus (4.15) is obtained.

We are now ready to prove (4.13). Let $\alpha^L = \arg\min_{\alpha\in\mathcal{A}} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha)$. By (4.15), we have

$$0 \le \operatorname*{plim}_{n\to\infty}\frac{\Gamma_{\mathrm{GIC}(\lambda)}(\alpha^L) - \Gamma_{\mathrm{GIC}(\lambda)}(\hat{\alpha}_{\mathrm{GIC}(\lambda)})}{\tau_n} = \operatorname*{plim}_{n\to\infty}\frac{L^{\mathrm{KL}}(\alpha^L;\boldsymbol{\theta}_{\alpha^L}) - L^{\mathrm{KL}}(\hat{\alpha}_{\mathrm{GIC}(\lambda)};\boldsymbol{\theta}_{\hat{\alpha}_{\mathrm{GIC}(\lambda)}})}{\tau_n} \le 0,$$

for some $\boldsymbol{\theta}_{\hat{\alpha}}, \boldsymbol{\theta}_{\alpha^L} \in \Theta$ where the first inequality follows from the definition of $\hat{\alpha}_{\mathrm{GIC}(\lambda)}$, the equality follows from (4.15) and the last inequality follows from the definition of α^L. It follows that

$$\operatorname*{plim}_{n\to\infty}\frac{L^{\mathrm{KL}}(\hat{\alpha}_{\mathrm{GIC}(\lambda);};\boldsymbol{\theta}_{\hat{\alpha}_{\mathrm{GIC}(\lambda)}}) - L^{\mathrm{KL}}(\alpha^L;\boldsymbol{\theta}_{\alpha^L})}{\tau_n} = 0. \tag{4.22}$$

In addition, by (A.1) and (4.16),

$$\operatorname*{plim}_{n\to\infty} \frac{2}{\tau_n} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha)) \;>\; \operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n} \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)\boldsymbol{\mu} > 0.$$

This together with (4.22) implies that $\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha^L;\boldsymbol{\theta}_\alpha)/L^{\mathrm{KL}}(\hat{\alpha}_{\mathrm{GIC}(\lambda)};\boldsymbol{\theta}_{\hat{\alpha}_{\mathrm{GIC}(\lambda)}}) = 1$. Then by (A.5),

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha^L;\hat{\boldsymbol{\theta}}(\alpha^L))/L^{\mathrm{KL}}(\hat{\alpha}_{\mathrm{GIC}(\lambda)};\hat{\boldsymbol{\theta}}(\hat{\alpha}_{\mathrm{GIC}(\lambda)})) = 1.$$

which gives (4.13). This completes the proof of (i).

(ii) We first prove (4.15) for $\mathcal{A}^c \neq \emptyset$. The proof is essentially the same as that in (i) except (4.18) needs to be shown as follows:

$$\begin{aligned}
&\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\quad= \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\quad= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\qquad -\left(\frac{\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}(\alpha)}{\tau_n^{1/2}}\right)\left(\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)\boldsymbol{X}(\alpha)}{\tau_n}\right)^{-1}\\
&\qquad \times\left(\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})}{\tau_n^{1/2}}\right)\\
&\quad= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + O_p(1)\\
&\quad= o_p(\tau_n),
\end{aligned}$$

where the second last equality follows similarly from the proof of (4.17) and the last equality follows from (4.21).

Second, we prove that $\lim_{n\to\infty} P\big(\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) > \Gamma_{\mathrm{GIC}(\lambda)}(\alpha^c)\big) = 1$, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$. By (A.4), we have for $\alpha \in \mathcal{A}^c$,

$$\begin{aligned}
\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) &= -2\ell(\boldsymbol{\theta}_\alpha;\alpha) + o_p(\tau_n)\\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta}_\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p(\tau_n)\\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_\alpha)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o_p(\tau_n),
\end{aligned} \tag{4.23}$$

where the first equality follows from $\lambda p = o(\tau_n)$ and the last equality follows from (4.17). Then, by (4.15) and (4.23), we have for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$,

$$\begin{aligned}
&\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) - \Gamma_{\mathrm{GIC}(\lambda)}(\alpha^c)\\
&\quad= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&\qquad - \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_{\alpha^c})) - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_{\alpha^c})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + o(\tau_n)\\
&\quad= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) - \mathrm{tr}\big(((\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}+\boldsymbol{\epsilon})' - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_{\alpha^c}) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\big)\big)\\
&\qquad -\big(\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_{\alpha^c})) - \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_{\alpha^c})) - n\big) + o_p(\tau_n)\\
&\quad= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha) + o_p(\tau_n) \;>\; 0,
\end{aligned}$$

as $n\to\infty$ with probability tending to 1, where the last equality follow from (4.12), (4.14) and (4.22). It follows that $\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} \in \mathcal{A}\setminus\mathcal{A}^c\big) = 0$.

Last, it remains to show that GIC achieves its minimum at α^c among $\alpha \in \mathcal{A}^c$. For $\alpha \in \mathcal{A}^c$, the ML estimate $\hat{\boldsymbol{\theta}}(\alpha)$ of $\boldsymbol{\theta}$ satisfies

$$\begin{aligned} -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) &= \inf_{\boldsymbol{\theta}\in\Theta} -2\ell(\boldsymbol{\theta};\alpha) \\ &= \inf_{\boldsymbol{\theta}\in\Theta} \big(n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big) \\ &= \inf_{\boldsymbol{\theta}\in\Theta} \big(n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big) + o_p(\lambda), \end{aligned}$$

where the last equality follows from (4.17). Hence, for $\alpha \in \mathcal{A}^c$ and $\lambda \to \infty$, we have

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\lambda}\Big| -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) - \inf_{\boldsymbol{\theta}\in\Theta}\big(n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big)\Big| = 0.$$

It then follows that

$$\begin{aligned} &\operatorname*{plim}_{n\to\infty} \frac{1}{\lambda}\big| -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\hat{\boldsymbol{\theta}}(\alpha^c);\alpha^c)\big| \\ &\leq \operatorname*{plim}_{n\to\infty} \frac{1}{\lambda}\Big(\big| -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) - \inf_{\boldsymbol{\theta}\in\Theta}\big(n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big)\big| \\ &\quad + \big| -2\ell(\hat{\boldsymbol{\theta}}(\alpha^c);\alpha^c) - \inf_{\boldsymbol{\theta}\in\Theta}\big(n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big)\big|\Big) \\ &= 0. \end{aligned}$$

Hence, for $\lambda \to \infty$,

$$\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) - \Gamma_{\mathrm{GIC}(\lambda)}(\alpha^c) = \lambda(p(\alpha) - p(\alpha^c)) + o_p(\lambda) \to \infty,$$

as $n \to \infty$ with probability tending to 1, which follows that $\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} \in \mathcal{A}^c, \hat{\alpha}_{\mathrm{GIC}(\lambda)} \neq \alpha^c\big) = 0$. This completes the proof of the theorem. □

Conditions (A.1)-(A.3) in Theorem 3 not only depend on explanatory variables but also depend on asymptotic frameworks. As shown in Theorem 7, those conditions are easier to be satisfied under the increasing domain asymptotic framework, particularly when the domain increases with the sample size in a faster rate. On the other hand, (A.1) may not be satisfied under the fixed domain asymptotic framework.

Theorem 3 is for fixed designs. A random design version is given in the following corollary.

Corollary 3 *(random design) Consider a class of models given by (3.1) with p fixed and $\boldsymbol{X}$ random, where $\boldsymbol{X}$ is independent of $(\boldsymbol{\eta}+\boldsymbol{\epsilon})$. Let Θ be a compact parameter space for $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_0 \in \Theta$ being the true parameter vector, and let $L^{KL}(\alpha)$ be the KL loss defined in (3.3). Suppose that for $\alpha \in \mathcal{A}$, $\ell(\boldsymbol{\theta};\alpha)$ defined in (4.9) is continuous in Θ, and there exists $\tau_n \to \infty$ such that*

(A.1') For $\boldsymbol{\theta} \in \Theta$, $\operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n} \inf_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} > 0$, where $\boldsymbol{A}(\alpha;\boldsymbol{\theta})$ is defined in (3.6),

(A.2') For $\boldsymbol{\theta} \in \Theta$, $\operatorname*{plim}_{n\to\infty} \frac{1}{\tau_n} \lambda_{\min}\big(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\big) > 0$,

(A.3') For $\boldsymbol{\theta} \in \Theta$, $\lim_{n\to\infty} \frac{1}{\tau_n} tr(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})E(\boldsymbol{X}_j\boldsymbol{X}_j')) < \infty$, *where* $\boldsymbol{X}_j$ *is the* j*th column of* $\boldsymbol{X}$,

and (A.4)-(A.5) are satisfied.

(i) For $\mathcal{A}^c = \emptyset$, *if* $\tau_n/\lambda \to \infty$, *(4.12) holds and*

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{1}{\tau_n} E\big(\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu}\big) < \infty,$$

for $\boldsymbol{\theta} \in \Theta$, *then GIC defined in (4.2) is asymptotically loss efficient:*

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\boldsymbol{\theta}}(\hat{\alpha}_{GIC(\lambda)}); \hat{\alpha}_{GIC(\lambda)}) \Big/ \min_{\alpha\in\mathcal{A}} L^{KL}(\hat{\boldsymbol{\theta}}(\alpha); \alpha) = 1.$$

(ii) For $\mathcal{A}^c \neq \emptyset$, *if* $\lambda \to \infty$, $\tau_n/\lambda \to \infty$, *(4.12) and (4.14) hold, then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Similar to (A.1)-(A.3) in Theorem 3 under fixed designs, (A.1')-(A.3') in Corollary 3 not only depend on explanatory variables but also depend on asymptotic frameworks. In contrast to fixed designs with smooth functions as explanatory variables, where (A.1) may not be satisfied (see Theorem 7) under the fixed domain asymptotic framework, condition (A.1') appear to be easier satisfied when random designs are considered (see some examples in Theorems 10 and 13).

Chapter 5

Exponential Covariance Models in One Dimension

In this chapter, we consider some examples in the one-dimensional space with $\eta(\cdot)$ of (2.2) generated from an exponential covariance function:

$$\mathrm{cov}(\eta(s), \eta(s)') = \sigma_\eta^2 \exp(-\kappa_\eta |s - s'|); \quad s, s' \in \mathbb{R}, \tag{5.1}$$

where $\sigma_\eta^2 > 0$ and $\kappa_\eta > 0$. Let $s_i = in^{-(1-\delta)}$ $i = 1, \ldots, n$, for some $\delta \in [0, 1)$. Then $\{\eta(s_1), \ldots \eta(s_n)\}$ can be expressed as an AR(1) process:

$$\eta(s_i) = \rho_n \eta(s_{i-1}) + \zeta_i, \tag{5.2}$$

where

$$\rho_n \equiv \exp(-\kappa_\eta n^{-(1-\delta)}), \tag{5.3}$$

$\eta(\boldsymbol{s}_1) \sim N(0, \sigma_\eta^2)$, $\zeta_i \sim N(0, \sigma_\eta^2(1 - \rho_n^2))$ is independent of $\eta(s_{i-1})$ for $i = 2, \ldots, n$, and $\eta(\boldsymbol{s}_1), \zeta_2, \ldots, \zeta_n$ are independent. Then the covariance parameter vector can be written as $\boldsymbol{\theta} \equiv (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)'$.

In what follows, we consider four examples corresponding to four different classes of explanatory variables in (3.1) with the exponential covariance model of (5.1) for $\eta(\cdot)$.

Example 1 (polynomials) Suppose that there are p explanatory variables, $x_j(s_i)$; $j = 1, \ldots, p$, sampled at $s_i = in^{-(1-\delta)}$; $i = 1, \ldots, n$, with $x_j(\cdot)$ given by

$$x_j(s) = s^j; \quad s \in \mathbb{R}, j = 1, \ldots, p, \tag{5.4}$$

where p is fixed and $\delta \in [0, 1)$.

Example 2 (polynomials varying with n) Suppose that there are p explanatory variables $x_j(s_i)$; $j = 1, \ldots, p$, sampled at $s_i = in^{-(1-\delta)}$; $i = 1, \ldots, n$, with $x_j(\cdot)$ given by

$$x_j(s) = (sn^{-\delta})^j; \quad s \in \mathbb{R}, j = 1, \ldots, p, \tag{5.5}$$

where p is fixed and $\delta \in [0, 1)$.

Example 3 (spatially dependent processes) Suppose that there are p explanatory variables $x_j(s_i)$; $j = 1, \ldots, p$, sampled at $s_i = in^{-(1-\delta)}$; $i = 1, \ldots, n$, where $x_1(\cdot), \ldots, x_p(\cdot)$ are independent zero-mean Gaussian spatial processes with covariance functions,

$$\mathrm{cov}(x_j(s), x_j(s')) = \sigma_j^2 \exp\{-\kappa_j |s - s'|\}; \quad s, s' \in \mathbb{R}, j = 1, \ldots, p, \tag{5.6}$$

p is fixed, $\delta \in [0, 1)$, and $\sigma_j^2, \kappa_j > 0$; $j = 1, \ldots, p$.

Example 4 (white noise processes) Suppose that there are p explanatory variables $x_j(s_i)$; $j = 1, \dots, p$, sampled at $s_i = in^{-(1-\delta)}$; $i = 1, \dots, n$, where $x_1(\cdot), \dots, x_p(\cdot)$ are independent white-noise processes with

$$x_j(s_i) \sim N(0, \sigma_j^2); \quad i = 1, \dots, n, \ j = 1, \dots, p, \tag{5.7}$$

p is fixed, $\delta \in [0, 1)$, and $\sigma_j^2 > 0$; $j = 1, \dots, p$.

We shall characterize the asymptotic behavior of GIC under both the fixed domain and the increasing domain frameworks with $\boldsymbol{\theta}$ being either known or estimated by ML. We shall also show how different generating mechanism of explanatory variables in the aforementioned examples affects the asymptotic behavior.

First, we introduce some notations and a number of technical lemmas regarding exponential covariance functions, which are crucial for developing the asymptotical results of GIC. Let

$$\boldsymbol{G}_k \equiv \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_n & 1 & 0 & \ddots & \vdots \\ 0 & -\rho_n & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho_n & 1 \end{pmatrix}_{k\times k}, \tag{5.8}$$

$$\boldsymbol{T}_k \equiv \begin{pmatrix} \sigma_\eta^2 + \sigma_\epsilon^2 & -\sigma_\epsilon^2\rho_n & 0 & \cdots & 0 \\ -\sigma_\epsilon^2\rho_n & f_1(\rho_n) & -\sigma_\epsilon^2\rho_n & \ddots & \vdots \\ 0 & -\sigma_\epsilon^2\rho_n & f_1(\rho_n) & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\sigma_\epsilon^2\rho_n \\ 0 & \cdots & 0 & -\sigma_\epsilon^2\rho_n & f_1(\rho_n) \end{pmatrix}_{k\times k}, \tag{5.9}$$

be $k \times k$ matrices, where

$$f_1(\rho_n) \equiv (1 - \rho_n^2)\sigma_\eta^2 + (1 + \rho_n^2)\sigma_\epsilon^2. \tag{5.10}$$

Lemma 4 *Consider $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_\eta$ defined in (3.2) and (5.1), where $s_i = in^{-(1-\delta)}$; $i = 1, \dots, n$, and $\delta \in [0, 1)$. Then*

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) = \boldsymbol{G}_n' \boldsymbol{T}_n^{-1} \boldsymbol{G}_n, \tag{5.11}$$

where $\boldsymbol{G}_n$ and $\boldsymbol{T}_n$ are given by (5.8) and (5.9), respectively.

Lemma 5 *For any $c > 0$ and $\delta \in [0, 1)$ with $n^{(1-\delta)/2+c} < n$, consider $\boldsymbol{T}_{j_n}$ defined in (5.9) with $n^{(1-\delta)/2+c} \le j_n \le n$. Let $C_{j_n}(k, \ell)$ be the (k, ℓ)th element of $\boldsymbol{T}_{j_n}^{-1}$. Then there exists a constant $\tau > 0$ such that*

$$\sigma_\epsilon^{-2j_n} \det(\boldsymbol{T}_{j_n}) = \frac{f_2^{j_n-1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}((\sigma_\eta^2 + \sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2) + o(\exp(-\tau n^{c/2})), \tag{5.12}$$

where ρ_n and $f_1(\rho_n)$ are given by (5.3) and (5.10), respectively, and

$$f_2(\rho_n) \equiv \frac{f_1(\rho_n) + (f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}{2\sigma_\epsilon^2}. \tag{5.13}$$

In addition,

$$C_{j_n}(1,\ell)=C_{j_n}(\ell,1) = \frac{\rho_n^{\ell-1}}{((\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n)-\rho_n^2\sigma_\epsilon^2)f_2^{\ell-2}(\rho_n)}+o(\exp(-\tau n^{c/2})); \quad 1\le \ell \le j_n - n^{(1-\delta+c)/2}, \tag{5.14}$$

$$C_{j_n}(j_n,\ell)=C_{j_n}(\ell,j_n) = \frac{\rho_n^{j_n-\ell}}{f_2^{j_n-\ell+1}(\rho_n)\sigma_\epsilon^2}+o(\exp(-\tau n^{c/2})); \quad n^{(1-\delta+c)/2}<\ell\le j_n, \tag{5.15}$$

$$\max_{1\le k,\ell\le j_n} C_{j_n}(k,\ell) = \frac{1}{(8\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}n^{(1-\delta)/2}+o(n^{-(1-\delta)}), \tag{5.16}$$

and

$$tr(\boldsymbol{T}_n^{-1}) = \frac{n^{(3-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}}+O(n^{1-\delta}). \tag{5.17}$$

Furthermore, let $\boldsymbol{T}_n^{(1)}$ be the matrix with $(\sigma_\eta^2,\kappa_\eta,\sigma_\epsilon^2)$ in $\boldsymbol{T}_n$ replaced by $(\sigma_\eta^{(1)2},\kappa_\eta^{(1)},\sigma_\epsilon^{(1)2})$. Then

$$tr(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(1)-1}) = \frac{n^{(5-3\delta)/2}}{2^{5/2}(\kappa_\eta\sigma_\eta^2\kappa_\eta^{(1)}\sigma_\eta^{(1)2})^{1/2}((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{(1)2})^{1/2}+(\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^2)^{1/2})} +O(n^{2-\delta}). \tag{5.18}$$

Notice that $\boldsymbol{T}_n$ defined in (5.9) corresponds to the variance-covariance matrix of a moving average (MA) process $\{\upsilon_1,\dots,\upsilon_n\}$ of order 1:

$$\text{var}(\boldsymbol{\upsilon}_n)=\boldsymbol{T}_n, \tag{5.19}$$

where $\boldsymbol{\upsilon}_n\equiv(\upsilon_1,\dots,\upsilon_n)$,

$$\upsilon_i=u_i-f_4(\rho_n)u_{i-1};\quad i=2,\dots,n, \tag{5.20}$$

with $u_1\sim N(0,(\sigma_\eta^2-\sigma_\epsilon^2-f_2(\rho_n)\sigma_\epsilon^2)f_4^{-2}(\rho_n))$ and $u_i\sim N(0,f_2(\rho_n)\sigma_\epsilon^2)$; $i=2,\dots,n$,

$$f_4(\rho_n)\equiv\rho_n/f_2(\rho_n), \tag{5.21}$$

and recall that $f_2(\rho_n)$ and ρ_n are defined in (5.13) and (5.3), respectively. Some asymptotic properties of $f_4(\rho_n)$ and $\boldsymbol{T}_n$ are given in the follow lemmas.

Lemma 6 *With $f_2(\rho_n)$ and $f_4(\rho_n)$ defined in (5.13) and (5.21), respectively, we have*

$$f_4(\rho_n) = 1-(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+O(n^{-(1-\delta)}), \tag{5.22}$$

$$f_2(\rho_n) = 1+(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+(\sigma_\eta^2-\sigma_\epsilon^2)\sigma_\epsilon^{-2}\kappa_\eta n^{-(1-\delta)}+O(n^{-3(1-\delta)/2}), \tag{5.23}$$

and

$$\log f_4(\rho_n)=-(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+O(n^{-(1-\delta)}). \tag{5.24}$$

In addition, for any $c>0$ and $\delta\in[0,1)$ with $n^{(1-\delta)/2+c}<n$, and any j_n with $n^{(1-\delta)/2+c}\le j_n\le n$, there exists a constant $\tau>0$ such that

$$f_4^{j_n}(\rho_n)=o(\exp(-\tau n^c)). \tag{5.25}$$

Lemma 7 *Consider $\boldsymbol{T}_n$ defined in (5.9). For any $c > 0$, $\delta \in [0, 1)$ with $n^{(1-\delta)/2+c} < n$, and any j_n with $n^{(1-\delta)/2+c} \le j_n \le n$, there exists a constant $\tau > 0$ such that*

$$\boldsymbol{T}_n^{-1} = \boldsymbol{\Omega}_n' \begin{pmatrix} \boldsymbol{\Lambda}_{j_n}^{-1} & \mathbf{0} \\ \mathbf{0} & (f_2(\rho_n)\sigma_\epsilon^2)^{-1}\boldsymbol{I}_{n-j_n} \end{pmatrix} \boldsymbol{\Omega}_n + o(\exp(-\tau n^{c/2})), \tag{5.26}$$

where

$$\boldsymbol{\Omega}_n \equiv \begin{pmatrix} 1 & 0 & \cdots & 0 \\ f_4(\rho_n) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ f_4^{n-1}(\rho_n) & \cdots & f_4(\rho_n) & 1 \end{pmatrix}, \tag{5.27}$$

$f_2(\rho_n)$ and $f_4(\rho_n)$ are given by (5.13) and (5.21), respectively, and

$$\boldsymbol{\Lambda}_k = \boldsymbol{\Omega}_k \boldsymbol{T}_k \boldsymbol{\Omega}_k'. \tag{5.28}$$

The following three lemmas are based on Lemmas 4-7, which are crucial in developing the asymptotical results of ML estimates in Sections 5.1-5.3.

Lemma 8 *Consider $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_\eta$ defined in (3.2) and (5.1), where $s_i = in^{-(1-\delta)}$; $i = 1, \ldots, n$, and $\delta \in [0, 1)$. Let $\boldsymbol{\Sigma}_\eta^{(j)}$ be the same as $\boldsymbol{\Sigma}_\eta$ except $(\sigma_\eta^2, \kappa_\eta)$ are replaced by $\big(\sigma_\eta^{(j)2}, \kappa_\eta^{(j)}\big)$. Define $\boldsymbol{\Sigma}^{(j)} \equiv \boldsymbol{\Sigma}_\eta^{(j)} + \sigma_\epsilon^{(j)2}$; $j = 1, 2, 3$. Then for $\delta \in [0, 1)$,*

$$\begin{aligned} \log(\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}))) &= n \log \sigma_\epsilon^2 + \left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2} n^{(1+\delta)/2} - \left(\frac{\kappa_\eta(\sigma_\eta^2 + \sigma_\epsilon^2)}{\sigma_\epsilon^2}\right) n^\delta \\ &\quad - \log n^{(1-\delta)/2} + o(n^\delta) + O(1), \end{aligned} \tag{5.29}$$

$$\begin{aligned} tr(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} n^{(1+\delta)/2} + \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}(\kappa_\eta - \kappa_\eta^{(1)})}{\kappa_\eta \sigma_\eta^2} n^\delta \\ &\quad + \frac{\sigma_\eta^{(1)2}(\kappa_\eta - \kappa_\eta^{(1)})^2}{2\kappa_\eta\sigma_\eta^2} n^\delta + o(n^\delta) + O(1), \end{aligned} \tag{5.30}$$

$$tr(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) = \frac{n}{\sigma_\epsilon^2} - \frac{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}{2\sigma_\epsilon^2} n^{(1+\delta)/2} + o(n^\delta) + O(1), \tag{5.31}$$

$$\begin{aligned} tr(\boldsymbol{\Sigma}^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{\sigma_\epsilon^{(1)2}}{\sigma_\epsilon^2} n - \frac{\sigma_\epsilon^{(1)2}}{2\sigma_\epsilon^2} (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2} n^{(1+\delta)/2} + \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} n^{(1+\delta)/2} \\ &\quad + \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}(\kappa_\eta - \kappa_\eta^{(1)})}{\kappa_\eta\sigma_\eta^2} n^\delta + \frac{\sigma_\eta^{(1)2}(\kappa_\eta - \kappa_\eta^{(1)})^2}{2\kappa_\eta\sigma_\eta^2} n^\delta \\ &\quad + o(n^\delta) + O(1), \end{aligned} \tag{5.32}$$

$$\begin{aligned} tr(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_\eta^{(2)} \boldsymbol{\Sigma}^{(3)-1}) &= \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}\sigma_\eta^{(2)2}\kappa_\eta^{(2)} n^{(1+\delta)/2}}{2^{1/2}(\kappa_\eta\sigma_\eta^2\kappa_\eta^{(3)}\sigma_\eta^{(3)2})^{1/2}((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{(3)2})^{1/2} + (\kappa_\eta^{(3)}\sigma_\eta^{(3)2}\sigma_\epsilon^2)^{1/2})} \\ &\quad + O(n^\delta), \end{aligned} \tag{5.33}$$

$$tr(\mathbf{\Sigma}_\eta^{(1)}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{\Sigma}^{(2)-1}) = \frac{\sigma_\eta^{(1)}\kappa_\eta^{(1)}}{\sigma_\epsilon^2\sigma_\epsilon^{(2)2}((2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}+(2\kappa_\eta^{(2)}\sigma_\eta^{(2)2}\sigma_\epsilon^{(2)-2})^{1/2})}+O(n^\delta), \tag{5.34}$$

$$\begin{aligned} tr(\mathbf{\Sigma}^{(1)-1}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{n}{\sigma_\epsilon^{(1)2}\sigma_\epsilon^2} - \frac{1}{\sigma_\epsilon^{(1)2}\sigma_\epsilon^2}\Bigg(\frac{(\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}+(\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2}}{2^{1/2}} \\ &\quad -\frac{(\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}(\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2}}{2^{1/2}((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}+(\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2})}\Bigg)n^{(1+\delta)/2} \\ &\quad +O(n^\delta). \end{aligned} \tag{5.35}$$

Lemma 9 *Consider $\mathbf{\Sigma}(\boldsymbol{\theta})$ and $\mathbf{\Sigma}_\eta$ defined in (3.2) and (5.1), where $s_i = in^{-(1-\delta)}$; $i = 1,\dots,n$, and $\delta \in [0,1)$. Let $\boldsymbol{\psi}_k \equiv n^{-k}(1^k,2^k,\dots,n^k)'$; $k \in \{0,1,\dots\}$. Then for any $k = 0,1,\dots$, $\ell = 1,2,\dots$, and any $\delta \in [0,1)$,*

$$\boldsymbol{\psi}_k'\mathbf{\Sigma}^{-1}\boldsymbol{\psi}_\ell = \frac{\kappa_\eta}{2\sigma_\eta^2(k+\ell+1)}n^\delta + \frac{1}{2\sigma_\eta^2} + \frac{k\ell}{2\kappa_\eta\sigma_\eta^2(k+\ell-1)}n^{-\delta}+o(n^\delta), \tag{5.36}$$

$$\boldsymbol{\psi}_0'\mathbf{\Sigma}^{-1}\boldsymbol{\psi}_0 = \frac{\kappa}{2\sigma_\eta^2}n^\delta + \frac{1}{\sigma_\eta^2}+o(n^\delta). \tag{5.37}$$

In addition, for $\delta \in [0,1)$, $k,\ell = 0,1,\dots,p$, and $\mathbf{\Sigma}^{(1)}$ defined in Lemma 8,

$$\boldsymbol{\psi}_k'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{(1)}\mathbf{\Sigma}^{-1}\boldsymbol{\psi}_\ell = O(n^\delta). \tag{5.38}$$

Lemma 10 *Consider $\mathbf{\Sigma}(\boldsymbol{\theta})$ and $\mathbf{\Sigma}_\eta$ defined in (3.2) and (5.1), where $s_i = in^{-(1-\delta)}$; $i = 1,\dots,n$, and $\delta \in [0,1)$. Let $\mathbf{\Sigma}_j = var((x_j(s_1),\dots,x_j(s_n))')$ with $x_j(s)$ defined in (5.6).*

(i) Suppose that $|\sigma_\epsilon^2-\sigma^2| = o(1)$ for some constant $\sigma^2>0$. Then

$$\log(\sigma_\epsilon^2)+\frac{\sigma^2}{\sigma_\epsilon^2}-\log(\sigma^2)-1 = \frac{1}{2\sigma^4}(\sigma_\epsilon^2-\sigma^2)^2+o((\sigma_\epsilon^2-\sigma^2)^2). \tag{5.39}$$

(ii) Suppose that $|\sigma_\epsilon^2-\sigma^2| = o(1)$ for some constants $\sigma^2>0$. Then for any κ_η, σ_η^2, $\tau>0$,

$$\left(\frac{1}{\sigma_\epsilon^2}\right)^{1/2}\left(1-\frac{\sigma^2}{2\sigma_\epsilon^2}+\frac{\tau}{2\kappa_\eta\sigma_\eta^2}\right)-\left(\frac{1}{\sigma^2}\right)^{1/2}\left(\frac{1}{2}+\frac{\tau}{2\kappa_\eta\sigma_\eta^2}\right) = o(\sigma_\epsilon^2-\sigma^2). \tag{5.40}$$

(iii) Suppose that $|\kappa_\eta\sigma_\eta^2-\tau| = o(1)$ for some constant $\tau>0$. Then for any $\sigma^2>0$,

$$\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma^2}\right)^{1/2}\left(\frac{1}{2}+\frac{\tau}{2\kappa_\eta\sigma_\eta^2}\right)-\left(\frac{2\tau}{\sigma^2}\right)^{1/2} = \frac{(\kappa_\eta\sigma_\eta^2-\tau)^2}{2^{5/2}\sigma\tau^{3/2}}+o((\kappa_\eta\sigma_\eta^2-\tau)^2). \tag{5.41}$$

(iv) Suppose that $|\sigma_\epsilon^2-\sigma^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2-\tau| = o(1)$ for some constants $\sigma^2,\tau>0$. Then for any $\kappa_j,\kappa_{j'},\sigma_j^2,\sigma_{j'}>0$,

$$\begin{aligned} &tr(\mathbf{\Sigma}_j(\mathbf{\Sigma}(\boldsymbol{\theta})-\mathbf{\Sigma}((\sigma_\eta^2,\kappa_\eta,\sigma^2)'))\mathbf{\Sigma}_j(\mathbf{\Sigma}((\sigma_\eta^2,\kappa_\eta,\sigma_\epsilon^2)')-\mathbf{\Sigma}((\sigma_\eta^2,\kappa_\eta,\sigma^2)'))) \\ &= \frac{\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2^{9/2}\tau^{3/2}\sigma^3}(\sigma_\epsilon^2-\sigma^2)^2n^{(1+\delta)/2}+o((\sigma_\epsilon^2-\sigma^2)^2n^{(1+\delta)/2})+O(n^\delta). \end{aligned} \tag{5.42}$$

(v) *Suppose that* $|\sigma_\epsilon^2 - \sigma^2| = o(1)$ *and* $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$ *for some constants* $\sigma^2, \tau > 0$. *Then for any* $\kappa_j, \sigma_j^2 > 0$,

$$\begin{aligned} & tr(\boldsymbol{\Sigma}_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)'))(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)'))) \\ = \; & \frac{5\kappa_j \sigma_j^2}{2^{9/2}\tau^{1/2}\sigma^7}(\sigma_\epsilon^2 - \sigma^2)^2 n^{(1+\delta)/2} + o((\sigma_\epsilon^2 - \sigma^2)^2 n^{(1+\delta)/2}) + O(n^\delta). \end{aligned} \tag{5.43}$$

(vi) *Suppose that* $|\sigma_\epsilon^2 - \sigma^2| = o(1)$ *and* $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$ *for some constants* $\sigma^2, \tau > 0$. *Then*

$$\begin{aligned} & tr((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)'))(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)'))) \\ = \; & \frac{1}{\sigma^8}(\sigma_\epsilon^2 - \sigma^2)^2 n + o((\sigma_\epsilon^2 - \sigma^2) n^{(1+\delta)/2}) + O(n^\delta). \end{aligned} \tag{5.44}$$

(vii) *Suppose that* $|\sigma_\epsilon^2 - \sigma^2| = o(n^{-(1-\delta)/2})$ *and* $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$ *for some constants* $\sigma^2, \tau > 0$. *Then for any* $\kappa_j, \kappa_{j'}, \sigma_j^2, \sigma_{j'}, c, d > 0$ *with* $cd = \tau$,

$$\begin{aligned} & tr(\boldsymbol{\Sigma}_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))\boldsymbol{\Sigma}_{j'}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))) \\ = \; & \frac{5\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{2^{9/2}\sigma\tau^{7/2}}(\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2}) + O(n^\delta). \end{aligned} \tag{5.45}$$

(viii) *Suppose that* $|\sigma_\epsilon^2 - \sigma^2| = o(n^{-(1-\delta)/2})$ *and* $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$ *for some constants* $\sigma^2, \tau > 0$. *Then for any* $\kappa_j, \sigma_j^2, c, d > 0$ *with* $cd = \tau$,

$$\begin{aligned} & tr(\boldsymbol{\Sigma}_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))) \\ = \; & \frac{\kappa_j \sigma_j^2}{2^{9/2}\sigma^3\tau^{5/2}}(\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2}) + O(n^\delta). \end{aligned} \tag{5.46}$$

(ix) *Suppose that* $|\sigma_\epsilon^2 - \sigma^2| = o(n^{-(1-\delta)/2})$ *and* $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$ *for some constants* $\sigma^2, \tau > 0$. *Then for any* $c, d > 0$ *with* $cd = \tau$,

$$\begin{aligned} & tr((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((c, d, \sigma^2)'))) \\ = \; & \frac{1}{2^{9/2}\tau^{5/2}}(\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta \sigma_\eta^2 - \tau)^2 n^{(1+\delta)/2}) + O(n^\delta). \end{aligned} \tag{5.47}$$

5.1 Polynomial Order Selection

In this section, we consider Examples 1 and 2 given by (5.4) and (5.5) for polynomial order selection. Note that in Example 1, the underlying true polynomial does not vary with the sample size, whereas in Example 2, the magnitude of the underlying true polynomial decreases as the sample size increases, making estimation and polynomial order selection more difficult. Let $\boldsymbol{V}_{j \times j'}$ be a $j \times j'$ matrix with the (k, ℓ)th element,

$$\frac{1}{k + \ell + 1}; \quad k = 1, \ldots, j,\; \ell = 1, \ldots, j'. \tag{5.48}$$

Note that when $j = j'$, the square matrix $\boldsymbol{V}_{j,j}$ is nonsingular (see Shibata 1981).

Proposition 4 *Consider a class of models given by (3.1) with p explanatory variables corresponding to p monomials defined in (5.5), where p is fixed. Let* $\mathcal{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_p\}$*, where* $\alpha_0 \equiv \emptyset$ *and* $\alpha_j = \{1, \ldots, j\}$*;* $j = 1, \ldots, p$*. Suppose that* $\mathcal{A}^c \neq \emptyset$ *and the data are sampled at* $s_i = in^{-(1-\delta)} \in [0, n^\delta]$*;* $i = 1, \ldots, n$*, for some* $\delta \in [0, 1)$*. Consider the exponential covariance model of (5.1) for* $\eta(\cdot)$*. Let* $\boldsymbol{\theta} = (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)' \in \Theta$*, and let* $\boldsymbol{\theta}_0 = (\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,0}^2)' \in \Theta$ *be the true parameter vector, where* $\Theta \equiv (0, \infty)^3$*. Then for any* $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$*,*

$$\lim_{n\to\infty} \frac{1}{n^\delta} \boldsymbol{\mu}' \boldsymbol{A}(\alpha; \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu} = \frac{\kappa_\eta}{2\sigma_\eta^2} \gamma(\alpha); \quad \text{if } \delta \in (0,1), \tag{5.49}$$

$$\limsup_{n\to\infty} \boldsymbol{\mu}' \boldsymbol{A}(\alpha; \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{A}(\alpha; \boldsymbol{\theta}) \boldsymbol{\mu} < \infty; \quad \text{if } \delta = 0, \tag{5.50}$$

where $\boldsymbol{A}(\alpha; \boldsymbol{\theta})$ *is defined in (3.6),*

$$\gamma(\alpha) \equiv \boldsymbol{\beta}' \boldsymbol{V}_{p\times p} \boldsymbol{\beta} - \boldsymbol{\beta}' \boldsymbol{V}_{p\times p(\alpha)} \boldsymbol{V}^{-1}_{p(\alpha)\times p(\alpha)} \boldsymbol{V}_{p(\alpha)\times p} \boldsymbol{\beta}, \tag{5.51}$$

and $\boldsymbol{V}_{j\times j'}$ *is defined in (5.48). In addition, the log-likelihood of (4.9) based on model* $\alpha \in \mathcal{A}$ *can be decomposed into the following:*

(i) For $\delta \in (0, 1)$*,*

$$\begin{aligned} -2\ell(\boldsymbol{\theta}; \alpha) &= n \log(2\pi) - \frac{1-\delta}{2} \log n + \left(\log \sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} \right) n \\ &\quad + \left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2} \right)^{1/2} \left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2} \right) n^{(1+\delta)/2} \\ &\quad + \left(-\frac{\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta \left(\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\kappa_\eta \sigma_\eta^2} - 1 \right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2} \right) n^\delta \\ &\quad + \frac{1}{2\kappa_\eta \sigma_\eta^2} \left((\sigma_{\eta,0}^2 + \gamma(\alpha))^{1/2} \kappa_\eta - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{(\sigma_{\eta,0}^2 + \gamma(\alpha))^{1/2}} \right)^2 n^\delta \\ &\quad + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2} \left(\kappa_{\eta,0} - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\eta,0}^2 + \gamma(\alpha)} \right) n^\delta + \xi(\boldsymbol{\theta}) + o_p(n^\delta), \end{aligned} \tag{5.52}$$

where

$$\xi(\boldsymbol{\theta}) = (\boldsymbol{\eta} + \boldsymbol{\epsilon})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) (\boldsymbol{\eta} + \boldsymbol{\epsilon}) - tr(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})). \tag{5.53}$$

(ii) For $\delta = 0$*,*

$$\begin{aligned} -2\ell(\boldsymbol{\theta}; \alpha) &= n \log(2\pi) - \frac{1-\delta}{2} \log n + \left(\log \sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} \right) n \\ &\quad + \left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2} \right)^{1/2} \left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2} \right) n^{1/2} + \xi(\boldsymbol{\theta}) + O_p(1). \end{aligned} \tag{5.54}$$

Equation (5.52) provides some guidance of applying GIC to distinguish between correct and incorrect models in polynomial order selection. For example, it follows from (5.52) and $\gamma(\alpha^c) = 0$ that for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$ and $\delta \in (0, 1)$,

$$-2\ell(\boldsymbol{\theta}; \alpha) + 2\ell(\boldsymbol{\theta}; \alpha^c) = \frac{\gamma(\alpha)\kappa_\eta}{2\sigma_\eta^2} n^\delta + o_p(n^\delta), \tag{5.55}$$

so that we can get rid of underfitted models if the penalty term has a smaller order than $O(n^{\delta})$. As to be demonstrated in Theorem 6, we can use (5.55) to find an appropriate penalty λ that leads to selection consistency. On the other hand, applying (5.54) to Example 2 under the fixed domain asymptotic framework (i.e., $\delta = 0$), we obtain

$$-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta};\alpha^c) = O_p(1), \tag{5.56}$$

indicating that consistency or asymptotic loss efficiency of GIC is almost impossible.

Additionally, we see from (5.54) that the likelihood value depends on κ_η and σ_η^2 mainly through their product, but not their individual values under the fixed domain asymptotic framework when $\delta = 0$. Consequently, variable selection based on GIC is expected to be not much affected by individual estimates of κ_η and σ_η^2 as long as the estimate of the microergodic parameter, $\kappa_\eta\sigma_\eta^2$, remains the same.

The following lemma provides consistency of the ML estimate of σ_ϵ^2 and the microergodic parameter, $\kappa_\eta\sigma_\eta^2$, under both the fixed domain and the increasing domain asymptotic frameworks with $\delta \in [0,1)$. The results are extended from Chen *et al.* (2000) who consider only $\alpha \in \mathcal{A}^c$ and $\delta = 0$.

Lemma 11 *Under the setup of Proposition 4, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat\sigma_\eta^2(\alpha), \hat\kappa_\eta(\alpha), \hat\sigma_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then for any $\delta \in [0,1)$,*

$$\hat\sigma_\epsilon^2(\alpha) = \sigma_{\epsilon,0}^2 + o_p(1), \tag{5.57}$$

$$\hat\kappa_\eta(\alpha)\hat\sigma_\eta^2(\alpha) = \kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(1). \tag{5.58}$$

The following theorem further provides the convergence rates for the ML estimates of κ_η, σ_η^2 and σ_ϵ^2. These results are also extended from Chen *et al.* (2000) who consider only $\alpha \in \mathcal{A}^c$ and $\delta = 0$, and are keys for establishing some asymptotic properties of GIC in Theorem 7.

Theorem 4 *Under the setup of Proposition 4, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat\sigma_\eta^2(\alpha), \hat\kappa_\eta(\alpha), \hat\sigma_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then*

(i) For $\delta \in (0,1)$,

$$\hat\sigma_\epsilon^2(\alpha) = \sigma_{\epsilon,0}^2 + o_p(n^{-(1-\delta)/2}); \quad \alpha \in \mathcal{A}, \tag{5.59}$$

$$\hat\kappa_\eta(\alpha)\hat\sigma_\eta^2(\alpha) = \kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(n^{-(1-\delta)/4}); \quad \alpha \in \mathcal{A}, \tag{5.60}$$

$$\hat\sigma_\eta^2(\alpha) = \begin{cases} \sigma_{\eta,0}^2 + o_p(1); & \text{if } \alpha \in \mathcal{A}^c, \\ \gamma(\alpha) + \sigma_{\eta,0}^2 + o_p(1); & \text{if } \alpha \in \mathcal{A}\setminus\mathcal{A}^c, \end{cases} \tag{5.61}$$

$$\hat\kappa_\eta(\alpha) = \begin{cases} \kappa_{\eta,0} + o_p(1); & \text{if } \alpha \in \mathcal{A}^c, \\ \kappa_{\eta,0}\sigma_{\eta,0}^2(\gamma(\alpha) + \sigma_{\eta,0}^2)^{-1} + o_p(1); & \text{if } \alpha \in \mathcal{A}\setminus\mathcal{A}^c, \end{cases} \tag{5.62}$$

where $\gamma(\alpha) > 0$ is a constant defined in (5.51) for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$.

(ii) For $\delta = 0$ and any $\alpha \in \mathcal{A}$,

$$\hat\sigma_\epsilon^2(\alpha) = \sigma_{\epsilon,0}^2 + O_p(n^{-1/2}), \tag{5.63}$$

$$\hat\kappa_\eta(\alpha)\hat\sigma_\eta^2(\alpha) = \kappa_{\eta,0}\sigma_{\eta,0}^2 + O_p(n^{-1/4}). \tag{5.64}$$

Proof. Denote $\sigma^2_{\eta,\alpha} \equiv \gamma(\alpha) + \sigma^2_{\eta,0}$ and $\kappa_{\eta,\alpha} \equiv \kappa_{\eta,0}\sigma^2_{\eta,0}/(\gamma(\alpha)+\sigma^2_{\eta,0})$, for $\alpha \in \mathcal{A}$, where $\gamma(\alpha) \equiv 0$ for $\alpha \in \mathcal{A}^c$. Note that $\kappa_{\eta,\alpha}\sigma^2_{\eta,\alpha} = \kappa_{\eta,0}\sigma^2_{\eta,0}$.

First, we prove (5.59). By (5.57) and (5.58), it suffices to show that for $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = o(1)$, $|\kappa_\eta\sigma^2_\eta - \kappa_{\eta,0}\sigma^2_{\eta,0}| = o(1)$ and any $\varepsilon > 0$,

$$\inf_{|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| \geq \varepsilon n^{-(1-\delta)/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})';\alpha)) \;\;>\;\; 0, \tag{5.65}$$

as $n \to \infty$ with probability tending to 1. By (5.52), we can write

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma^2_\epsilon + \frac{\sigma^2_{\epsilon,0}}{\sigma^2_\epsilon}\right)n \\ &+ \left(\frac{2\kappa_\eta\sigma^2_\eta}{\sigma^2_\epsilon}\right)^{1/2}\left(1 - \frac{\sigma^2_{\epsilon,0}}{2\sigma^2_\epsilon} + \frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{2\kappa_\eta\sigma^2_\eta}\right)n^{(1+\delta)/2} \\ &+ \left(-\frac{\kappa_\eta\sigma^2_\eta}{\sigma^2_\epsilon} + \kappa_\eta\left(\frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{\kappa_\eta\sigma^2_\eta} - 1\right) - \frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{2\kappa_\eta\sigma^2_\eta}\right)n^\delta + \frac{\sigma^2_{\eta,\alpha}}{2\kappa_\eta\sigma^2_\eta}(\kappa_\eta - \kappa_{\eta,\alpha})^2 n^\delta \\ &+ \frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{2\kappa_\eta\sigma^2_\eta}(\kappa_{\eta,0} - \kappa_{\eta,\alpha})n^\delta + \xi(\boldsymbol{\theta}) + o_p(n^\delta), \end{aligned} \tag{5.66}$$

where $\xi(\boldsymbol{\theta})$ is given in (5.53). Then for $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = o(1)$ and $|\kappa_\eta\sigma^2_\eta - \kappa_{\eta,0}\sigma^2_{\eta,0}| = o(1)$, we have

$$\begin{aligned} &-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})';\alpha) \\ &= \left(\log\sigma^2_\epsilon + \frac{\sigma^2_{\epsilon,0}}{\sigma^2_\epsilon} - \log\sigma^2_{\epsilon,0} - 1\right)n \\ &\quad + \left\{\left(\frac{2\kappa_\eta\sigma^2_\eta}{\sigma^2_\epsilon}\right)^{1/2}\left(1 - \frac{\sigma^2_{\epsilon,0}}{2\sigma^2_\epsilon} + \frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{2\kappa_\eta\sigma^2_\eta}\right) - \left(\frac{2\kappa_\eta\sigma^2_\eta}{\sigma^2_{\epsilon,0}}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma^2_{\eta,0}}{2\kappa_\eta\sigma^2_\eta}\right)\right\}n^{(1+\delta)/2} \\ &\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})') + o_p(n^\delta) \\ &= \frac{1}{2\sigma^4_{\epsilon,0}}(\sigma^2_\epsilon - \sigma^2_{\epsilon,0})^2 n + \xi(\boldsymbol{\theta}) - \xi((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})') + o_p(n^\delta), \end{aligned}$$

where the last equality follows from (5.39) and (5.40). Therefore, for (5.65) to hold, it remains to show that

$$\xi(\boldsymbol{\theta}) - \xi((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})') = o_p(\max((\sigma^2_\epsilon - \sigma^2_{\epsilon,0})^2 n,\, n^\delta)). \tag{5.67}$$

We can decompose $\xi(\boldsymbol{\theta}) - \xi((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')$ into the following three parts:

$$\begin{aligned} &\xi(\boldsymbol{\theta}) - \xi((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})') \\ &= \boldsymbol{\eta}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')\big)\boldsymbol{\eta} - \mathrm{tr}\big(\boldsymbol{\Sigma}_\eta(\boldsymbol{\theta}_0)(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})'))\big) \\ &\quad + 2\boldsymbol{\eta}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')\big)\boldsymbol{\epsilon} \\ &\quad + \boldsymbol{\epsilon}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')\big)\boldsymbol{\epsilon} - \sigma^2_{\epsilon,0}\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')\big). \end{aligned} \tag{5.68}$$

Applying Chebyshev's inequality on each of the three parts and using the following three moment conditions given from (5.42)-(5.44) on (5.68):

$$\begin{aligned} \mathrm{var}\big(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})'))\boldsymbol{\eta}\big) &= O\big(\max((\sigma^2_\epsilon - \sigma^2_{\epsilon,0})^2 n, n^\delta)\big), \\ \mathrm{var}\big(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})'))\boldsymbol{\epsilon}\big) &= O\big(\max((\sigma^2_\epsilon - \sigma^2_{\epsilon,0})^2 n, n^\delta)\big), \\ \mathrm{var}\big(\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})'))\boldsymbol{\epsilon}\big) &= O\big(\max((\sigma^2_\epsilon - \sigma^2_{\epsilon,0})^2 n, n^\delta)\big), \end{aligned}$$

we obtain (5.67). This completes the proof of (5.59).

Second, we prove (5.60). By (5.58) and (5.59), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$ and any $\varepsilon > 0$,

$$\inf_{|\sigma_\eta^2\kappa_\eta - \sigma_{\eta,0}^2\kappa_{\eta,0}| \geq \varepsilon n^{-(1-\delta)/4}} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha)\big) \;>\; 0, \tag{5.69}$$

as $n \to \infty$ with probability tending to 1. By (5.66), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\right\} n^{(1+\delta)/2} \\
&\quad + \frac{\sigma_{\eta,\alpha}^2}{2\kappa_{\eta,0}\sigma_{\eta,0}^2}(\kappa_\eta - \kappa_{\eta,\alpha})^2 n^\delta + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + o_p(n^\delta) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + \frac{\sigma_{\eta,\alpha}^2(\kappa_\eta - \kappa_{\eta,\alpha})n^\delta}{2\kappa_{\eta,0}\sigma_{\eta,0}^2} + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') \\
&\quad + o_p(n^\delta),
\end{aligned} \tag{5.70}$$

where the first equality follows from (5.39) and the second equality follows from (5.41). Therefore, for (5.69) to hold, it remains to show that

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') \;=\; o_p(\max((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}, n^\delta)), \tag{5.71}$$

which can be obtained from a decomposition similar to (5.68) in addition to the following three moment conditions given from (5.45)-(5.47):

$$\begin{aligned}
\mathrm{var}\big(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\eta}\big) &= O(\max((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}, n^\delta)), \\
\mathrm{var}\big(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\epsilon}\big) &= O(\max((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}, n^\delta)), \\
\mathrm{var}\big(\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\epsilon}\big) &= O(\max((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}, n^\delta)).
\end{aligned}$$

Thus (5.69) is obtained. This completes the proof of (5.60).

Third, we prove (5.61) and (5.62). By (5.70) and (5.71), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, $|\sigma_\eta^2\kappa_\eta - \sigma_{\eta,0}^2\kappa_{\eta,0}| = o(n^{-(1-\delta)/4})$ and any $\varepsilon > 0$, we have

$$\inf_{|\kappa_\eta - \kappa_{\eta,\alpha}| \geq \varepsilon} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \;=\; \frac{\sigma_{\eta,\alpha}^2}{2\kappa_{\eta,0}\sigma_{\eta,0}^2}\varepsilon^2 n^\delta + o_p(n^\delta) \;>\; 0,$$

as $n \to \infty$ with probability tending to 1, which gives (5.62). This together with (5.60) gives (5.61).

Fourth, we prove (5.63). By (5.57) and (5.58), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| \geq Mn^{-1/2}} \big\{-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha)\big\} \;>\; 0, \tag{5.72}$$

as $n \to \infty$ with probability tending to 1. By (5.54), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} - \log\sigma_{\epsilon,0}^2 - 1\right)n \\
&\quad + \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)\right\}n^{1/2} \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + O_p(1) \\
&= \frac{1}{2\sigma_{\epsilon,0}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + o_p\big((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n\big) + O_p(1) \\
&= \frac{1}{2\sigma_{\epsilon,0}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n + o_p\big((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n\big) + O_p(1),
\end{aligned}$$

where the second equality follows from (5.39) and (5.40), and the last equality follows from

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') = o_p\big((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n\big) + O_p(1), \tag{5.73}$$

which can be obtained in a way similar to (5.67). Consequently, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| \geq Mn^{-1/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha)) = \frac{M^2}{2\sigma_{\epsilon,0}^4} + O_p(1) > 0,$$

as $n \to \infty$ with probability tending to 1. Thus, we obtain (5.72), and hence the proof of (5.63) is complete.

Finally, we prove (5.64). By (5.58) and (5.63), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,\alpha}\sigma_{\eta,\alpha}^2| = o(1)$, there exists $M > 0$ such that

$$\inf_{|\sigma_\eta^2\kappa_\eta - \sigma_{\eta,0}^2\kappa_{\eta,0}| \geq Mn^{-1/4}} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha)\big) > 0, \tag{5.74}$$

as $n \to \infty$ with probability tending to 1, where $\kappa_{\eta,\alpha}\sigma_{\eta,\alpha}^2 = \kappa_{\eta,0}\sigma_{\eta,0}^2$. By (5.54), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,\alpha}\sigma_{\eta,\alpha}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,\alpha}\sigma_{\eta,\alpha}^2}{2\kappa_\eta\sigma_\eta^2}\right)\right\}n^{1/2} \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') \\
&\quad + o_p((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}) + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + o_p((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}) + O_p(1),
\end{aligned} \tag{5.75}$$

where the first equality follows from $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, the second equality follows from (5.41), and the last equality follows from

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') = o_p\big((\kappa_\eta \sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}\big) + O_p(1), \tag{5.76}$$

which can be obtained in a way similar to (5.71). Thus, (5.74), and hence (5.64) are obtained. This completes the proof. □

Note that a special case of Theorem 4 for which $\delta = 0$ and $\boldsymbol{\beta} = \mathbf{0}$, can be found in Zhang and Zimmerman (2005), where they consider no regressor, and hence consider no underfitted model.

Corollary 4 *Under the setup of Theorem 4, let*

$$\boldsymbol{\theta}_\alpha^{(1)} = (\gamma(\alpha) + \sigma_{\eta,0}^2, \kappa_{\eta,0}\sigma_{\eta,0}^2(\gamma(\alpha) + \sigma_{\eta,0}^2)^{-1}, \sigma_{\epsilon,0}^2)'; \quad \alpha \in \mathcal{A}, \tag{5.77}$$

where $\gamma(\alpha) \equiv 0$ for $\alpha \in \mathcal{A}^c$. Then

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n^\delta}(-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha)) = 0; \quad \text{if } \delta \in (0,1), \tag{5.78}$$

$$-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha) = O_p(1); \quad \text{if } \delta = 0. \tag{5.79}$$

In addition, for $L^{KL}(\alpha;\boldsymbol{\theta})$ defined in (3.3) and $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$,

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\boldsymbol{\theta}}(\alpha);\alpha)/L^{KL}(\boldsymbol{\theta}_\alpha^{(1)};\alpha) = 1; \quad \text{if } \delta \in (0,1), \tag{5.80}$$

$$L^{KL}(\hat{\boldsymbol{\theta}}(\alpha);\alpha) - L^{KL}(\boldsymbol{\theta}_\alpha^{(1)};\alpha) = O_p(1); \quad \text{if } \delta = 0. \tag{5.81}$$

Note that from Theorem 4, we have $\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\theta}}(\alpha) = \boldsymbol{\theta}_\alpha^{(1)}$ for $\delta \in (0,1)$, which immediately gives (5.78). On the other hand, (5.79) is somewhat surprising, because $\hat{\boldsymbol{\theta}}(\alpha)$ generally does not converge to $\boldsymbol{\theta}_\alpha^{(1)}$ for $\delta = 0$.

Theorem 5 *Consider a class of models given by (3.1) with $x_j(s) = s^j$; $j = 1,\dots,p$, and $cov(\eta(s),\eta(s')) = \sigma_\eta^2 \exp(-\kappa_\eta|s-s'|)$, where $\sigma_\eta^2 > 0$, $\kappa_\eta > 0$ and $\sigma_\epsilon^2 > 0$ are known, and p is fixed. Suppose that $\mathcal{A} = \{\alpha_0, \alpha_1, \dots, \alpha_p\}$, where $\alpha_0 = \emptyset$, $\alpha_j = \{1,\dots,j\}$ for $j = 1,\dots,p$, and $\mathcal{A}^c \neq \emptyset$. In addition, suppose that the data are collected at $s_i = in^{-(1-\delta)}$; $i = 1,\dots,n$, for some $\delta \in [0,1)$.*

(i) For $\delta = 0$ and any $\lambda > 0$,

$$\lim_{n\to\infty} P\big(\alpha^c = \arg\min_{\alpha\in\mathcal{A}} L^{KL}(\alpha)\big) < 1. \tag{5.82}$$

In addition, if $\lambda \to \infty$, then

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha_0\big) = 1, \tag{5.83}$$

where $\hat{\alpha}_{GIC(\lambda)}$ is defined in (4.2).

(ii) For $\delta \in (0,1)$, if $\lambda \to \infty$ and $n^{(2p(\alpha^c)+1)\delta}/\lambda \to \infty$ as $n \to \infty$, then

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. (i) For $\delta = 0$, by (3.8),

$$L^{\text{KL}}(\alpha) \;=\; \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}),$$

where

$$\begin{aligned}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &\sim \chi^2(p(\alpha)), \qquad (5.84)\end{aligned}$$

with $\chi^2(k)$ denoting the chi-square distribution with k degrees of freedom. Similarly,

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'(\boldsymbol{M}(\alpha^c)-\boldsymbol{M}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \sim \chi^2(p(\alpha^c)-p(\alpha)).$$

By (5.50), for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, we have $\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} = O(1)$. Hence, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$,

$$\begin{aligned}&\lim_{n\to\infty} P\big(L^{\text{KL}}(\alpha^c) - L^{\text{KL}}(\alpha) > 0\big)\\ &= \lim_{n\to\infty} P\big((\boldsymbol{\eta}+\boldsymbol{\epsilon})'(\boldsymbol{M}(\alpha^c)-\boldsymbol{M}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} > 0\big) > 0.\end{aligned}$$

Thus (5.82) is obtained.

For (5.83), by (4.6) with $\lambda \to \infty$,

$$\begin{aligned}\Gamma_{\text{GIC}(\lambda)}(\alpha) &= (\boldsymbol{Z}-\hat{\boldsymbol{\mu}}(\alpha))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\hat{\boldsymbol{\mu}}(\alpha)) + \lambda p(\alpha)\\ &= \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &\quad +\lambda p(\alpha)\\ &= 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \lambda p(\alpha) + o_p(\lambda),\end{aligned}$$

where the last equality follows from (5.50) and (5.84). In addition, by Chebyshev's inequality and the following moment condition:

$$\operatorname{var}(\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon})) = \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} = O(1),$$

we have $\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1)$. Therefore, for $\alpha \in \mathcal{A}\setminus\{\alpha_0\}$,

$$\Gamma_{\text{GIC}(\lambda)}(\alpha) - \Gamma_{\text{GIC}(\lambda)}(\alpha_0) = \lambda(p(\alpha)-p(\alpha_0)) + o_p(\lambda),$$

which is greater than zero with probability tending to 1. Thus (5.83) is obtained.

(ii) It suffices to show that $\lim_{n\to\infty} \mathrm{E}L^{\text{KL}}(\alpha)/\lambda = \infty$ for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$ by (4.7). First, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$,

$$\begin{aligned}&\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu}\\ &= \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha)(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)\boldsymbol{\Sigma}^{-1})\boldsymbol{X}\boldsymbol{\beta}\\ &= \boldsymbol{\beta}^{*'}\boldsymbol{X}^{*'}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}^*(\alpha)(\boldsymbol{X}^*(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^*(\alpha))^{-1}\boldsymbol{X}^*(\alpha)\boldsymbol{\Sigma}^{-1})\boldsymbol{X}^*\boldsymbol{\beta}^*\\ &= \boldsymbol{\beta}^{*'}(\boldsymbol{V}_{p,p} - \boldsymbol{V}_{p,p(\alpha)}\boldsymbol{V}^{-1}_{p(\alpha),p(\alpha)}\boldsymbol{V}_{p(\alpha),p})\boldsymbol{\beta}^* + o(n^{(2p(\alpha^c)+1)\delta})\\ &= \beta^2_{p(\alpha^c)}\boldsymbol{e}'_{p(\alpha^c)}(\boldsymbol{V}_{p,p} - \boldsymbol{V}_{p,p(\alpha)}\boldsymbol{V}^{-1}_{p(\alpha),p(\alpha)}\boldsymbol{V}_{p(\alpha),p})\boldsymbol{e}_{p(\alpha^c)}n^{(2p(\alpha^c)+1)n^\delta} + o(n^{(2p(\alpha^c)+1)\delta}),\end{aligned}$$

where $\boldsymbol{e}_j$ is the jth column of $\boldsymbol{I}_p$, $\boldsymbol{\beta}^*(\alpha) = \boldsymbol{D}(\alpha)\boldsymbol{\beta}(\alpha)$, $\boldsymbol{X}^*(\alpha) = \boldsymbol{D}^{-1}(\alpha)\boldsymbol{X}(\alpha)$ with

$$\boldsymbol{D}(\alpha) = \begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & n^\delta & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & n^{p(\alpha)\delta}\end{pmatrix},$$

$\boldsymbol{V}_{j\times j'}$ is defined in (5.48), and $\boldsymbol{e}'_{p(\alpha^c)}(\boldsymbol{V}_{p,p} - \boldsymbol{V}_{p,p(\alpha)}\boldsymbol{V}^{-1}_{p(\alpha),p(\alpha)}\boldsymbol{V}_{p(\alpha),p})\boldsymbol{e}_{p(\alpha^c)}$ is a constant, which is bounded away from 0 by Theorem 3.1 of Shibata (1981). It follows from (3.9) and $\lim_{n\to\infty} \lambda/n^{(2p(\alpha)+1)\delta} = 0$ that

$$\lim_{n\to\infty} \frac{\mathrm{E}L^{\mathrm{KL}}(\alpha)}{\lambda} = \lim_{n\to\infty} \frac{\mathrm{E}L^{\mathrm{KL}}(\alpha)/n^{(2p(\alpha^c)+1)\delta}}{\lambda/n^{(2p(\alpha^c)+1)\delta}} = \infty.$$

This completes the proofs. □

Theorem 6 *Consider the same setup as in Theorem 5 except* $x_j(s) = (sn^{-\delta})^j$; $j = 1,\ldots,p$.

(i) *For* $\delta = 0$ *and any* $\lambda > 0$,

$$\lim_{n\to\infty} P\big(\alpha^c = \arg\min_{\alpha\in\mathcal{A}} L^{KL}(\alpha)\big) < 1.$$

In addition, if $\lambda \to \infty$, *then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha_0\big) = 1,$$

where $\hat{\alpha}_{GIC(\lambda)}$ *is defined in (4.2).*

(ii) *For* $\delta \in (0,1)$, *if* $\lambda \to \infty$ *and* $n^\delta/\lambda \to \infty$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. (i). See (i) in Proof of Theorem 5.

(ii). From (ii) of Theorem 2, it suffices to show that $\lim_{n\to\infty} \mathrm{E}L^{\mathrm{KL}}(\alpha)/\lambda = \infty$ for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$. By (5.49),

$$\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} = \gamma(\alpha)n^\delta + o(n^\delta),$$

where $\gamma(\alpha)$ is a constant, which is bounded away from 0 by Theorem 3.1 of Shibata (1981). It follows from (3.9) and $\lim_{n\to\infty} \lambda/n^\delta = 0$ that

$$\lim_{n\to\infty} \frac{\mathrm{E}L^{\mathrm{KL}}(\alpha)}{\lambda} = \lim_{n\to\infty} \frac{\mathrm{E}L^{\mathrm{KL}}(\alpha)/n^\delta}{\lambda/n^\delta} = \infty.$$

This completes the proof. □

Theorem 7 *Under the setup of Theorem 6, suppose that* $\boldsymbol{\theta} = (\sigma^2_\eta, \kappa_\eta, \sigma^2_\epsilon)' \in \Theta$ *is unknown, where* $\Theta \subset (0,\infty)^3$ *is a compact set such that* $\boldsymbol{\theta}_0 \in \Theta$. *Let* $\hat{\boldsymbol{\theta}}(\alpha)$ *be the ML estimate of* θ . *For* $\delta = 0$, *if* $\lambda \to \infty$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha_0\big) = 1.$$

For $\delta \in (0,1)$, *if* $\lambda \to \infty$ *and* $\lambda/n^\delta \to 0$ *as* $n \to \infty$, *then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. First, for $\delta = 0$, we prove

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} = \alpha_0\big) = 1.$$

By (4.10) and by (5.79), for $\alpha_0 = \emptyset$ and $\boldsymbol{\theta}_\alpha^{(1)}$ defined in (5.77), we have

$$\begin{aligned}\Gamma_{\mathrm{GIC}(\lambda)}(\alpha) - \Gamma_{\mathrm{GIC}(\lambda)}(\alpha_0) &= -2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha) + 2\ell(\boldsymbol{\theta}_{\alpha_0}^{(1)};\alpha_0) + \lambda(p(\alpha) - p(\alpha_0)) + O_p(1)\\ &= \lambda(p(\alpha) - p(\alpha_0)) + \xi(\boldsymbol{\theta}_\alpha^{(1)}) - \xi(\boldsymbol{\theta}_{\alpha_0}^{(1)}) + O_p(1)\\ &= \lambda(p(\alpha) - p(\alpha_0)) + O_p(1) \;>\; 0,\end{aligned}$$

as $n \to \infty$ with probability tending to 1, where the second equality follows from (5.54) and the third equality follows from (5.76).

Second, for $\delta \in (0, 1)$, we prove

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} = \alpha^c\big) = 1.$$

It suffices to show that the conditions in Theorem 3 are satisfied. First, by (5.36) and (5.37), we have

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X} \;=\; \frac{\kappa_\eta}{2\sigma_\eta^2}\boldsymbol{V}_{p\times p}n^\delta + o(n^\delta),$$

where $\boldsymbol{V}_{p\times p}$ is defined in (5.48) and is nonsingular. Then (A.2) is satisfied. Second, by (5.38), (A.3) is satisfied trivially. Third, (A.4)-(A.5) are followed by (5.78) and (5.80) for $\tau_n = n^\delta$ and $\boldsymbol{\theta}_\alpha = \boldsymbol{\theta}_\alpha^{(1)}$ defined in (5.77). Fourth, (A.1) holds by (5.49). Fifth, for $\xi(\boldsymbol{\theta})$ defined in (5.53), by (5.71), we have

$$\xi(\boldsymbol{\theta}_0) - \xi(\boldsymbol{\theta}_\alpha^{(1)}) \;=\; o_p(n^\delta).$$

Hence, (4.12) holds. Last, for $\alpha \in \mathcal{A}^c$, $\boldsymbol{\theta}_\alpha^{(1)} = \boldsymbol{\theta}_0$, (4.14) holds trivially. This completes the proof. □

5.2 Spatially Dependent Regressors

In this section, we consider Example 3 with explanatory variables generated independently from spatially dependent processes with exponential covariance functions of (5.6). This example considers spatial dependence not only for the response but also for explanatory variables.

Proposition 5 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7) and $\mathrm{cov}(\eta(s), \eta(s')) = \sigma_\eta^2 \exp(-\kappa_\eta|s-s'|)$, where $\mathcal{A}^c \neq \emptyset$ and p is fixed. Suppose that the data are collected at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1,\ldots,n$ for some $\delta \in [0,1)$. Let $\boldsymbol{\theta} = (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)' \in \Theta$ and let $\boldsymbol{\theta}_0 = (\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,0}^2)' \in \Theta$ be the true parameter vector, where $\Theta = (0,\infty)^3$. Define*

$$\theta_{\eta,\alpha} \equiv \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2\kappa_j + \kappa_{\eta,0}\sigma_{\eta,0}^2, \tag{5.85}$$

Then the log-likelihood of (4.9) based on model $\alpha \in \mathcal{A}$ can be decomposed into the following.

(i) For $\delta \in (0,1)$,

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\
&\quad + \left(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\left(\frac{\theta_{\eta,\alpha}}{\kappa_\eta\sigma_\eta^2} - 1\right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^\delta \\
&\quad + \frac{\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\left(\kappa_\eta - \frac{\theta_{\eta,\alpha}}{\left(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)}\right)^2 n^\delta \\
&\quad + \frac{1}{2\kappa_\eta\sigma_\epsilon^2}\left(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\kappa_j^2 + \sigma_{\eta,0}^2\kappa_{\eta,0}^2 - \frac{\theta_{\eta,\alpha}^2}{\left(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)}\right)n^\delta \\
&\quad + \xi^{(2)}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta), \qquad (5.86)
\end{aligned}$$

where

$$\begin{aligned}
\xi^{(2)}(\alpha;\boldsymbol{\theta}) &= \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}A(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} - \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2 tr(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad -2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}), \qquad (5.87)
\end{aligned}$$

$\boldsymbol{A}(\alpha;\boldsymbol{\theta})$ is defined in (3.6), $\boldsymbol{\Sigma}_j = var(\boldsymbol{X}_j)$ with $\boldsymbol{X}_j$ being the jth column of $\boldsymbol{X}$, and $\xi(\boldsymbol{\theta})$ is defined in (5.53).

(ii) For $\delta = 0$,

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} + \xi^{(2)}(\alpha;\boldsymbol{\theta}) + O_p(1). \quad (5.88)
\end{aligned}$$

By (5.86), it can be seen that for any $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$ and $\delta \in (0,1)$,

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta};\alpha^c) &= (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{-1/2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\kappa_j n^{(1+\delta)/2} + \xi^{(2)}(\alpha;\boldsymbol{\theta}) - \xi^{(2)}(\alpha^c;\boldsymbol{\theta}) \\
&\quad + o_p(n^\delta) \\
&= (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{-1/2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\kappa_j n^{(1+\delta)/2} + o_p(n^{(1+\delta)/2}), \qquad (5.89)
\end{aligned}$$

where the last equality holds because by (5.87), $\xi^{(2)}(\alpha;\boldsymbol{\theta}) - \xi^{(2)}(\alpha^c;\boldsymbol{\theta}) = o_p(n^{(1+\delta)/2})$. Similarly, by (5.88) for $\delta = 0$,

$$-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta};\alpha^c) = (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{-1/2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\kappa_j n^{1/2} + o_p(n^{1/2}).$$

As to be demonstrated in Theorem 9, we can use (5.89) to find an appropriate penalty λ that leads to selection consistency.

The following lemma shows that $\kappa_\eta\sigma_\eta^2$ is over-estimated by ML asymptotically when $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$ under both the fixed domain and the increasing domain asymptotic frameworks.

Lemma 12 *Under the setup of Proposition 5, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat{\sigma}_\eta^2(\alpha), \hat{\kappa}_\eta(\alpha), \hat{\sigma}_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then, for $\delta \in [0,1)$ and $\alpha \in \mathcal{A}$,*

$$\begin{aligned}\hat{\sigma}_\epsilon^2(\alpha) &= \sigma_{\epsilon,0}^2 + o_p(1), && (5.90)\\ \hat{\kappa}_\eta(\alpha)\hat{\sigma}_\eta^2(\alpha) &= \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2\kappa_j + \kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(1). && (5.91)\end{aligned}$$

The following theorem further provides the convergence rates for the ML estimates of κ_η, σ_η^2 and σ_ϵ^2. These results are keys for establishing some asymptotic properties of GIC in Theorem 10.

Theorem 8 *Under the setup of Proposition 5, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat{\sigma}_\eta^2(\alpha), \hat{\kappa}_\eta(\alpha), \hat{\sigma}_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then*

(i) For $\delta \in (0,1)$ and $\theta_{\eta,\alpha}$ defined in (5.85),

$$\begin{aligned}\hat{\sigma}_\epsilon^2(\alpha) &= \sigma_{\epsilon,0}^2 + o_p(n^{-(1-\delta)/2}); \quad \alpha\in\mathcal{A}, && (5.92)\\ \hat{\kappa}_\eta(\alpha)\hat{\sigma}_\eta^2(\alpha) &= \begin{cases}\kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(n^{-(1-\delta)/4}); & \text{if } \alpha\in\mathcal{A}^c,\\ \theta_{\eta,\alpha} + o_p(n^{-(1-\delta)/4}); & \text{if } \alpha\in\mathcal{A}\setminus\mathcal{A}^c,\end{cases} && (5.93)\\ \hat{\sigma}_\eta^2(\alpha) &= \begin{cases}\sigma_{\eta,0}^2 + o_p(1); & \text{if } \alpha\in\mathcal{A}^c,\\ \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2 + o_p(1); & \text{if } \alpha\in\mathcal{A}\setminus\mathcal{A}^c,\end{cases} && (5.94)\\ \hat{\kappa}_\eta(\alpha) &= \begin{cases}\kappa_{\eta,0} + o_p(1); & \text{if } \alpha\in\mathcal{A}^c,\\ \theta_{\eta,\alpha}\big(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\big)^{-1} + o_p(1); & \text{if } \alpha\in\mathcal{A}\setminus\mathcal{A}^c.\end{cases} && (5.95)\end{aligned}$$

(ii) For $\delta = 0$,

$$\begin{aligned}\hat{\sigma}_\epsilon^2(\alpha) &= \sigma_{\epsilon,0}^2 + O_p(n^{-1/2}); \quad \alpha\in\mathcal{A}, && (5.96)\\ \hat{\kappa}_\eta(\alpha)\hat{\sigma}_\eta^2(\alpha) &= \begin{cases}\kappa_{\eta,0}\sigma_{\eta,0}^2 + O_p(n^{-1/4}); & \text{if } \alpha\in\mathcal{A}^c,\\ \theta_{\eta,\alpha} + O_p(n^{-1/4}); & \text{if } \alpha\in\mathcal{A}\setminus\mathcal{A}^c.\end{cases} && (5.97)\end{aligned}$$

Proof. Let $\sigma_{\eta,\alpha}^2 \equiv \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2$ and $\kappa_{\eta,\alpha} \equiv \theta_{\eta,\alpha}/\sigma_{\eta,\alpha}^2$, for $\alpha\in\mathcal{A}$, where $\theta_{\eta,\alpha}$ is defined in (5.85). In addition, for $\alpha\in\mathcal{A}\setminus\mathcal{A}^c$, let

$$\begin{aligned}\xi_1(\boldsymbol{\theta};\alpha) &= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha)\\ &\quad - \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})),\\ \xi_2(\boldsymbol{\theta};\alpha) &= -2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}). && (5.98)\end{aligned}$$

In advance, we prove a simpler expansion of (5.87),

$$\xi^{(2)}(\boldsymbol{\theta};\alpha) = \xi_1(\boldsymbol{\theta};\alpha) + \xi_2(\boldsymbol{\theta};\alpha) + \xi(\boldsymbol{\theta}) + O_p(1). \qquad (5.99)$$

By (5.87), we have

$$\begin{aligned}\xi^{(2)}(\boldsymbol{\theta};\alpha) &= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha)\\ &\quad -\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha)\\ &\quad -2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &\quad +2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}),\end{aligned}$$

where $\boldsymbol{M}(\alpha;\boldsymbol{\theta})$ is defined in (3.5). Therefore, for (5.99) to hold, it remains to show that

$$\begin{aligned}\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) &= O_p(1), \qquad (5.100)\\ \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= O_p(1).\end{aligned}$$

It follows easily from

$$n^{(1+\delta)/2}\big(\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha)\big)^{-1} = O_p(1), \qquad (5.101)$$

$$\frac{1}{n^{(1+\delta)/4}}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1), \qquad (5.102)$$

which follows from (5.33), (5.34) and Chebyshev's inequality by checking the following moment conditions:

$$\begin{aligned}\mathrm{var}\big(\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_{j'}\big) &= \mathrm{tr}\big(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big) = O(n^{(1+\delta)/2}),\\ \mathrm{var}\big(\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big) &= \mathrm{tr}\big(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big) = O(n^{(1+\delta)/2}).\end{aligned}$$

Thus, (5.99) is obtained.

First, we prove (5.92). By (5.90) and (5.91), it suffices to show that for $|\sigma_\epsilon^2-\sigma_{\epsilon,0}^2| = o(1)$, $|\kappa_\eta\sigma_\eta^2-\theta_{\eta,\alpha}| = o(1)$, and any $\varepsilon>0$,

$$\inf_{|\sigma_\epsilon^2-\sigma_{\epsilon,0}^2|\geq\varepsilon n^{-(1-\delta)/2}}\big(-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,0}^2)';\alpha)\big) > 0, \qquad (5.103)$$

as $n\to\infty$ with probability tending to 1. By (5.86),

$$\begin{aligned}-2\ell(\boldsymbol{\theta};\alpha) = {}& n\log(2\pi)-\frac{1-\delta}{2}\log n+\Big(\log\sigma_\epsilon^2+\frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\Big)n\\ &+\Big(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\Big)^{1/2}\Big(1-\frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2}+\frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\Big)n^{(1+\delta)/2}\\ &+\Big(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}+\kappa_\eta\Big(\frac{\theta_{\eta,\alpha}}{\kappa_\eta\sigma_\eta^2}-1\Big)-\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\Big)n^\delta\\ &+\frac{\sigma_{\eta,\alpha}^2}{2\kappa_\eta\sigma_\eta^2}\big(\kappa_\eta-\kappa_{\eta,\alpha}\big)^2n^\delta+\frac{1}{2\kappa_\eta\sigma_\eta^2}\Big(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\kappa_j^2+\sigma_{\eta,0}^2\kappa_{\eta,0}^2-\theta_{\eta,\alpha}\kappa_{\eta,\alpha}\Big)n^\delta\\ &+\xi_1(\boldsymbol{\theta};\alpha)+\xi_2(\boldsymbol{\theta};\alpha)+\xi(\boldsymbol{\theta})+o_p(n^\delta), \qquad (5.104)\end{aligned}$$

where $\xi(\boldsymbol{\theta})$ is defined in (5.53) and the last equality follows from (5.99). Then, for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} - \log\sigma_{\epsilon,0}^2 - 1\right)n \\
&\quad + \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right) - \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)\right\}n^{(1+\delta)/2}\right. \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + o_p(n^\delta) \\
&= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n}{2\sigma_{\epsilon,0}^4} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + o((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n) + o_p(n^\delta) \\
&= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n}{2\sigma_{\epsilon,0}^4} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + o((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n) + o_p(n^\delta),
\end{aligned}$$

where the second equality follows from (5.39) and (5.40) and the last equality follows from (5.67). Therefore, for (5.103) to hold, it remains to show that

$$\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) = o_p\big(\max\big((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n,\ n^\delta\big)\big), \tag{5.105}$$

$$\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) = o_p\big(\max\big((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n,\ n^\delta\big)\big). \tag{5.106}$$

By (5.98), we have

$$\begin{aligned}
\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) &= -2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')\big)\boldsymbol{\eta} \\
&\quad -2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')\big)\boldsymbol{\epsilon}.
\end{aligned}$$

Then, (5.105)-(5.106) follow from Chebyshev's inequality and using the following three moments conditions given from (5.42)-(5.43):

$$\begin{aligned}
\mathrm{var}(\boldsymbol{X}_j'((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')))\boldsymbol{X}_{j'}) &= O(\max((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n, n^\delta)), \\
\mathrm{var}(\boldsymbol{X}_j'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)'))\boldsymbol{\eta}) &= O(\max((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n, n^\delta)), \\
\mathrm{var}(\boldsymbol{X}_j'((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')))\boldsymbol{\epsilon}) &= O(\max((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n, n^\delta)),
\end{aligned}$$

we obtain (5.105)-(5.106). This completes the proof of (5.92).

Second, we prove (5.93). By (5.91) and (5.92), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, $|\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$ and any $\varepsilon > 0$,

$$\inf_{|\sigma_\eta^2\kappa_\eta - \theta_{\eta,\alpha}| \geq \varepsilon n^{-(1-\delta)/4}} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha)\big) > 0, \tag{5.107}$$

as $n \to \infty$ with probability tending to 1. By (5.104), we have for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$ and $|\kappa_\eta \sigma_\epsilon^2 - \theta_{\eta,\alpha}| = o(1)$,

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\theta_{\eta,\alpha}}{\sigma_{\epsilon,0}^2}\right)^{1/2}\right\} n^{(1+\delta)/2} + \frac{1}{2\kappa_{\eta,\alpha}}(\kappa_\eta - \kappa_{\eta,\alpha})^2 n^\delta \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + o_p(n^\delta) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}}{2^{5/2}\sigma_{\epsilon,0}\theta_{\eta,\alpha}^{3/2}} + \frac{1}{2\kappa_{\eta,\alpha}}(\kappa_\eta - \kappa_{\eta,\alpha})^2 n^\delta + o((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}) \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + o_p(n^\delta),
\end{aligned} \tag{5.108}$$

where the first equality follows from $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, the second equality follows from (5.41). Therefore, for (5.107) to hold, it suffices to show that

$$\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) = o_p(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)), \tag{5.109}$$

$$\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) = o_p(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)), \tag{5.110}$$

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') = o_p(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)), \tag{5.111}$$

which can be obtained from the following three moment conditions given from (5.45)-(5.47):

$$\begin{aligned}
\mathrm{var}(\boldsymbol{X}_j'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)'))\boldsymbol{\eta}) &= O(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)), \\
\mathrm{var}(\boldsymbol{X}_j'((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')))\boldsymbol{\epsilon}) &= O(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)), \\
\mathrm{var}(\boldsymbol{\epsilon}'((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')))\boldsymbol{\epsilon}) &= O(\max((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}, n^\delta)),
\end{aligned}$$

Thus, (5.107) is obtained. This completes the proof of (5.93).

Third, we prove (5.94) and (5.95). By (5.108), we have for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, $|\sigma_\eta^2\kappa_\eta - \theta_{\eta,\alpha}| = o(n^{-(1-\delta)/4})$ and any $\varepsilon > 0$,

$$\inf_{|\kappa_\eta - \kappa_{\eta,\alpha}| \geq \varepsilon} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) = \frac{1}{2\kappa_{\eta,\alpha}}\varepsilon^2 n^\delta + o_p(n^\delta) > 0,$$

as $n \to \infty$ with probability tending to 1, which gives (5.95). This together with (5.93) gives (5.94).

Fourth, we prove (5.96). By (5.90) and (5.91), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$, $|\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| \geq Mn^{-1/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha)) > 0, \tag{5.112}$$

as $n \to \infty$ with probability tending to 1. By (5.88) and (5.99),

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) + \xi_2(\boldsymbol{\theta};\alpha) + \xi(\boldsymbol{\theta}) + O_p(1).
\end{aligned} \tag{5.113}$$

Then, for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ and $|\kappa_\eta \sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left(\log \sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} - \log \sigma_{\epsilon,0}^2 - 1\right) n \\
&\quad + \left\{\left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta \sigma_\eta^2}\right)\right\} n^{1/2} \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + O_p(1) \\
&= \frac{1}{2\sigma_{\epsilon,0}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)') + o_p\big(((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n)\big) + O_p(1) \\
&= \frac{1}{2\sigma_{\epsilon,0}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n + o_p\big(((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n)\big) + O_p(1),
\end{aligned}$$

where the second equality follows from (5.39) and (5.40), and the last equality follows from (5.73) and

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) &= o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n) + O_p(1), \\
\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) &= o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n) + O_p(1),
\end{aligned}$$

which can be obtained in a way similar to (5.105)-(5.106). Consequently, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| \geq M n^{-1/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha)) = \frac{M^2}{2\sigma_{\epsilon,0}^4}\varepsilon^2 + O_p(1) > 0,$$

as $n \to \infty$ with probability tending to 1. Thus, we obtain (5.112) and hence the proof of (5.96) is complete.

Finally, we prove (5.97). By (5.91) and (5.96), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$, $|\kappa_\eta \sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$ and there exist $M > 0$ such that

$$\inf_{|\sigma_\eta^2 \kappa_\eta - \theta_{\eta,\alpha}| \geq M n^{-1/4}} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha)\big) > 0, \tag{5.114}$$

as $n \to \infty$ with probability tending to 1, where $\kappa_{\eta,\alpha}\sigma_{\eta,\alpha}^2 = \theta_{\eta,\alpha}$. By (5.113), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$ and $|\kappa_\eta \sigma_\eta^2 - \theta_{\eta,\alpha}| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\theta_{\eta,\alpha}}{\sigma_{\epsilon,0}^2}\right)^{1/2}\right\} n^{1/2} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}}{2^{5/2}\theta_{\eta,\alpha}^{3/2}} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&\quad + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') \\
&\quad + o((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}) + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}}{2^{5/2}\theta_{\eta,\alpha}^{3/2}} + o((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}) + O_p(1),
\end{aligned} \tag{5.115}$$

where the first equality follows from $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$, the second equality follows from (5.41) and the last equality follows from

$$\begin{aligned} \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})';\alpha) &= o_p((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}) + O_p(1), \\ \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})';\alpha) &= o_p((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}) + O_p(1), \quad (5.116) \\ \xi(\boldsymbol{\theta}) - \xi((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})') &= o_p((\kappa_\eta\sigma_\eta^2 - \theta_{\eta,\alpha})^2 n^{1/2}) + O_p(1), \quad (5.117) \end{aligned}$$

which can be obtained in a way similar to (5.109)-(5.111). Thus, (5.114) and hence (5.97) are obtained. This completes the proof. □

Corollary 5 *Under the setup of Theorem 8, let*

$$\boldsymbol{\theta}_\alpha^{(2)} = \left(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2, \theta_{\eta,\alpha}\left(\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)^{-1}, \sigma_{\epsilon,0}^2\right)', \quad (5.118)$$

where $\theta_{\eta,\alpha}$ is defined in (5.85). For the log-likelihood defined in (2.9),

$$\begin{aligned} \operatorname*{plim}_{n\to\infty}\frac{1}{n^\delta}(-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(2)};\alpha)) &= 0; \quad \text{if } \delta\in(0,1), \quad (5.119) \\ -2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(2)};\alpha) &= O_p(1); \quad \text{if } \delta = 0. \quad (5.120) \end{aligned}$$

In addition, for $L^{KL}(\boldsymbol{\theta};\alpha)$ defined in (3.3),

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\boldsymbol{\theta}}(\alpha);\alpha)/L^{KL}(\boldsymbol{\theta}_\alpha^{(2)};\alpha)) = 0; \quad \text{if } \delta\in[0,1). \quad (5.121)$$

Note that from Theorem 8, $\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\theta}}(\alpha) = \boldsymbol{\theta}_\alpha^{(2)}$ for $\delta\in(0,1)$, which immediately implies (5.119). On the other hand, (5.79) is somewhat surprising, because $\hat{\boldsymbol{\theta}}(\alpha)$ generally does not converge to $\boldsymbol{\theta}_\alpha^{(2)}$ for $\delta = 0$. However, selection consistency and asymptotic loss efficiency are possible for geostatistical model selection even if some covariance parameters cannot be consistently estimated under the fixed domain asymptotic framework (see Theorem 10).

Theorem 9 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from zero-mean spatial processes having exponential covariance functions of (5.6) and $cov(\eta(s),\eta(s')) = \sigma_\eta^2\exp(-\kappa_\eta|s-s'|)$, where $\mathcal{A}^c \neq \emptyset$ and p is fixed. Suppose that $\sigma_\eta^2 > 0, \kappa_\eta > 0$ and $\sigma_\epsilon^2 > 0$ are known. In addition, suppose that the data are collected at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1,\ldots,n$ for some $\delta\in[0,1)$. If $\lambda\to\infty$ and $\lambda/n\to 0$, then*

$$L^{KL}(\hat{\alpha}_{GIC(\lambda)})/\min_{\alpha\in\mathcal{A}} L^{KL}(\alpha) \xrightarrow{p} 1, \quad \text{as } n\to\infty.$$

In addition,

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. By Theorem 2, it suffices to show that $\lim_{n\to\infty} \mathrm{E}(L^{\mathrm{KL}}(\alpha))/\lambda = \infty$ for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$ by (4.7). First, for $\delta[0,1)$,

$$\begin{aligned}
\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} &= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) \\
&\quad -\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) \\
&= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) + O_p(1) \\
&= \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2 \mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}) + o_p(n^{(1+\delta)/2}) \\
&= \sum_{j\in\alpha^c\setminus\alpha} \frac{\beta_j^2\sigma_j^2\kappa_j}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} n^{(1+\delta)/2} + o_p(n^{(1+\delta)/2}), \qquad (5.122)
\end{aligned}$$

where $\boldsymbol{M}(\alpha)$ is defined in (3.5), $\boldsymbol{\Sigma}_j = \mathrm{var}(\boldsymbol{X}_j)$, the second equality follows from (5.100), the third equality follows if

$$\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) = \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2 \mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}) + o_p(n^{(1+\delta)/2}), \qquad (5.123)$$

and the last equality follows from (5.30). By Chebyshev's inequality, (5.123) holds by

$$\mathrm{var}(\boldsymbol{X}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{j'}) = \boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1} = O(n^{(1+\delta)/2}),$$

which follows from (5.33). It follows from (3.9) and $\lim_{n\to\infty} \lambda/n^{(1+\delta)/2} = 0$ that for $\delta \in [0,1)$.

$$\lim_{n\to\infty} \frac{E(L(\alpha))}{\lambda} \geq \lim_{n\to\infty} \frac{\mathrm{E}(\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu})/n^{(1+\delta)/2}}{\lambda/n^{(1+\delta/2)}} = \infty.$$

This completes the proof. □

¿From Theorems 5, 6 and 9, we see that the behavior of GIC highly depends on the variables to be selected under the fixed domain asymptotic framework (i.e., $\delta = 0$). When the variables to be selected are polynomials, GIC fails to select α^c unless $\alpha^c = \emptyset$. In contrast, when the variables to be selected are generated from some spatial processes, GIC is consistent as long as $\lambda \to \infty$ and $\lambda/n \to 0$. Generally speaking, GIC has better ability to distinguish among variables that are less smooth, which is somewhat expected, because less smooth variables tends to produce less smooth mean structure and hence is less confounded with the spatial process $\eta(\cdot)$.

Theorem 10 *Under the setup of Theorem 9, suppose that $\boldsymbol{\theta} = (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)' \in \Theta$ is unknown, where $\Theta \subset (0,\infty)^3$ is a compact set such that $\boldsymbol{\theta}_0 \in \Theta$. Let $\hat{\boldsymbol{\theta}}(\alpha)$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. If $\delta \in [0,1)$, $\lambda \to \infty$ and $\lambda/n^{(1+\delta)/2} \to 0$, then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. For the consistency, it suffices to show that the conditions in Corollary 3 are satisfied with $\tau_n = n^{(1+\delta)/2}$. First, by (5.122), we have for any $\boldsymbol{\theta} \in \Theta$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n^{(1+\delta)/2}} \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{\theta})(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} = \sum_{j\in\alpha^c\setminus\alpha} \frac{\beta_j^2\sigma_j^2\kappa_j}{(2\kappa_\eta\kappa_\sigma^2\sigma_\epsilon^2)^{1/2}},$$

where Θ is the covariance parameter space. Hence, (A.1') is satisfied. Second, by (5.30) and (5.33), we have for any $\boldsymbol{\theta} \in \Theta$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n^{(1+\delta)/2}} \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X} = \boldsymbol{D}(\boldsymbol{\theta}),$$

where $\boldsymbol{D}(\boldsymbol{\theta})$ is a $p \times p$ diagonal matrix with diagonals $\kappa_j \sigma_j^2/(2\kappa_\eta \sigma_\eta \sigma_\epsilon^2)^{1/2}$, $j = 1, \dots, p$. Hence, (A.2') holds. Third, by (5.33) and (5.34), (A.3') holds trivially. Fourth, by (5.119)-(5.121), (A.4) and (A.5) hold trivially for $\tau_n = n^{(1+\delta)/2}$ and $\boldsymbol{\theta}_\alpha^{(2)}$ defined in (5.118). Fifth, for $\xi(\boldsymbol{\theta})$ defined in (5.53), by (5.111) and (5.117), we have for $\delta \in [0, 1)$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n^{(1+\delta)/2}} \big(\xi(\boldsymbol{\theta}_0) - \xi(\boldsymbol{\theta}_\alpha^{(2)})\big) \;=\; 0.$$

Hence, (4.12) is satisfied. Last, for $\alpha \in \mathcal{A}^c$, $\boldsymbol{\theta}_\alpha^{(2)} = \boldsymbol{\theta}_0$, (4.14) holds trivially. Then, for $\mathcal{A}^c \neq \emptyset$, $\lambda \to \infty$ and $\lambda = o(n^{(1+\delta)/2})$, we have

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} = \alpha^c\big) = 1,$$

which completes the proof. □

5.3 White Noise Regressors

In this section, we consider explanatory variables generated independently from Gaussian white noise processes of (5.7).

Proposition 6 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7) and $\mathrm{cov}(\eta(s), \eta(s')) = \sigma_\eta^2 \exp(-\kappa_\eta |s - s'|)$, where $\mathcal{A}^c \neq \emptyset$ and p is fixed. Suppose that $\sigma_\eta^2 > 0$, $\kappa_\eta > 0$ and $\sigma_\epsilon^2 > 0$ are known. In addition, suppose that the data are collected at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1, \dots, n$ for some $\delta \in [0, 1)$. Let $\boldsymbol{\theta} = (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)' \in \Theta$ and $\boldsymbol{\theta}_0 = (\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,0}^2)' \in \Theta$ be the true parameter vector, where $\Theta = (0, \infty)^3$. Then the log-likelihood of (4.9) based on $\alpha \in \mathcal{A}$ can be decomposed into the following:*

(i) For $\delta \in (0, 1)$,

$$\begin{aligned} -2\ell(\boldsymbol{\theta}; \alpha) \;=\;& n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\ &+\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\ &+\left(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\left(\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\kappa_\eta\sigma_\eta^2} - 1\right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^\delta \\ &+\frac{\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\left(\kappa_\eta - \kappa_{\eta,0}\right)^2 n^\delta + \xi^{(3)}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta), \end{aligned} \qquad (5.124)$$

where

$$\begin{aligned} \xi^{(3)}(\alpha;\boldsymbol{\theta}) \;=\;& \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} - \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\, tr(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\ &+\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &-tr(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})), \end{aligned} \qquad (5.125)$$

and $\boldsymbol{A}(\alpha;\boldsymbol{\theta})$ is defined in (3.6).

(ii) For $\delta = 0$,

$$
\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} \\
&\quad + \xi^{(3)}(\alpha;\boldsymbol{\theta}) + O_p(1). \qquad (5.126)
\end{aligned}
$$

By (5.124), for $\delta \in (0,1)$ and any $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$,

$$
\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta};\alpha^c) &= \sigma_\epsilon^{-2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 n + \xi^{(3)}(\alpha;\boldsymbol{\theta}) - \xi^{(2)}(\alpha^c;\boldsymbol{\theta}) + o_p(n) \\
&= \sigma_\epsilon^{-2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 n + o_p(n), \qquad (5.127)
\end{aligned}
$$

where the last equality holds because by (5.125), $\xi^{(3)}(\alpha;\boldsymbol{\theta}) - \xi^{(3)}(\alpha^c;\boldsymbol{\theta}) = o_p(n)$. Similarly by (5.126) for $\delta = 0$,

$$
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta};\alpha^c) = \sigma_\epsilon^{-2}\sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 n + o_p(n).
$$

As to be demonstrated in Theorem 12, we can use (5.127) to find an appropriate penalty λ that leads to selection consistency.

The following lemma shows that σ_ϵ^2 is over-estimated by ML asymptotically when $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$ under both the fixed domain and the increasing domain asymptotic frameworks.

Lemma 13 *Under the setup of Proposition 6, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat\sigma_\eta^2(\alpha), \hat\kappa_\eta(\alpha), \hat\sigma_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then for $\delta \in [0,1)$ and $\alpha \in \mathcal{A}$,*

$$
\begin{aligned}
\hat\sigma_\epsilon^2(\alpha) &= \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2 + o_p(1), \qquad (5.128) \\
\hat\kappa_\eta(\alpha)\hat\sigma_\eta^2(\alpha) &= \kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(1). \qquad (5.129)
\end{aligned}
$$

The following theorem further provides the convergence rates for the ML estimates of κ_η, σ_η^2 and σ_ϵ^2. These results are keys for establishing some asymptotic properties of GIC in Theorem 13.

Theorem 11 *Under the setup of Proposition 6, let $\Theta \subset (0,\infty)^3$ be a compact set and let $\hat{\boldsymbol{\theta}}(\alpha) = (\hat\sigma_\eta^2(\alpha), \hat\kappa_\eta(\alpha), \hat\sigma_\epsilon^2(\alpha))'$ be the ML estimate of $\boldsymbol{\theta}$ based on model α. Then*

(i) For $\delta \in (0,1)$,

$$
\begin{aligned}
\hat\sigma_\epsilon^2(\alpha) &= \begin{cases} \sigma_{\epsilon,0}^2 + o_p(n^{-(1-\delta)/2}); & \text{if } \alpha \in \mathcal{A}^c, \\ \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2 + o_p(n^{-(1-\delta)/2}); & \text{if } \alpha \in \mathcal{A}\setminus\mathcal{A}^c, \end{cases} \qquad (5.130) \\
\hat\kappa_\eta(\alpha)\hat\sigma_\eta^2(\alpha) &= \kappa_{\eta,0}\sigma_{\eta,0}^2 + o_p(n^{-(1-\delta)/4}), \qquad (5.131) \\
\hat\sigma_\eta^2(\alpha) &= \sigma_{\eta,0}^2 + o_p(1); \quad \alpha \in \mathcal{A}, \qquad (5.132) \\
\hat\kappa_\eta(\alpha) &= \kappa_{\eta,0} + o_p(1); \quad \alpha \in \mathcal{A}. \qquad (5.133)
\end{aligned}
$$

(ii) *For* $\delta = 0$,

$$\hat{\sigma}_\epsilon^2(\alpha) = \begin{cases} \sigma_{\epsilon,0}^2 + O_p(n^{-1/2}); & \text{if } \alpha \in \mathcal{A}^c, \\ \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2 + O_p(n^{-1/2}); & \text{if } \alpha \in \mathcal{A}\setminus\mathcal{A}^c, \end{cases} \quad (5.134)$$

$$\hat{\kappa}_\eta(\alpha)\hat{\sigma}_\eta^2(\alpha) = \kappa_{\eta,0}\sigma_{\eta,0}^2 + O_p(n^{-1/4}). \quad (5.135)$$

Proof. Let $\sigma_{\epsilon,\alpha}^2 \equiv \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2$ for $\alpha \in \mathcal{A}$. In addition, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, let

$$\begin{aligned} \xi_1(\boldsymbol{\theta};\alpha) &= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) \\ &\quad - \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})), \end{aligned} \quad (5.136)$$

$$\xi_2(\boldsymbol{\theta};\alpha) = -2\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}). \quad (5.137)$$

Then, it can be obtained in a way similar to (5.99) that

$$\xi^{(3)}(\boldsymbol{\theta};\alpha) = \xi_1(\boldsymbol{\theta};\alpha) + \xi_2(\boldsymbol{\theta};\alpha) + \xi(\boldsymbol{\theta}) + O_p(1). \quad (5.138)$$

First, we prove (5.130). By (5.128) and (5.129), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(1)$, $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, and any $\varepsilon > 0$,

$$\inf_{|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|\geq\varepsilon n^{-(1-\delta)/2}} \left(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha)\right) > 0, \quad (5.139)$$

as $n\to\infty$ with probability tending to 1. By (5.124),

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,\alpha}^2}{\sigma_\epsilon^2}\right)n \\ &\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,\alpha}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\ &\quad + \left(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\left(\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\kappa_\eta\sigma_\eta^2} - 1\right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^\delta + \frac{\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\big(\kappa_\eta - \kappa_{\eta,0}\big)^2 n^\delta \\ &\quad + \xi_1(\boldsymbol{\theta};\alpha) + \xi_2(\boldsymbol{\theta};\alpha) + \xi(\boldsymbol{\theta}) + o_p(n^\delta), \end{aligned} \quad (5.140)$$

where $\xi(\boldsymbol{\theta})$ is defined in (5.53) and the equality follows from (5.138). Then, for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned} &-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) \\ &= \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,\alpha}^2}{\sigma_\epsilon^2} - \log\sigma_{\epsilon,\alpha}^2 - 1\right)n \\ &\quad + \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,\alpha}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,\alpha}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,\alpha}^2}{2\kappa_\eta\sigma_\eta^2}\right)\right\}n^{(1+\delta)/2} \\ &\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) \\ &\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') + o_p(n^\delta) \\ &= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n}{2\sigma_{\epsilon,\alpha}^4} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) \\ &\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') + o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + o_p(n^\delta) \\ &= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n}{2\sigma_{\epsilon,\alpha}^4} + o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + o_p(n^\delta), \end{aligned}$$

where the second equality follows from $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(1)$, (5.39) and (5.40), and the last equality follows form (5.67) and

$$\begin{aligned}\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(\max((\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2)^2n,\,n^\delta)), &(5.141)\\ \xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(\max((\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2)^2n,\,n^\delta)), &(5.142)\end{aligned}$$

which can be obtained in a way similar to (5.105)-(5.106) where the moment conditions are given from (5.43)-(5.44) in this case. Thus, (5.139) is obtained. This completes the proof of (5.130).

Second, we prove (5.131). By (5.129) and (5.130), it suffices to show that for $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(n^{-(1-\delta)/2})$, $|\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2|=o(1)$ and any $\varepsilon>0$,

$$\inf_{|\sigma_\eta^2\kappa_\eta-\kappa_{\eta,0}\sigma_{\eta,0}^2|\geq\varepsilon n^{-(1-\delta)/4}}\big(-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\big) > 0, \tag{5.143}$$

as $n\to\infty$ with probability tending to 1. By (5.140), we have for $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(n^{-(1-\delta)/2})$ and $|\kappa_\eta\sigma_\epsilon^2-\kappa_{\eta,0}\sigma_{\eta,0}^2|=o(1)$,

$$\begin{aligned}&-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\ &= \Big\{\Big(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,\alpha}^2}\Big)^{1/2}\Big(\frac{1}{2}+\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\Big)-\frac{(2\kappa_{\eta,0}\sigma_{\eta,0}^2)^{1/2}}{\sigma_{\epsilon,\alpha}}\Big\}n^{(1+\delta)/2}+\frac{1}{2\kappa_{\eta,\alpha}}(\kappa_\eta-\kappa_{\eta,\alpha})^2n^\delta\\ &\quad+\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)+\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\ &\quad+\xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)')+o_p(n^\delta)\\ &= \frac{(\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}}+\frac{1}{2\kappa_{\eta,0}}(\kappa_\eta-\kappa_{\eta,0})^2n^\delta+\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\ &\quad+\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)+\xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)')\\ &\quad+o((\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2})+o_p(n^\delta)\\ &= \frac{(\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}}+\frac{(\kappa_\eta-\kappa_{\eta,0})^2n^\delta}{2\kappa_{\eta,0}}+o((\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2})\\ &\quad+o_p(n^\delta),\end{aligned} \tag{5.144}$$

where the first equality follows from $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(n^{-(1-\delta)/2})$, the second equality follows from (5.41), and the last equality follows from

$$\begin{aligned}\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(\max((\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2},n^\delta)), &(5.145)\\ \xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(\max((\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2},n^\delta)), &(5.146)\\ \xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)') &= o_p(\max((\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2},n^\delta)), &(5.147)\end{aligned}$$

which can be obtained in a way similar to (5.109)-(5.111). Thus, (5.143) is obtained. This completes the proof of (5.131).

Third, we prove (5.132) and (5.133). By (5.144), we have for $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(n^{-(1-\delta)/2})$, $|\sigma_\eta^2\kappa_\eta-\kappa_{\eta,0}\sigma_{\eta,0}^2|=o(n^{-(1-\delta)/4})$ and any $\varepsilon>0$,

$$\inf_{|\kappa_\eta-\kappa_{\eta,0}|\geq\varepsilon}-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha) = \frac{1}{2\kappa_{\eta,0}}\varepsilon^2n^\delta+o_p(n^\delta) > 0,$$

as $n\to\infty$ with probability tending to 1, which gives (5.133). This together with (5.131) gives (5.132).

Fourth, we prove (5.134). By (5.128) and (5.129), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(1)$, $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|\geq Mn^{-1/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha)) \;>\; 0, \qquad (5.148)$$

as $n \to \infty$ with probability tending to 1. By (5.126) and (5.138),

$$\begin{aligned}-2\ell(\boldsymbol{\theta};\alpha) \;=\; & n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,\alpha}^2}{\sigma_\epsilon^2}\right)n \\ & + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,\alpha}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} \\ & + \xi_1(\boldsymbol{\theta};\alpha) + \xi_2(\boldsymbol{\theta};\alpha) + \xi(\boldsymbol{\theta}) + O_p(1). \qquad (5.149)\end{aligned}$$

Then, for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned}& -2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) \\ = \;& \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,\alpha}^2}{\sigma_\epsilon^2} - \log\sigma_{\epsilon,\alpha}^2 - 1\right)n \\ & + \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,\alpha}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,\alpha}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)\right\}n^{1/2} \\ & + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) \\ & + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)') + O_p(1) \\ = \;& \frac{1}{2\sigma_{\epsilon,\alpha}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) \\ & + \xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)') + o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + O_p(1) \\ = \;& \frac{1}{2\sigma_{\epsilon,\alpha}^4}(\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n + o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + O_p(1),\end{aligned}$$

where the second equality follows from (5.39) and (5.40) with $\sigma = \sigma_{\epsilon,\alpha}$ and $\tau = \kappa_{\eta,0}\sigma_{\eta,0}^2$, and the last equality follows from (5.73) and

$$\begin{aligned}\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) \;&=\; o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + O_p(1), \\ \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha) \;&=\; o_p((\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n) + O_p(1),\end{aligned}$$

which can be obtained in a way similar to (5.105)-(5.106). Consequently, there exists $M > 0$ such that

$$\inf_{|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|\geq Mn^{-1/2}} (-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,\alpha}^2)';\alpha)) \;=\; \frac{M^2}{2\sigma_{\epsilon,\alpha}^4}\varepsilon^2 + O_p(1) > 0,$$

as $n \to \infty$ with probability tending to 1. Thus, (5.148) is obtained. This completes the proof of (5.134).

Finally, we prove (5.135). By (5.129) and (5.134), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = O(n^{-1/2})$, $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$ and there exist $M > 0$ such that

$$\inf_{|\sigma_\eta^2\kappa_\eta-\kappa_{\eta,0}\sigma_{\eta,0}^2|\geq Mn^{-1/4}} \Big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha)\Big) \;>\; 0, \qquad (5.150)$$

as $n \to \infty$ with probability tending to 1. By (5.149), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = O(n^{-1/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,\alpha}^2}\right)^{1/2}\left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right) - \left(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,\alpha}^2}\right)\right\}n^{1/2} \\
&\quad + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)') + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) \\
&\quad + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)') \\
&\quad + o\big((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}\big) + O_p(1) \\
&= \frac{(\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}}{2^{5/2}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + o\big((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}\big) + O_p(1), \qquad (5.151)
\end{aligned}$$

where the first equality follows from $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = O(n^{-1/2})$, the second equality follows from (5.41), and the last equality follows from

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}) + O_p(1), \\
\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}) + O_p(1), \quad (5.152) \\
\xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)') &= o_p((\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{1/2}) + O_p(1), \quad (5.153)
\end{aligned}$$

which can be obtained in a way similar to (5.109)-(5.111). Thus, (5.150) is obtained. This completes the proof of (5.135). □

Corollary 6 *Under the setup of Theorem 11, let*

$$\boldsymbol{\theta}_\alpha^{(3)} = \left(\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2\right). \qquad (5.154)$$

Then for $\ell(\boldsymbol{\theta};\alpha)$ defined in (2.9),

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n^\delta}(-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha)) = 0; \quad \text{if } \delta \in (0,1), \qquad (5.155)$$

$$-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha) = O_p(1); \quad \text{if } \delta = 0. \qquad (5.156)$$

In addition, for $L^{KL}(\boldsymbol{\theta};\alpha)$ defined in (3.3),

$$\operatorname*{plim}_{n\to\infty} L^{KL}(\hat{\boldsymbol{\theta}}(\alpha);\alpha)/L^{KL}(\boldsymbol{\theta}_\alpha^{(3)};\alpha)) = 0; \quad \text{if } \delta \in [0,1). \qquad (5.157)$$

Note that from Theorem 11, $\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\theta}}(\alpha) = \boldsymbol{\theta}_\alpha^{(3)}$ for $\delta \in (0,1)$, which immediately implies (5.155). On the other hand, (5.156) is somewhat surprising, because $\hat{\boldsymbol{\theta}}(\alpha)$ generally does not converge to $\boldsymbol{\theta}_\alpha^{(2)}$ for $\delta = 0$. However, selection consistency and asymptotic loss efficiency are possible for geostatistical model selection even if some covariance parameters cannot be consistently estimated under the fixed domain asymptotic framework (see Theorem 13).

Theorem 12 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7) and* $cov(\eta(s),\eta(s')) = \sigma_\eta^2\exp(-\kappa_\eta|s-s'|)$, *where* $\mathcal{A}^c \neq \emptyset$ *and p is fixed. Suppose that* $\sigma_\eta^2 > 0, \kappa_\eta > 0$ *and* $\sigma_\epsilon^2 > 0$ *are known. In addition, suppose that the data are collected at* $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1, \dots, n$ *for some* $\delta \in [0,1)$. *If* $\lambda \to \infty$ *and* $\lambda/n \to 0$, *then*

$$L^{KL}(\hat{\alpha}_{GIC(\lambda)})/\min_{\alpha\in\mathcal{A}} L^{KL}(\alpha) \xrightarrow{p} 1, \quad \text{as } n \to \infty.$$

In addition,

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. By Corollary 2, it suffices to show that

$$\lim_{n\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}^{-1})/\lambda = \infty, \tag{5.158}$$

which follows from (5.31) and $\lambda = o(n)$. This completes the proof. □

Theorem 13 *Under the setup of Theorem 12, suppose that* $\boldsymbol{\theta} = (\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)'$ *is unknown, where* $\Theta \subset (0,\infty)^3$ *is a compact set such that* $\boldsymbol{\theta}_0 \in \Theta$. *Let* $\hat{\boldsymbol{\theta}}(\alpha)$ *be the ML estimate of* $\boldsymbol{\theta}$ *based on model* α. *For* $\delta \in [0,1)$, *if* $\lambda \to \infty$ *and* $\lambda/n \to 0$, *then*

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1.$$

Proof. For the consistency, it suffices to show that the conditions in Corollary 3 are satisfied with $\tau_n = n$. First, for $\delta \in [0,1)$,

$$\begin{aligned}
\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} &= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) \\
&\quad -\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) \\
&= \boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) + O_p(1) \\
&= \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + o_p(n) \\
&= \sum_{j\in\alpha^c\setminus\alpha} \frac{\beta_j^2\sigma_j^2}{\sigma_\epsilon^2} n + o_p(n),
\end{aligned} \tag{5.159}$$

where the second equality is obtained in a way similar to (5.100), the third equality follows from

$$\boldsymbol{\beta}(\alpha^c\setminus\alpha)'\boldsymbol{X}(\alpha^c\setminus\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha^c\setminus\alpha)\boldsymbol{\beta}(\alpha^c\setminus\alpha) = \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + o_p(n),$$

which can be obtained by (5.35), Chebyshev's inequality and using the following moment condition:

$$\mathrm{var}(\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_{j'}) = \sigma_j^2\sigma_{j'}^2\mathrm{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) = O(n),$$

and the last equality follows from (5.31). Hence, (A.1') is satisfied. Second, by (5.31) and (5.35), we have for any $\boldsymbol{\theta} \in \Theta$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X} = \boldsymbol{D}(\boldsymbol{\theta}),$$

where $\boldsymbol{D}(\boldsymbol{\theta})$ is a $p \times p$ diagonal matrix with diagonals $\sigma_j^2/\sigma_\epsilon^2$, $j = 1, \ldots, p$. Hence, (A.2') holds. Third, by (5.34) and (5.35), (A.3') holds. Fourth, by (5.155)-(5.157), (A.4) and (A.5) hold trivially for $\tau_n = n$ and $\boldsymbol{\theta}_\alpha^{(3)}$ defined in (5.154). Fifth, for $\xi(\boldsymbol{\theta})$ defined in (5.53), by (5.147) and (5.153), we have for $\delta \in [0, 1)$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\big(\xi(\boldsymbol{\theta}_0) - \xi(\boldsymbol{\theta}_\alpha^{(3)})\big) \;=\; 0.$$

Hence, (4.12) holds. Last, for $\alpha \in \mathcal{A}^c$, $\boldsymbol{\theta}_\alpha^{(3)} = \boldsymbol{\theta}_0$, (4.14) holds trivially. Then, for $\mathcal{A}^c \neq \emptyset$, $\lambda \to \infty$ and $\lambda = o(n)$, we have

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} = \alpha^c\big) = 1,$$

which completes the proof. □

Comparing among Theorems 7, 10 and 13, we see that GIC is easiest to be consistent when the variables to be selected are from white-noise processes, but is most difficult to be so when the variables to be selected are polynomials.

Chapter 6

Conditional Generalized Information Criterion

If we are interested to find the asymptotic optimal properties of (3.14) throughout some selection procedure, it is somehow difficult to prove the asymptotic properties directly from GIC we introduce above. Another criterion is needed. Vaida and Blanchard (2005) suggest a suitable criterion when we are interesting in spatial process prediction which is named conditional AIC (CAIC). Here we will also suggest a conditional generalized information criterion (CGIC) which includes CAIC as a special case. In the following sections, we are going to introduce the asymptotic theory of CGIC in geostatistical model selection problems.

6.1 Conditional Akaike's Information Criterion

Consider the loss, $L(\alpha)$ defined in (3.14) with estimators, $\hat{\boldsymbol{S}}(\alpha)$ defined in (3.15). It's difficult to find the optimal properties of $L(\alpha)$ directly from the criterion (4.6). Vaida and Blanchard (2005) suggested a conditional AIC (CAIC) selection procedure for the linear mixed models which is an unbiased estimator of $\mathrm{E}(L(\alpha))$ shown in (3.17). They suggested when focus on the mean function estimate, the AIC in (4.3) is good to be a selection procedure. When focus on both the mean function estimate and the spatial process prediction, CAIC is much adequate than AIC to be a selection procedure. That is for $\alpha \in \mathcal{A}$,

$$\Gamma_{\mathrm{CAIC}}(\alpha) = \|\boldsymbol{Z} - \hat{\boldsymbol{S}}(\alpha)\|^2 + 2\mathrm{tr}(\boldsymbol{H}(\alpha))\sigma_\epsilon^2, \tag{6.1}$$

where $\hat{\boldsymbol{S}}(\alpha) = \boldsymbol{H}(\alpha)\boldsymbol{Z}$ with $\boldsymbol{H}(\alpha)$ defined in (3.16). Let

$$\hat{\alpha}_{\mathrm{CAIC}} = \arg\min_{\alpha \in \mathcal{A}} \Gamma_{\mathrm{CAIC}}(\alpha). \tag{6.2}$$

Then we have the following theorem.

Theorem 14 *Consider a class of models given by (3.1). Suppose that*

$$\lim_{n\to\infty} \sum_{\alpha \in \mathcal{A}} \frac{1}{E(L(\alpha))} = 0, \tag{6.3}$$

where $L(\alpha)$ is defined in (3.14). Then the criterion $\Gamma_{CAIC}(\alpha)$ defined in (6.1) is asymptotically loss efficient:

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}_{CAIC}) \Big/ \inf_{\alpha \in \mathcal{A}} L(\alpha) = 1.$$

Proof. Here, we first expand the CAIC defined in (6.1). It is

$$\begin{aligned}
\Gamma_{\text{CAIC}}(\alpha) &= (\boldsymbol{Z}-\hat{\boldsymbol{S}}(\alpha))'(\boldsymbol{Z}-\hat{\boldsymbol{S}}(\alpha)) + 2\sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha)) \\
&= (\boldsymbol{S}-\hat{\boldsymbol{S}}(\alpha)+\boldsymbol{\epsilon})'(\boldsymbol{S}-\hat{\boldsymbol{S}}(\alpha)+\boldsymbol{\epsilon}) + 2\sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha)) \\
&= L(\alpha) + 2\boldsymbol{\epsilon}'(\boldsymbol{S}-\hat{\boldsymbol{S}}(\alpha)) + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + 2\sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha)) \\
&= L(\alpha) + 2\boldsymbol{\epsilon}'(\boldsymbol{I}-\boldsymbol{H}(\alpha))\boldsymbol{S} - 2\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + 2\sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha)) \\
&= L(\alpha) + 2\sigma_\epsilon^2\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{S} - 2\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + 2\sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha)) \\
&= L(\alpha) + 2\sigma_\epsilon^2\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu} + 2\sigma_\epsilon^2\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\eta} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\
&\quad -2\big(\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha))\big), \qquad (6.4)
\end{aligned}$$

where the third equality follows from (3.14) and the second last equality follows from

$$\boldsymbol{I}-\boldsymbol{H}(\alpha) = \boldsymbol{A}(\alpha) - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha) = \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha).$$

It then needs to show that for $\alpha \in \mathcal{A}$,

$$\Gamma_{\text{CAIC}}(\alpha) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + L(\alpha) + o_p(L(\alpha)), \qquad (6.5)$$

which suffices to show that

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu}|}{\text{E}(L(\alpha))} &= 0, \qquad (6.6) \\
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\eta}|}{\text{E}(L(\alpha))} &= 0, \qquad (6.7) \\
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\text{tr}(\boldsymbol{H}(\alpha))|}{\text{E}(L(\alpha))} &= 0, \qquad (6.8) \\
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \left|\frac{L(\alpha)}{\text{E}(L(\alpha))} - 1\right| &= 0. \qquad (6.9)
\end{aligned}$$

Hence, by (6.5), for $\hat{\alpha}_{\text{CAIC}}$ defined in (6.2) and $\alpha^L = \arg\min_{\alpha\in\mathcal{A}} L(\alpha)$, we can easily conclude that

$$\begin{aligned}
\Gamma_{\text{CAIC}}(\hat{\alpha}_{\text{CAIC}}) &= \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + L(\hat{\alpha}_{\text{CAIC}}) + o_p(L(\hat{\alpha}_{\text{CAIC}})), \\
\Gamma_{\text{CAIC}}(\alpha^L) &= \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + L(\alpha^L) + o_p(L(\alpha^L)).
\end{aligned}$$

It follows that

$$0 \le \frac{\Gamma_{\text{CAIC}}(\alpha^L) - \Gamma_{\text{CAIC}}(\hat{\alpha}_{\text{CAIC}})}{L(\hat{\alpha}_{\text{CAIC}})} = \frac{L(\alpha^L) - L(\hat{\alpha}_{\text{CAIC}})}{L(\hat{\alpha}_{\text{CAIC}})} + o_p(1),$$

and then

$$\operatorname*{plim}_{n\to\infty} \frac{L(\alpha^L) - L(\hat{\alpha}_{\text{CAIC}})}{L(\hat{\alpha}_{\text{CAIC}})} = 0,$$

which gives $\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}_{\text{CAIC}}) \big/ \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1$.

Here, we start to prove (6.6)-(6.9) one by one. First, any $\varepsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\Big(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big) &\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\Big(\frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big)\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{1}{\sigma_\epsilon^2\varepsilon^2\mathrm{E}(L(\alpha))}\\
&= 0,
\end{aligned}$$

which gives (6.6), where the second last inequality follows from

$$\sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} \leq \mathrm{E}(L(\alpha)),$$

by (3.17) and the last equality follows from (6.3).

Second, for any $\varepsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\Big(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big) &\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\Big(\frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big)\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\mathrm{tr}(\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&= \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{1}{\sigma_\epsilon^2\varepsilon^2\mathrm{E}(L(\alpha))}\\
&= 0,
\end{aligned}$$

where the third inequality follows from $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$, the second last inequality follows from

$$\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\Sigma}_\eta) \leq \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \leq E(L(\alpha)),$$

by (3.17) and the last equality follows from (6.3).

Third, for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\Big(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{H}(\alpha))|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big)\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\Big(\frac{|\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{H}(\alpha))|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\Big)\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_1\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{H}(\alpha)\boldsymbol{H}(\alpha)')}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_1\sigma_\epsilon^4\big(\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) + 3\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\big)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{3c_1\sigma_\epsilon^2}{\varepsilon^2\mathrm{E}(L(\alpha))} = 0,
\end{aligned}$$

where the second inequality is an application of Theorem 2 of Whittle (1960) for some $c_1 > 0$, the third equality follows from

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{H}(\alpha)\boldsymbol{H}(\alpha)') &= \mathrm{tr}\big(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} + \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} + \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))'\big) \\
&= \mathrm{tr}\big(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta + \sigma_\epsilon^2\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} + \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta \\
&\quad +\sigma_\epsilon^4\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\big) \\
&\leq \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) + 3\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \\
&\leq 3\sigma_\epsilon^{-2}\mathrm{E}(L(\alpha)),
\end{aligned}$$

by

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta) &= \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} - \sigma_\epsilon^2\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta) \\
&\leq \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \\
\mathrm{tr}(\sigma_\epsilon^2\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}) &= \mathrm{tr}(\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} - \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1})) \\
&\leq \mathrm{tr}(\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1})),
\end{aligned}$$

and

$$\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)) \leq \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) = \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)).$$

Last, it remains to show (6.9). Here, we first expand $L(\alpha)$ defined in (3.14). That is

$$\begin{aligned}
L(\alpha) &= (\boldsymbol{S} - \hat{\boldsymbol{S}}(\alpha))'(\boldsymbol{S} - \hat{\boldsymbol{S}}(\alpha)) \\
&= \|(\boldsymbol{I} - \boldsymbol{H}(\alpha))\boldsymbol{\mu} + (\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}+\boldsymbol{\epsilon})) - \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\|^2 \\
&= \|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + (\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) - \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\|^2 \\
&= \sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\|^2 - 2\sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad +\sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + 2\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) \\
&\quad -2\sigma_\epsilon^2(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}). \qquad (6.10)
\end{aligned}$$

It then follows together with (3.17),

$$\begin{aligned}
L(\alpha) - \mathrm{E}(L(\alpha)) &= \|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\|^2 - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \\
&\quad +\sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \\
&\quad +2\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) - 2\sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad -2\sigma_\epsilon^2(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}).
\end{aligned}$$

Then, to show (6.9), it suffices to show that

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{\big|\|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\|^2 - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})\big|}{\mathrm{E}(L(\alpha))} = 0, \qquad (6.11)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} = 0, \qquad (6.12)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})|}{\mathrm{E}(L(\alpha))} = 0, \qquad (6.13)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L(\alpha))} = 0, \qquad (6.14)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L(\alpha))} = 0. \qquad (6.15)$$

Now, we start to prove (6.11)-(6.15) one by one.

For (6.11), we have

$$
\begin{aligned}
&\big|\|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}-\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\|^2-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})\big| \\
&\quad = \big|\sigma_\epsilon^4\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})+\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2})\boldsymbol{\epsilon}+2\sigma_\epsilon^2\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\big| \\
&\quad \le \big|\sigma_\epsilon^4\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})\big|+\big|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2})\boldsymbol{\epsilon}\big|+2\sigma_\epsilon^2\big|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\big|.
\end{aligned}
$$

Hence, to show (6.11), it suffices to show

$$
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}}\frac{\big|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.16}
$$

$$
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}}\frac{\big|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.17}
$$

$$
\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}}\frac{\big|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\big|}{\mathrm{E}(L(\alpha))} = 0. \tag{6.18}
$$

First, (6.16) can be established in a similar manner by Theorem 2 of Whittle. It is for any $\varepsilon>0$,

$$
\begin{aligned}
\lim_{n\to\infty}P\bigg(\sup_{\alpha\in\mathcal{A}}\frac{\big|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\bigg) &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}P\bigg(\frac{\big|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\bigg) \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2\sigma_\epsilon^4(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_2}{\varepsilon^2\sigma_\epsilon^6\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}
$$

for some $c_2>0$, where the third inequality follows from

$$
\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2})\le\sigma_\epsilon^{-4}\mathrm{tr}(\boldsymbol{\Sigma}^2\boldsymbol{\Sigma}^{-2})\le\sigma_\epsilon^{-4}\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}), \tag{6.19}
$$

by $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\le\boldsymbol{I}$ and $\boldsymbol{\Sigma}_\eta^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^{1/2}\le\boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta\le\boldsymbol{\Sigma}$, and the last equality follows from (4.5). Second, (6.17) is also established by Theorem 2 of Whittle. It is for any $\varepsilon>0$,

$$
\begin{aligned}
&\lim_{n\to\infty}P\bigg(\sup_{\alpha\in\mathcal{A}}\frac{\big|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\bigg) \\
&\quad\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}P\bigg(\frac{\big|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}-\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2})\big|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\bigg) \\
&\quad\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_3\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\quad\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_3\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\quad\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{c_3\sigma_\epsilon^2}{\varepsilon^2\mathrm{E}(L(\alpha))} \\
&\quad= 0,
\end{aligned}
$$

for some $c_3 > 0$, where third inequality follows from

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}) \le \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta^2\boldsymbol{\Sigma}^{-1}) \le \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}),$$

by $\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta \le \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta^2 \le \boldsymbol{\Sigma}^2$ and the last equality follows from (6.3). Third, similarly from the proof of (6.7). It is for any $\varepsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{\left|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right|}{\mathrm{E}(L(\alpha))} \ge \varepsilon\right) &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} P\left(\frac{\left|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right|}{\mathrm{E}(L(\alpha))} \ge \varepsilon\right) \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2\sigma_\epsilon^4(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{1}{\sigma_\epsilon^2\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}$$

where the third equality follows from

$$\begin{aligned}
\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) &\le \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \\
&\le \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \\
&\le \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}),
\end{aligned}$$

by $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \le \boldsymbol{I}$ and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1/2} \le \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta \le \boldsymbol{\Sigma}$, and the last equality follows from (6.3). It then gives (6.11).

For (6.12), it can be established by Theorem 2 of Whittle. It is for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} \ge \varepsilon\right) \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} P\left(\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} \ge \varepsilon\right) \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{c_4\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{c_4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2\sigma_\epsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}} \frac{c_4}{\sigma_\epsilon^4\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}$$

for some $c_4 > 0$, where the third inequality follows from

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)) &= \mathrm{tr}(\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}) \\
&\le \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}) \\
&\le \sigma_\epsilon^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)),
\end{aligned}$$

by $\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}'(\alpha)\boldsymbol{\Sigma}^{-1} = \boldsymbol{M}(\alpha)$, $\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha) \le \boldsymbol{\Sigma}^{-1}$ and $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \le \boldsymbol{I}$, and the last equality follows from (6.3).

For (6.13), we have

$$\begin{aligned}|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}-\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})| &= |\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}-\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}|\\ &\le |\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}|+|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}|.\end{aligned}$$

Hence, to show (6.13), it suffices to show that

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.20}$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))} = 0. \tag{6.21}$$

First, (6.20) can be show similarly from (6.6). It is for any $\varepsilon>0$,

$$\begin{aligned}\lim_{n\to\infty}P\left(\sup_{\alpha\in\mathcal{A}}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\right) &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}P\left(\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\right)\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{\sigma_\epsilon^{-2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{1}{\sigma_\epsilon^6\varepsilon^2\mathrm{E}(L(\alpha))}\\ &= 0,\end{aligned}$$

where the third inequality follows from

$$\begin{aligned}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} &\le \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-3}\boldsymbol{A}(\alpha)\boldsymbol{\mu}\\ &\le \sigma_\epsilon^{-2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu},\end{aligned}$$

by $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1/2}\le\boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta\le\boldsymbol{\Sigma}$ and $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\le\boldsymbol{I}$, and the last equality follows from (6.3). Second, (6.21) is similar to (6.20). It is for any $\varepsilon>0$,

$$\begin{aligned}\lim_{n\to\infty}P\left(\sup_{\alpha\in\mathcal{A}}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\right) &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}P\left(\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))}\ge\varepsilon\right)\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2}\\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}}\frac{1}{\sigma_\epsilon^2\varepsilon^2\mathrm{E}(L(\alpha))}\\ &= 0,\end{aligned}$$

where the third inequality follows from $\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta\le\boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta^2\le\boldsymbol{\Sigma}^2$, and the last equality follows from (6.3). It then gives (6.13).

For (6.14), we have for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\bigg(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\bigg) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\bigg(\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\bigg) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_5\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&= \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_5\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_5\sigma_\epsilon^{-2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_5}{\sigma_\epsilon^6\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}$$

for some $c_5 > 0$, where the third equality follows from $\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}'(\alpha)\boldsymbol{\Sigma}^{-1} = \boldsymbol{M}(\alpha)$, the fourth inequality follows from $\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha) \leq \boldsymbol{\Sigma}^{-1}$ and $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$, and the last equality follows from (6.3). It then gives (6.14).

For (6.15), we have

$$\begin{aligned}
&|(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})| \\
&= |\sigma_\epsilon^2\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)) \\
&\quad -\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)| \\
&\leq |\sigma_\epsilon^2\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))| \\
&\quad +|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)| \\
&\leq |\sigma_\epsilon^2\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\eta} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))| + |\sigma_\epsilon^2\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon}| \\
&\quad +|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\eta}| + |\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)|.
\end{aligned}$$

Then, to show (6.15), it suffices to show that

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\eta} - \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.22}$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.23}$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} = 0, \tag{6.24}$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)|}{\mathrm{E}(L(\alpha))} = 0. \tag{6.25}$$

Now, we start to show (6.22)-(6.25) one by one. First, (6.22) can be established by

Theorem 2 of Whittle. That is for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\eta} - \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\quad \leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\left(\frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\eta} - \mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\quad \leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_6\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\quad \leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_6\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\quad \leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_6\sigma_\epsilon^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\quad \leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_6}{\sigma_\epsilon^4\varepsilon^2\mathrm{E}(L(\alpha))} \\
&\quad = 0,
\end{aligned}$$

for some $c_6 > 0$, where the third inequality follows from

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha),$$

and the fourth inequality follows from $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta \leq \boldsymbol{\Sigma}$ and $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$, and the last equality follows from (6.3). Second, (6.23) is similarly to (6.18). It is

$$\begin{aligned}
\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) &\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\left(\frac{|\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{1}{\sigma_\epsilon^4\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}$$

where the third inequality follows from $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)$, and the fourth inequality follows from $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta \leq \boldsymbol{\Sigma}$ and $\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$, and the last equality follows from (6.3). Third, (6.24) is similar to (6.23). It is for any $\varepsilon > 0$,

$$
\begin{aligned}
&\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\left(\frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\eta}|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{1}{\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}
$$

where the third inequality follows from

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha),$$

and the fourth inequality follows from $\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta \leq \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta^2 \leq \boldsymbol{\Sigma}^2$, and the last equality follows from (6.3). Last, (6.25) can be established by Theorem 2 of Whittle. It is for any $\varepsilon > 0$,

$$
\begin{aligned}
&\lim_{n\to\infty} P\left(\sup_{\alpha\in\mathcal{A}} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} P\left(\frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta)|}{\mathrm{E}(L(\alpha))} \geq \varepsilon\right) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_7\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_7\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta)}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_7\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}} \frac{c_7}{\varepsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}
$$

for some $c_7 > 0$, where the third equality follows from

$$\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha),$$

and the fourth inequality follows from $\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}_\eta \leq \boldsymbol{I}$ by $\boldsymbol{\Sigma}_\eta^2 \leq \boldsymbol{\Sigma}^2$, and the last equality follows from (6.3). Thus, we ends the proof of (6.9), which completes the proof. □

Note that (6.3) holds in general. Here, we consider an example where (6.3) is satisfied.

Corollary 7 *Consider a class of models given by (3.1) with p fixed and any arbitrary explanatory variables. Suppose that the data are collected at* $s_i = in^{-(1-\delta)} \in [0, n^{\delta}]$*;* $i = 1, \ldots, n$ *for some* $\delta \in [0, 1)$*. Consider the exponential covariance model of (5.1) for* $\eta(\cdot)$*. Let* $\hat{\alpha}_{CAIC}$ *be the model selected by CAIC as defined in (6.2). Then,*

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}_{CAIC}) / \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1.$$

Further, if $\mathcal{A}^c \neq \emptyset$*, then for any model selection procedure* $\hat{\alpha}$*, such that* $\lim_{n\to\infty} P(\hat{\alpha} \in \mathcal{A}^c) = 1$*,*

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}) / \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1.$$

It is shown in (3.17) that $\mathrm{E}(L(\alpha))$ is lower bounded by dominated by $\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})$ for $\alpha \in \mathcal{A}$, which is often a dominated term of $\mathrm{E}(L(\alpha))$. In addition, for $\alpha \in \mathcal{A}^c$, $\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})$ is the dominated term of $\mathrm{E}(L(\alpha))$. Hence, it might suggests us that whatever correct model we select, it will be always satisfied the asymptotic loss efficiency. Further, in the following example, $\mathrm{E}L((\alpha))$ are dominated by $\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})$ for $\alpha \in \mathcal{A}$. In such case, every candidate model achieves the asymptotic loss efficiency.

Corollary 8 *Consider a class of models given by (3.1) with* $x_j(s) = (sn^{-\delta})^j$*;* $j = 1, \ldots, p$*, and* $cov(\eta(s), \eta(s')) = \sigma_\eta^2 \exp(-\kappa_\eta |s - s'|)$*, where p fixed and* $\mathcal{A}^c \neq \emptyset$*. Suppose that the data are collected at* $s_i = in^{-(1-\delta)} \in [0, n^{\delta}]$*;* $i = 1, \ldots, n$ *for some* $\delta \in [0, 1)$*. Let* $\hat{\alpha}_{CAIC}$ *be the model selected by CAIC as defined in (6.2). Then*

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}_{CAIC}) / \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1.$$

Further, for any model selection procedure $\hat{\alpha}$*,*

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}) / \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1.$$

¿From (7) and (8), it might suggest us that the variable selection is somehow unnecessary for the asymptotic loss efficiency of $L(\alpha)$ in those cases. Here, we consider the strongly asymptotic loss efficiency of $L(\alpha)$ defined in (3.21).

Theorem 15 *Consider a class of models given by (3.1) and the universal kriging predictor* $\hat{\boldsymbol{S}}(\alpha)$ *of* $\boldsymbol{S}$ *defined in (3.15). Suppose*

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{1}{E(L(\alpha)) - \sigma_\epsilon^2 tr(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})} = 0, \tag{6.26}$$

where $L(\alpha)$ *is defined in (3.14). If* $|\mathcal{A}^c| \leq 1$ *and* α^c *is fixed, then* $\hat{\alpha}_{CAIC}$ *of (6.2) is strongly asymptotic loss efficient:*

$$\operatorname*{plim}_{n\to\infty} \frac{L(\hat{\alpha}_{CAIC}) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2}{\inf_{\alpha\in\mathcal{A}} L(\alpha) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2} = 1.$$

Proof. Here, we first suppose that $\mathcal{A}^c = \emptyset$. Now, we expand the CAIC defined in (6.1) from (6.4). It is

$$\begin{aligned}\Gamma_{\text{CAIC}}(\alpha) &= L(\alpha) + 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu} + 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\eta} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\ &\quad -2\big(\boldsymbol{\epsilon}'\boldsymbol{H}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \text{tr}(\boldsymbol{H}(\alpha))\big) \\ &= L(\alpha) + 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu} + 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'\boldsymbol{M}(\alpha)\boldsymbol{\eta} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\ &\quad -2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon} - \sigma_\epsilon^2 \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\big) - 2\sigma_\epsilon^2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\big),\end{aligned} \tag{6.27}$$

where the last equality follows from $\boldsymbol{H}(\alpha) = \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1} + \sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)$ by (3.16). Note that $2\sigma_\epsilon^2\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} - 2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon} - \sigma_\epsilon^2\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\big)$ is constant in variable selection. It then needs to show that for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$,

$$\Gamma_{\text{CAIC}}(\alpha) = constant + L^*(\alpha) + o_p(L^*(\alpha)), \tag{6.28}$$

where $L^*(\alpha) = L(\alpha) - \|\boldsymbol{S} - \text{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2$, which suffices to show that

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha))\boldsymbol{\mu}|}{\text{E}(L^*(\alpha))} = 0, \tag{6.29}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\boldsymbol{\eta}|}{\text{E}(L^*(\alpha))} = 0, \tag{6.30}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\text{E}(L^*(\alpha))} = 0, \tag{6.31}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \left|\frac{L^*(\alpha)}{\text{E}(L^*(\alpha))} - 1\right| = 0. \tag{6.32}$$

Hence, by (6.28), for $\hat{\alpha}_{\text{CAIC}}$ defined in (6.2) and $\alpha^L = \arg\min_{\alpha\in\mathcal{A}} L^*(\alpha)$, we can easily conclude that

$$\begin{aligned}\Gamma_{\text{CAIC}}(\hat{\alpha}_{\text{CAIC}}) &= constant + L^*(\hat{\alpha}_{\text{CAIC}}) + o_p(L^*(\hat{\alpha}_{\text{CAIC}})), \\ \Gamma_{\text{CAIC}}(\alpha^L) &= constant + L^*(\alpha^L) + o_p(L^*(\alpha^L)).\end{aligned}$$

It follows that

$$0 \le \frac{\Gamma_{\text{CAIC}}(\alpha^L) - \Gamma_{\text{CAIC}}(\hat{\alpha}_{\text{CAIC}})}{L^*(\hat{\alpha}_{\text{CAIC}})} = \frac{L^*(\alpha^L) - L^*(\hat{\alpha}_{\text{CAIC}})}{L^*(\hat{\alpha}_{\text{CAIC}})} + o_p(1),$$

and then

$$\operatorname*{plim}_{n\to\infty} \frac{L^*(\alpha^L) - L^*(\hat{\alpha}_{\text{CAIC}})}{L^*(\hat{\alpha}_{\text{CAIC}})} = 0,$$

which gives $\operatorname*{plim}_{n\to\infty} L^*(\hat{\alpha}_{\text{CAIC}})/\inf_{\alpha\in\mathcal{A}} L^*(\alpha) = 1$ when $\mathcal{A}^c = \emptyset$.

Here, we first calculate $\text{E}L^*(\alpha)$. By (9.1), we have

$$\begin{aligned}\text{E}(L^*(\alpha)) &= \text{E}(L(\alpha)) - \text{E}\|\boldsymbol{S} - \text{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 \\ &= \text{E}(L(\alpha)) - \sigma_\epsilon^2\text{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}) \\ &= \sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \sigma_\epsilon^4\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)),\end{aligned} \tag{6.33}$$

by (3.17). Now, we start to prove (6.29)-(6.32) one by one. For (6.29), the proof can be followed from the proof of (6.6) by replacing $\text{E}(L(\alpha))$ with (6.33).

For (6.30), we have for any $\varepsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\Big(\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\boldsymbol{\eta}|}{\mathrm{E}(L^*(\alpha))} \geq \varepsilon\Big) &\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} P\Big(\frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\boldsymbol{\eta}|}{\mathrm{E}(L^*(\alpha))} \geq \varepsilon\Big) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L^*(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L^*(\alpha)))^2} \\
&= \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L^*(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{1}{\sigma_\epsilon^2\varepsilon^2\mathrm{E}(L^*(\alpha))} \\
&= 0,
\end{aligned}$$

where the third inequality follows from $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1/2} \leq \boldsymbol{I}$, the second last inequality follows from (6.33) and the last equality follows from (6.26).

For (6.31), we have for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\Big(\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L^*(\alpha))} \geq \varepsilon\Big) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} P\Big(\frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L^*(\alpha))} \geq \varepsilon\Big) \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{c_1\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha))}{\varepsilon^2(\mathrm{E}(L^*(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{c_1\sigma_\epsilon^2\big(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\big)}{\varepsilon^2(\mathrm{E}(L^*(\alpha)))^2} \\
&\leq \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{c_1}{\varepsilon^2\sigma_\epsilon^2\mathrm{E}(L(\alpha))} \\
&= 0,
\end{aligned}$$

where the second inequality is an application of Theorem 2 of Whittle (1960) for some $c_1 > 0$, and the third and fourth inequality follows from

$$\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)) \leq \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) = \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \leq \sigma_\epsilon^{-2}\mathrm{E}(L^*(\alpha)),$$

and the last equality follows from (6.26).

Now, it remains to show (6.32). Here, we first expand $L^*(\alpha)$ from (6.10). That is

$$\begin{aligned}
L^*(\alpha) &= L(\alpha) - \|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 \\
&= L(\alpha) - \|\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\|^2 \\
&= \sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad +2\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) - 2\sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad -2\sigma_\epsilon^2(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}), \qquad (6.34)
\end{aligned}$$

the second equality follows from (9.1). It then follows together with (3.17),

$$\begin{aligned}L^*(\alpha) - \mathrm{E}(L^*(\alpha)) &= \sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\\ &\quad +2\sigma_\epsilon^2\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) - 2\sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\ &\quad -2\sigma_\epsilon^2(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}).\end{aligned}$$

Equation (6.32) can then be followed by

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L^*(\alpha))} = 0, \quad (6.35)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})|}{\mathrm{E}(L^*(\alpha))} = 0, \quad (6.36)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L^*(\alpha))} = 0, \quad (6.37)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L^*(\alpha))} = 0. \quad (6.38)$$

Note that the proofs of (6.35)-(6.38) can be followed from the proofs of (6.12)-(6.15) by replacing $\mathrm{E}L(\alpha)$ with $\mathrm{E}(L^*(\alpha))$. Hence, (6.32) is then followed, which completes the proof when $\mathcal{A}^c = \emptyset$.

Not, we suppose that $\mathcal{A}^c = \{\alpha^c\}$. To show that the CAIC is still asymptotically loss efficient, it remains to show that for fixed α^c,

$$L^*(\alpha^c) = o_p(L^*(\alpha)); \quad \text{if } \alpha\in\mathcal{A}\setminus\mathcal{A}^c, \quad (6.39)$$

$$\Gamma_{\mathrm{CAIC}}(\alpha^c) = \mathit{constant} + L^*(\alpha^c) + o_p(L^*(\alpha)). \quad (6.40)$$

Hence, by (6.39) and (6.40), we can easily conclude that

$$\lim_{n\to\infty} P\big(\alpha^L = \alpha^c\big) = 1,$$

$$\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{CAIC}} = \alpha^c\big) = 1,$$

which gives $\operatorname*{plim}_{n\to\infty} L^*(\hat{\alpha}_{\mathrm{CAIC}})\big/\inf_{\alpha\in\mathcal{A}} L^*(\alpha) = 1$ if $|\mathcal{A}^c| \le 1$.

Now, we start to prove (6.39). Equation (6.39) can be followed by (6.32) and

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{L^*(\alpha^c)}{\mathrm{E}(L^*(\alpha))} = 0. \quad (6.41)$$

By (6.34), we have

$$\begin{aligned}L^*(\alpha^c) &= \sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha^c)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha^c)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))\\ &\quad +\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)) - 2\sigma_\epsilon^2(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)(\boldsymbol{\eta}+\boldsymbol{\epsilon}).\end{aligned}$$

Equations (6.41) can then be followed by

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha^c)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha^c)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\mathrm{E}(L^*(\alpha))} = 0, \quad (6.42)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))|}{\mathrm{E}(L^*(\alpha))} = 0, \quad (6.43)$$

$$\operatorname*{plim}_{n\to\infty}\sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c}\frac{|(\sigma_\epsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta} - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\mathrm{E}(L^*(\alpha))} = 0. \quad (6.44)$$

Note that (6.42) can be followed similarly from the proof of (6.35) and (6.43) is trivial since $\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)) \leq p(\alpha^c) < \infty$, and (6.44) can be followed similarly from the proof of (6.38). It then gives (6.39).

Now we start to prove (6.40). By (6.27), we have

$$\Gamma_{\mathrm{CAIC}}(\alpha^c) \;=\; constant + L^*(\alpha^c) - 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'\boldsymbol{M}(\alpha^c)\boldsymbol{\eta} - 2\sigma_\epsilon^2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))\big).$$

Equation (6.40) can then be followed by

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{M}(\alpha^c)\boldsymbol{\eta}|}{\mathrm{E}(L^*(\alpha))} = 0,$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))|}{\mathrm{E}(L^*(\alpha))} = 0,$$

which can be followed easily from (6.30) and (6.31). It then gives (6.40). This completes the proof. □

An example is given here for the Theorem 15.

Corollary 9 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7), where p fixed and $\mathcal{A}^c = \{\alpha^c\}$. If $\lim_{n\to\infty} tr(\boldsymbol{\Sigma}^{-2}) = \infty$, then*

$$\operatorname*{plim}_{n\to\infty} \frac{L(\hat{\alpha}_{CAIC}) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2}{\inf_{\alpha\in\mathcal{A}} L(\alpha) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2} = 1.$$

The model with smallest value of $L(\alpha)$ might not exist for $|\mathcal{A}^c| \geq 2$. If there are at least two correct models with fixed dimensions in $\mathcal{A}^c$, there will be no asymptotic optimal properties under the level of loss comparison. We are then interested to ask if the model selection procedure still has some optimal properties on $\mathrm{E}(L(\alpha))$ in the cases of $|\mathcal{A}^c| \geq 2$. Hence, we need a much more heavily penalty on model dimension to select α^c among $\mathcal{A}^c$.

6.2 Conditional Generalized Information Criterion

We have the criterion $\Gamma_{\mathrm{CAIC}}(\alpha)$ in (6.1) is weakly asymptotic loss efficient. Further, It also suggests that the asymptotic loss efficiency may not exist in the case of $|\mathcal{A}^c| \geq 2$. Hence, we might ask if there has any optimal properties which is defined on the risk (3.17). In the following, we will introduce the conditional generalized information criterion (CGIC) which includes CAIC as a special case,

$$\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha) = \|\boldsymbol{Z} - \hat{S}(\alpha)\|^2 + \lambda\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{H}(\alpha)),$$

for any $\alpha \in \mathcal{A}$. Let

$$\hat{\alpha}_{\mathrm{CGIC}(\lambda)} = \arg\min_{\alpha\in\mathcal{A}} \Gamma_{\mathrm{CGIC}(\lambda)}(\alpha). \tag{6.45}$$

Theorem 16 *Consider a class of models given by (3.1). Consider the loss function $L(\alpha)$ defined in (3.14) and $\hat{\alpha}_{CGIC(\lambda)}$ defined in (6.45).*

(i) For $\mathcal{A}^c = \emptyset$, if

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\lambda p}{E(L(\alpha)) - \sigma_\epsilon^2 tr(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})} = 0, \tag{6.46}$$

then

$$\operatorname*{plim}_{n\to\infty} \frac{L(\hat{\alpha}_{CGIC}) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2}{\inf_{\alpha\in\mathcal{A}} L(\alpha) - \|\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{Z})\|^2} = 1.$$

(ii) For $\mathcal{A}^c \neq \emptyset$ with fixed α, if $\lambda \to \infty$, (6.46) holds and

$$\lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} \frac{1}{tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))} < \infty, \tag{6.47}$$

then $\lim_{n\to\infty} P\big(\hat{\alpha}_{CGIC(\lambda)} = \alpha^c\big) = 1.$

Proof. First, we suppose that $\mathcal{A}^c = \emptyset$. By (6.40), for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$,

$$\begin{aligned}
\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha) &= constant + L^*(\alpha) + (\lambda - 2)\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{H}(\alpha)) + o_p(L^*(\alpha)) \\
&= constant + L^*(\alpha) + (\lambda - 2)\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1}) + (\lambda - 2)\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) + o_p(L^*(\alpha)) \\
&= constant + L^*(\alpha) + (\lambda - 2)\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) + o_p(L^*(\alpha)) \\
&= constant + L^*(\alpha) + o_p(L^*(\alpha)),
\end{aligned}$$

where $L^*(\alpha) = L(\alpha) - \|\boldsymbol{S} - \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2$, the second last equality follows from $(\lambda - 2)\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1})$ is independent of α and the last equality follows from

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{\lambda \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\mathrm{E}(L^*(\alpha))} = 0,$$

by (6.32) and $\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \leq p(\alpha)$ and (6.46). Hence, for $\hat{\alpha}_{\mathrm{CGIC}(\lambda)}$ defined in (6.45) and $\alpha^L = \arg\min_{\alpha\in\mathcal{A}} L^*(\alpha)$, we can easily conclude that

$$\begin{aligned}
\Gamma_{\mathrm{CGIC}(\lambda)}(\hat{\alpha}_{\mathrm{CGIC}(\lambda)}) &= constant + L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)}) + o_p(L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})), \\
\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha^L) &= constant + L^*(\alpha^L) + o_p(L^*(\alpha^L)).
\end{aligned}$$

It follows that

$$0 \leq \frac{\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha^L) - \Gamma_{\mathrm{CGIC}(\lambda)}(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})}{L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})} = \frac{L^*(\alpha^L) - L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})}{L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})} + o_p(1),$$

and then

$$\operatorname*{plim}_{n\to\infty} \frac{L^*(\alpha^L) - L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})}{L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)})} = 0,$$

which gives $\operatorname*{plim}_{n\to\infty} L^*(\hat{\alpha}_{\mathrm{CGIC}(\lambda)}) \big/ \inf_{\alpha\in\mathcal{A}} L^*(\alpha) = 1$ when $\mathcal{A}^c = \emptyset$.

Now, we suppose that $\mathcal{A}^c \neq \emptyset$. To show the consistency, it suffices to show that

$$\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha^c) = constant + o_p(L^*(\alpha)); \quad \text{if } \alpha \in \mathcal{A} \setminus \mathcal{A}^c, \tag{6.48}$$

$$\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha) = constant + \lambda R_2(\alpha) + o_p(\lambda R_2(\alpha)); \quad \text{if } \alpha \in \mathcal{A}^c, \tag{6.49}$$

where $R_2(\alpha) = \sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))$ is defined in (3.18) and $\alpha^c = \arg\min_{\alpha \in \mathcal{A}^c} R_2(\alpha)$. Hence, by (6.48), it follows that $\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{CGIC}(\lambda)} \in \mathcal{A} \setminus \mathcal{A}^c\big) = 0$. In addition, by (6.49), it follows that $\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{CGIC}(\lambda)} \in \mathcal{A}^c, \hat{\alpha}_{\mathrm{CGIC}(\lambda)} = \alpha^c\big) = 1$. Thus, it completes the proof of the consistency.

Now, we first start to prove (6.48). By (6.27), we have

$$\begin{aligned}\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha^c) &= \mathit{constant} + L^*(\alpha^c) - 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'\boldsymbol{M}(\alpha^c)\boldsymbol{\eta} + (\lambda-2)\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \\ &\quad -2\sigma_\epsilon^2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))\big).\end{aligned}$$

Equation (6.48) can then be followed by (6.32) and

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{M}(\alpha^c)\boldsymbol{\eta}|}{\mathrm{E}(L^*(\alpha))} = 0, \tag{6.50}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))|}{\mathrm{E}(L^*(\alpha))} = 0, \tag{6.51}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{(\lambda-2)\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c))}{\mathrm{E}(L^*(\alpha))} = 0, \tag{6.52}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}\setminus\mathcal{A}^c} \frac{L^*(\alpha^c)}{\mathrm{E}(L^*(\alpha))} = 0. \tag{6.53}$$

Equations (6.50) and (6.51) can be followed easily from (6.30) and (6.31). Equation (6.52) holds trivially for $\lambda\sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)) \le \sigma_\epsilon^2\lambda p$. Equation (6.53) is the same to (6.41).

Now, we start to prove (6.49). By (6.27), we have that for $\alpha \in \mathcal{A}^c$,

$$\begin{aligned}\Gamma_{\mathrm{CGIC}(\lambda)}(\alpha) &= \mathit{constant} + L^*(\alpha) - 2\sigma_\epsilon^2 \boldsymbol{\epsilon}'\boldsymbol{M}(\alpha)\boldsymbol{\eta} + (\lambda-2)\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \\ &\quad -2\sigma_\epsilon^2\big(\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\big).\end{aligned}$$

Equation (6.49) can then be followed from (6.32) and

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{M}(\alpha)\boldsymbol{\eta}|}{\lambda R_2(\alpha)} = 0, \tag{6.54}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)\boldsymbol{\epsilon} - \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\lambda R_2(\alpha)} = 0, \tag{6.55}$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|L^*(\alpha)|}{\lambda R_2(\alpha)} = 0. \tag{6.56}$$

For (6.54), we have for any $\varepsilon > 0$,

$$\begin{aligned}\lim_{n\to\infty} P\bigg(\sup_{\alpha\in\mathcal{A}^c} \frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\boldsymbol{\eta}|}{\lambda R_2(\alpha)} \ge \varepsilon\bigg) &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}^c} P\bigg(\frac{|\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))\boldsymbol{\eta}|}{\lambda R_2(\alpha)} \ge \varepsilon\bigg) \\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}^c} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\lambda R_2(\alpha))^2} \\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}^c} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\lambda R_2(\alpha))^2} \\ &= \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}^c} \frac{\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2(\lambda R_2(\alpha))^2} \\ &\le \lim_{n\to\infty}\sum_{\alpha\in\mathcal{A}^c} \frac{1}{\sigma_\epsilon^2\varepsilon^2\lambda^2 R_2(\alpha)} \\ &= 0,\end{aligned}$$

where the third inequality follows from $\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}_{\eta}\mathbf{\Sigma}^{-1/2} \le \boldsymbol{I}$, the second last inequality follows from (6.33) and the last equality follows from $\lambda \to \infty$ and (6.47). Equation (6.55) can then be followed by (6.31) in a similar way.

Now we start to prove (6.56). By (6.34) for $\alpha \in \mathcal{A}^c$,

$$\begin{aligned} L^*(\alpha) &= \sigma_\epsilon^4(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \sigma_\epsilon^4 \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \\ &\quad + \sigma_\epsilon^4 \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha)) - 2\sigma_\epsilon^2(\sigma_\epsilon^2\mathbf{\Sigma}^{-1}\boldsymbol{\eta} - \mathbf{\Sigma}_\eta\mathbf{\Sigma}^{-1}\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}). \end{aligned}$$

Equations (6.56) can then be followed by

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\lambda R_2(\alpha)} = 0, \quad (6.57)$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|\mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\lambda R_2(\alpha)} = 0, \quad (6.58)$$

$$\operatorname*{plim}_{n\to\infty} \sup_{\alpha\in\mathcal{A}^c} \frac{|(\sigma_\epsilon^2\mathbf{\Sigma}^{-1}\boldsymbol{\eta} - \mathbf{\Sigma}_\eta\mathbf{\Sigma}^{-1}\boldsymbol{\epsilon})'\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})|}{\lambda R_2(\alpha)} = 0. \quad (6.59)$$

Equation (6.57) can be followed from (6.12). That is for any $\varepsilon > 0$,

$$\begin{aligned} &\lim_{n\to\infty} P\bigg(\sup_{\alpha\in\mathcal{A}^c} \frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\lambda R_2(\alpha)} \ge \varepsilon\bigg) \\ &\le \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} P\bigg(\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha))|}{\lambda R_2(\alpha)} \ge \varepsilon\bigg) \\ &\le \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} \frac{c\,\mathrm{tr}(\mathbf{\Sigma}\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha)\mathbf{\Sigma}\boldsymbol{M}(\alpha)'\mathbf{\Sigma}^{-2}\boldsymbol{M}(\alpha))}{\varepsilon^2(\lambda R_2(\alpha))^2} \\ &\le \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} \frac{c\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha))}{\varepsilon^2\sigma_\epsilon^2(\lambda R_2(\alpha))^2} \\ &\le \lim_{n\to\infty} \sum_{\alpha\in\mathcal{A}^c} \frac{c}{\sigma_\epsilon^4\varepsilon^2\lambda^2 R_2(\alpha)} \\ &= 0, \end{aligned}$$

where the second inequality is an application of Theorem 2 of Whittle (1960) for some $c > 0$, the last equality follows from (6.47). Equation (6.58) holds trivially for $\lambda \to \infty$. Equation (6.59) can also be followed from (6.15) in a similar way. It then completes the proof. □

Similar to the conditions given by (4.7) and (4.8) in Theorem 2, Equation (6.46) provides a condition for risks associated with incorrect models and (6.47) is a weak technique condition that holds trivially when p is fixed. Here, we give an example for Theorem 16.

Corollary 10 *Consider a class of models given by (3.1) with $x_j(s)$'s independently generated from white-noise processes of (5.7). If $\lambda \to \infty$, $\lim_{n\to\infty} tr(\mathbf{\Sigma}^{-2})/\lambda = \infty$ and*

$$\lim_{n\to\infty} tr(\mathbf{\Sigma}^{-2})/tr(\mathbf{\Sigma}^{-1}) > 0,$$

then $\lim_{n\to\infty} P\big(\hat{\alpha}_{GIC(\lambda)} = \alpha^c\big) = 1$.

Chapter 7

Simulations

In this chapter, we consider three simulation experiments (Experiments I-III) in the following three sections corresponding to Examples 2-4 given at the beginning of Chapter 5. We shall examine their finite sample behaviors and compare them with their asymptotic results developed in Chapter 5.

7.1 Experiment I: Polynomial Order Selection

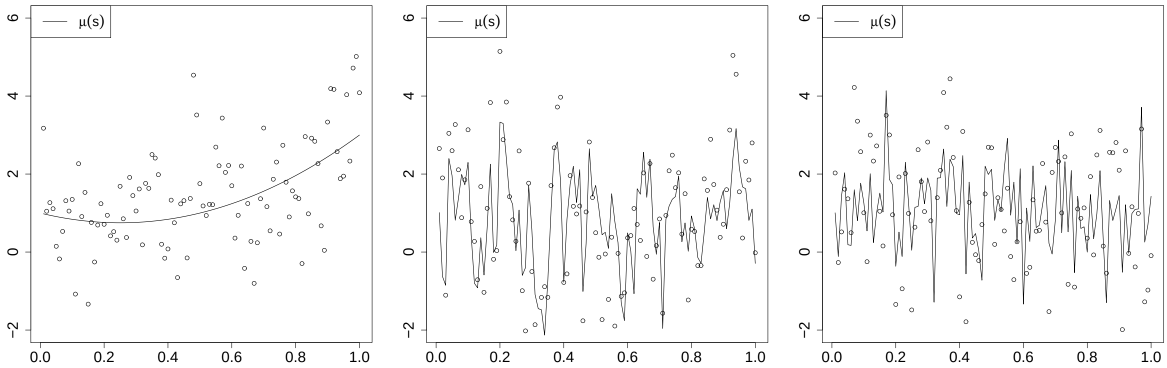
In this experiment, we consider $p = 3$ monomials, $x_j(s) = (sn^{-\delta})^j$; $j = 1, \ldots, 3$, of (5.5) as the explanatory variables. We generated the data at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1, \ldots, n$, for some $\delta \in [0, 1)$ according to the following:

$$Z(s_i) = x_0(s_i) - 2x_1(s_i) + 4x_2(s_i) + \eta(s_i) + \epsilon(s_i); \quad i = 1, \ldots, n,$$

where $\mathrm{cov}(\eta(s_i), \eta(s_j)) = \sigma^2_{\eta,0} \exp(-\kappa_{\eta,0}|s_i - s_j|)$, $\epsilon(\cdot) \sim N(0, \sigma^2_{\epsilon,0})$ and the parameter values are chosen as $(\sigma^2_{\eta,0}, \kappa_{\eta,0}, \sigma^2_{\epsilon,0})' = (0.5, 1, 0.5)'$. Denote the collection of candidate models as $\mathcal{A} = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$, where $\alpha_0 \equiv \emptyset$ and $\alpha_j = \{1, \ldots, j\}$; $j = 1, 2, 3$. Note that, $x_0(\cdot) = 1$ is always included in the model and the smallest correct model is $\alpha^c = \alpha_2$. We consider two δ values: $\delta = 0$ and $\delta = 0.75$, corresponding to the fixed domain asymptotic and increasing domain asymptotic frameworks. For each case, we consider six different sample sizes (n = 100, 500, 1000, 5000, 10000, 50000). Figure 7.1a shows the mean, $x_0(\cdot) - 2x_1(\cdot) + 4x_2(\cdot)$, and a typical realization of data for $\delta = 0$ and $n = 100$.

The results are shown in Table 7.1, and Figures 7.2 and 7.3. Table 7.1 shows the frequencies of models selected by BIC and GIC with $\lambda = 2 \log n$ for Experiment I based on 100 simulation replicates. Basically, BIC tends to select the smallest model α_0 when $\delta = 0$, and tends to select α^c when $\delta = 0.75$ particularly when n is large regardless of whether the covariance parameters are known or unknown, which is consistent with the theoretical results developed in Theorems 6 and 7. Similar results can be seen for GIC with $\lambda = 2 \log n$. Notice that when $\delta = 0$, a very large sample size is needed for BIC to achieve the asymptotic result of selecting only α_0, which is especially the case when the covariance parameters are unknown. On the other hand, GIC with $\lambda = 2 \log n$ requires a much smaller sample size to achieve the same asymptotic result.

Figures 7.2 and 7.3 show the probability density functions for $\hat{\sigma}^2_\epsilon(\alpha)$, $\hat{\kappa}_\eta(\alpha)$ and $\hat{\sigma}^2_\eta(\alpha)$ under $\delta = 0$ and 0.75 based on 100 simulation replicates. As expected from Theorem 4, we see that both $\hat{\sigma}^2_\epsilon(\alpha)$ and $\hat{\kappa}_\eta(\alpha)\hat{\sigma}^2_\eta(\alpha)$ tend to $\sigma^2_{\epsilon,0}$ and $\kappa_{\eta,0}\sigma^2_{\eta,0}$ for all cases except for α_0 and α_1 with $\delta = 0.75$, where the converge rate of the two estimates for a larger δ tends to be slower (see Theorem 4). When $\delta = 0.75$, we see from Figure 7.3 that both



(a) (b) (c)

Figure 7.1: Mean functions and simulated data from (a) Experiment I, (b) Experiment II, and (c) Experiment III for $\delta = 0$ and $n = 100$.

Table 7.1: Frequencies of models selected by GIC with two tuning parameter values of λ for Experiment I based on 100 simulation replicates.

λ	n	known $(\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)$								unknown $(\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)$							
		$\delta = 0$				$\delta = 0.75$				$\delta = 0$				$\delta = 0.75$			
		α_0	α_1	α_2	α_3	α_0	α_1	α_2	α_3	α_0	α_1	α_2	α_3	α_0	α_1	α_2	α_3
$\log(n)$	100	22	37	41	0	10	59	29	2	3	37	59	1	83	16	1	0
	500	39	18	31	12	0	29	68	3	51	28	17	3	0	28	70	2
	1000	40	29	24	7	0	10	90	0	60	29	10	1	0	9	91	0
	5000	54	23	22	1	0	0	99	1	48	18	31	2	0	0	98	2
	10000	65	23	12	0	0	0	100	0	32	18	45	5	0	0	100	0
	50000	69	22	9	0	0	0	100	0	45	16	38	1	0	0	100	0
$2\log(n)$	100	84	11	5	0	48	48	4	0	74	21	5	0	83	16	1	0
	500	83	11	4	2	0	74	26	0	96	4	0	0	29	54	17	0
	1000	86	9	4	1	0	49	51	0	98	2	0	0	2	57	41	0
	5000	90	7	3	0	0	0	100	0	96	3	1	0	0	0	100	0
	10000	95	4	1	0	0	0	100	0	93	7	0	0	0	0	100	0
	50000	100	0	0	0	0	0	100	0	100	0	0	0	0	0	100	0

$\hat{\sigma}_\eta^2(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ tend to their theoretical convergence values as n increases. In contrast, when $\delta = 0$, both $\hat{\sigma}_\eta^2(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ display no clear convergence pattern, because neither of them converges to a degenerate distribution (Chen *et al.* 2000).

7.2 Experiment II: Spatially Dependent Regressors

In this experiment, we consider $p = 3$ spatially dependent processes, $x_j(s)$; $j = 1, 2, 3$, of (5.6) with $\sigma_j^2 = \kappa_j = 0.75$ as the explanatory variables. We generated the data at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1, \ldots, n$, for some $\delta \in [0, 1)$ according to the following:

$$Z(s_i) = x_0(s_i) + x_1(s_i) + x_2(s_i) + \eta(s_i) + \epsilon(s_i); \quad i = 1, \ldots, n, \tag{7.1}$$

where $\mathrm{cov}(\eta(s_i), \eta(s_j)) = \sigma_{\eta,0}^2 \exp(-\kappa_{\eta,0}|s_i - s_j|)$, $\epsilon(\cdot) \sim N(0, \sigma_{\epsilon,0}^2)$ and the parameter values are chosen as $(\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,0}^2)' = (0.5, 1, 0.5)'$. We consider exhausted search over all possible models with $\mathcal{A} = 2^{\{1,2,3\}}$, where α_j's are defined in Table 7.2. Note that, $x_0(\cdot) = 1$ is always included in the model and the smallest true model is $\alpha^c = \alpha_3$. Similar

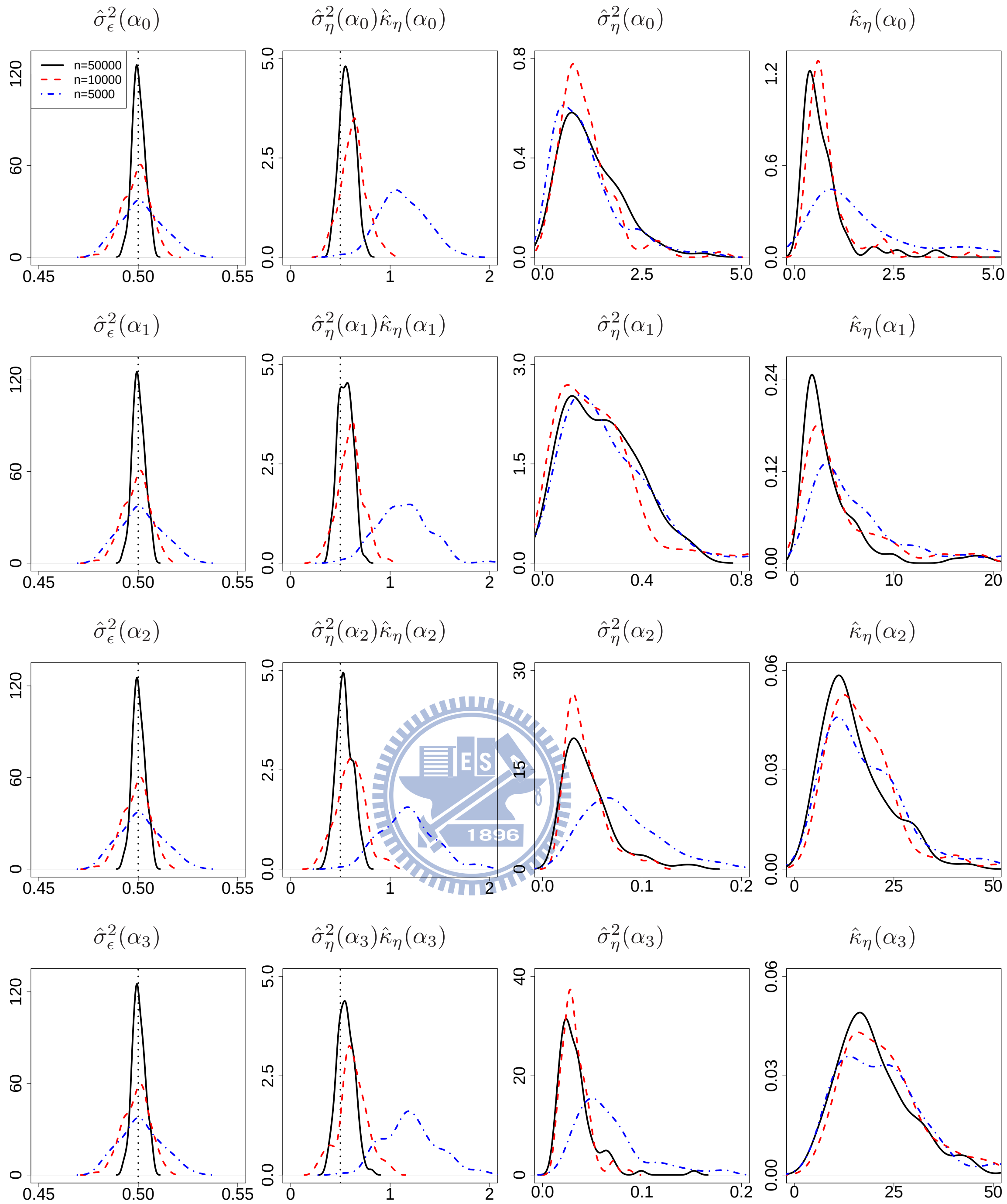


Figure 7.2: Probability density functions for the ML estimates of covariance parameters in Experiment I with $\delta = 0$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

to Experiment I, we consider two δ values ($\delta = 0, 0.75$) combined with five different sample sizes ($n = 100, 500, 1000, 5000, 10000$). Figure 7.1b shows a realization of the mean, $x_0(\cdot) + x_1(\cdot) + x_2(\cdot)$, and a typical realization of data for $\delta = 0$ and $n = 100$.

The results are shown in Tables 7.3 and 7.4, and Figures 7.4 and 7.5. Tables 7.3 and 7.4 show the frequencies of models selected by BIC for Experiment II based on 100 simulation

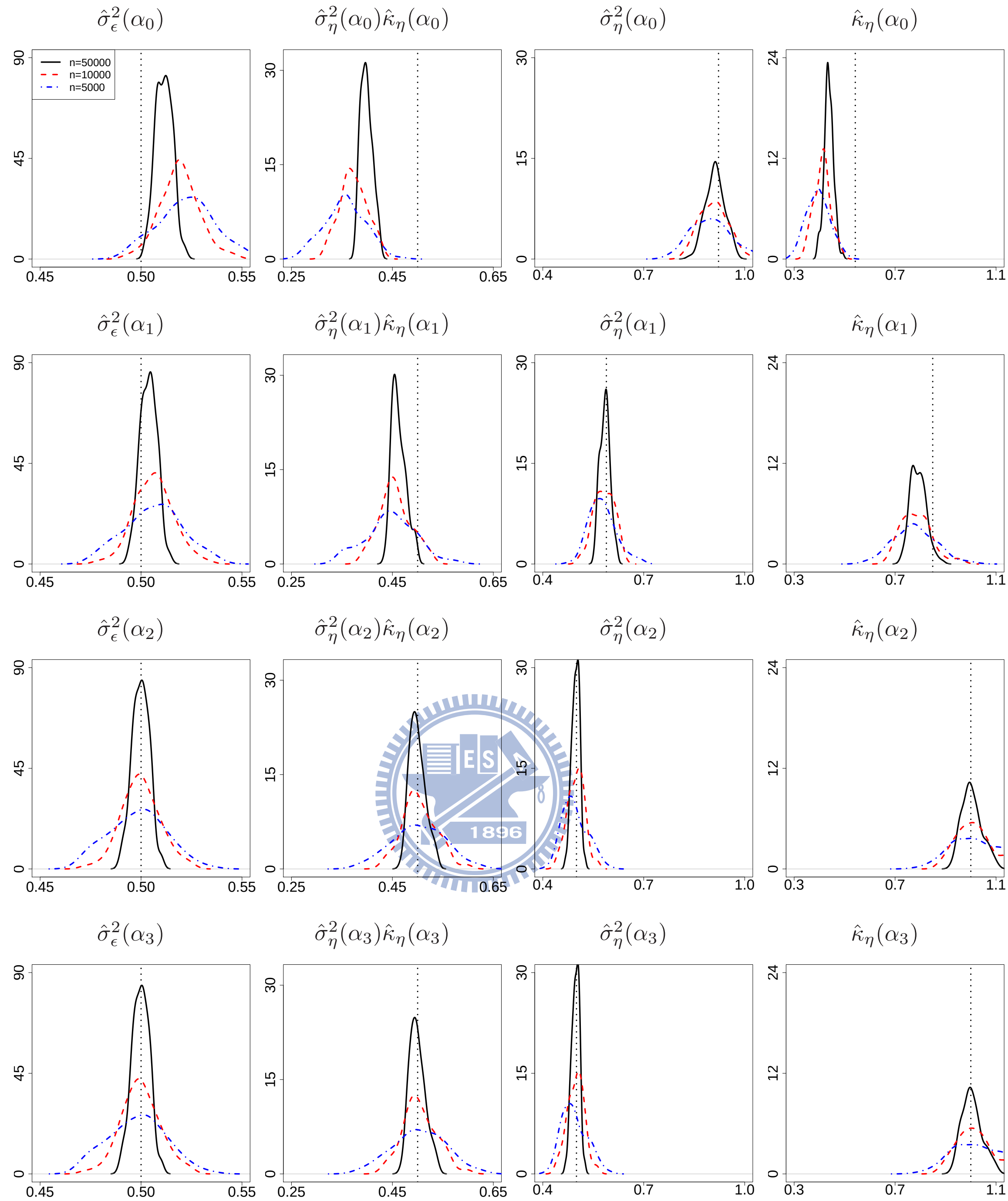


Figure 7.3: Probability density functions for the ML estimates of covariance parameters in Experiment I with $\delta = 0.75$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

replicates. Basically, BIC tends to select $\alpha_3 = \alpha^c$ when $\delta = 0$ and $\delta = 0.75$ regardless of whether the covariance parameters are known or unknown, which is consistent with the theoretical results developed in Theorems 9 and 10.

Figures 7.4 and 7.5 show the probability density functions for $\hat{\sigma}^2_\epsilon(\alpha)$, $\hat{\kappa}_\eta(\alpha)$ and $\hat{\sigma}^2_\eta(\alpha)$ under $\delta = 0$ and $\delta = 0.75$ based on 100 replicates. We show only a nested sequence

Table 7.2: Candidate models for Experiments II and III.

Models	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
indices	$\emptyset$	{1}	{2}	{1,2}	{3}	{1,3}	{2,3}	{1,2,3}

Table 7.3: Frequencies of models selected by BIC for Experiment II with $\{\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2\}$ known based on 100 simulation replicates.

n	$\delta = 0$								$\delta = 0.75$							
	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
100	0	0	0	86	0	0	0	14	0	0	0	98	0	0	0	2
500	0	0	0	82	0	0	0	18	0	0	0	99	0	0	0	1
1000	0	0	0	89	0	0	0	11	0	0	0	99	0	0	0	1
5000	0	0	0	98	0	0	0	2	0	0	0	100	0	0	0	0
10000	0	0	0	98	0	0	0	2	0	0	0	100	0	0	0	0

of models, $\alpha_0 \subset \alpha_1 \subset \alpha_3 \subset \alpha_7$, including two incorrect models α_0 and α_1, the smallest correct model α_3, and the full model α_7. As expected from Theorem 8, we see that both $\hat{\sigma}_\epsilon^2(\alpha)$ and $\hat{\kappa}_\eta(\alpha)\hat{\sigma}_\eta^2(\alpha)$ tend to $\sigma_{\epsilon,0}^2$ and $\theta_{\eta,\alpha}$ defined in (5.118) for all cases. When $\delta = 0.75$, we see from Figure 7.5 that both $\hat{\sigma}_\eta^2(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ tend to their theoretical convergence values as n increases. In contrast, when $\delta = 0$, both $\hat{\sigma}_\eta^2(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ display no clear convergence pattern, because neither of them converges to a degenerate distribution (Chen *et al.* 2000).

7.3 Experiment III: White Noise Regressors

In this experiment, we consider $p = 3$ white-noise processes, $x_j(s)$; $j = 1, \ldots, 3$, of (5.7) with $\sigma_j^2 = 0.6$ as the explanatory variables. We generated the data at $s_i = in^{-(1-\delta)} \in [0, n^\delta]$; $i = 1, \ldots, n$, for some $\delta \in [0, 1)$ according to (7.1) with the parameter values chosen as $(\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,0}^2)' = (0.5, 1, 0.5)'$. We consider exhausted search over all possible models with $\mathcal{A} = 2^{\{1,2,3\}}$, where α_j's are defined in Table 7.2. Note that, $x_0(\cdot) = 1$ is always included in the model and the smallest true model is $\alpha^c = \alpha_3$. Similar to Experiments I and II, we consider two δ values ($\delta = 0,\ 0.75$) combined with five different sample sizes ($n = 100,\ 500,\ 1000,\ 5000,\ 10000$). Figure 7.1c shows a realization of the mean, $x_0(\cdot) + x_1(\cdot) + x_2(\cdot)$, and a typical realization of data for $\delta = 0$ and $n = 100$.

The results are shown in Tables 7.5 and 7.6, and Figures 7.6 and 7.7. Tables 7.5 and 7.6 show the frequencies of models selected by BIC for Experiment III based on 100 simulation

Table 7.4: Frequencies of models selected by BIC for Experiment II with $\{\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2\}$ unknown based on 100 simulation replicates.

n	$\delta = 0$								$\delta = 1$							
	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
100	0	0	0	94	0	0	0	6	0	0	0	94	0	0	0	6
500	0	0	0	99	0	0	0	1	0	0	0	99	0	0	0	1
1000	0	0	0	99	0	0	0	1	0	0	0	99	0	0	0	1
5000	0	0	0	97	0	0	0	3	0	0	0	100	0	0	0	0
10000	0	0	0	98	0	0	0	2	0	0	0	100	0	0	0	0

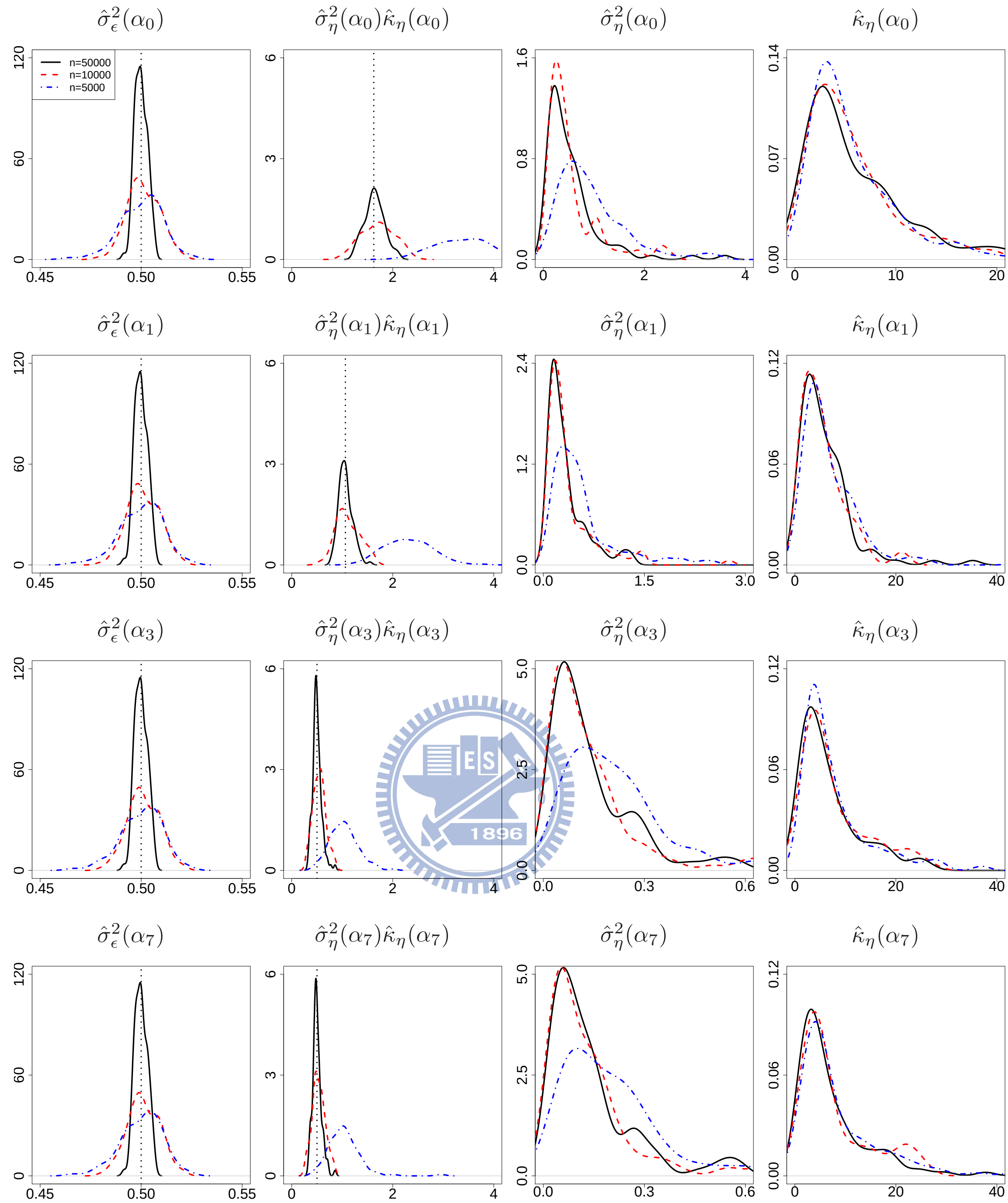


Figure 7.4: Probability density functions for the ML estimates of covariance parameters in Experiment II with $\delta = 0$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

replicates. Basically, BIC tends to select $\alpha_3 = \alpha^c$ when $\delta = 0$ and $\delta = 0.75$ regardless of whether the covariance parameters are known or unknown, which is consistent with the theoretical results developed in Theorems 12 and 13.

Figures 7.6 and 7.7 show the probability density functions for $\hat{\sigma}^2_\epsilon(\alpha)$, $\hat{\kappa}_\eta(\alpha)$ and $\hat{\sigma}^2_\eta(\alpha)$ under $\delta = 0$ and $\delta = 0.75$ based on 100 replicates. As in Experiment II, we show only

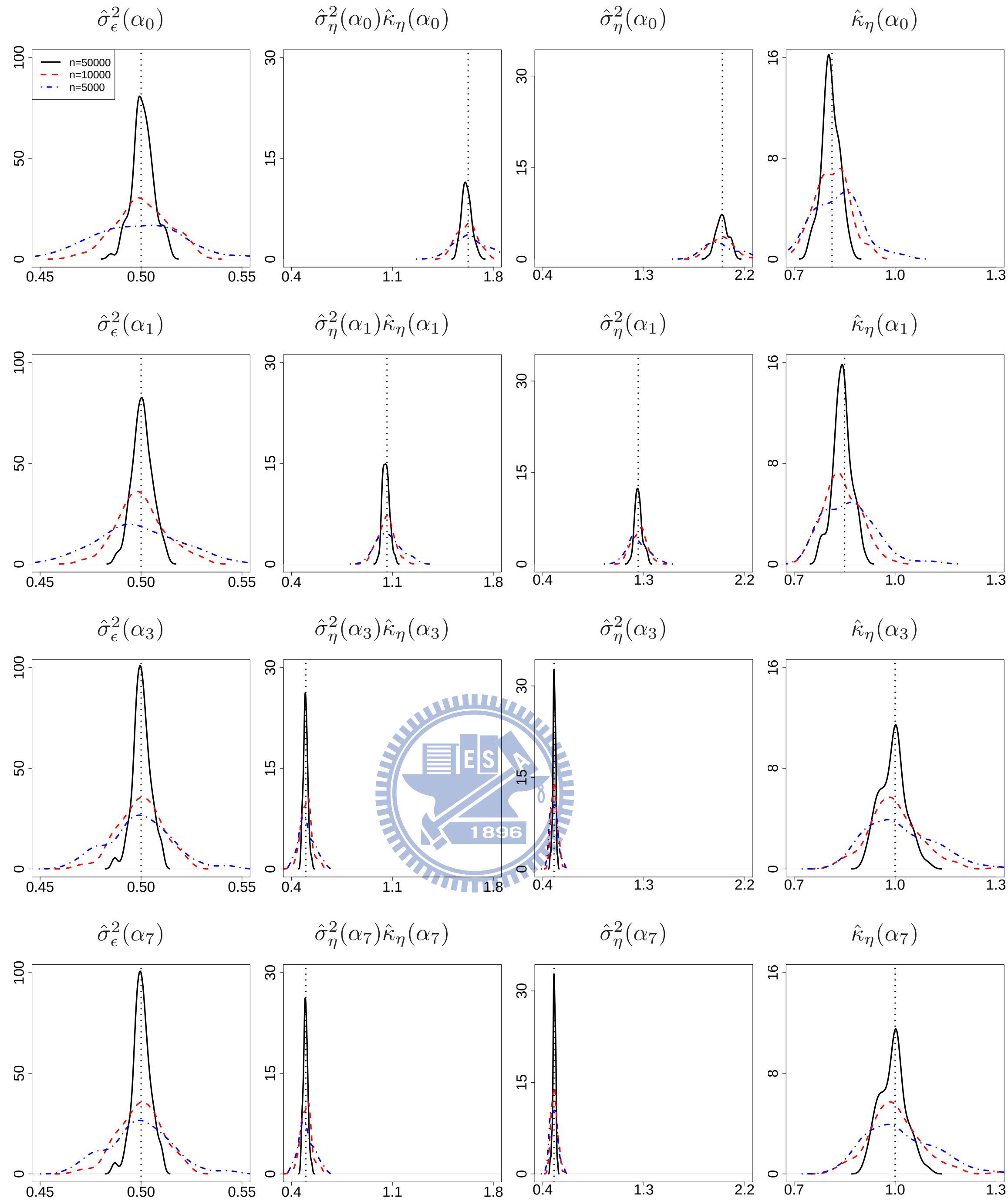


Figure 7.5: Probability density functions for the ML estimates of covariance parameters in Experiment II with $\delta = 0.75$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

a nested sequence of models, $\alpha_0 \subset \alpha_1 \subset \alpha_3 \subset \alpha_7$. As expected from Theorem 11, we see that both $\hat{\sigma}^2_\epsilon(\alpha)$ and $\hat{\kappa}_\eta(\alpha)\hat{\sigma}^2_\eta(\alpha)$ tend to their theoretical values given in the theorem for all cases. When $\delta = 0.75$, we see from Figure 7.7 that both $\hat{\sigma}^2_\eta(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ tend to $\sigma^2_{\eta,0}$ and $\kappa_{\eta,0}$ as n increases. In contrast, when $\delta = 0$, both $\hat{\sigma}^2_\eta(\alpha)$ and $\hat{\kappa}_\eta(\alpha)$ display no clear convergence pattern, because neither of them converges to a degenerate distribution

Table 7.5: Frequencies of models selected by BIC for Experiment III with $\{\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2\}$ known based on 100 simulation replicates.

n	$\delta = 0$								$\delta = 0.75$							
	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
100	0	0	0	90	0	0	0	10	0	0	0	97	0	0	0	3
500	0	0	0	96	0	0	0	4	0	0	0	100	0	0	0	0
1000	0	0	0	100	0	0	0	0	0	0	0	100	0	0	0	0
5000	0	0	0	98	0	0	0	2	0	0	0	100	0	0	0	0
10000	0	0	0	98	0	0	0	2	0	0	0	100	0	0	0	0

Table 7.6: Frequencies of models selected by BIC for Experiment III with $\{\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2\}$ unknown based on 100 simulation replicates.

n	$\delta = 0$								$\delta = 0.75$							
	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
100	0	0	0	96	0	0	0	4	0	0	0	97	0	0	0	3
500	0	0	0	99	0	0	0	1	0	0	0	100	0	0	0	0
1000	0	0	0	100	0	0	0	0	0	0	0	100	0	0	0	0
5000	0	0	0	100	0	0	0	0	0	0	0	100	0	0	0	0
10000	0	0	0	99	0	0	0	1	0	0	0	100	0	0	0	0

(Chen *et al.* 2000).

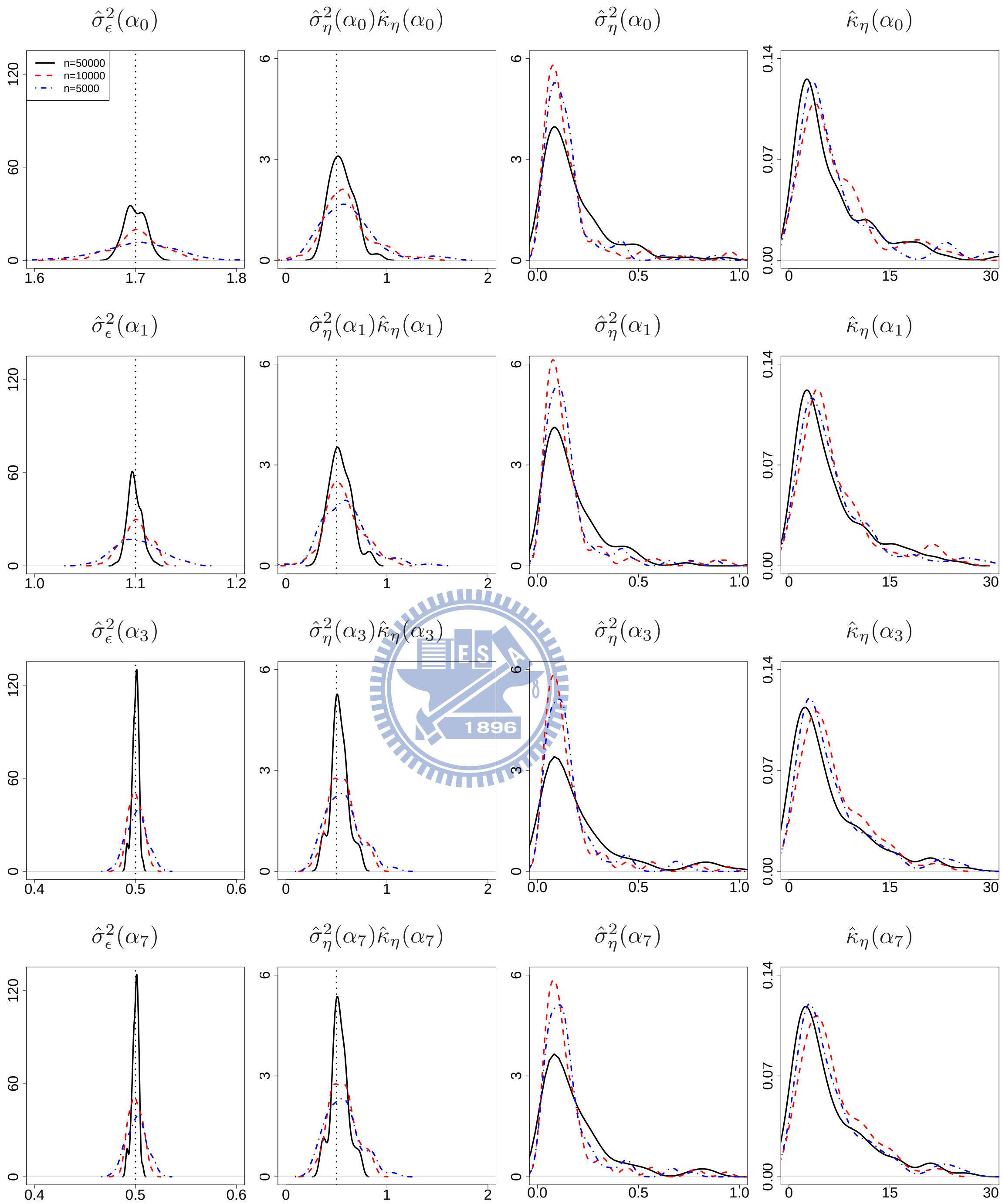


Figure 7.6: Probability density functions for the ML estimates of covariance parameters in Experiment III with $\delta = 0$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

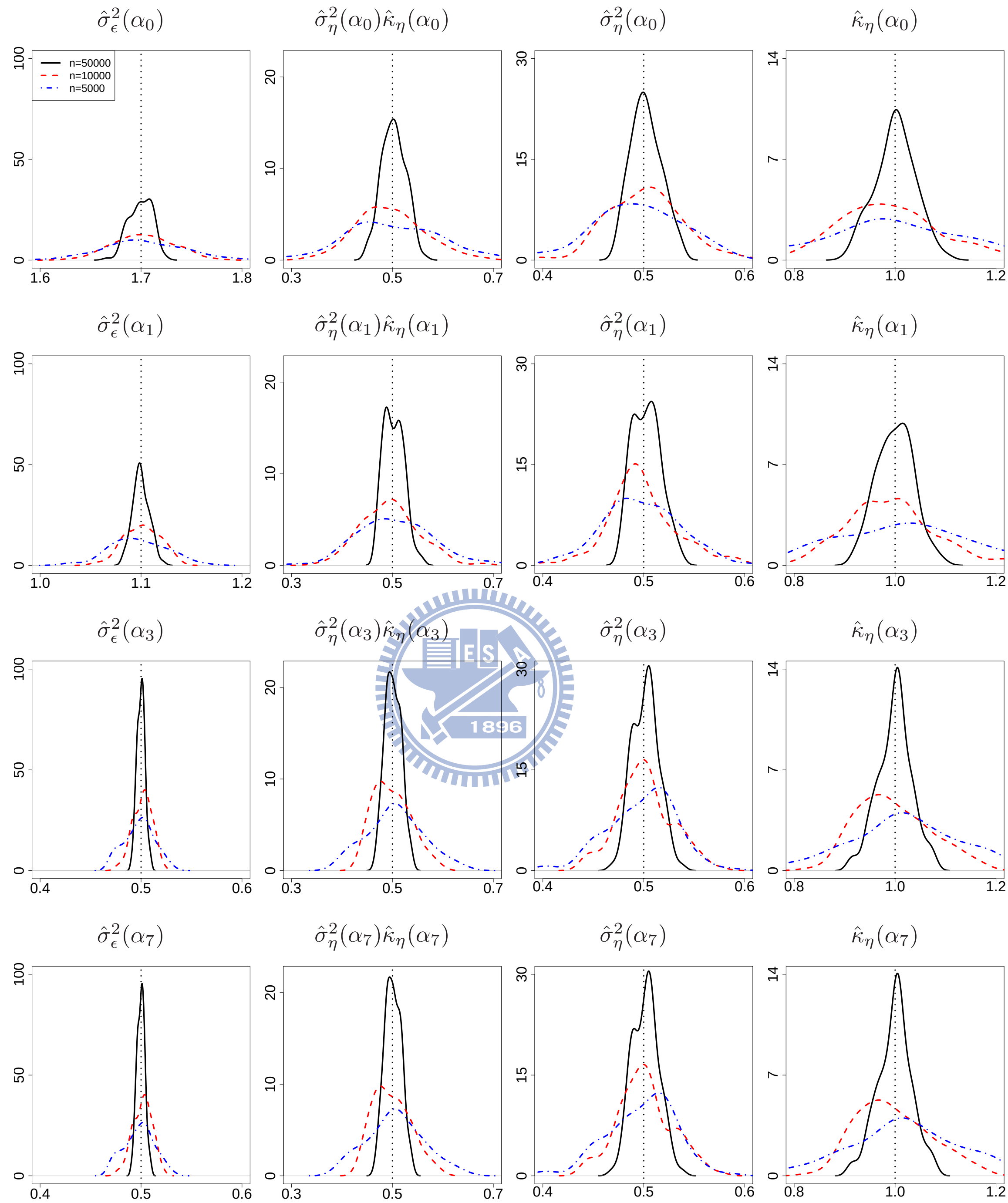


Figure 7.7: Probability density functions for the ML estimates of covariance parameters in Experiment III with $\delta = 0.75$ based on 100 simulation replicates, where the solid lines, the dashed lines, and the dot-dashed lines correspond to $n = 50000$, 10000, 5000, respectively, and the vertical dotted lines correspond to the convergence values of the ML estimates. Some density functions have support outside displayed regions.

Chapter 8

Summary and Discussion

In this thesis, we study asymptotic properties of geostatistical model selection. We find that asymptotic behaviors of GIC and CGIC depend not only on asymptotic frameworks but also on the smoothness of explanatory processes in space. For example, if the domain does not grow fast enough, GIC may select the smallest model asymptotically under some situation, and may possess different asymptotic properties if the domain grows in different rates. In addition, we find that the convergence rates of the ML estimates of covariance parameters also depend on the growth rate of the domain. In particular, we show that some covariance parameters are overestimated and some are underestimated by ML when fitting an incorrect model. These results are interesting and somewhat unique in geostatistics and geostatistical model selection.

The following are some topics we consider for further research.

8.1 Zeros of Covariance Parameters

In Chapter 5, we assume that $\{\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2\}$ are all positive. How if some of them is zero? When $\sigma_\eta^2 = 0$, the model of (3.1) reduces to the traditional regression model. However, it is of interest to study GIC to cover either $\sigma_\epsilon^2 = 0$ or $\kappa_\eta = 0$, which require modifications of theorems and their proofs to avoid singularity.

8.2 Other Covariance Structures

In Chapter 5, we consider the exponential covariance function class. There are many other covariance function classes that can be considered for $\eta(\cdot)$. For example, we may consider the Gaussian covariance function class defined in (2.4) or the Matérn class defined in (2.5), and study their asymptotic behavior for GIC or CGIC.

8.3 Sampling Designs

In this article, we focus on regular designs for the sampling locations in a one-dimensional domain. Asymptotic properties of GIC and CGIC for higher-dimensional spaces and some other commonly used spatial sampling designs, such as simple random sampling and stratified sampling, are of interest and require further research.

8.4 Continuous Functions as Explanatory Variables

In Chapter 5, we consider polynomial order selection. It is of interest to extend the polynomial variables to continuous or smooth functions. We conjecture that the asymptotic results similar to Theorems 6 and 7 can be extended from polynomials to functions of bounded variation.

Chapter 9

Appendix: Proofs

Proof of Lemma 1

By (3.8), we have $\mathrm{E}\big((\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)(\boldsymbol{\eta}+\boldsymbol{\epsilon})\big) = p(\alpha)$, which gives (3.9).

Proof of Lemma 2

First, by (3.8), we have for $\alpha \in \mathcal{A}^c$,

$$\begin{aligned} L^{\mathrm{KL}}(\alpha) - L^{\mathrm{KL}}(\alpha^c) &= \frac{1}{2}(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\big(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha) - \boldsymbol{M}(\alpha^c)'\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha^c)\big)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &= \chi^2(p(\alpha)-p(\alpha^c)) > 0, \end{aligned}$$

with $\chi^2(p(\alpha)-p(\alpha^c))$ denoting the chi-square distribution with $p(\alpha)-p(\alpha^c)$ degrees of freedom. Hence we obtain (3.10). Second, (3.11) follows trivially from (3.9). Last, by (3.8) and (3.12), for any $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, we have

$$L^{\mathrm{KL}}(\alpha) - L^{(\mathrm{KL})}(\alpha^c) = \boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + O_p(1) \to \infty,$$

as $n\to\infty$ with probability tending to 1. Hence (3.13) follows. This completes the proof.

Proof of Lemma 3

Since $\boldsymbol{S}$ and $\boldsymbol{Z}$ are jointly Gaussian,

$$\begin{aligned} \mathrm{E}(\boldsymbol{S}|\boldsymbol{Z}) &= \boldsymbol{\mu} + \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu}), \\ \mathrm{E}\|\boldsymbol{S}-\mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 &= \mathrm{tr}(\mathrm{var}(\boldsymbol{S}|\boldsymbol{Z})) = \mathrm{tr}\big(\boldsymbol{\Sigma}_\eta - \boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\eta\big) = \sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}). \end{aligned} \tag{9.1}$$

In addition, from (3.16),

$$\begin{aligned} \mathrm{E}\|\hat{\boldsymbol{S}}(\alpha)-\mathrm{E}(\boldsymbol{S}|\boldsymbol{Z})\|^2 &= \mathrm{E}\|(\boldsymbol{H}(\alpha)-\boldsymbol{I})\boldsymbol{\mu} + (\boldsymbol{H}(\alpha)-\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\|^2 \\ &= \mathrm{E}\|(\boldsymbol{H}(\alpha)-\boldsymbol{I})\boldsymbol{\mu}\|^2 + \mathrm{E}\|(\boldsymbol{H}(\alpha)-\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\|^2 \\ &= \sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \sigma_\epsilon^4\mathrm{tr}(\boldsymbol{M}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{M}(\alpha)\boldsymbol{\Sigma}) \\ &= \sigma_\epsilon^4\boldsymbol{\mu}'\boldsymbol{A}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{A}(\alpha)\boldsymbol{\mu} + \sigma_\epsilon^4\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)). \end{aligned}$$

Therefore, we obtain (3.17). This completes the proof.

Proof of Corollary 1

By (4.5) and (3.9), it suffices to show that for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \boldsymbol{\mu}' \boldsymbol{A}(\alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{A}(\alpha) \boldsymbol{\mu} = \sum_{j \in \alpha^c \setminus \alpha} \beta_j^2 \sigma_j^2 > 0. \tag{9.2}$$

Since

$$\begin{aligned} \frac{\boldsymbol{\mu}' \boldsymbol{A}(\alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{A}(\alpha) \boldsymbol{\mu}}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} &= \boldsymbol{\beta}(\alpha^c \setminus \alpha)' \left(\frac{\boldsymbol{X}(\alpha^c \setminus \alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}(\alpha^c \setminus \alpha)}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \right) \boldsymbol{\beta}(\alpha^c \setminus \alpha) \\ &\quad - \boldsymbol{\beta}'(\alpha^c \setminus \alpha)' \left(\frac{\boldsymbol{X}(\alpha^c \setminus \alpha) \boldsymbol{\Sigma}^{-1} \boldsymbol{X}(\alpha)}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \right) \left(\frac{\boldsymbol{X}(\alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}(\alpha)}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \right)^{-1} \\ &\quad \times \left(\frac{\boldsymbol{X}(\alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}(\alpha^c \setminus \alpha)}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \right) \boldsymbol{\beta}(\alpha^c \setminus \alpha), \end{aligned}$$

it is enough to show that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \boldsymbol{X}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{j'} = \begin{cases} \sigma_j^2; & \text{if } j = j', \\ 0; & \text{if } j \neq j', \end{cases}$$

where $\boldsymbol{X}_j$ is the jth column of $\boldsymbol{X}$. The desired result then follows from

$$\frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \mathrm{E}\big(\boldsymbol{X}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{j'} \big) = \begin{cases} \sigma_j^2; & \text{if } j = j', \\ 0; & \text{if } j \neq j', \end{cases}$$

and

$$\begin{aligned} \mathrm{var}\left(\frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \boldsymbol{X}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{j'} \right) &= \frac{1}{(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}))^2} \sigma_j^2 \sigma_{j'}^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) \\ &\leq \frac{1}{\sigma_\epsilon^2 (\mathrm{tr}(\boldsymbol{\Sigma}^{-1}))^2} \sigma_j^2 \sigma_{j'}^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1}) \\ &= \frac{1}{\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \sigma_j^2 \sigma_{j'}^2 \to 0, \end{aligned}$$

as $n \to \infty$ for $1 \leq j \leq j' \leq p$ by applying Chebyshev's inequality. This completes the proof. □

Proof of Corollary 2

Since (4.8) holds trivially, it suffices to check (4.7), which follows from (9.2) and the assumption of $\lim_{n\to\infty} \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\big)/\lambda = \infty$. This completes the proof. □

Proof of Corollary 3

The proof is essentially the same as that for Theorem 3 except (A.2) and (A.3) are now replaced by (A.2') and (A.3') in proving the corresponding statements. For example, (4.21) holds because $\mathrm{E}(\boldsymbol{X}_j' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta} + \boldsymbol{\epsilon})) = 0$ and

$$\lim_{n\to\infty} \frac{1}{\tau_n} \mathrm{var}(\boldsymbol{X}_j' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha)(\boldsymbol{\eta} + \boldsymbol{\epsilon})) = \lim_{n\to\infty} \frac{1}{\tau_n} \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_\alpha) \mathrm{E}(\boldsymbol{X}_j \boldsymbol{X}_j') \big) < \infty.$$

Details of the proof is omitted. □

Proof of Lemma 4

By (5.2), $\boldsymbol{G}_n\boldsymbol{\eta} = (\eta(s_1), \xi_2, \ldots, \xi_n)'$, which gives $\boldsymbol{\eta} = \boldsymbol{G}_n^{-1}(\eta(s_1), \xi_2, \ldots, \xi)'$. Taking the variance on both sides, we have

$$\boldsymbol{\Sigma}_\eta = \boldsymbol{G}_n^{-1}\boldsymbol{D}_n(\boldsymbol{G}_n^{-1})',$$

where

$$\boldsymbol{D}_n = \mathrm{var}\begin{pmatrix} \eta(s_1) \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} = \sigma_\eta^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1-\rho_n^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1-\rho_n^2 \end{pmatrix}. \tag{9.3}$$

It follows that

$$\boldsymbol{\Sigma} = \boldsymbol{G}_n^{-1}(\boldsymbol{D}_n + \sigma_\epsilon^2\boldsymbol{G}_n\boldsymbol{G}_n')(\boldsymbol{G}_n^{-1}),$$

and hence

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{G}_n'(\boldsymbol{D}_n + \sigma_\epsilon^2\boldsymbol{G}_n\boldsymbol{G}_n')^{-1}\boldsymbol{G}_n = \boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n,$$

where the second equality follows from $\boldsymbol{D}_n + \sigma_\epsilon^2\boldsymbol{G}_n\boldsymbol{G}_n' = \boldsymbol{T}_n$ obtained by direct computation. This completes the proof. □

Proof of Lemma 5

For (5.12), we start by writing $\boldsymbol{T}_k$ in terms of

$$\boldsymbol{B}_k = \begin{pmatrix} f_1(\rho_n) & -\rho_n\sigma_\epsilon^2 & 0 & \cdots & 0 \\ -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) & -\rho_n\sigma_\epsilon^2 & \ddots & \vdots \\ 0 & -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\rho_n\sigma_\epsilon^2 \\ 0 & \cdots & 0 & -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) \end{pmatrix}_{k\times k}. \tag{9.4}$$

Then

$$\det(\boldsymbol{T}_k) = (\sigma_\eta^2 + \sigma_\epsilon^2)\det(\boldsymbol{B}_{k-1}) - \rho_n^2\sigma_\epsilon^4\det(\boldsymbol{B}_{k-2}), \tag{9.5}$$

and

$$\det(\boldsymbol{B}_{k-1}) = f_1(\rho_n)\det(\boldsymbol{B}_{k-2}) - \rho_n^2\sigma_\epsilon^4\det(\boldsymbol{B}_{k-3}), \tag{9.6}$$

for $k \geq 3$, where $\det(\boldsymbol{B}_0) \equiv 1$. Solving the difference equation of (9.6), we have

$$\det(\boldsymbol{B}_{k-1}) = \frac{\sigma_\epsilon^{2k}(f_2^k(\rho_n) - f_3^k(\rho_n))}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}, \tag{9.7}$$

where

$$f_2(\rho_n) \equiv (f_1(\rho_n) + (f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2})(2\sigma_\epsilon^2)^{-1},$$

and

$$f_3(\rho_n) \equiv (f_1(\rho_n) - (f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2})(2\sigma_\epsilon^2)^{-1}.$$

Hence by (9.5)

$$\sigma_\epsilon^{-2j_n}\det(\boldsymbol{T}_{j_n}) = \frac{(\sigma_\eta^2+\sigma_\epsilon^2)f_2^{j_n}(\rho_n) - \rho_n^2\sigma_\epsilon^2 f_2^{j_n-1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} - \frac{(\sigma_\eta^2+\sigma_\epsilon^2)f_3^{j_n}(\rho_n) - \rho_n^2\sigma_\epsilon^2 f_3^{j_n-1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}.$$

Consequently, for (5.12) to hold, we remain to show the following:

$$\frac{f_3^{j_n}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} = o(\exp(-\tau n^{c/2})), \tag{9.8}$$

which can be obtained if the following two equations are satisfied:

$$(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2} = 2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}n^{-(1-\delta)/2} + O(n^{-3(1-\delta)/2}), \tag{9.9}$$

$$f_3^{j_n}(\rho_n) = o(\exp(-\tau n^c)), \quad \text{for some constant } \tau > 0. \tag{9.10}$$

For (9.9), we note by (5.3),

$$\rho_n^k = 1 - k\kappa_\eta n^{-(1-\delta)} + O(n^{-2(1-\delta)}); \quad k \in \mathbb{N}. \tag{9.11}$$

It follows that

$$\begin{aligned} f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4 &= ((1-\rho_n^2)\sigma_\eta^2 + (1+\rho_n^2)\sigma_\epsilon^2)^2 - 4\rho_n^2\sigma_\epsilon^4 \\ &= (1-\rho_n^2)^2\sigma_\eta^4 + 2(1-\rho_n^4)\sigma_\eta^2\sigma_\epsilon^2 + (1+\rho_n^2)^2\sigma_\epsilon^4 - 4\rho_n^2\sigma_\epsilon^4 \\ &= (1-\rho_n^2)^2\sigma_\eta^4 + 2(1-\rho_n^4)\sigma_\eta^2\sigma_\epsilon^2 + (1-\rho_n^2)^2\sigma_\epsilon^4 \\ &= 8\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2 n^{-(1-\delta)} + O(n^{-2(1-\delta)}). \end{aligned}$$

For (9.10), we note by (9.11),

$$\begin{aligned} f_1(\rho_n) &= (1-\rho_n^2)\sigma_\eta^2 + (1+\rho_n^2)\sigma_\epsilon^2 \\ &= 2\sigma_\epsilon^2 + (1-\rho_n^2)(\sigma_\eta^2 - \sigma_\epsilon^2) = 2\sigma_\epsilon^2 + 2\kappa_\eta(\sigma_\eta^2 - \sigma_\epsilon^2)n^{-(1-\delta)} + O(n^{-2(1-\delta)}). \end{aligned}$$

It follows that

$$\begin{aligned} f_3(\rho_n) &= \frac{f_1(\rho_n) - (f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}{2\sigma_\epsilon^2} \\ &= \frac{2\sigma_\epsilon^2 + 2(\sigma_\eta^2 - \sigma_\epsilon^2)\kappa_\eta n^{-(1-\delta)} - 2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}n^{-(1-\delta)/2}}{2\sigma_\epsilon^2} + O(n^{-3(1-\delta)/2}) \\ &= 1 - (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2} + O(n^{-(1-\delta)}). \end{aligned} \tag{9.12}$$

Since it is not difficult to show that $f_3(\rho_n) < 1$, we have

$$\begin{aligned} \log f_3^{j_n}(\rho_n) &\leq n^{(1-\delta)/2+c}\log(f_3(\rho_n)) \\ &= -(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^c + O(n^{-(1-\delta)/2+c}), \end{aligned}$$

where the equality follows by (9.12) and $\log(1-x) = -x + O(x^2)$ as $x \to 0$. Taking $\tau = (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}$, we obtain (9.10). This completes the proof of (5.12).

For (5.14) and (5.15), we first define $\boldsymbol{W}_{j_n}(k,\ell)$ to be the $(j_n-1)\times(j_n-1)$ matrix resulting from deleting row k and column ℓ of $\boldsymbol{T}_{j_n}$, for $1 \leq k,\ell \leq j_n$, then

$$C_{j_n}(k,\ell) = (-1)^{k+\ell}(\det(\boldsymbol{T}_{j_n}))^{-1}\det(\boldsymbol{W}_{j_n}(k,\ell)). \tag{9.13}$$

Note that for $1 \leq k \leq \ell \leq j_n$,

$$\boldsymbol{W}_{j_n}(k,\ell) = \begin{pmatrix} \boldsymbol{T}_{k-1} & * & * \\ \boldsymbol{0} & \boldsymbol{P}_{\ell-k} & * \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_{j_n-\ell} \end{pmatrix}, \tag{9.14}$$

where $\boldsymbol{T}_{k-1}$ and $\boldsymbol{B}_{j_n-\ell}$ are defined in (5.9) and (9.4), respectively,

$$\boldsymbol{P}_m \equiv \begin{pmatrix} -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) & -\rho_n\sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) & -\rho_n\sigma_\epsilon^2 & \ddots & \vdots \\ 0 & 0 & -\rho_n\sigma_\epsilon^2 & f_1(\rho_n) & \ddots & 0 \\ 0 & 0 & 0 & -\rho_n\sigma_\epsilon^2 & \ddots & -\rho_n\sigma_\epsilon^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & f_1(\rho_n) \\ 0 & 0 & \dots & 0 & 0 & -\rho_n\sigma_\epsilon^2 \end{pmatrix},$$

is an $m \times m$ matrix, and $\boldsymbol{P}_0 = \boldsymbol{T}_0 = \boldsymbol{B}_0 \equiv \emptyset$. Similarly, for $1 \le \ell \le k \le j_n$,

$$\boldsymbol{W}_{j_n}(k,\ell) = \begin{pmatrix} \boldsymbol{T}_{\ell-1} & \boldsymbol{0} & \boldsymbol{0} \\ * & \boldsymbol{P}'_{k-\ell} & \boldsymbol{0} \\ * & * & \boldsymbol{B}_{j_n-k} \end{pmatrix}. \tag{9.15}$$

It follows from (9.14) and (9.15) for $1 \le k,\ell \le j_n$ that

$$\det(\boldsymbol{W}_{j_n}(k,\ell)) = \det(\boldsymbol{T}_{\min(k,\ell)-1})\det(\boldsymbol{P}_{|k-\ell|})\det(\boldsymbol{B}_{j_n-\max(k,\ell)}). \tag{9.16}$$

Hence, by (5.12), (9.7), (9.8), (9.13) and (9.16),

$$\begin{aligned} C_{j_n}(1,\ell) &= \frac{(-1)^{\ell+1}(-\rho_n\sigma_\epsilon^2)^{\ell-1}\det(\boldsymbol{B}_{j_n-\ell})}{\det(\boldsymbol{T}_{j_n})} \\ &= \frac{f_2(\rho_n)}{(\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n)-\rho_n^2\sigma_\epsilon^2}\left(\frac{\rho_n}{f_2(\rho_n)}\right)^{\ell-1} + o(\tau\exp(-n^{c/2})), \end{aligned}$$

for $1 \le \ell \le j_n - n^{(1-\delta+c)/2}$, and

$$\begin{aligned} C_{j_n}(j_n,\ell) &= \frac{(-1)^{j_n+\ell}(-\rho_n\sigma_\epsilon^2)^{j_n-\ell}\det(\boldsymbol{T}_{\ell-1})}{\det(\boldsymbol{T}_{j_n})} \\ &= \frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\left(\frac{\rho_n}{f_2(\rho_n)}\right)^{j_n-\ell} + o(\tau\exp(-n^{c/2})), \end{aligned}$$

for $n^{(1-\delta+c)/2} \le \ell \le j_n$.

For (5.16), it is not difficult to show that $f_2(\rho_n) > 1$. Hence, by (5.12), (9.7) and (9.16), we have $\max_{1\le k,\ell\le j_n} C_{j_n}(k,\ell) = C_{j_n}(j_n/2, j_n/2)$. Hence,

$$\begin{aligned} \max_{1\le k,\ell\le j_n} C_{j_n}(k,\ell) &= \frac{\det(\boldsymbol{T}_{j_n/2-1})}{\det(\boldsymbol{T}_{j_n})}\det(\boldsymbol{B}_{j_n/2}) \\ &= \frac{1}{\sigma_\epsilon^2\sigma_\epsilon^{j_n} f_2^{j_n/2+1}(\rho_n)}\det(\boldsymbol{B}_{j_n/2}) + o(\exp(-\tau n^{c/4})) \\ &= \frac{1}{\sigma_\epsilon^2\sigma_\epsilon^{j_n} f_2^{j_n/2+1}(\rho_n)}\frac{\sigma_\epsilon^2\sigma_\epsilon^{j_n} f_2^{j_n/2+1}(\rho_n)}{(f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/4})) \\ &= \frac{1}{(8\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}n^{(1-\delta)/2} + o(n^{-(1-\delta)}), \end{aligned}$$

where the first equality follows from (9.13) and (9.16), the second equality follows (5.12), the third equality follows from (9.7) and (9.8), and the last equality follows from (9.9). Thus, we obtain (5.16).

For (5.17), by (5.12), (9.7) and (9.16), let $\phi = ((\sigma_\eta^2 + \sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2)/((\sigma_\eta^2 + \sigma_\epsilon^2)f_3(\rho_n) - \rho_n^2\sigma_\epsilon^2)$, we have for $i = 1, \dots, n$,

$$\begin{aligned}
C_n(i,i) &= \left(\frac{f_2^{i-2}(\rho_n) - \phi f_3^{i-2}(\rho_n)}{f_2^{n-1}(\rho_n)}\right)\left(\frac{f_2^{n-i+1}(\rho_n) - f_3^{n-i+1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}\right) + o(\exp(-\tau n^{c/2})) \\
&= \left(1 - \frac{\phi f_3^{i-2}(\rho_n)}{f_2^{i-2}(\rho_n)} - \frac{f_3^{n-i+1}(\rho_n)}{f_2^{n-i+1}(\rho_n)} + \frac{\phi f_3^{n-1}(\rho_n)}{f_2^{n-1}(\rho_n)}\right)\frac{1}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/2})) \\
&= \left(1 - \frac{\phi f_3^{i-2}(\rho_n)}{f_2^{i-2}(\rho_n)} - \frac{f_3^{n-i+1}(\rho_n)}{f_2^{n-i+1}(\rho_n)}\right)\frac{1}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/2})),
\end{aligned}$$

where the last equality follows from (9.10). Hence, we have

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{T}_n^{-1}) &= \sum_{i=1}^n C_n(i,i) \\
&= \left(n - \phi\sum_{i=1}^n \frac{f_3^{i-2}(\rho_n)}{f_2^{i-2}(\rho_n)} - \sum_{i=1}^n \frac{f_3^{n-i+1}(\rho_n)}{f_2^{n-i+1}(\rho_n)}\right)\frac{1}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/3})) \\
&= \left(n - \frac{\phi f_2^3(\rho_n) + f_3^2(\rho_n)}{f_2(\rho_n)f_3(\rho_n)(f_2(\rho_n) - f_3(\rho_n))}\right)\frac{1}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/3})) \\
&= \frac{n^{(3-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} + O(n^{1-\delta}),
\end{aligned}$$

where the second equality follows from

$$\begin{aligned}
\sum_{i=1}^n \frac{f_3^{i-2}(\rho_n)}{f_2^{i-2}(\rho_n)} &= \frac{f_2^2(\rho_n)}{f_3(\rho_n)(f_2(\rho_n) - f_3(\rho_n))} + o(\exp(-\tau n^c)), \\
\sum_{i=1}^n \frac{f_3^{n-i+1}(\rho_n)}{f_2^{n-i+1}(\rho_n)} &= \frac{f_3^2(\rho_n)}{f_2(\rho_n)(f_2(\rho_n) - f_3(\rho_n))} + o(\exp(-\tau n^c)),
\end{aligned}$$

and the last equality follows from $\phi = 1 + O(n^{-(1-\delta)/2})$, (9.9) and (9.10) that

$$\frac{1}{f_2(\rho_n) - f_3(\rho_n)} = \frac{n^{(1-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}} + O(1),$$

and

$$\frac{1}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} = \frac{n^{(1-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} + O(n^{-(1-\delta)/2}). \tag{9.17}$$

Thus, (5.17) is obtained.

Finally, we prove (5.18). By (5.12), (9.7) and (9.16), for $k < \ell$, we have

$$\begin{aligned}
C_n(k,\ell) &= \left(\frac{f_2^{k-2}(\rho_n) - \phi f_3^{k-2}(\rho_n)}{f_2^{n-1}(\rho_n)}\right)\rho_n^{\ell-k}\left(\frac{f_2^{n-\ell+1}(\rho_n) - f_3^{n-\ell+1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}\right) + o(\exp(-\tau n^{c/2})) \\
&= \left(f_2^{k-\ell}(\rho_n) - \frac{\phi f_3^{k-2}(\rho_n)}{f_2^{\ell-2}(\rho_n)} - \frac{f_3^{n-\ell+1}(\rho_n)}{f_2^{n-k+1}(\rho_n)}\right)\frac{\rho_n^{\ell-k}}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}} + o(\exp(-\tau n^{c/2})).
\end{aligned}$$

Hence, let $C_n^{(1)}(k,\ell)$ be the (k,ℓ)th element of $\boldsymbol{T}_n^{(1)-1}$, we have for $k,\ell = 1,\ldots,n$,

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(1)-1}) &= \sum_{k=1}^{n} C_n(k,k)C_n^{(1)}(k,k) + 2\sum_{k<\ell} C_n(k,\ell)C_n^{(1)}(k,\ell) \\
&= 2\sum_{k<\ell}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{\ell-k}\left(\frac{n^{1-\delta}}{4(2\kappa_\eta\sigma_\eta\sigma_\epsilon^2)^{1/2}(2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)2})^{1/2}}\right) + O(n^{2-\delta}) \\
&= \frac{n^{(5-3\delta)/2}}{2^{5/2}(\kappa_\eta\sigma_\eta^2\kappa_\eta^{(1)}\sigma_\eta^{(1)2})^{1/2}((\kappa_\eta\sigma_\eta^2)^{1/2} + (\kappa_\eta^{(1)}\sigma_\eta^{(1)2})^{1/2})} + O(n^{2-\delta}),
\end{aligned}$$

where the second equality follows from (9.17),

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{\ell=k}^{n}\frac{f_3^{k-2}(\rho_n)}{f_2^{\ell-2}(\rho_n)}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n^{(1)})}\right)^{\ell-k} &= \sum_{k=1}^{n}\left(\frac{f_3(\rho_n)}{f_2(\rho_n)}\right)^{k-2}\sum_{\ell=k}^{n}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{\ell-k} \\
&= O\left(n^{(1-\delta)/2}\sum_{k=1}^{n}\left(\frac{f_3(\rho_n)}{f_2(\rho_n)}\right)^{k}\right) \\
&= O(n^{1-\delta}),
\end{aligned}$$

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{\ell=k}^{n}\frac{f_3^{n-\ell+1}(\rho_n)}{f_2^{n-k+1}(\rho_n)}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n^{(1)})}\right)^{\ell-k} &= \sum_{\ell=1}^{n}\left(\frac{f_3(\rho_n)}{f_2(\rho_n)}\right)^{n-\ell+1}\sum_{k=1}^{\ell}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n^{(1)})}\right)^{\ell-k} \\
&= O\left(n^{(1-\delta)/2}\sum_{\ell=1}^{n}\left(\frac{f_3(\rho_n)}{f_2(\rho_n)}\right)^{n-\ell+1}\right) \\
&= O(n^{1-\delta}),
\end{aligned}$$

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{\ell=k}^{n}\frac{(f_3(\rho_n)f_3(\rho_n^{(1)}))^{k-2}}{(f_2(\rho_n)f_2(\rho_n^{(1)}))^{\ell-2}}(\rho_n\rho_n^{(1)})^{\ell-k} &= \sum_{k=1}^{n}\left(\frac{f_3(\rho_n)f_3(\rho_n^{(1)})}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{k-2}\sum_{\ell=k}^{n}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{\ell-k} \\
&= O\left(n^{(1-\delta)/2}\sum_{k=1}^{n}\left(\frac{f_3(\rho_n)f_3(\rho_n^{(1)})}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{k-2}\right) \\
&= O(n^{1-\delta}),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{\ell=k}^{n}\frac{(f_3(\rho_n)f_3(\rho_n^{(1)}))^{n-\ell+1}}{(f_2(\rho_n)f_2(\rho_n^{(1)}))^{n-k+1}}(\rho_n\rho_n^{(1)})^{\ell-k} &= \sum_{\ell=1}^{n}\left(\frac{f_3(\rho_n)f_3(\rho_n^{(1)})}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{n-\ell+1}\sum_{k=1}^{\ell}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{\ell-k} \\
&= O\left(n^{(1-\delta)/2}\sum_{\ell=1}^{n}\left(\frac{f_3(\rho_n)f_3(\rho_n^{(1)})}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{n-\ell+1}\right) \\
&= O(n^{1-\delta}),
\end{aligned}$$

and the last equality follows from

$$\begin{aligned}
&\sum_{k=1}^{n}\sum_{\ell=k+1}^{n}\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{\ell-k} \\
&= (n-1)\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)+\cdots+\left(\frac{\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})}\right)^{n-1} \\
&= \frac{n\rho_n\rho_n^{(1)}}{f_2(\rho_n)f_2(\rho_n^{(1)})-\rho_n\rho_n^{(1)}}-\frac{f_2(\rho_n)f_2(\rho_n^{(1)})\rho_n\rho_n^{(1)}}{(f_2(\rho_n)f_2(\rho_n^{(1)})-\rho_n\rho_n^{(1)})^2}+o(\exp(-\tau n^c)) \\
&= \frac{nn^{(1-\delta)/2}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}+(2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)2})^{1/2}}+O(n^{1-\delta}).
\end{aligned}$$

This completes the proof. □

Proof of Lemma 6

We first prove (5.22)-(5.24). By (5.10), (9.9), and (9.11), we have

$$\begin{aligned}
f_4(\rho_n) &= \frac{f_1(\rho_n)-(f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2}}{2\rho_n\sigma_\epsilon^2} \\
&= \frac{(1-\rho_n^2)\sigma_\eta^2+(1+\rho_n^2)\sigma_\epsilon^2-(8\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}n^{-(1-\delta)/2}}{2\rho_n\sigma_\epsilon^2}+o(n^{-(1-\delta)}) \\
&= \frac{2\kappa_\eta n^{-(1-\delta)}\sigma_\eta^2+2\sigma_\epsilon^2-2\kappa_\eta n^{-(1-\delta)}\sigma_\epsilon^2-(8\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}n^{-(1-\delta)/2}}{2\rho_n\sigma_\epsilon^2}+o(n^{-(1-\delta)}) \\
&= 1-(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+O(n^{-(1-\delta)}).
\end{aligned}$$

That is, we obtain (5.22). By (9.11) and (5.22), we have

$$\begin{aligned}
f_2(\rho_n) &= \frac{f_1(\rho_n)+(f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2}}{2\sigma_\epsilon^2} \\
&= \frac{2\sigma_\epsilon^2+2(\sigma_\eta^2-\sigma_\epsilon^2)\kappa_\eta n^{-(1-\delta)}+2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}n^{-(1-\delta)/2}}{2\sigma_\epsilon^2}+O(n^{-3(1-\delta)/2}) \\
&= 1+(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+(\sigma_\eta^2-\sigma_\epsilon^2)\sigma_\epsilon^{-2}\kappa_\eta n^{-(1-\delta)}+O(n^{-(1-\delta)}),
\end{aligned}$$

and hence (5.23) holds. Applying $\log(1-x)=-x+O(x^2)$ as $x\to 0$ to (5.22), we have

$$\log f_4(\rho_n)=f_4(\rho_n)-1+O((f_4(\rho_n)-1)^2)=-(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2}+O(n^{-(1-\delta)}).$$

Thus (5.24) is obtained.

We remain to show (5.25). Applying $\log(1-x)=-x+O(x^2)$ as $x\to 0$, we have

$$\begin{aligned}
\log f_4^{j_n}(\rho_n) &\le n^{(1-\delta)/2+c}\log f_4(\rho_n) \\
&= n^{(1-\delta)/2+c}\log(1-(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2})+O(n^{-(1-\delta)/2+c}) \\
&= -(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^c+O(n^{-(1-\delta)/2+c}).
\end{aligned}$$

Taking the exponential on both sides of the above equation with $\tau=(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}$ yields (5.25). This completes the proof. □

Proof of Lemma 7

Let $\boldsymbol{F}_n = (F_1, \ldots, F_n) = \boldsymbol{\Omega}_n \boldsymbol{\upsilon}_n$, where $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_n)'$ is given in (5.20). Since $\boldsymbol{T}_n = \mathrm{var}(\boldsymbol{\upsilon})$ by (5.19), we have $\mathrm{var}(\boldsymbol{F}_n) = \boldsymbol{\Omega}_n \boldsymbol{T}_n \boldsymbol{\Omega}_n'$, and hence

$$\boldsymbol{T}_n^{-1} = \boldsymbol{\Omega}_n'(\mathrm{var}(\boldsymbol{F}_n))^{-1}\boldsymbol{\Omega}_n.$$

Therefore, to prove (5.26), it is enough to show that

$$(\mathrm{var}(\boldsymbol{F}_n))^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{j_n}^{-1} & \mathbf{0} \\ \mathbf{0} & (f_2(\rho_n)\sigma_\epsilon^2)^{-1}\boldsymbol{I}_{n-j_n} \end{pmatrix} + o(\exp(-\tau n^{2c/3})). \tag{9.18}$$

By (5.20) for any $m \in \mathbb{N}$,

$$F_m = \sum_{k=0}^{m-1} f_4^k(\rho_n)\upsilon_{m-k} = \sum_{k=0}^{m-1} f_4^k(\rho_n)(u_{m-k} - f_4(\rho_n)u_{m-(k+1)}) = u_m - f_4^m(\rho_n)u_0.$$

It follows that

$$\boldsymbol{F}_n = (\boldsymbol{F}_{j_n}', ((u_{j_n+1}, \ldots, u_n) + f_4^{j_n}(\rho_n)u_0\boldsymbol{f}_{n-j_n})')',$$

where $\boldsymbol{f}_{n-j_n} = (1, f_4(\rho_n), \ldots, f_4^{n-j_n}(\rho_n))'$. It follows that

$$\mathrm{var}(\boldsymbol{F}_n) \;=\; \begin{pmatrix} \boldsymbol{\Lambda}_{j_n} & \mathbf{0} \\ \mathbf{0} & f_2(\rho_n)\sigma_\epsilon^2\boldsymbol{I}_{n-j_n} \end{pmatrix} + o(\exp(-\tau n^{4c/5})), \tag{9.19}$$

which follows from (5.25), (5.28) and $\mathrm{cov}(\boldsymbol{\upsilon}_{j_n}, u_0) = (-f_4(\rho_n)\mathrm{var}(u_0), 0, \ldots, 0)'$, by (5.20). Since

$$\boldsymbol{\Omega}_k^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -f_4(\rho_n) & 1 & 0 & \ddots & \vdots \\ 0 & -f_4(\rho_n) & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -f_4(\rho_n) & 1 \end{pmatrix},$$

it follows from (5.16) that all elements of $\boldsymbol{\Lambda}_{j_n}^{-1} = (\boldsymbol{\Omega}_{j_n}')^{-1}\boldsymbol{T}_{j_n}^{-1}\boldsymbol{\Omega}_{j_n}^{-1}$ are less than or equal to $4n^{(1-\delta)/2}$. This together with (9.19) give (9.18). Thus (5.26) is obtained.

Proof of Lemma 8

First, we prove (5.29). By (5.8), we have $\det(\boldsymbol{G}) = 1$ and hence by (5.11) and (5.12),

$$\begin{aligned} \det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) &= (\det(\boldsymbol{G}_n'))^{-1}\det(\boldsymbol{T}_n)(\det(\boldsymbol{G}_n^{-1}))^{-1} \\ &= \det(\boldsymbol{T}_n) \\ &= \sigma_\epsilon^{2n}\frac{f_2^{n-1}(\rho_n)}{(f_1^2(\rho_n) - 4\rho_n^2\sigma_\epsilon^4)^{1/2}}((\sigma_\eta^2 + \sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2) + o(\exp(-\tau n^{c/2})). \end{aligned}$$

It follows from $\log(x+\Delta_x) = \log x + O(\Delta_x/x)$ as $\Delta_x \to 0$ that

$$\begin{aligned}
&\log(\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}))) \\
&= \log\left(\sigma_\epsilon^{2n}\frac{f_2^{n-1}(\rho_n)}{(f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2}}((\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n)-\rho_n^2\sigma_\epsilon^2)\right)+o(\exp(-\tau n^{c/2})) \\
&= n\log\sigma_\epsilon^2+(n-1)\log(f_2(\rho_n))+\log((\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n)-\rho_n^2\sigma_\epsilon^2) \\
&\quad -\log((f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2})+o(\exp(-\tau n^{c/2})) \\
&= n\log\sigma_\epsilon^2+(n-1)\log(f_2(\rho_n))+\log(\sigma_\eta^2)-\log((f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2})+O(n^{-(1-\delta)/2}) \\
&= n\log\sigma_\epsilon^2+(n-1)\log(f_2(\rho_n))-\log n^{(1-\delta)/2}+O(1) \\
&= n\log\sigma_\epsilon^2+(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{(1+\delta)/2}-(\kappa_\eta(\sigma_\eta^2+\sigma_\epsilon^2)\sigma_\epsilon^{-2})n^\delta-\log n^{(1-\delta)/2}+o(n^\delta)+O(1),
\end{aligned}$$

where the third equality follows from (9.11) and (5.23), the fourth equality follows from (9.9) that

$$\log((f_1^2(\rho_n)-4\rho_n^2\sigma_\epsilon^4)^{1/2}) = \frac{1-\delta}{2}\log n + O(1),$$

and the last equality follows from a Taylor expansion of $\log(f_2(\rho_n))$ at $f_2(\rho_n)=1$ and together with (5.23) that

$$\log f_2(\rho_n) = (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{-(1-\delta)/2} - (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})n^{-(1-\delta)} + \frac{\sigma_\eta^2-\sigma_\epsilon^2}{\sigma_\epsilon^2}\kappa_\eta n^{-(1-\delta)} + O(n^{-3(1-\delta)/2}).$$

Hence, (5.29) is obtained.

Second, we prove (5.30). By (5.11),

$$\boldsymbol{\Sigma}_\eta^{(1)} = \boldsymbol{G}_n^{(1)}\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)\prime})^{-1},$$

where $\boldsymbol{G}_n^{(1)}$ and $\boldsymbol{D}_n^{(1)}$ are given in (5.8) and (9.3) with σ_η^2, κ_η and ρ_n are replaced by $\sigma_\eta^{(1)2}$, $\kappa_\eta^{(1)}$ and $\rho_n^{(1)} = \exp(-\kappa_\eta^{(1)}n^{-(1-\delta)})$, respectively. It follows together with (5.11),

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \mathrm{tr}(\boldsymbol{G}_n^{(1)-1}\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)\prime})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n) \\
&= \mathrm{tr}(\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)\prime})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}) \\
&= \sigma_\eta^{(1)2}(1-\rho_n^{(1)2})\mathrm{tr}((\boldsymbol{G}_n^{(1)\prime})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}) \\
&\quad +\sigma_\eta^{(1)2}\rho_n^{(1)2}\mathrm{tr}(\boldsymbol{e}_1\boldsymbol{e}_1'(\boldsymbol{G}_n^{(1)\prime})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}),
\end{aligned}$$

where $\boldsymbol{e}_1 \equiv (1,0,\ldots,0)'$. Therefore, for (5.30) to hold, it remains to show that

$$\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 = O(1), \tag{9.20}$$

and

$$\begin{aligned}
\mathrm{tr}((\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}) &= \frac{n^{(3-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}} + \frac{(\kappa_\eta-\kappa_\eta^{(1)})n}{2\kappa_\eta\sigma_\eta^2} + \frac{(\kappa_\eta-\kappa_\eta^{(1)})^2 n}{4\kappa_\eta^{(1)}\kappa_\eta\sigma_\eta^2} \\
&\quad +o(n)+O(n^{1-\delta}).
\end{aligned} \tag{9.21}$$

Before proving (9.20) and (9.21), we compute some matrices that are used very often in the followings. First,

$$\boldsymbol{G}_n^{(1)-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \rho_n^{(1)} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \rho_n^{(1)n-1} & \cdots & \rho_n^{(1)} & 1 \end{pmatrix},$$

and hence

$$
\begin{aligned}
\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1} &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho_n^{(1)}-\rho_n & 1 & 0 & \cdots & 0 \\ \rho_n^{(1)}(\rho_n^{(1)}-\rho_n) & \rho_n^{(1)}-\rho_n & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \rho_n^{(1)n-2}(\rho_n^{(1)}-\rho_n) & \cdots & \rho_n^{(1)}(\rho_n^{(1)}-\rho_n) & \rho_n^{(1)}-\rho_n & 1 \end{pmatrix} \\
&= \boldsymbol{I}_n + (\rho_n^{(1)}-\rho_n)\boldsymbol{L}_n^{(1)}, \qquad (9.22)
\end{aligned}
$$

where

$$
\boldsymbol{L}_n^{(1)} \equiv \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \rho_n^{(1)} & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \rho_n^{(1)n-2} & \cdots & \rho_n^{(1)} & 1 & 0 \end{pmatrix}. \qquad (9.23)
$$

Second, for $\boldsymbol{\Omega}$ defined in (5.27), we have

$$
\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ g_1(\rho_n) & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ g_{n-2}(\rho_n) & \cdots & g_1(\rho_n) & 1 & 0 \end{pmatrix}, \qquad (9.24)
$$

where for $i = 1, \ldots, n-2$,

$$
\begin{aligned}
g_i(\rho_n) &= f_4^i(\rho_n) + f_4^{i-1}(\rho_n)\rho_n^{(1)} + \cdots + f_4(\rho_n)\rho_n^{(1)i-1} + \rho_n^{(1)i} \\
&= \frac{\rho_n^{(1)i+1} - f_4^{i+1}(\rho_n)}{\rho_n^{(1)} - f_4(\rho_n)}. \qquad (9.25)
\end{aligned}
$$

Now, we prove (9.20). By (9.22), it is enough to show

$$
\begin{aligned}
\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 &= O(1), \qquad (9.26) \\
(\rho_n^{(1)}-\rho_n)^2\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 &= O(1). \qquad (9.27)
\end{aligned}
$$

For (9.26), it follows easily from (5.14) that $C_n(1,1) = O(1)$. For (9.27), by (5.26) with some $c > 0$ such that $n^* = n^{(1-\delta)/2+c} < n$, we have

$$
\begin{aligned}
(\rho_n^{(1)}-\rho_n)^2\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 &= (\rho_n^{(1)}-\rho_n)^2\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)\prime}\begin{pmatrix} \boldsymbol{T}_{n^*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 + o(\exp(-\tau n^{c/3}) \\
&\quad + \frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)\prime}\boldsymbol{\Omega}_n'\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\
&= \frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)\prime}\boldsymbol{\Omega}_n'\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 + o(1) \\
&= O\left(n^{-2(1-\delta)}\sum_{i=n^*}^{n} g_i^2(\rho_n)\right) \\
&= O(1),
\end{aligned}
$$

where the second equality follows from

$$\begin{aligned}
\boldsymbol{e}_1' \boldsymbol{L}_n^{(1)'} \begin{pmatrix} \boldsymbol{T}_{n^*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{L}_n^{(1)} \boldsymbol{e}_1 &= O(\boldsymbol{1}_{n^*}' \boldsymbol{T}_{n^*} \boldsymbol{1}_{n^*}) \\
&= O(n^{(1-\delta)/2} n^{*2}) = o(n^{2(1-\delta)}),
\end{aligned}$$

and

$$\rho_n^{(1)} - \rho_n = (\kappa_\eta - \kappa_\eta^{(1)}) n^{-(1-\delta)} + O(n^{-2(1-\delta)}), \tag{9.28}$$

the second last equality follows from (9.24) and (9.28), and the last equality follows from (9.25) that

$$\sum_{i=1}^n g_i^2(\rho_n) = O\left(n^{1-\delta}\left(\frac{1}{1-\rho_n^{(1)2}} + \frac{1}{1-f_4^2(\rho_n)} - \frac{2}{1-\rho_n^{(1)} f_4(\rho_n)}\right)\right) = O(n^{2(1-\delta)}).$$

Thus, (9.20) is obtained. Next, we prove (9.21). By (9.22), we have

$$\begin{aligned}
&\mathrm{tr}((\boldsymbol{G}_n \boldsymbol{G}_n^{(1)-1})' \boldsymbol{T}_n^{-1} \boldsymbol{G}_n \boldsymbol{G}_n^{(1)-1}) \\
&= \mathrm{tr}(\boldsymbol{T}_n^{-1}) + 2(\rho_n^{(1)} - \rho_n) \mathrm{tr}(\boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) + (\rho_n^{(1)} - \rho_n)^2 \mathrm{tr}(\boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) \\
&= 2^{-1} (2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^2)^{-1/2} n^{(3-\delta)/2} + 2(\rho_n^{(1)} - \rho_n) \mathrm{tr}(\boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) \\
&\quad + (\rho_n^{(1)} - \rho_n)^2 \mathrm{tr}(\boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) + O(n^{1-\delta}),
\end{aligned}$$

where the second last equality follows from (5.17). Therefore, for (9.21) to hold, it remains to show that

$$(\rho_n^{(1)} - \rho_n) \mathrm{tr}(\boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) = \frac{(\kappa_\eta - \kappa_\eta^{(1)})}{4\kappa_\eta \sigma_\eta^2} n + o(n), \tag{9.29}$$

$$(\rho_n^{(1)} - \rho_n)^2 \mathrm{tr}(\boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) = \frac{(\kappa_\eta - \kappa_\eta^{(1)})^2 n}{4\kappa_\eta^{(1)} \kappa_\eta \sigma_\eta^2} + o(n) + O(n^{1-\delta}). \tag{9.30}$$

For (9.29), we have

$$\begin{aligned}
(\rho_n^{(1)} - \rho_n) \mathrm{tr}(\boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)}) &= (\rho_n^{(1)} - \rho_n) \mathrm{tr}\left(\begin{pmatrix} \boldsymbol{T}_{n^*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{L}_n^{(1)}\right) \\
&\quad + \frac{\rho_n^{(1)} - \rho_n}{f_2(\rho_n) \sigma_\epsilon^2} \mathrm{tr}\left(\boldsymbol{\Omega}_n' \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)}\right) + o(\exp(-\tau n^{c/2})) \\
&= \frac{\rho_n^{(1)} - \rho_n}{f_2(\rho_n) \sigma_\epsilon^2} \mathrm{tr}\left(\boldsymbol{\Omega}_n' \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)}\right) + o(n^{1-\delta}),
\end{aligned}$$

where the first equality follows from (5.26) and the last equality follows from (5.16) that,

$$\begin{aligned}
(\rho_n^{(1)} - \rho_n) \mathrm{tr}\left(\begin{pmatrix} \boldsymbol{T}_{n^*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{L}_n^{(1)}\right) &= O(n^{-(1-\delta)/2}(1 + 2 + \cdots + n^*)) \\
&= O(n^{-(1-\delta)/2} n^{*2}) \\
&= o(n^{1-\delta}).
\end{aligned}$$

Therefore, for (9.29) to hold, it remains to show that

$$
\begin{aligned}
&\frac{\rho_n^{(1)}-\rho_n}{f_2(\rho_n)\sigma_\epsilon^2}\mathrm{tr}\left(\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\right)\\
&= \frac{(\rho_n^{(1)}-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*}^{n}(g_i(\rho_n)f_4^{i+1}(\rho_n)+g_{i-1}(\rho_n)f_4^{i}(\rho_n)+\ldots+f_4(\rho_n))\\
&= \frac{(\rho_n^{(1)}-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2(\rho_n^{(1)}-f_4(\rho_n))}\sum_{i=n^*}^{n}\left(\frac{f_4(\rho_n)\rho_n^{(1)}}{1-f_4(\rho_n)\rho_n^{(1)}}-\frac{f_4^2(\rho_n)}{1-f_4^2(\rho_n)}\right)+o(\exp(-\tau n^{c/2}))\\
&= \frac{(\kappa_\eta-\kappa_\eta^{(1)})n^{-(1-\delta)}}{\sigma_\epsilon^2}\frac{n^{(1-\delta)/2}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}\frac{n^{(1-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}(n-n^*)+o(n)\\
&= \frac{(\kappa_\eta-\kappa_\eta^{(1)})}{4\kappa_\eta\sigma_\eta^2}n+o(n), \qquad (9.31)
\end{aligned}
$$

where the first equality follows from (5.27) and (9.24), the second equality follows from (5.25) and (9.25) that

$$
\begin{aligned}
&(\rho_n^{(1)}-f_4(\rho_n))(g_i(\rho_n)f_4^{i+1}(\rho_n)+g_{i-1}(\rho_n)f_4^{i}(\rho_n)+\ldots+f_4(\rho_n))\\
&= (\rho_n^{(1)}f_4(\rho_n)+\cdots+(\rho_n^{(1)}f_4(\rho_n))^{i+1})-(f_4^2(\rho_n)+\cdots+f_4^{2(i+1)}(\rho_n))\\
&= \frac{\rho_n^{(1)}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)}-\frac{f_4^2(\rho_n)}{1-f_4^2(\rho_n)}+o(\exp(-\tau n^{c})),
\end{aligned}
$$

for $i=n^*,\ldots,n$, and the third equality follows from (9.28),

$$\frac{1}{\rho_n^{(1)}-f_4(\rho_n)}=\frac{n^{(1-\delta)/2}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}+o(n^{(1-\delta)/2}), \qquad (9.32)$$

$$\frac{\rho_n^{(1)}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)}=\frac{n^{(1-\delta)/2}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}+o(n^{(1-\delta)/2}), \qquad (9.33)$$

and

$$\frac{f_4^2(\rho_n)}{1-f_4^2(\rho_n)}=\frac{n^{(1-\delta)/2}}{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}+o(n^{(1-\delta)/2}). \qquad (9.34)$$

Thus, (9.29) is obtained. For (9.30), we have

$$
\begin{aligned}
(\rho_n^{(1)}-\rho_n)^2\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}) &= (\rho_n^{(1)}-\rho_n)^2\mathrm{tr}\left(\boldsymbol{L}_n^{(1)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\right)\\
&\quad+\frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\mathrm{tr}\left(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\right)\\
&\quad+o(\exp(-\tau n^{c/2}))\\
&= \frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\mathrm{tr}\left(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\right)+o(n),
\end{aligned}
$$

where the first equality follows from (5.26), the last equality follows from (5.16) and (9.23) that

$$
\begin{aligned}
(\rho_n^{(1)}-\rho_n)^2\mathrm{tr}\left(\boldsymbol{L}_n^{(1)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\right) &= O\left(n^{-3(1-\delta)/2}\sum_{i=1}^{n^*}(i-1)^2\right)\\
&= o(n).
\end{aligned}
$$

Therefore, for (9.30) to hold, it remains to show that

$$\begin{aligned}
&\frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\text{tr}\left(\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\right)\\
&= \frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*+1}^{n}(1+g_1^2(\rho_n)+\cdots+g_i^2(\rho_n))\\
&= \frac{(\rho_n^{(1)}-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2(\rho_n^{(1)}-f_4(\rho_n))^2}\sum_{i=n^*+1}^{n}\left(\frac{\rho_n^{(1)2}-\rho_n^{(1)2i+4}}{1-\rho_n^{(1)2}}-\frac{2\rho_n^{(1)}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)}+\frac{f_4^2(\rho_n)}{1-f_4^2(\rho_n)}\right)\\
&\quad +o(\exp(-\tau n^{c/3}))\\
&= \frac{(\kappa_\eta-\kappa_\eta^{(1)})^2 n^{-2(1-\delta)}}{\sigma_\epsilon^2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})n^{-(1-\delta)}}\frac{n^{1-\delta}}{2\kappa_\eta^{(1)}}(n-n^*)+o(n)+O(n^{1-\delta})\\
&= \frac{(\kappa_\eta-\kappa_\eta^{(1)})^2}{4\kappa_\eta^{(1)}\kappa_\eta\sigma_\eta^2}n+o(n)+O(n^{1-\delta}), \qquad (9.35)
\end{aligned}$$

where the first equality follows from (9.24) and (9.25), where the second equality follows from

$$\begin{aligned}
&(\rho_n^{(1)}-f_4(\rho_n))^2(1+g_1^2(\rho_n)+\cdots+g_i^2(\rho_n))\\
&\quad= (\rho_n^{(1)}-f_4(\rho_n))^2+(\rho_n^{(1)2}-f_4^2(\rho_n))^2+\cdots+(\rho_n^{(1)i+1}-f_4^{i+1}(\rho_n))^2\\
&\quad= (\rho_n^{(1)2}+\cdots+\rho_n^{(1)2(i+1)})-2(\rho_n^{(1)}f_4(\rho_n)+\cdots+\rho_n^{(1)i+1}f_4^{i+1}(\rho_n))\\
&\qquad +(f_4^2(\rho_n)+\cdots+f_4^{2(i+1)}(\rho_n))\\
&\quad= \frac{\rho_n^{(1)2}-\rho_n^{(1)2i+4}}{1-\rho_n^{(1)2}}-\frac{2\rho_n^{(1)}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)}+\frac{f_4^2(\rho_n)}{1-f_4^2(\rho_n)}+o(\exp(-\tau n^{c/2})),
\end{aligned}$$

for $i=n^*,\ldots,n$, and the third equality follows from (9.32), (9.33), (9.34) and

$$\begin{aligned}
\frac{\rho_n^{(1)2}}{1-\rho_n^{(1)2}} &= \frac{n^{1-\delta}}{2\kappa_\eta^{(1)}}+o(n^{1-\delta}),\\
\sum_{i=n^*}^{n}\frac{\rho_n^{(1)2i+4}}{1-\rho_n^{(1)2}} &= O(n^{2(1-\delta)}).
\end{aligned}$$

Thus, (9.30) and hence (9.21) are obtained. It completes the proof of (5.30).

Second, we prove (5.31). By (5.30), we have

$$\text{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))=\frac{\sigma_\eta^2\kappa_\eta}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}}n^{(1+\delta)/2}+o(n^\delta)+O(1),$$

and

$$\text{tr}(\boldsymbol{I}_n) = \text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))=\sigma_\epsilon^2\text{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))+\text{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})).$$

Then, we have

$$\begin{aligned}
\text{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{1}{\sigma_\epsilon^2}\text{tr}(\boldsymbol{I}_n)-\frac{1}{\sigma_\epsilon^2}\text{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))\\
&= \frac{n}{\sigma_\epsilon^2}-\frac{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}{2\sigma_\epsilon^2}n^{(1+\delta)/2}+o(n^\delta).
\end{aligned}$$

Thus, (5.31) is obtained.

Third, we prove (5.32). By (5.31), we have

$$\sigma_\epsilon^{(1)}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) = \frac{\sigma_\epsilon^{(1)2}}{\sigma_\epsilon^2}n - \frac{\sigma_\epsilon^{(1)2}}{2\sigma_\epsilon^2}(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{(1+\delta)/2} + o(n^\delta) + O(1).$$

Then, by (5.30),

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Sigma}^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \sigma_\epsilon^{(1)}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))\\ &= \frac{\sigma_\epsilon^{(1)2}}{\sigma_\epsilon^2}n - \frac{\sigma_\epsilon^{(1)2}}{2\sigma_\epsilon^2}(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{(1+\delta)/2} + \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}}n^{(1+\delta)/2}\\ &\quad + \frac{\sigma_\eta^{(1)2}\kappa_\eta^{(1)}(\kappa_\eta - \kappa_\eta^{(1)})}{\kappa_\eta\sigma_\eta^2}n^\delta + \frac{\sigma_\eta^{(1)2}(\kappa_\eta - \kappa_\eta^{(1)})^2}{2\kappa_\eta\sigma_\eta^2}n^\delta + o(n^\delta) + O(1).\end{aligned}$$

Thus, (5.32) is obtained.

Fourth, before proving (5.33), we need several equations that are helpful in the following. First, for some $c > 0$ such that $n^* = n^{(1-\delta)/2+c} < n$ and $\boldsymbol{\Omega}_n^{(3)}$ defined in (5.27) with ρ_n replaced by $\rho_n^{(3)} = \exp(-\kappa_\eta^{(3)}n^{-(1-\delta)})$,

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\Omega}_n^{(3)\prime} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{(3)} \end{pmatrix} + o(\exp(-\tau n^c)), \tag{9.36}$$

where

$$\boldsymbol{Q}_{n-n^*}^{(3)} = \frac{1}{1 - f_4(\rho_n)f_4(\rho_n^{(3)})}\begin{pmatrix} 1 & f_4(\rho_n^{(3)}) & \cdots & f_4^{n-n^*-1}(\rho_n^{(3)}) \\ f_4(\rho_n) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & f_4(\rho_n^{(3)}) \\ f_4^{n-n^*-1}(\rho_n) & \cdots & f_4(\rho_n) & 1 \end{pmatrix},$$

which follows from

$$\boldsymbol{\omega}_i'\boldsymbol{\omega}_j^{(3)} = f_4^{|i-j|}(\rho_n^*)\sum_{k=0}^{\min(i,j)} f_4^k(\rho_n)f_4^k(\rho_n^{(3)}) = \frac{f_4^{|i-j|}(\rho_n^*)}{1 - f_4(\rho_n)f_4(\rho_n^{(3)})} + o(\exp(-\tau n^{c/2})),$$

for $n^* \le i, j \le n$, by (5.25), where $\rho_n^* = \rho_n I_{\{i\le j\}} + \rho_n^{(3)} I_{\{i>j\}}$ and $\boldsymbol{\omega}_i$ is the ith column of $\boldsymbol{\Omega}_n'$.

Second, for $\boldsymbol{L}_n^{(1)}$ defined in (9.23),

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{\Omega}_n^{(3)\prime} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{*(1)} \end{pmatrix} + O(n^{-(1-\delta)/2}), \tag{9.37}$$

where

$$\boldsymbol{Q}_{n-n^*}^{*(1)} = O\left(n^{1-\delta}\begin{pmatrix} f_4(\rho_n^{(3)}) & f_4^2(\rho_n^{(3)}) & \cdots & f_4^{n-n^*}(\rho_n^{(3)}) \\ \rho_n^{(1)} & f_4(\rho_n^{(3)}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & f_4^2(\rho_n^{(3)}) \\ \rho_n^{(1)n-n^*-1} & \cdots & \rho_n^{(1)} & f_4(\rho_n^{(3)}) \end{pmatrix}\right),$$

which follows from (5.25), (9.24), (9.25) that for $n^* + 1 \le i \le j \le n$,

$$
\begin{aligned}
&(g_{i-2}(\rho_n), \cdots, g_1(\rho_n), 1, 0, \cdots, 0)\boldsymbol{\omega}_j^{(3)} \\
&= f_4^{j-i+1}(\rho_n^{(3)}) \sum_{k=1}^{i-1} g_{k-1}(\rho_n) f_4^{k-1}(\rho_n^{(3)}) \\
&= \frac{f_4^{j-i+1}(\rho_n^{(3)})}{\rho_n^{(1)} - f_4(\rho_n)} \left(\frac{\rho_n^{(1)}}{1 - \rho_n^{(1)} f_4(\rho_n^{(3)})} - \frac{f_4(\rho_n)}{1 - f_4(\rho_n) f_4(\rho_n^{(3)})} \right) + o(\exp(-\tau n^{c/2})) \\
&= \frac{f_4^{j-i+1}(\rho_n^{(3)})}{(1 - \rho_n^{(1)} f_4(\rho_n^{(3)}))(1 - f_4(\rho_n) f_4(\rho_n^{(3)}))} + o(\exp(-\tau n^{c/2})) \\
&= O(n^{1-\delta} f_4^{j-i+1}(\rho_n^{(3)})),
\end{aligned}
$$

and for $n^* + 1 \le j < i \le n$,

$$
\begin{aligned}
&(g_{i-2}(\rho_n), \cdots, g_1(\rho_n), 1, 0, \cdots, 0)\boldsymbol{\omega}_j^{(3)} \\
&= \sum_{k=0}^{j-1} g_{i-j-1+k}(\rho_n) f_4^k(\rho_n^{(3)}) \\
&= \frac{1}{\rho_n^{(1)} - f_4(\rho_n)} \left(\frac{\rho_n^{(1)i-j} - \rho_n^{(1)i-1}}{1 - \rho_n^{(1)} f_4(\rho_n^{(3)})} - \frac{f_4^{i-j}(\rho_n) - f_4^{i-1}(\rho_n) f_4^{j-1}(\rho_n^{(3)})}{1 - f_4(\rho_n) f_4(\rho_n^{(3)})} \right) \\
&= O(n^{1-\delta} \rho_n^{(1)i-j}).
\end{aligned}
$$

Note that the result is similar when $\boldsymbol{L}_n^{(1)}$ is replaced by $\boldsymbol{L}_n^{(1)'}$. Third, by (9.24) and (9.25), we have

$$
\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)'} \boldsymbol{\Omega}_n^{(3)'} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(1)} \end{pmatrix}, \tag{9.38}
$$

where

$$
\boldsymbol{Q}_{n-n^*}^{\dagger(1)} = O\left(n^{2(1-\delta)} \begin{pmatrix} 1 & \rho_n^{(1)} & \cdots & \rho_n^{(1)n-n^*-1} \\ \rho_n^{(1)} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_n^{(1)} \\ \rho_n^{(1)n-n^*-1} & \cdots & \rho_n^{(1)} & 1 \end{pmatrix} \right),
$$

which follows from (9.24) and (9.25) that for $n^*+1 \le i \le j \le n$,

$$
\begin{aligned}
&(g_{i-2}(\rho_n),\cdots,g_1(\rho_n),1,0,\cdots,0)(g_{j-2}(\rho_n^{(3)}),\cdots,g_1(\rho_n^{(3)}),1,0,\cdots,0)' \\
&= \sum_{k=1}^{i-1} g_{k-1}(\rho_n)g_{j-i-1+k}(\rho_n^{(3)}) \\
&= \frac{1}{(\rho_n^{(1)}-f_4(\rho_n))(\rho_n^{(1)}-f_4(\rho_n^{(3)}))} \\
&\quad \times \left(\frac{\rho_n^{(1)j-i+1}-\rho_n^{(1)i+j-2}}{1-\rho_n^{(1)2}} - \frac{\rho_n^{(1)}f_4^{j-i}(\rho_n^{(3)})}{1-\rho_n^{(1)}f_4(\rho_n^{(3)})} - \frac{\rho_n^{(1)j-i}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)} + \frac{f_4^{j-i}(\rho_n^{(3)})f_4(\rho_n)}{1-f_4(\rho_n)f_4(\rho_n^{(3)})}\right) \\
&\quad +o(\exp(-\tau n^{c/2}) \\
&= \frac{1}{(\rho_n^{(1)}-f_4(\rho_n))(\rho_n^{(1)}-f_4(\rho_n^{(3)}))} \\
&\quad \times \left(\frac{\rho_n^{(1)j-i+1}-\rho_n^{(1)i+j-2}}{1-\rho_n^{(1)2}} - \frac{\rho_n^{(1)j-i}f_4(\rho_n)}{1-\rho_n^{(1)}f_4(\rho_n)} - \frac{f_4^{j-i}(\rho_n^{(3)})(\rho_n^{(1)}-f_4(\rho_n))}{(1-\rho_n^{(1)}f_4(\rho_n^{(3)}))(1-f_4(\rho_n)f_4(\rho_n^{(3)}))}\right) \\
&\quad +o(\exp(-\tau n^{c/2}) \\
&= O(n^{2(1-\delta)}\rho_n^{(1)j-i+1}).
\end{aligned}
$$

Fourth, by (5.16),

$$
\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = O\left(n^{(1-\delta)/2}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\right). \tag{9.39}
$$

Last, by (5.25) and (5.27),

$$
\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} = o(\exp(-\tau n^c)). \tag{9.40}
$$

Now, we prove (5.33). By (5.11), we have

$$
\boldsymbol{\Sigma}_\eta^{(2)} = \boldsymbol{G}_n^{(2)-1}\boldsymbol{D}_n^{(2)}(\boldsymbol{G}_n^{(2)'})^{-1},
$$

where $\boldsymbol{G}_n^{(2)}$ and $\boldsymbol{D}_n^{(2)}$ are given in (5.8) and (9.3) with σ_η^2, κ_η and ρ_n are replaced by $\sigma_\eta^{(2)2}$, $\kappa_\eta^{(2)}$ and $\rho_n^{(2)} = \exp(-\kappa_\eta^{(2)}n^{-(1-\delta)})$, respectively. It follows together with (5.11),

$$
\begin{aligned}
&\mathrm{tr}(\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_\eta^{(2)}\boldsymbol{\Sigma}^{(3)-1}) \\
&= \mathrm{tr}(\boldsymbol{G}_n^{(1)-1}\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)'})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}\boldsymbol{D}_n^{(2)}(\boldsymbol{G}_n^{(2)'})^{-1}\boldsymbol{G}_n^{(3)'}\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}) \\
&= \sigma_\eta^{(1)2}(1-\rho_n^{(1)2})\mathrm{tr}((\boldsymbol{G}_n^{(1)'})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}\boldsymbol{D}_n^{(2)}(\boldsymbol{G}_n^{(2)'})^{-1}\boldsymbol{G}_n^{(3)'}\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}) \\
&\quad +\sigma_\eta^{(1)2}\rho_n^{(1)2}\mathrm{tr}(\boldsymbol{e}_1\boldsymbol{e}_1'(\boldsymbol{G}_n^{(1)'})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}\boldsymbol{D}_n^{(2)}(\boldsymbol{G}_n^{(2)'})^{-1}\boldsymbol{G}_n^{(3)'}\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}),
\end{aligned}
$$

where $\boldsymbol{e}_1 = (1,0,\ldots,0)'$. Therefore, for (5.33) to hold, it remains to show that

$$
\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}\boldsymbol{e}_1\boldsymbol{e}_1'(\boldsymbol{G}_n^{(2)'})^{-1}\boldsymbol{G}_n^{(3)'}\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 = O(1), \tag{9.41}
$$

$$
(1-\rho_n^{(1)2})\mathrm{tr}((\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}\boldsymbol{e}_1\boldsymbol{e}_1'(\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(2)-1})\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}) = O(n^\delta), \tag{9.42}
$$

and

$$
\begin{aligned}
&\mathrm{tr}((\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(2)-1})'\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}) \\
&= \frac{n^{(5-3\delta)/2}}{2^{5/2}(\kappa_\eta\sigma_\eta^2\kappa_\eta^{(3)}\sigma_\eta^{(3)2})^{1/2}((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{(3)2})^{1/2}+(\kappa_\eta^{(3)}\sigma_\eta^{(3)2}\sigma_\epsilon^2)^{1/2})} + O(n^{2-\delta}),
\end{aligned} \tag{9.43}
$$

For (9.41), it follows easily from (9.20). For (9.42), by (5.26), it is enough to show that

$$\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 = o(n), \tag{9.44}$$

and

$$\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1$$
$$= O(n). \tag{9.45}$$

For (9.44), by (9.22) and (9.39), we have

$$\begin{aligned} &\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\begin{pmatrix} \boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 \\ &= O\left(n^{2c}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}\right) \\ &= O(n^{2c}n^*) = O(n^{(1-\delta)/2+3c}) = o(n). \end{aligned}$$

For (9.45), we have

$$\boldsymbol{e}_1'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} = o(\exp(-\tau n^{c/2})),$$

$$\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n = O\left(\frac{1}{(1-f_4(\rho_n))^2}\mathbf{1}_n\right) = O(n^{1-\delta}\mathbf{1}_n),$$

and hence by (9.22) that,

$$\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1} = O(n^{1-\delta}\mathbf{1}_n).$$

It then follows

$$\begin{aligned} &\boldsymbol{e}_1'(\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1})'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 \\ &= (\rho_n^{(1)}-\rho_n)^2\boldsymbol{e}_1'\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1})'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\ &\quad +o(\exp(-\tau n^{c/2})) \\ &= O((\rho_n^{(1)}-\rho_n)^2 n^{2(1-\delta)}n) \\ &= O(n). \end{aligned}$$

Thus, (9.45) and hence (9.42) are obtained. Last, we prove (9.43). By (9.22),

$$
\begin{aligned}
&\mathrm{tr}((\boldsymbol{G}_n^{(1)\prime})^{-1}\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(2)-1}(\boldsymbol{G}_n^{(2)\prime})^{-1}\boldsymbol{G}_n^{(3)\prime}\boldsymbol{T}_n^{(3)-1}\boldsymbol{G}_n^{(3)}\boldsymbol{G}_n^{(1)-1}) \\
&= \mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}) + ((\rho_n^{(1)}-\rho_n^{(3)}) + (\rho_n^{(1)}-\rho_n))\mathrm{tr}(\boldsymbol{L}^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +((\rho_n^{(2)}-\rho_n^{(3)}) + (\rho_n^{(2)}-\rho_n))\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{L}^{(2)\prime}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +((\rho_n^{(1)}-\rho_n)(\rho_n^{(2)}-\rho_n) + (\rho_n^{(2)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n^{(3)}))\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +((\rho_n^{(1)}-\rho_n)(\rho_n^{(2)}-\rho_n^{(3)}) + (\rho_n^{(2)}-\rho_n)(\rho_n^{(1)}-\rho_n^{(3)}))\mathrm{tr}(\boldsymbol{L}_n^{(1)}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +(\rho_n^{(1)}-\rho_n)(\rho_n^{(1)}-\rho_n^{(3)})\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) \\
&\quad +(\rho_n^{(2)}-\rho_n)(\rho_n^{(2)}-\rho_n^{(3)})\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +(\rho_n^{(1)}-\rho_n)(\rho_n^{(2)}-\rho_n)(\rho_n^{(2)}-\rho_n^{(3)})\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}) \\
&\quad +(\rho_n^{(2)}-\rho_n)(\rho_n^{(2)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n^{(3)})\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) \\
&\quad +(\rho_n^{(2)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n)\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) \\
&\quad +(\rho_n^{(1)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n)(\rho_n^{(2)}-\rho_n)\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) \\
&\quad +(\rho_n^{(1)}-\rho_n)(\rho_n^{(2)}-\rho_n)(\rho_n^{(2)}-\rho_n^{(3)})(\rho_n^{(1)}-\rho_n^{(3)})\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}). \qquad (9.46)
\end{aligned}
$$

Therefore, (9.43) to hold, it is enough to show that

$$
\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}) = \frac{n^{(5-3\delta)/2}}{2^{5/2}(\kappa_\eta\sigma_\eta^2\kappa_\eta^{(3)}\sigma_\eta^{(3)2})^{1/2}((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{(3)2})^{1/2} + (\kappa_\eta^{(3)}\sigma_\eta^{(3)2}\sigma_\epsilon^2)^{1/2})} + O(n^{2-\delta}), \qquad (9.47)
$$

$$
\begin{aligned}
\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) &= O(n^{3-\delta}), &(9.48)\\
\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{T}_n^{(3)-1}) &= O(n^{4-3\delta}), &(9.49)\\
\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}) &= O(n^{4-3\delta}), &(9.50)\\
\mathrm{tr}(\boldsymbol{L}_n^{(1)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1}) &= O(n^{5-4\delta}), &(9.51)\\
\mathrm{tr}\left(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right) &= O(n^{6-5\delta}). &(9.52)
\end{aligned}
$$

For (9.47), it follows from (5.18). For (9.48), we have

$$
\begin{aligned}
&\mathrm{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{T}_n^{(3)-1}\boldsymbol{L}_n^{(1)}) \\
&= \mathrm{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\right) \\
&\quad + \frac{1}{f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(1)}\right) \\
&\quad + \frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\mathrm{tr}\left(\boldsymbol{\Omega}_n'\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\right) \\
&\quad + \frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(1)}\boldsymbol{\Omega}_n'\right) \\
&= \frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ * & \boldsymbol{Q}_{n-n^*}^{(3)}\end{pmatrix}\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(1)}\end{pmatrix}\right) + o(n^{3-2\delta}) \\
&= O(n^{3-2\delta}),
\end{aligned}
$$

where the first equality follows from (5.26), the second equality follows (9.36), (9.37),

$$\begin{aligned}
&\mathrm{tr}\left(\boldsymbol{L}_n^{(1)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&= O\left(\mathrm{tr}\left(\boldsymbol{L}_n^{(1)}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}' n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\right)\right)\\
&= O\left(\mathrm{tr}(\left(n^{1-\delta}n^*\mathbf{1}n^*\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\right)\right)\\
&= O(n^{1-\delta}n^{*3}) = O(n^{5(1-\delta)/2+3c}) = o(n^{3-2\delta}),
\end{aligned}$$

and by (9.39) and (9.40) that

$$\begin{aligned}
&\mathrm{tr}\left(\boldsymbol{L}_n^{(1)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\right)\\
&= O\left(\mathrm{tr}\left(\boldsymbol{L}_n^{(1)}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\right)\right)\\
&= O(n^{2-\delta}),
\end{aligned}$$

and the last equality follows from (9.36) and (9.37) that

$$\begin{aligned}
\mathrm{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{(3)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(1)}\end{pmatrix}\right) &= O\left(n^{(1-\delta)/2}n^{1-\delta}n^{(1-\delta)/2}n\right)\\
&= O(n^{3-2\delta}).
\end{aligned}$$

For (9.49), we have

$$\begin{aligned}
&\mathrm{tr}(\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{T}_n^{(3)-1})\\
&= \mathrm{tr}\left(\boldsymbol{L}_n^{(1)'}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&\quad+\frac{1}{f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\boldsymbol{L}_n^{(1)'}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{\Omega}_n^{(3)'}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\\
&\quad+\frac{1}{f_2(\rho_n)\sigma_\epsilon^2)}\mathrm{tr}\left(\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n^{(3)'}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(2)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&\quad+\frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(2)}\boldsymbol{\Omega}_n^{(3)'}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\\
&= \frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\mathrm{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(2)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(1)}\end{pmatrix}\right)+o(n^{2-\delta})\\
&= O(n^{4-3\delta}),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.37), (9.39) that

$$\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{L}_n^{(1)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&\quad= O\left(\operatorname{tr}\left(\boldsymbol{L}_n^{(1)\prime}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}\boldsymbol{L}_n^{(2)}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}\right)\right)\\
&\quad= O\left(\operatorname{tr}\left(n^{1-\delta}n^*\mathbf{1}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}n^*\mathbf{1}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}\right)\right)\\
&\quad= O(n^{(1-\delta)}n^{*4}) = O(n^{3(1-\delta)+4c}) = O(n^{4-3\delta}),
\end{aligned}$$

by (9.39) and (9.40) that

$$\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{L}_n^{(1)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\\
&\quad= O\left(\operatorname{tr}\left(\boldsymbol{L}_n^{(1)\prime}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}\boldsymbol{L}_n^{(2)}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\right)\\
&\quad= O\left(\operatorname{tr}\left(n^{*2}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}^{\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\right)\\
&\quad= O(n^{4-3\delta}),
\end{aligned}$$

and the last equality follows easily from (9.37) that

$$\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(2)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(1)}\end{pmatrix}\right) = O(n^{1-\delta}n^{1-\delta}n^{1-\delta}n) = O(n^{4-3\delta}).$$

For (9.50), we have

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{(3)-1})\\
&= \operatorname{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&\quad+\frac{1}{f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\\
&\quad+\frac{1}{f_2(\rho_n)\sigma_\epsilon^{2}}\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right)\\
&\quad+\frac{1}{f_2(\rho_n)\sigma_\epsilon^{2}f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{\Omega}_n^{\prime}\right)\\
&= \frac{1}{f_2(\rho_n)\sigma_\epsilon^{2}f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{(3)}\end{pmatrix}\right)+o(n^{2-\delta})\\
&= O(n^{4-3\delta}),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.36), (9.38), (9.39) that

$$
\begin{aligned}
&\operatorname{tr}\left(\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) \\
&\quad = O\left(\operatorname{tr}\left(n^{1-\delta} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\right)\right) \\
&\quad = O\left(\operatorname{tr}\left(n^{1-\delta} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} n^* \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' n^* \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\right)\right) \\
&\quad = O(n^{1-\delta} n^{*4}) = O(n^{4-3\delta}),
\end{aligned}
$$

by (9.39) and (9.40) that,

$$
\begin{aligned}
&\operatorname{tr}\left(\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n^{(3)\prime} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n^{(3)}\right) \\
&\quad = O\left(\operatorname{tr}\left(n^{(1-\delta)/2} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n\right)\right) \\
&\quad = O(n^{4-3\delta}),
\end{aligned}
$$

and the last equality follows easily from (9.36), (9.38) that

$$
\operatorname{tr}\left(\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{(3)} \end{pmatrix}\right) = O(n^{2(1-\delta)} n^{(1-\delta)/2} n^{(1-\delta)} n) = O(n^{4-3\delta}).
$$

For (9.51),

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{L}_n^{(1)\prime} \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{T}_n^{(3)-1}\right) \\
&= \operatorname{tr}\left(\boldsymbol{L}_n^{(1)} \begin{pmatrix} \boldsymbol{T}_{n^*}^{(-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) \\
&\quad + \frac{1}{f_2(\rho_n^{(3)}) \sigma_\epsilon^{(3)2}} \operatorname{tr}\left(\boldsymbol{L}_n^{(1)} \begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n^{(3)\prime} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n^{(3)}\right) \\
&\quad + \frac{1}{f_2(\rho_n) \sigma_\epsilon^2} \operatorname{tr}\left(\boldsymbol{L}_n^{(1)} \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) \\
&\quad + \frac{1}{f_2(\rho_n) \sigma_\epsilon^2 f_2(\rho_n^{(3)}) \sigma_\epsilon^{(3)2}} \operatorname{tr}\left(\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n^{(3)\prime} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n^{(3)} \boldsymbol{L}_n^{(1)\prime} \boldsymbol{\Omega}_n'\right) \\
&= \frac{1}{f_2(\rho_n) \sigma_\epsilon^2 f_2(\rho_n^{(3)}) \sigma_\epsilon^{(3)2}} \operatorname{tr}\left(\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{*(1)} \end{pmatrix}\right) + O(n^{5-4\delta}) \\
&= O(n^{5-4\delta}),
\end{aligned}
$$

where the first equality follows from (5.26), the second equality follows from (9.37), (9.38) and (9.39) that

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(1)} \begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) \\
&\quad = O\left(\operatorname{tr}\left(n^{1-\delta} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \boldsymbol{L}_n^{(1)} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\right)\right) \\
&\quad = O\left(\operatorname{tr}\left(n^{1-\delta} n^{*3} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\right)\right) \\
&\quad = O(n^{1-\delta} n^{*5}) = O(n^{5-4\delta}),
\end{aligned}
$$

by (9.39) and (9.40) that

$$
\begin{aligned}
& \operatorname{tr}\left(\boldsymbol{L}_n^{(1)}\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\\
&\quad = O\left(\operatorname{tr}\left(\boldsymbol{L}_n^{(1)}n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\right)\\
&\quad = O\left(\operatorname{tr}\left(n^*n^{(1-\delta)/2}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\right)\right)\\
&\quad = O(n^{5-4\delta}),
\end{aligned}
$$

and the last equality follows easily from (9.37), (9.38) that

$$
\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{*(1)}\end{pmatrix}\right) = O(n^{3(1-\delta)}n^{1-\delta}n) = O(n^{5-4\delta}).
$$

For (9.52), we have

$$
\begin{aligned}
& \operatorname{tr}\left(\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\\
&\quad = \operatorname{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\\
&\qquad +\frac{1}{f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\\
&\qquad +\frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\operatorname{tr}\left(\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\\
&\qquad +\frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\boldsymbol{\Omega}_n^{(3)\prime}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n^{(3)}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\boldsymbol{\Omega}_n'\right)\\
&\quad = \frac{1}{f_2(\rho_n)\sigma_\epsilon^2 f_2(\rho_n^{(3)})\sigma_\epsilon^{(3)2}}\operatorname{tr}\left(\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(1)}\end{pmatrix}\right)+o(n^{(5-3\delta)/2})\\
&\quad = O(n^{6-5\delta}),
\end{aligned}
$$

where the first equality follows from (5.26), the second equality follows from (9.38), (9.39) that

$$
\begin{aligned}
& \operatorname{tr}\left(\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\begin{pmatrix}\boldsymbol{T}_{n^*}^{(3)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\\
&\quad = O\left(\left(n^{1-\delta}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\boldsymbol{L}_n^{(2)}\boldsymbol{L}_n^{(2)\prime}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)\prime}\right)\right)\\
&\quad = O\left(\left(n^{1-\delta}n^{*4}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{1}_{n^*}\\ \mathbf{0}\end{pmatrix}'\right)\right)\\
&\quad = O(n^{6-5\delta}),
\end{aligned}
$$

by (9.39) that

$$
\begin{aligned}
& tr\left(\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n^{(3)\prime} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n^{(3)} \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)\prime}\right) \\
&= O\left(\operatorname{tr}\left(n^{(1-\delta)/2} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \boldsymbol{L}_n^{(2)} \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)\prime}\right)\right) \\
&= O\left(\operatorname{tr}\left(n^{(1-\delta)/2} n^{*2} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}' \boldsymbol{L}_n^{(2)\prime} \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)}\right)\right) \\
&= O\left(\operatorname{tr}\left(n^{(1-\delta)/2} n^{*4} \mathbf{1}' \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \mathbf{1}\right)\right) \\
&= O(n^{(1-\delta)/2} n^{*4} n^{1-\delta} n) = O(n^{6-5\delta}),
\end{aligned}
$$

and the last equality follows easily from (9.38) that

$$
\operatorname{tr}\left(\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(2)} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ * & \boldsymbol{Q}_{n-n^*}^{\dagger(1)} \end{pmatrix}\right) = O(n^{4(1-\delta)} n^{1-\delta} n) = O(n^{6-5\delta}).
$$

Thus, (5.33) is obtained.

Fifth, we prove (5.34). We have

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}^{(2)} \boldsymbol{\Sigma}^{(2)-1}) \\
&= \sigma_\epsilon^{(2)2} \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}^{(2)-1}) + \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_\eta^{(2)} \boldsymbol{\Sigma}^{(2)-1}).
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
& \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}^{(2)-1}) \\
&= \frac{1}{\sigma_\epsilon^{(2)2}} \left(\operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_\eta^{(2)} \boldsymbol{\Sigma}^{(2)-1})\right) \\
&= \frac{1}{\sigma_\epsilon^{(2)2}} \left(\frac{\sigma_\eta^{(1)2} \kappa_\eta^{(1)}}{(2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^2)^{1/2}} - \frac{8\sigma_\eta^{(1)2} \kappa_\eta^{(1)} \sigma_\eta^{(2)2} \kappa_\eta^{(2)}}{\sigma_\epsilon^2 \sigma_\epsilon^{(2)2} ((2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^{-2})^{1/2} + (2\kappa_\eta^{(2)} \sigma_\eta^{(2)2} \sigma_\epsilon^{(2)-2})^{1/2})^3} \right) + O(n^\delta) \\
&= \frac{\sigma_\eta^{(1)} \kappa_\eta^{(1)}}{\sigma_\epsilon^2 \sigma_\epsilon^{(2)2}} \left(\frac{1}{(2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^{-2})^{1/2}} - \frac{8\sigma_\eta^{(2)2} \kappa_\eta^{(2)}}{\sigma_\epsilon^{(2)2} ((2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^{-2})^{1/2} + (2\kappa_\eta^{(2)} \sigma_\eta^{(2)2} \sigma_\epsilon^{(2)-2})^{1/2})^3} \right) + O(n^\delta),
\end{aligned}
$$

which gives (5.34).

Finally, we prove (5.35). We have

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \operatorname{tr}(\boldsymbol{\Sigma}^{(1)} \boldsymbol{\Sigma}^{(1)-1} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&= \sigma_\epsilon^{(1)2} \operatorname{tr}(\boldsymbol{\Sigma}^{(1)} \boldsymbol{\Sigma}^{(1)-1} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}_\eta^{(1)} \boldsymbol{\Sigma}^{(1)-1} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})).
\end{aligned}
$$

It then follows that

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}^{(1)-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{1}{\sigma_\epsilon^{(1)2}}\big(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - \mathrm{tr}(\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{(1)-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))\big) \\
&= \frac{1}{\sigma_\epsilon^{(1)2}}\left(\frac{n}{\sigma_\epsilon^2} - \left(\frac{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}{2\sigma_\epsilon^2} + \frac{(2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2}}{2\sigma_\epsilon^2}\right)n^{(1+\delta)/2}\right. \\
&\quad \left. + \frac{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})(2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})}{\sigma_\epsilon^2((2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2} + (2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2})^3}n^{(1+\delta)/2}\right) + O(n^\delta) \\
&= \frac{n}{\sigma_\epsilon^{(1)2}\sigma_\epsilon^2} - \frac{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2} + (2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2}}{2\sigma_\epsilon^{(1)2}\sigma_\epsilon^2}n^{(1+\delta)/2} \\
&\quad + \frac{2(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})(2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})}{\sigma_\epsilon^{(1)2}\sigma_\epsilon^2((2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2} + (2\kappa_\eta^{(1)}\sigma_\eta^{(1)2}\sigma_\epsilon^{(1)-2})^{1/2})^3}n^{(1+\delta)/2} + O(n^\delta),
\end{aligned}$$

where the second equality follows from (5.31) and (5.34). Thus, we completes the proof. □

Proof of Lemma 9

First, we prove (5.37). By (5.8) and (5.11), we have

$$\begin{aligned}
\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} &= \mathbf{1}'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\mathbf{1} \\
&= (1-\rho_n)^2\mathbf{1}'\boldsymbol{T}_n^{-1}\mathbf{1} + 2\rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 + \rho_n^2\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1,
\end{aligned}$$

where the second equality follows from

$$\boldsymbol{G}_n\mathbf{1} = (1-\rho_n)\mathbf{1} + \rho_n\boldsymbol{e}_1, \tag{9.53}$$

with $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$. Therefore, for (5.37) to hold, it remains to show that

$$\begin{aligned}
\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 &= \frac{1}{\sigma_\eta^2} + o(1), && (9.54) \\
\rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 &= o(1), && (9.55) \\
(1-\rho_n)^2\mathbf{1}'\boldsymbol{T}_n^{-1}\mathbf{1} &= \frac{\kappa_\eta}{2\sigma_\eta^2}n^\delta + o(1). && (9.56)
\end{aligned}$$

For (9.54), by (5.14), we have

$$\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 = C_n(1,1) = \frac{f_2(\rho_n)}{(\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2} + o(1) = \frac{1}{\sigma_\eta^2} + o(1).$$

For (9.55), by (5.14) with some $c > 0$ such that $n^* = n^{(1-\delta)/2+c} < n$, we have

$$\begin{aligned}
\rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 &= 2\rho_n(1-\rho_n)\mathbf{1}'(C_n(1,1), \ldots, C_n(1,n))' \\
&= \rho_n(1-\rho_n)\frac{f_2(\rho_n)}{(\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2}\sum_{i=1}^{n-n^*}\left(\frac{\rho_n}{f_2(\rho_n)}\right)^i + o(\exp(-\tau n^{c/3})) \\
&= \rho_n(1-\rho_n)\frac{f_2(\rho_n)}{(\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n) - \rho_n^2\sigma_\epsilon^2}\frac{\rho_n}{f_2(\rho_n)-\rho_n} + o(\exp(-\tau n^{c/4})) \\
&= O(n^{-(1-\delta)}n^{(1-\delta)/2}) = o(1),
\end{aligned}$$

where the last equality follows from (5.13) and (9.11),

$$f_2(\rho_n) - \rho_n = (2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^{-2})^{1/2} n^{-(1-\delta)/2} + \kappa_\eta \sigma_\eta^2 n^{-(1-\delta)} + O(n^{-3(1-\delta)/2}).$$

For (9.56), we have

$$\begin{aligned}
\mathbf{1}'\boldsymbol{T}_n^{-1}\mathbf{1} &= \mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{1} + \frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} + o(\exp(-\tau n^{c/3})) \\
&= \frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} + o(1)) \\
&= \frac{1}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} \\
&= n^{(1-\delta)/2} n^{*2} + \frac{1}{(1-f_4(\rho_n))^2}(n-n^*) \\
&= \frac{n^{2-\delta}}{2\kappa_\eta\sigma_\eta^2} + o(n^{2(1-\delta)}),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) and the third equality follows from (5.22) that

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} = \frac{1}{1-f_4(\rho_n)}\begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{n-n^*} \end{pmatrix} + o(\exp(-\tau n^{c/3})), \tag{9.57}$$

and the last equality follows from (5.21). Thus, (9.56) and hence (5.37) are obtained.

Second, we prove (5.36). We first go a step further to show that for $k \geq 1$,

$$\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\psi}_k = \frac{\kappa_\eta}{2\sigma_\eta^2(k+1)}n^\delta + \frac{1}{2\sigma_\eta^2} + o(n^\delta). \tag{9.58}$$

For $k \geq 1$, we have

$$\begin{aligned}
\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\psi}_k &= \mathbf{1}'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k \\
&= ((1-\rho_n)\mathbf{1} + \rho_n\boldsymbol{e}_1)'\boldsymbol{T}_n^{-1}(1-\rho_n)\boldsymbol{\psi}_k + \rho_n(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) \\
&= (1-\rho_n)^2\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k + \rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) \\
&\quad + \rho_n(1-\rho_n)\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k - \rho_n^2\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b),
\end{aligned}$$

where the first equality follows from (5.11) and the second equality follows from (9.53) and

$$\boldsymbol{G}_n\boldsymbol{\psi}_k = \boldsymbol{\psi}_k - \rho_n\boldsymbol{\psi}_k^b = (1-\rho_n)\boldsymbol{\psi}_k + \rho_n(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b), \tag{9.59}$$

with $\boldsymbol{\psi}_k^b \equiv n^{-k}(0, 1^k, \ldots, (n-1)^k)$. Therefore, for (9.58) to hold, it remains to show that

$$(1-\rho_n)^2\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k = \frac{\kappa_\eta}{2\sigma_\eta^2(k+1)}n^\delta + o(n^\delta), \tag{9.60}$$

$$\rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) = \frac{1}{2\sigma_\eta^2} + o(1), \tag{9.61}$$

$$\rho_n(1-\rho_n)\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k = o(1), \tag{9.62}$$

$$\rho_n^2\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) = o(1). \tag{9.63}$$

For (9.60),

$$
\begin{aligned}
&(1-\rho_n)^2\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k \\
&= (1-\rho_n)^2\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{\psi}_k + \frac{(1-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\psi}_k + o(\exp(-\tau n^{c/2})) \\
&= \frac{(1-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*+1}^{n}(\boldsymbol{\omega}_i'\mathbf{1})(\boldsymbol{\omega}_i'\boldsymbol{\psi}_k) + o(1) \\
&= \frac{(1-\rho_n)^2}{(f_2(\rho_n)\sigma_\epsilon^2)(1-f_4(\rho_n))^2}\sum_{i=n^*+1}^{n}\left(\frac{i}{n}\right)^k + o(1) \\
&= \kappa_\eta^2 n^{-2(1-\delta)}\frac{n^{1-\delta}}{2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2}\frac{1}{\sigma_\epsilon^2}\frac{n}{k+1} + o(n^\delta) \\
&= \frac{\kappa_\eta}{2\sigma_\eta^2(k+1)}n^\delta + o(n^\delta),
\end{aligned}
$$

where the first equality follows from (5.26), the second equality follows from (9.39) that

$$(1-\rho_n)^2\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{\psi}_k = O(n^{-2(1-\delta)}n^{(1-\delta)/2}n^{*2}) = o(1),$$

ω_i is the ith column of $\boldsymbol{\Omega}_n$ and the third equality follows from

$$\boldsymbol{\omega}_i'\boldsymbol{\psi}_k = \frac{1}{1-f_4(\rho_n)}\left(\frac{i}{n}\right)^k + O(n^{-\delta}), \tag{9.64}$$

which will be proved later, and the fourth equality follows from that for $k \geq 1$,

$$\sum_{i=n^*}^{n}\left(\frac{i}{n}\right)^k = (k+1)^{-1}n + o(n). \tag{9.65}$$

We now prove (9.64). It follows from

$$
\begin{aligned}
\boldsymbol{\omega}_i'\boldsymbol{\psi}_k - f_4(\rho_n)\boldsymbol{\omega}_i'\boldsymbol{\psi}_k &= \sum_{j=1}^{i} f_4^{i-j}(\rho_n)\left(\frac{i}{n}\right)^k - f_4(\rho_n)\sum_{j=1}^{i} f_4^{i-j}(\rho_n)\left(\frac{j}{n}\right)^k \\
&= \left(\frac{i}{n}\right)^k - f_4^i(\rho_n)\left(\frac{1}{n}\right)^k + \sum_{j=1}^{i-1} f_4^{i-j}(\rho_n)\left(\left(\frac{j}{n}\right)^k - \left(\frac{j-1}{n}\right)^k\right) \\
&= \left(\frac{i}{n}\right)^k + \sum_{j=1}^{i-1} f_4^{i-j}(\rho_n)\left(\left(\frac{j}{n}\right)^k - \left(\frac{j-1}{n}\right)^k\right) + o(\exp(-\tau n^c)) \\
&= \left(\frac{i}{n}\right)^k + \frac{k}{n}\sum_{j=1}^{i-1} f_4^{i-j}(\rho_n)\left(\frac{j}{n}\right)^{k-1} + o(n^{-1}) \\
&= \left(\frac{i}{n}\right)^k + O(n^{-(1+\delta)/2}),
\end{aligned}
$$

where the third equality follows from the Taylor's expansion directly,

$$\left(\frac{j}{n}\right)^k - \left(\frac{j-1}{n}\right)^k = \frac{k}{n}\left(\frac{j}{n}\right)^{k-1} + O(n^{-2}), \tag{9.66}$$

and the second last equality follows from

$$\frac{k}{n}\sum_{j=1}^{i-1} f_4^{i-j}(\rho_n)\left(\frac{j}{n}\right)^{k-1} \leq \frac{k}{n}\sum_{j=1}^{i-1} f_4^{i-j}(\rho_n) = \frac{k}{n}\frac{f_4(\rho_n)}{1-f_4(\rho_n)} + o(\exp(-\tau n^c)) = O(n^{-(1+\delta)/2}).$$

For (9.61), we have

$$\begin{aligned}
\rho_n(1-\rho_n)\mathbf{1}'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) &= \rho_n(1-\rho_n)\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \\
&\quad + \frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n(\boldsymbol{\psi}_{k,1}-\boldsymbol{\psi}_{k,1}^b) \\
&\quad + o(\exp(-\tau n^{c/3})) \\
&= \frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*}^{n}(\boldsymbol{\omega}_i'\mathbf{1})(\boldsymbol{\omega}_i'(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)) + o(1) \\
&= \frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2(1-f_4(\rho_n))^2}\frac{k}{n}\sum_{i=n^*}^{n}\left(\frac{i}{n}\right)^{k-1} + o(1) \\
&= \kappa_\eta n^{-(1-\delta)}\frac{n^{1-\delta}}{2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2}}\frac{1}{\sigma_\epsilon^2}\frac{k}{n}\frac{n}{k} + o(1) \\
&= \frac{1}{2\sigma_\eta^2} + o(1),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) and $\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b = O(n^{-1})$ that

$$\rho_n(1-\rho_n)\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) = O(n^{-(1-\delta)}n^{(1-\delta)/2}n^{-1}n^{2*}) = o(1),$$

the third equality follows from (9.64) and (9.66) that

$$\begin{aligned}
\boldsymbol{\omega}_i'(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) &= \frac{k}{n}\boldsymbol{\omega}_j\boldsymbol{\psi}_{k-1} + o(n^{-(3+\delta)/2}) \\
&= \frac{k}{n}\frac{1}{1-f_4(\rho_n)}\left(\frac{i}{n}\right)^{k-1} + o(n^{-(1+\delta)}), \qquad (9.67)
\end{aligned}$$

and the second last equality follows from (9.65). For (9.62), by (5.14), and (9.11),

$$\begin{aligned}
\rho_n(1-\rho_n)\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k &= \rho_n(1-\rho_n)(C_n(1,1),\ldots,C_n(1,n))\boldsymbol{\psi}_k \\
&= \rho_n(1-\rho_n)\frac{f_2(\rho_n)}{(\sigma_\eta^2+\sigma_\epsilon^2)f_2(\rho_n)-\rho_n^2\sigma_\epsilon^2}\sum_{i=1}^{n-n^*} f_4^{i-1}(\rho_n)\left(\frac{i}{n}\right)^k \\
&\quad + o(\exp(-\tau n^{c/2})) \\
&= o(1),
\end{aligned}$$

where the last equality follows from

$$\sum_{i=1}^{n-n^*} f_4(\rho_n)^{i-1}\left(\frac{i}{n}\right)^k \leq \sum_{i=1}^{n-n^*} f_4(\rho_n)^{i-1} = \frac{1}{1-f_4(\rho_n)} + o(\exp(-\tau n^c)).$$

For (9.63), we have

$$\rho_n^2 \boldsymbol{e}_1' \boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) = \rho_n^2 (C_n(1,1), \ldots, C_n(1,n))(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b) = O(n^{-(1+\delta)/2}).$$

Hence, (9.58) is obtained. It now remains to show that (5.36) for $k, \ell \geq 1$. For $k, \ell \geq 1$, we have

$$\begin{aligned}
\boldsymbol{\psi}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\psi}_\ell &= \boldsymbol{\psi}_k' \boldsymbol{G}_n' \boldsymbol{T}_n^{-1} \boldsymbol{G}_n \boldsymbol{\psi} \\
&= ((1-\rho_n)\boldsymbol{\psi}_k + \rho_n(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b))' \boldsymbol{T}_n^{-1} ((1-\rho_n)\boldsymbol{\psi}_\ell + \rho_n(\boldsymbol{\psi}_\ell - \boldsymbol{\psi}_\ell^b)) \\
&= (1-\rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_\ell + \rho_n(1-\rho_n) \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_\ell - \boldsymbol{\psi}_\ell^b) \\
&\quad + \rho_n(1-\rho_n)(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b)' \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_\ell + \rho_n^2 (\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b)' \boldsymbol{T}_n^{-1} (\boldsymbol{\psi}_\ell - \boldsymbol{\psi}_\ell^b),
\end{aligned}$$

where the first equality follows from (5.11) and the second equality follows from (9.59). Therefore, for (5.36) to hold, it remains to show that

$$(1-\rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_\ell = \frac{\kappa_\eta}{2\sigma_\eta^2 (k+\ell+1)} n^\delta + o(n^\delta), \tag{9.68}$$

$$\rho_n(1-\rho_n) \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_\ell - \boldsymbol{\psi}_\ell^b) = \frac{1}{2\sigma_\epsilon^2} \frac{k}{k+\ell} + o(1), \tag{9.69}$$

$$\rho_n^2 (\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^b)' \boldsymbol{T}_n^{-1} (\boldsymbol{\psi}_\ell - \boldsymbol{\psi}_\ell^b) = \frac{k\ell n^{-\delta}}{2\kappa_\eta \sigma_\eta^2 (k+\ell-1)} + o(1). \tag{9.70}$$

For (9.68), we have

$$\begin{aligned}
(1-\rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_\ell &= (1-\rho_n)^2 \boldsymbol{\psi}_k' \begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\psi}_\ell + \frac{(1-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2} \boldsymbol{\psi}_k' \boldsymbol{\Omega}_n' \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{\psi}_\ell \\
&\quad + o(\exp(-\tau n^{c/3})) \\
&= \frac{(1-\rho_n)^2}{f_2(\rho_n)\sigma_\epsilon^2} \sum_{j=n^*+1}^{n} (\boldsymbol{\omega}_j' \boldsymbol{\psi}_k)(\boldsymbol{\omega}_j' \boldsymbol{\psi}_\ell) + o(1) \\
&= \frac{(1-\rho_n)^2}{(f_2(\rho_n)\sigma_\epsilon^2)(1-f_4(\rho_n))^2} \sum_{i=n^*+1}^{n} \left(\frac{i}{n}\right)^{k+\ell} + o(1) \\
&= \kappa_\eta^2 n^{2\delta-2} \frac{n^{1-\delta}}{2\kappa_\eta \sigma_\eta^2 \sigma_\epsilon^{-2}} \frac{1}{\sigma_\epsilon^2} \frac{n}{k+\ell+1} + o(n^\delta) \\
&= \frac{\kappa_\eta}{2\sigma_\eta^2 (k+\ell+1)} n^\delta + o(n^\delta),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) that

$$(1-\rho_n)^2 \boldsymbol{\psi}_k' \begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\psi}_\ell = O(n^{-2(1-\delta)} n^{-(1-\delta)/2} n^{*2}) = o(1),$$

the third equality follows from (9.64), and the fourth equality follows from (5.22), (9.11) and (9.65). For (9.69), we have

$$\begin{aligned}
&\rho_n(1-\rho_n)\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_\ell-\boldsymbol{\psi}_\ell^b)\\
&= \rho_n(1-\rho_n)\boldsymbol{\psi}_k'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)+\frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\boldsymbol{\psi}_k'\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\psi}_\ell\\
&\quad +o(\exp(-\tau n^{c/3}))\\
&= \frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*+1}^{n}(\boldsymbol{\omega}_i'\boldsymbol{\psi}_k)(\boldsymbol{\omega}_i'(\boldsymbol{\psi}_\ell-\boldsymbol{\psi}_\ell^b))+o(1)\\
&= \frac{\rho_n(1-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2(1-f_4(\rho_n))^2}\frac{k}{n}\sum_{i=n^*+1}^{n}\left(\frac{i}{n}\right)^{\ell+k-1}+o(1)\\
&= \kappa_\eta n^{-(1-\delta)}\frac{1}{\sigma_\epsilon^2}\frac{n^{1-\delta}}{2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2}}\frac{k}{n}\frac{n}{k+\ell}+o(1)\\
&= \frac{1}{2\sigma_\epsilon^2}\frac{k}{k+\ell}+o(1),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) and $\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b=O(n^{-1})$ that

$$\rho_n(1-\rho_n)\boldsymbol{\psi}_k'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{-(1-\delta)}n^{-(1-\delta)/2}n^{-1}n^{*2})=o(1),$$

the third equality follows from (9.64), (9.67) and the fourth equality follows from (5.22), (9.11) and (9.65). For (9.70), we have

$$\begin{aligned}
&\rho_n^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_\ell-\boldsymbol{\psi}_\ell^b)\\
&= \rho_n^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)\\
&\quad +\frac{\rho_n^2}{f_2(\rho_n)\sigma_\epsilon^2}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)+o(\exp(-\tau n^{c/3}))\\
&= \frac{\rho_n^2}{f_2(\rho_n)\sigma_\epsilon^2}\sum_{i=n^*}^{n}(\boldsymbol{\omega}_i'(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))(\boldsymbol{\omega}_i'(\boldsymbol{\psi}_\ell-\boldsymbol{\psi}_\ell^b))+o(1)\\
&= \frac{\rho_n^2}{f_2(\rho_n)\sigma_\epsilon^2}\frac{k\ell}{n^2}\sum_{i=n^*+1}^{n}\left(\frac{i}{n}\right)^{k+\ell-2}+o(1)\\
&= n^{-2}\frac{n^{1-\delta}}{2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2}}\frac{1}{\sigma_\epsilon^2}\frac{nk\ell}{k+\ell-1}+o(1)\\
&= \frac{k\ell n^{-\delta}}{2\kappa_\eta\sigma_\eta^2(k+\ell-1)}+o(1),
\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) and $\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b=O(m^{-1})$ that

$$\rho_n^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{-2}n^{(1-\delta)/2}n^{*2})=o(1),$$

the third equality follows from (9.67), and fourth equality follows from (5.22), (9.11) and (9.65). Hence, (5.36) is obtained.

Finally, we prove (5.38). It follows easily that

$$\boldsymbol{\psi}_k'\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})\boldsymbol{\psi}_k = O(\boldsymbol{\psi}_k'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\psi}_k).$$

It then suffices to show that

$$\boldsymbol{\psi}_k\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\psi}_\ell = O(n^\delta). \tag{9.71}$$

For $k \geq 0$, we have

$$\begin{aligned}\boldsymbol{\psi}_k'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_\eta^{(1)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\psi}_k &= \boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k \\ &= \sigma_\eta^{(1)2}(1-\rho_n^{(1)2})\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k \\ &\quad +\sigma_\eta^{(1)2}\rho_n^{(1)2}\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1\boldsymbol{e}_1'(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k,\end{aligned}$$

where the first equality follows from (5.11) and

$$\boldsymbol{\Sigma}_\eta^{(1)} = \boldsymbol{G}_n^{(1)-1}\boldsymbol{D}_n^{(1)}(\boldsymbol{G}_n^{(1)-1})',$$

$\boldsymbol{G}_n^{(1)}$ and $\boldsymbol{D}_n^{(1)}$ are given in (5.8) and (9.3) with σ_η^2, κ_η and ρ_n are replaced by $\sigma_\eta^{(1)2}$, $\kappa_\eta^{(1)}$ and $\rho_n^{(1)} = \exp(-\kappa_\eta^{(1)}n^{-(1-\delta)})$, respectively. Therefore, for (9.71) to hold, it remains to show that

$$\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 = O(1), \tag{9.72}$$

$$(1-\rho_n^{(1)2})\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k = O(n^\delta). \tag{9.73}$$

For (9.72), by (9.22) and (9.59), we have

$$\begin{aligned}\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}\boldsymbol{e}_1 &= ((1-\rho_n)\boldsymbol{\psi}_k+\rho_n(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))'\boldsymbol{T}_n^{-1}(\boldsymbol{e}_1+(\rho_n^{(1)}-\rho_n)\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1) \\ &= (1-\rho_n)\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1+\rho_n(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))'\boldsymbol{T}_n^{-1}\boldsymbol{e}_1 \\ &\quad +(1-\rho_n)(\rho_n^{(1)}-\rho_n)\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\ &\quad +\rho_n(\rho_n^{(1)}-\rho_n)(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\ &= (1-\rho_n)(\rho_n^{(1)}-\rho_n)\boldsymbol{1}'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1+o(1),\end{aligned}$$

where the last equality follows from (9.62), (9.63) and by (9.55) that

$$(\rho_n^{(1)}-\rho_n)\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \leq (\rho_n^{(1)}-\rho_n)\boldsymbol{e}_1'\boldsymbol{T}_n^{-1}\boldsymbol{1} = o(1).$$

Therefore, for (9.72) to hold, it is enough to show that

$$\begin{aligned}&(1-\rho_n)(\rho_n^{(1)}-\rho_n)\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\ &= (1-\rho_n)(\rho_n^{(1)}-\rho_n)\boldsymbol{\psi}_k'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \\ &\quad +\frac{(1-\rho_n)(\rho_n^{(1)}-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\boldsymbol{\psi}_k'\boldsymbol{\Omega}_n'\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1+o(\exp(-\tau n^{c/3})) \\ &= \frac{(1-\rho_n)(\rho_n^{(1)}-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2}\boldsymbol{\psi}_k'\boldsymbol{\Omega}_n'\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1+o(1) \\ &= \frac{(1-\rho_n)(\rho_n^{(1)}-\rho_n)}{f_2(\rho_n)\sigma_\epsilon^2(1-f_4(\rho_n))}\sum_{i=n^*}^{n}\left(\frac{i}{n}\right)^k g_i(\rho_n)+o(1) \\ &= O\left(n^{-2(1-\delta)}n^{(1-\delta)/2}\sum_{i=n^*}^{n}g_i(\rho_n)\right) \\ &= O(1),\end{aligned}$$

where the first equality follows from (5.26), the second equality follows from (9.39) that

$$(1-\rho_n)(\rho_n^{(1)}-\rho_n)\boldsymbol{\psi}_k'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{e}_1 \;=\; O(n^{-2(1-\delta)}n^{(1-\delta)/2}n^{*2})=o(1),$$

the third equality follows from (9.64) and (9.24), and the last equality follows from (9.25) that

$$\sum_{i=n^*}^{n} g_i(\rho_n)=O(n^{3(1-\delta)/2}).$$

Hence, (9.72) is obtained. For (9.73), we have

$$\begin{aligned}
&(1-\rho_n^{(1)2})\boldsymbol{\psi}_k'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{\psi}_k\\
&\quad= (1-\rho_n^{(1)2})((1-\rho_n)\boldsymbol{\psi}_k+\rho_n(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\\
&\qquad\times((1-\rho_n)\boldsymbol{\psi}_k+\rho_n(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b))\\
&\quad= (1-\rho_n^{(1)2})(1-\rho_n)^2\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k\\
&\qquad+2\rho_n(1-\rho_n^{(1)2})(1-\rho_n)\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)\\
&\qquad+\rho_n^2(1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b).
\end{aligned}$$

Therefore, for (9.73) to hold, it suffices to show that

$$(1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{\delta}), \tag{9.74}$$

$$(1-\rho_n^{(1)2})(1-\rho_n)^2\boldsymbol{\psi}_k'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_k \;=\; O(n^{\delta}). \tag{9.75}$$

For (9.74), by (9.22), we have

$$\begin{aligned}
&(1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{G}_n\boldsymbol{G}_n^{(1)-1}(\boldsymbol{G}_n^{(1)-1})'\boldsymbol{G}_n'\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)\\
&\quad= (1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-2}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)\\
&\qquad+2(1-\rho_n^{(1)2})(\rho_n^{(1)}-\rho_n)(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)\\
&\qquad+(1-\rho_n^{(1)2})(\rho_n^{(1)}-\rho_n)^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b).
\end{aligned}$$

Therefore, for (9.74) to hold, it remains to show that

$$(1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-2}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{\delta}), \tag{9.76}$$

$$(1-\rho_n^{(1)2})(\rho_n^{(1)}-\rho_n)^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{\delta}). \tag{9.77}$$

For (9.76), by (9.66), we have

$$(1-\rho_n^{(1)2})(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-2}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \;=\; O(n^{-(3-\delta)}\boldsymbol{\psi}_{k-1}'\boldsymbol{T}_n^{-2}\boldsymbol{\psi}_{k-1})=O(n^{-(3-\delta)}\mathbf{1}'\boldsymbol{T}_n^{-2}\mathbf{1}).$$

Hence, for (9.76) to hold, it is enough to show that

$$\begin{aligned}
\mathbf{1}'\boldsymbol{T}_n^{-2}\mathbf{1} \;&=\; (1-\rho_n^{(1)2})(1-\rho_n)^2\mathbf{1}'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{1}\\
&\quad+\frac{2}{f_4(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\begin{pmatrix}\boldsymbol{T}_{n^*}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1}\\
&\quad+\frac{1}{(f_4(\rho_n)\sigma_\epsilon^2)^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\Omega}_n'\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-n^*}\end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1}+o(\exp(-\tau n^{c/3}))\\
&=\; O(n^{3-2\delta}),
\end{aligned} \tag{9.78}$$

where the first equality follows from (5.26) and the last equality follows from (9.39) that

$$\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{1} = O(n^{1-\delta}n^{*2}) = O(n^{2(1-\delta+c)}),$$

and by (9.57), (9.36) with $\rho_n^{(1)} = \rho_n$ that,

$$\begin{aligned}
&\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} \\
&= \frac{1}{(f_4(\rho_n)\sigma_\epsilon^2)^2(1-f_4(\rho_n))^2}\mathbf{1}'\boldsymbol{Q}_{n-n^*}^{(3)}\mathbf{1} + o(1) \\
&= O\left(n^{1-\delta}\frac{1}{1-f_4^2(\rho_n)}\left(2\sum_{i=0}^{n-n^*}(n-n^*-i)f_4^i(\rho_n) - (n-n^*)\right)\right) \\
&= O\left(n^{1-\delta}\frac{1}{1-f_4^2(\rho_n)}\left(\frac{n-n^*}{1-f_4(\rho_n)} - \frac{f_4(\rho_n)}{(1-f_4(\rho_n))^2}\right)\right) \\
&= O(n^{3-2\delta}).
\end{aligned}$$

Thus, (9.76) is obtained. For (9.77), by (9.66), we have

$$\begin{aligned}
&(1-\rho_n^{(1)2})(\rho_n^{(1)}-\rho_n)^2(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b)'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}(\boldsymbol{\psi}_k-\boldsymbol{\psi}_k^b) \\
&= O(n^{-3(1-\delta)}n^{-2}\boldsymbol{\psi}_{k-1}'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}\boldsymbol{\psi}_{k-1}) \\
&= O(n^{-(5-3\delta)}\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}\mathbf{1}).
\end{aligned}$$

Thus, for (9.77) to hold, it is enough to show that

$$\begin{aligned}
&\mathbf{1}'\boldsymbol{T}_n^{-1}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{T}_n^{-1}\mathbf{1} \\
&= \mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{1} \\
&\quad + \frac{2}{f_2(\rho_n)\sigma_\epsilon^2}\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} \\
&\quad + \frac{2}{(f_2(\rho_n)\sigma_\epsilon^2)^2}\mathbf{1}'\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\boldsymbol{\Omega}_n'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix}\boldsymbol{\Omega}_n\mathbf{1} + o(\exp(-\tau n^{c/3})) \\
&= O(n^{5-4\delta}), \qquad (9.79)
\end{aligned}$$

where the first equality follows from(5.26) and the last equality follows from (9.39) and (9.23) that

$$\begin{aligned}
&\mathbf{1}'\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\begin{pmatrix} \boldsymbol{T}_{n^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{1} \\
&= O\left(n^{(1-\delta)}\mathbf{1}'\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\boldsymbol{L}_n^{(1)}\boldsymbol{L}_n^{(1)'}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{1}_{n^*} \\ \mathbf{0} \end{pmatrix}'\mathbf{1}\right) \\
&= O(n^{(1-\delta)}n^{*2}(n^*, n^*-1, \cdots, 1)'(n^*, n^*-1, \cdots, 1)) \\
&= O(n^{(1-\delta)}n^{*5}) = O(n^{5-4\delta}),
\end{aligned}$$

and by (9.57) and (9.38) with $\rho_n^{(3)} = \rho_n$ that,

$$
\begin{aligned}
& \mathbf{1}'\boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)'} \boldsymbol{\Omega}_n' \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-n^*} \end{pmatrix} \boldsymbol{\Omega}_n \mathbf{1} \\
= \; & O\Big(\frac{1}{(1 - f_4(\rho_n))^2} \mathbf{1} \boldsymbol{Q}_{n-n^*}^{\dagger(1)} \mathbf{1} \Big) \\
= \; & O\Big(n^{3(1-\delta)} \Big(2 \sum_{i=0}^{n-n^*} (n - n^* - i)\rho_n^i - (n - n^*) \Big) \Big) \\
= \; & O\Big(n^{3(1-\delta)} \Big(\frac{n - n^*}{1 - \rho_n} - \frac{1}{(1 - \rho_n)^2} \Big) \Big) = O(n^{5-4\delta}).
\end{aligned}
$$

Thus, (9.74) is obtained. For (9.75), we have

$$
\begin{aligned}
& (1 - \rho_n^{(1)2})(1 - \rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{G}_n \boldsymbol{G}_n^{(1)-1} (\boldsymbol{G}_n^{(1)-1})' \boldsymbol{G}_n' \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_k \\
= \; & (1 - \rho_n^{(1)2})(1 - \rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-2} \boldsymbol{\psi}_k + 2(1 - \rho_n^{(1)2})(1 - \rho_n)^2 (\rho_n^{(1)} - \rho_n) \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)} \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_k \\
& + (1 - \rho_n^{(1)2})(1 - \rho_n)^2 (\rho_n^{(1)} - \rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_k \\
= \; & O(n^{\delta}),
\end{aligned}
$$

where the first equality follows from (9.22) and the last equality follows from (9.78) that

$$
(1 - \rho_n^{(1)2})(1 - \rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-2} \boldsymbol{\psi}_k = O(n^{-3(1-\delta)} \mathbf{1}' \boldsymbol{T}_n^{-2} \mathbf{1}) = O(n^{-3(1-\delta)} n^{3-2\delta}) = O(n^{\delta}),
$$

and by (9.79) that

$$
\begin{aligned}
(1 - \rho_n^{(1)2})(1 - \rho_n)^2 (\rho_n^{(1)} - \rho_n)^2 \boldsymbol{\psi}_k' \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \boldsymbol{\psi}_k &= O(n^{-5(1-\delta)} \mathbf{1}' \boldsymbol{T}_n^{-1} \boldsymbol{L}_n^{(1)} \boldsymbol{L}_n^{(1)'} \boldsymbol{T}_n^{-1} \mathbf{1}) \\
&= O(n^{-5(1-\delta)} n^{5-4\delta}) = O(n^{\delta}).
\end{aligned}
$$

Thus (9.75) and hence (9.71). This completes the proof. □

Proof of Lemma 10

For (5.39) and (5.40), it follows by applying Taylor expansions on the lefthand sides of (5.39) and (5.40) at $\sigma_\epsilon^2 = \sigma^2$. Similarly, (5.41) follows by applying a Taylor expansion on the lefthand side of (5.41) at $\kappa_\eta \sigma_\eta^2 = \tau$.

For (5.42), we have

$$
\begin{aligned}
& \operatorname{tr}(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{j'} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)') \boldsymbol{\Sigma}_{j'} \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)')) \\
& \quad - \operatorname{tr}(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{j'} \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)')) - \operatorname{tr}(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}^{-1}((\sigma_\eta^2, \kappa_\eta, \sigma^2)') \boldsymbol{\Sigma}_{j'} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
= \; & \left(\frac{\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{\sigma_\epsilon (2\kappa_\eta \sigma_\eta^2)^{3/2}} + \frac{\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{\sigma (2\kappa_\eta \sigma_\eta^2)^{3/2}} - \frac{4\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{(\sigma_\epsilon + \sigma)(2\kappa_\eta \sigma_\eta^2)^{3/2}} \right) n^{(1+\delta)/2} + O(n^{\delta}) \\
= \; & \frac{\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{2^{3/2} (\kappa_\eta \sigma_\eta)^{3/2}} \left(\frac{(\sigma_\epsilon - \sigma)^2}{\sigma_\epsilon \sigma (\sigma_\epsilon + \sigma)} \right) n^{(1+\delta)/2} + O(n^{\delta}) \\
= \; & \frac{\kappa_j \sigma_j^2 \kappa_{j'} \sigma_{j'}^2}{2^{9/2} \tau^{3/2} \sigma^3} (\sigma_\epsilon^2 - \sigma^2)^2 n^{(1+\delta)/2} + o((\sigma_\epsilon^2 - \sigma^2)^2 n^{(1+\delta)/2}) + O(n^{\delta}),
\end{aligned}
$$

where the second equality follows from (5.33) and the last equality follows from $|\sigma_\epsilon^2 - \sigma^2| = o(1)$ and $|\kappa_\eta \sigma_\eta^2 - \tau| = o(1)$.

For (5.43), we have

$$
\begin{aligned}
&\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-2}((\sigma_\eta^2,\kappa_\eta,\sigma^2)')) - \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma^2)')) \\
&\qquad -\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma^2)')\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad = \left(\frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma_\epsilon^3(\kappa_\eta\sigma_\eta^2)^{1/2}} + \frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma^3(\kappa_\eta\sigma_\eta^2)^{1/2}} - \frac{2\kappa_j\sigma_j^2}{2^{1/2}\sigma_\epsilon\sigma(\sigma_\epsilon+\sigma)(\kappa_\eta\sigma_\eta^2)^{1/2}}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{\kappa_j\sigma_j^2}{2^{3/2}(\kappa_\eta\sigma_\eta^2)^{1/2}}\left(\frac{(\sigma_\epsilon^2-\sigma^2)^2+\sigma_\epsilon\sigma(\sigma_\epsilon-\sigma)^2}{\sigma_\epsilon^3\sigma^3(\sigma_\epsilon+\sigma)}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{5\kappa_j\sigma_j^2}{2^{9/2}\tau^{1/2}\sigma^7}(\sigma_\epsilon^2-\sigma^2)^2 n^{(1+\delta)/2} + o((\sigma_\epsilon^2-\sigma^2)^2 n^{(1+\delta)/2}) + O(n^\delta),
\end{aligned}
$$

where the second equality follows from (5.34) and the last equality follows from $|\sigma_\epsilon^2-\sigma^2| = o(1)$ and $|\kappa_\eta\sigma_\eta^2-\tau| = o(1)$.

For (5.44), we have

$$
\begin{aligned}
&\operatorname{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}^{-2}((\sigma_\eta^2,\kappa_\eta,\sigma^2)')) - 2\operatorname{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma^2)')) \\
&\quad = \left(\frac{n}{\sigma_\epsilon^4} - \frac{3(2\kappa_\eta\sigma_\eta^2)^{1/2}n^{(1+\delta)/2}}{4\sigma_\epsilon}\right) + \left(\frac{n}{\sigma^4} - \frac{3(2\kappa_\eta\sigma_\eta^2)^{1/2}n^{(1+\delta)/2}}{4\sigma}\right) \\
&\qquad -2\left(\frac{n}{\sigma_\epsilon^2\sigma^2} - \frac{2(2\kappa_\eta\sigma_\eta^2)^{1/2}(\sigma_\epsilon^2+\sigma^2+\sigma_\epsilon\sigma)n^{(1+\delta)/2}}{2\sigma_\epsilon\sigma(\sigma_\epsilon+\sigma)}\right) + O(n^\delta) \\
&\quad = \frac{(\sigma_\epsilon^2-\sigma^2)^2}{\sigma_\epsilon^4\sigma^4}n + \frac{(2\kappa_\eta\sigma_\eta^2)^{1/2}}{4}\left(\frac{(\sigma_\epsilon-\sigma)^2}{\sigma_\epsilon\sigma(\sigma_\epsilon+\sigma)}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{1}{\sigma^8}(\sigma_\epsilon^2-\sigma^2)^2 n + o((\sigma_\epsilon^2-\sigma^2)^2 n) + o((\sigma_\epsilon^2-\sigma^2)n^{(1+\delta)/2}) + O(n^\delta),
\end{aligned}
$$

where the second equality follows from (5.35) and the last equality follows from $|\sigma_\epsilon^2-\sigma^2| = o(1)$.

For (5.45), we have

$$
\begin{aligned}
&\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')) \\
&\qquad -\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')) - \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)'))\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad = \left(\frac{2\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2\sigma_\epsilon(2\kappa_\eta\sigma_\eta^2)^{3/2}} + \frac{2\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2\sigma(2\tau)^{3/2}} - \frac{4\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{\sigma_\epsilon\sigma(2\kappa_\eta\sigma_\eta^2)^{1/2}(2\tau)^{1/2}((2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}+(2\tau\sigma^{-2})^{1/2})}\right) \\
&\qquad \times n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2^{3/2}\sigma}\left(\frac{1}{(\kappa_\eta\sigma_\eta^2)^{3/2}} + \frac{1}{\tau^{3/2}} - \frac{4}{(\kappa_\eta\sigma_\eta^2)^{1/2}\tau^{1/2}((\kappa_\eta\sigma_\eta^2)^{1/2}+\tau^{1/2})}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2^{3/2}\sigma}\left(\frac{(\kappa_\eta\sigma_\eta-\tau)^2+(\kappa_\eta\sigma_\eta)^{1/2}\tau^{1/2}((\kappa_\eta\sigma_\eta)^{1/2}-\tau^{1/2})^2}{(\kappa_\eta\sigma_\eta)^{3/2}\tau^{3/2}((\kappa_\eta\sigma_\eta)^{1/2}+\tau^{1/2})}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&\quad = \frac{5\kappa_j\sigma_j^2\kappa_{j'}\sigma_{j'}^2}{2^{9/2}\sigma\tau^{7/2}}(\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2}) + O(n^\delta),
\end{aligned}
$$

where the second equality follows from (5.33), the third equality follows from $|\sigma_\epsilon^2-\sigma^2| = o(n^{-(1-\delta)/2})$, and the last equality follows from $|\kappa_\eta\sigma_\eta^2-\tau| = o(1)$.

For (5.46), we have

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-2}((c,d,\sigma^2)')) - \operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')) \\
&\quad -\operatorname{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&= \left(\frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma_\epsilon^3(\kappa_\eta\sigma_\eta^2)^{1/2}} + \frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma^3(\tau)^{1/2}} - \frac{2\kappa_j\sigma_j^2}{2^{1/2}\sigma_\epsilon^2\sigma^2((\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}+(\tau\sigma^{-2})^{1/2})}\right)n^{(1+\delta)/2} \\
&\quad +O(n^\delta) \\
&= \frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma^3}\left(\frac{1}{(\kappa_\eta\sigma_\eta^2)^{1/2}} + \frac{1}{\tau^{1/2}} - \frac{4}{((\kappa_\eta\sigma_\eta^2)^{1/2}+\tau^{1/2})}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&= \frac{\kappa_j\sigma_j^2}{2^{3/2}\sigma^3}\left(\frac{((\kappa_\eta\sigma_\eta^2)^{1/2}-\tau^{1/2})^2}{(\kappa_\eta\sigma_\eta^2)^{1/2}\tau^{1/2}((\kappa_\eta\sigma_\eta^2)^{1/2}+\tau^{1/2})}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&= \frac{\kappa_j\sigma_j^2}{2^{9/2}\sigma^3\tau^{5/2}}(\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2}) + O(n^\delta),
\end{aligned}$$

where the second equality follows from (5.34), the third equality follows from $|\sigma_\epsilon^2 - \sigma^2| = o(n^{-(1-\delta)/2})$ and the last equality follows from $|\kappa_\eta\sigma_\eta^2 - \tau| = o(1)$.

For (5.47), we have

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \operatorname{tr}(\boldsymbol{\Sigma}^{-2}((c,d,\sigma^2)')) - 2\operatorname{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}((c,d,\sigma^2)')) \\
&= \left(\frac{n}{\sigma_\epsilon^4} - \frac{3(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})}{4\sigma_\epsilon^4(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}n^{(1+\delta)/2}\right) + \left(\frac{n}{\sigma^4} - \frac{3(2\tau\sigma^{-2})}{4\sigma^4(2\tau\sigma^{-2})^{1/2}}n^{(1+\delta)/2}\right) \\
&\quad -2\left(\frac{n}{\sigma_\epsilon^2\sigma^2} - \frac{2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2} + (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}(2\tau\sigma^{-2})^{1/2} + 2\tau\sigma^{-2}}{2\sigma_\epsilon^2\sigma^2((2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2} + (2\tau\sigma^{-2})^{1/2})}n^{(1+\delta)/2}\right) + O(n^\delta) \\
&= -\frac{1}{2^{3/2}\sigma}\left(3(\kappa_\eta\sigma_\eta^2)^{1/2} + 3\tau^{1/2} - \frac{4(\kappa_\eta\sigma_\eta^2 + (\kappa_\eta\sigma_\eta^2)^{1/2}\tau^{1/2} + \tau)}{(\kappa_\eta\sigma_\eta^2)^{1/2} + \tau^{1/2}}\right)n^{(1+\delta)/2} + O(n^\delta) \\
&= \frac{1}{2^{3/2}\sigma}\frac{((\kappa_\eta\sigma_\eta^2)^{1/2} - \tau^{1/2})^2}{(\kappa_\eta\sigma_\eta^2)^{1/2} + \tau^{1/2}}n^{(1+\delta)/2} + O(n^\delta) \\
&= \frac{1}{2^{9/2}\tau^{5/2}}(\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2} + o((\kappa_\eta\sigma_\eta^2-\tau)^2 n^{(1+\delta)/2}) + O(n^\delta),
\end{aligned}$$

where the second equality follows from (5.35), the third equality follows from $|\sigma_\epsilon^2 - \sigma^2| = o(n^{-(1-\delta)/2})$, and the last equality follows from $|\kappa_\eta\sigma_\eta^2 - \tau| = o(1)$. □

Proof of Proposition 4

We first prove (5.49) and (5.50). Let $\boldsymbol{X}_j$ be the jth column of $\boldsymbol{X}$. For $\delta \in (0,1)$, by (5.36) and (5.37),

$$\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_{j'} = \frac{\kappa_\eta}{2\sigma_\eta^2(j+j'+1)}n^\delta + o(n^\delta); \quad j, j' = 1, \ldots, p, \tag{9.80}$$

whereas for $\delta = 0$, by (5.38),

$$\boldsymbol{X}_j'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_{j'} = O(1); \quad j, j' = 1, \ldots, p. \tag{9.81}$$

Hence, for $\boldsymbol{A}(\alpha;\boldsymbol{\theta})$ defined in (3.6), and $\delta \in (0,1)$, we obtain

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n^\delta}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} \\
&= \lim_{n\to\infty} \frac{1}{n^\delta}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} \\
&= \lim_{n\to\infty} \frac{1}{n^\delta}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\boldsymbol{\beta} \\
&\quad - \lim_{n\to\infty} \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha)}{n^\delta}\left(\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\alpha)}{n^\delta}\right)^{-1}\frac{\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\boldsymbol{\beta}}{n^\delta} \\
&= \gamma(\alpha),
\end{aligned}$$

where $\gamma(\alpha)$ is defined in (5.51) and the last equality follows from (9.80). It then gives (5.49). Similarly, it also gives (5.50) by (9.81).

Next, we prove (5.52). By (4.9), we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{Z}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{Z} \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} \\
&\quad -2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad +\xi(\boldsymbol{\theta}),
\end{aligned} \tag{9.82}$$

where the last equality follows from (5.53). In addition, by (5.29),

$$\begin{aligned}
\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) &= n\log\sigma_\epsilon^2 - \frac{1-\delta}{2}\log n + (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{(1+\delta)/2} - (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})n^\delta \\
&\quad +(\sigma_\eta^2-\sigma_\epsilon^2)\sigma_\epsilon^{-2}\kappa_\eta n^\delta + o(n^\delta) + O(1),
\end{aligned}$$

by (5.32),

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}n + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(-\frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\sigma_{\eta,0}^2\kappa_{\eta,0}}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\
&\quad +\frac{\sigma_{\eta,0}^2\kappa_{\eta,0}(\kappa_\eta-\kappa_{\eta,0})}{\kappa_\eta\sigma_\eta^2}n^\delta + \frac{\sigma_{\eta,0}^2(\kappa_\eta-\kappa_{\eta,0})^2}{2\kappa_\eta\sigma_\eta^2}n^\delta + o(n^\delta) + O(1),
\end{aligned}$$

and by (5.49),

$$\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} = \frac{\kappa_\eta}{2\sigma_\eta^2}\gamma(\alpha)n^\delta + o(n^\delta).$$

It then follows together with (9.82), for $\delta \in (0,1)$,

$$
\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\
&\quad + \left(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\left(\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\kappa_\eta\sigma_\eta^2} - 1\right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^\delta \\
&\quad + \frac{\sigma_{\eta,0}^2 + \gamma(\alpha)}{2\kappa_\eta\sigma_\eta^2}\left(\kappa_\eta - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\eta,0}^2 + \gamma(\alpha)}\right)^2 n^\delta + \frac{\kappa_{\eta,0}\sigma_{\epsilon,0}^2}{2\kappa_\eta\sigma_\eta^2}\left(\kappa_{\eta,0} - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\eta,0}^2 + \gamma(\alpha)}\right)n^\delta \\
&\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad + \xi(\boldsymbol{\theta}) + o_p(n^\delta) + O(1). \qquad (9.83)
\end{aligned}
$$

Therefore, for (5.52) to hold, it remains to show that for $\delta \in [0,1)$,

$$
\begin{aligned}
(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= O_p(1), \qquad (9.84) \\
\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= o_p(n^\delta), \qquad (9.85)
\end{aligned}
$$

which are enough to show

$$
\begin{aligned}
n^\delta(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X})^{-1} &= O(1), \qquad (9.86) \\
\frac{1}{n^{\delta/2}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) &= O_p(1). \qquad (9.87)
\end{aligned}
$$

For (9.86), it follows from (5.36), (5.37). For (9.87), by Chebyshev's inequality, it is enough to show

$$
\mathrm{var}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})) = \mathrm{tr}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{X}) = O(n^\delta),
$$

which follows from (5.38). Then, by (9.86) and (9.87), it then gives (9.84) and (9.85). It then completes the proof of (5.52).

Finally, we prove (5.54). For $\delta = 0$, it can be followed easily from (9.83) that

$$
\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} \\
&\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&\quad + \xi(\boldsymbol{\theta}) + O_p(1) \\
&= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\
&\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} + \xi(\boldsymbol{\theta}) + O_p(1),
\end{aligned}
$$

where the last equality follows from (9.84) and (9.85). □

Proof of Lemma 11

Denote $\sigma_{\eta,\alpha}^2 \equiv \gamma(\alpha) + \sigma_{\eta,0}^2$ and $\kappa_{\eta,\alpha} \equiv \kappa_{\eta,0}\sigma_{\eta,0}^2/(\gamma(\alpha) + \sigma_{\eta,0}^2)$, for $\alpha \in \mathcal{A}$, where $\gamma(\alpha)$ is defined in (5.51) and $\gamma(\alpha) = 0$ for $\alpha \in \mathcal{A}^c$.

First, we prove (5.57). It suffices to show that for $\delta \in [0,1)$ and any $\varepsilon > 0$,

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| \geq \varepsilon} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha)\big) \;>\; 0, \tag{9.88}$$

as $n \to \infty$ with probability tending to 1. By (5.54) and (5.66) and for $\delta \in [0,1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)';\alpha) \\
&= \left(\log \sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2} - \log \sigma_{\epsilon,0}^2 - 1\right) n \\
&\quad + \left\{ \left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2} \left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2}\right) - \left(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,0}^2}\right)^{1/2} \right\} n^{(1+\delta)/2} + o_p(n) \\
&= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2)^2 n}{2\sigma_{\epsilon,0}^4} + o_p(n),
\end{aligned}$$

where the first equality follows by $\xi(\boldsymbol{\theta}) = o_p(n)$ that

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_\eta^2, \kappa_\eta, \sigma_\epsilon^2)') = o_p(n),$$

and the last equality follows from (5.39) and (5.40). Thus, (9.88) and hence (5.57) are obtained.

Second, we prove (5.58). By (5.57), it suffices to show that for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$, $\delta \in [0,1)$ and any $\varepsilon > 0$,

$$\inf_{|\kappa_\eta \sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| \geq \varepsilon} \big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha)\big) \;>\; 0, \tag{9.89}$$

as $n \to \infty$ with probability tending to 1. By (5.54) and (5.66), for $\delta \in [0,1)$ and $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)';\alpha) \\
&\geq \left\{ \left(\frac{2\kappa_\eta \sigma_\eta^2}{\sigma_{\epsilon,0}^2}\right)^{1/2} \left(\frac{1}{2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta \sigma_\eta^2}\right) - \left(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,0}^2}\right)^{1/2} \right\} n^{(1+\delta)/2} \\
&\quad + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + o_p(n^{(1+\delta)/2}) \\
&= \frac{(\kappa_\eta \sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2)^2 n^{(1+\delta)/2}}{2^{5/2}\sigma_{\epsilon,0}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}} + \xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') + o_p(n^{(1+\delta)/2}),
\end{aligned}$$

where the first inequality follows from $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ and the lefthand side of (5.39) is positive for $\sigma_\epsilon > 0$, and the second equality follows from (5.41). Therefore, for (9.89) to hold, it remains to show that

$$\xi(\boldsymbol{\theta}) - \xi((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)') \;=\; o_p(n^{(1+\delta)/2}),$$

which follows from (5.68) and Chebyshev's inequality, by to checking the following moment conditions:

$$\begin{aligned}
\mathrm{var}(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\eta}) &= o(n), \\
\mathrm{var}(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\epsilon}) &= o(n), \\
\mathrm{var}(\boldsymbol{\epsilon}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)'))\boldsymbol{\epsilon}) &= o(n),
\end{aligned}$$

where the previous two equations follow trivially by (5.33) and (5.34) and the last equation follows from (5.35) and $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(1)$ that

$$\begin{aligned}
&\mathrm{var}\big(\boldsymbol{\epsilon}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)')\big)\boldsymbol{\epsilon}\big)/(2\sigma_{\epsilon,0}^4) \\
&\quad = \mathrm{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}^{-2}((\sigma_\eta^2, \kappa_\eta, \sigma_{\epsilon,0}^2)')) - 2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}((\sigma_{\eta,\alpha}^2, \kappa_{\eta,\alpha}, \sigma_{\epsilon,0}^2)')) \\
&\quad = \frac{n}{\sigma_\epsilon^4} + \frac{n}{\sigma_{\epsilon,0}^4} - \frac{2n}{\sigma_\epsilon^2\sigma_{\epsilon,0}^2} + O(n^{(1+\delta)/2}) \\
&\quad = o(n).
\end{aligned}$$

Thus, (9.89) and hence (5.58) are obtained. This completes the proof. □

Proof of Corollary 4

Let $\sigma_{\eta,\alpha}^2 \equiv \gamma(\alpha) + \sigma_{\eta,0}^2$ and $\kappa_{\eta,\alpha} \equiv \kappa_{\eta,0}\sigma_{\eta,0}^2/\sigma_{\eta,\alpha}^2$, for $\alpha \in \mathcal{A}$, where $\gamma(\alpha)$ is defined in (5.51).

First, we prove (5.78). By (5.70) and (5.71), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(n^{-(1-\delta)/4})$, we have

$$\sup_{|\kappa_\eta - \kappa_{\eta,\alpha}| = o(1)} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha) = o(n^\delta).$$

It then gives (5.78).

Second, we prove (5.79). By (5.75), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$, we have

$$\sup_{|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = O(n^{-1/4})} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha) = O_p(1).$$

It then gives (5.79).

Third, we prove (5.80). By (9.82) and (3.3), we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0) \\
&\quad +2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0) \\
&\quad +2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta), \qquad (9.90)
\end{aligned}$$

where $\xi(\boldsymbol{\theta})$ is defined in (5.53), the second equality follows from (3.3) and

$$\begin{aligned}
(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})) &= \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} \\
&\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}),
\end{aligned}$$

and the last equality follows from (9.84) and (9.85). Then for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$, $|\sigma_\epsilon^2 - \sigma_{\epsilon,0}^2| = o(n^{-(1-\delta)/2})$, $|\sigma_\eta^2\kappa_\eta - \sigma_{\eta,0}^2\kappa_{\eta,0}| = o(n^{-(1-\delta)/4})$ and $|\kappa_\eta - \kappa_{\eta,\alpha}| = o(1)$, we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(1)};\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(1)}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_\alpha^{(1)}) + o_p(n^\delta) \\
&= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(1)}) + o_p(n^\delta) \\
&= o_p(n^\delta),
\end{aligned}$$

where the second equality follows from (5.71) and the last equality follows from (5.78). In addition, by (5.49) and (9.84), we have

$$\begin{aligned} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} + O_p(1) \\ &= \frac{1}{2}\gamma(\alpha)n^{\delta} + o_p(n^{\delta}) \ > \ 0, \end{aligned}$$

as $n \to \infty$ with probability tending to 1. Then, we have

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})/L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_{\alpha}^{(1)}) = 1,$$

which gives (5.80).

Finally, we prove (5.81). Similar to (9.90), we have

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0) \\ &\quad +2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) + O_p(1). \end{aligned}$$

Hence, for $\alpha \in \mathcal{A}$, $|\sigma_{\epsilon}^2 - \sigma_{\epsilon,0}^2| = O(n^{-1/2})$ and $|\sigma_{\eta}^2\kappa_{\eta} - \sigma_{\eta,0}^2\kappa_{\eta,0}| = O(n^{-1/4})$,

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_{\alpha}^{(1)};\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_{\alpha}^{(1)}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_{\alpha}^{(1)}) + O_p(1) \\ &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_{\alpha}^{(1)}) + O_p(1) \\ &= O_p(1), \end{aligned}$$

where the second equality follows from (5.76) and the last equality follows from (5.79). Then,

$$L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_{\alpha}^{(1)}) = O_p(1),$$

which gives (5.81). This completes the proof. □

Proof of Proposition 5

First, we prove (5.86). By (4.9), we have

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{Z}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{Z} \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\ &\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(2)}(\alpha;\boldsymbol{\theta}), \end{aligned} \tag{9.91}$$

where the last equality follows from (5.87). In addition, by (5.29),

$$\begin{aligned} \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) &= n\log\sigma_{\epsilon}^2 - \frac{1-\delta}{2}\log n + (2\kappa_{\eta}\sigma_{\eta}^2\sigma_{\epsilon}^{-2})^{1/2}n^{(1+\delta)/2} - (2\kappa_{\eta}\sigma_{\eta}^2\sigma_{\epsilon}^{-2})n^{\delta} \\ &\quad +(\sigma_{\eta}^2 - \sigma_{\epsilon}^2)\sigma_{\epsilon}^{-2}\kappa_{\eta}n^{\delta} + o(n^{\delta}) + O(1), \end{aligned}$$

by (5.32),

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}n + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon}\right)^{1/2}\left(-\frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\sigma_{\eta,0}^2\kappa_{\eta,0}}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\ &\quad + \frac{\sigma_{\eta,0}^2\kappa_{\eta,0}(\kappa_\eta - \kappa_{\eta,0})}{\kappa_\eta\sigma_\eta^2}n^\delta + \frac{\sigma_{\eta,0}^2(\kappa_\eta - \kappa_{\eta,0})^2}{2\kappa_\eta\sigma_\eta^2}n^\delta + o(n^\delta) + O(1),\end{aligned}$$

and by (5.30),

$$\begin{aligned}\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \sum_{j\in\alpha^c\backslash\alpha}\frac{\beta_j^2\sigma_j^2\kappa_j}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}}n^{(1+\delta)/2} + \sum_{j\in\alpha^c\backslash\alpha}\frac{\beta_j^2\sigma_j^2\kappa_j(\kappa_\eta - \kappa_j)}{\kappa_\eta\sigma_\eta^2}n^\delta \\ &\quad + \sum_{j\in\alpha^c\backslash\alpha}\frac{\beta_j^2\sigma_j^2(\kappa_\eta - \kappa_j)^2}{2\kappa_\eta\sigma_\eta^2}n^\delta + o(n^\delta) + O(1).\end{aligned}$$

It then follows together with and (9.91), for $\delta \in (0,1)$,

$$\begin{aligned}-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\ &\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\ &\quad + \left(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\left(\frac{\theta_{\eta,\alpha}}{\kappa_\eta\sigma_\eta^2} - 1\right) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)n^\delta \\ &\quad + \frac{1}{2\kappa_\eta\sigma_\eta^2}\left(\left(\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)^{1/2}\kappa_\eta - \frac{\theta_{\eta,\alpha}}{\left(\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)^{1/2}}\right)^2 n^\delta \\ &\quad + \frac{1}{2\kappa_\eta\sigma_\epsilon^2}\left(\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2\kappa_j^2 + \sigma_{\eta,0}^2\kappa_{\eta,0}^2 - \frac{\theta_{\eta,\alpha}^2}{\left(\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\eta,0}^2\right)}\right)n^\delta \\ &\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(2)}(\alpha;\boldsymbol{\theta}) + o(n^\delta) + O(1). \end{aligned} \tag{9.92}$$

Therefore, for (5.86) to hold, it remains to show that for $\delta \in [0,1)$,

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1), \tag{9.93}$$

which follows in a similar way as (9.84). This completes the proof of (5.86).

Finally, we prove (5.88). For $\delta = 0$, it can be followed easily from (9.92) that

$$\begin{aligned}-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\ &\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} \\ &\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(2)}(\alpha;\boldsymbol{\theta}) + O_p(1) \\ &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \left(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\right)n \\ &\quad + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\right)^{1/2}\left(1 - \frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma_\eta^2}\right)n^{1/2} + \xi^{(2)}(\alpha;\boldsymbol{\theta}) + O_p(1),\end{aligned}$$

where the last equality follows from (9.93). □

Proof of Lemma 12

Let $\sigma^2_{\eta,\alpha} \equiv \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma^2_{\eta,0}$ and $\kappa_{\eta,\alpha} \equiv \theta_{\eta,\alpha}/\sigma^2_{\eta,\alpha}$, for $\alpha\in\mathcal{A}$, where $\theta_{\eta,\alpha}$ is defined in (5.85).

First, we prove (5.90). It suffices to show that for $\delta\in[0,1)$ and any $\varepsilon>0$,

$$\inf_{|\sigma^2_\epsilon-\sigma^2_{\epsilon,0}|\geq\varepsilon}\big(-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})';\alpha)\big) \;>\; 0, \tag{9.94}$$

as $n\to\infty$ with probability tending to 1. By (5.104) and (5.113), we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})';\alpha)\\
&= \left(\log\sigma^2_\epsilon+\frac{\sigma^2_{\epsilon,0}}{\sigma^2_\epsilon}-\log\sigma^2_{\epsilon,0}-1\right)n\\
&\quad+\left\{\left(\frac{2\kappa_\eta\sigma^2_\eta}{\sigma^2_\epsilon}\right)^{1/2}\left(1-\frac{\sigma^2_{\epsilon,0}}{2\sigma^2_\epsilon}+\frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma^2_\eta}\right)-\left(\frac{2\kappa_{\eta,0}\sigma^2_{\eta,0}}{\sigma^2_{\epsilon,0}}\right)^{1/2}\right\}n^{(1+\delta)/2}\\
&\quad+\xi^{(2)}(\alpha;\boldsymbol{\theta})-\xi^{(2)}(\alpha;(\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})')+o_p(n^{(1+\delta)/2})\\
&= \frac{(\sigma^2_\epsilon-\sigma^2_{\epsilon,0})^2n}{2\sigma^4_{\epsilon,0}}+\xi^{(2)}(\alpha;\boldsymbol{\theta})-\xi^{(2)}(\alpha;(\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})')+o_p(n),
\end{aligned}$$

where the last equality follows from (5.39) and (5.40). Therefore, for (9.94) to hold, it remains to show that

$$\xi^{(2)}(\alpha;\boldsymbol{\theta})-\xi^{(2)}(\alpha;(\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})') \;=\; o_p(n),$$

which follows from (5.67), (5.99) and

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})';\alpha) &= o_p(n),\\
\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})';\alpha) &= o_p(n).
\end{aligned}$$

by (5.33) and (5.35) that

$$\begin{aligned}
\mathrm{var}(\boldsymbol{X}'_j((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})')))\boldsymbol{X}_{j'}) &= O(n),\\
\mathrm{var}(\boldsymbol{X}'_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})'))\boldsymbol{\eta}) &= O(n),\\
\mathrm{var}(\boldsymbol{X}'_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma^2_\eta,\kappa_\eta,\sigma^2_{\epsilon,0})'))\boldsymbol{\epsilon}) &= O(n).
\end{aligned}$$

Thus, (9.94) and hence (5.90) are obtained.

Second, we prove (5.91). By (5.90), it suffices to show that for $|\sigma^2_\epsilon-\sigma^2_{\epsilon,0}|=o(1)$, $\delta\in[0,1)$ and any $\varepsilon>0$

$$\inf_{|\kappa_\eta\sigma^2_\eta-\theta_{\eta,\alpha}|\geq\varepsilon}\big(-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha)+2\ell((\sigma^2_{\eta,\alpha},\kappa_{\eta,\alpha},\sigma^2_{\epsilon,0})';\alpha)\big) \;>\; 0, \tag{9.95}$$

as $n\to\infty$ with probability tending to 1. By (5.104) and (5.113), for $|\sigma^2_\epsilon-\sigma^2_{\epsilon,0}|=o(1)$

and any $\varepsilon > 0$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) \\
\geq\ & \left\{ \left(\frac{2\kappa_\eta \sigma^2_\eta}{\sigma^2_{\epsilon,0}}\right)^{1/2} \left(\frac{1}{2} + \frac{\theta_{\eta,\alpha}}{2\kappa_\eta\sigma^2_\eta}\right) - \left(\frac{2\theta_{\eta,\alpha}}{\sigma^2_{\epsilon,0}}\right)^{1/2} \right\} n^{(1+\delta)/2} \\
&+\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) + \xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) \\
&+\xi(\boldsymbol{\theta}) - \xi((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})') + o_p(n^{(1+\delta)/2}) \\
=\ & \frac{(\kappa_\eta\sigma^2_\eta - \theta_{\eta,\alpha})^2 n^{(1+\delta)/2}}{2^{5/2}\sigma_{\epsilon,0}\theta^{3/2}_{\eta,\alpha}} + \xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) \\
&+\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) + \xi(\boldsymbol{\theta}) - \xi((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})') + o_p(n^{(1+\delta)/2}),
\end{aligned}$$

where the first inequality follows from $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = o(1)$ and the lefthand side of (5.39)is positive for $\sigma_\epsilon > 0$, the first equality follows from (5.41). Therefore, for (9.95) to hold, it remains to show that

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha) - \xi_1((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) &= o_p(n^{(1+\delta)/2}), \\
\xi_2(\boldsymbol{\theta};\alpha) - \xi_2((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})';\alpha) &= o_p(n^{(1+\delta)/2}), \\
\xi(\boldsymbol{\theta}) - \xi((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})') &= o_p(n^{(1+\delta)/2}),
\end{aligned}$$

where the previous two equations can be obtained in a way similarly to (5.105)- (5.106) by using the following moments conditions given by (5.33), (5.34) and (5.35):

$$\begin{aligned}
\mathrm{var}(\boldsymbol{X}'_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})'))\boldsymbol{X}_{j'}) &= o(n), \\
\mathrm{var}(\boldsymbol{X}'_j(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})'))\boldsymbol{\epsilon}) &= o(n),
\end{aligned}$$

and for $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = o(1)$,

$$\begin{aligned}
&\mathrm{var}\big(\boldsymbol{\epsilon}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})')\big)\boldsymbol{\epsilon}\big)/(2\sigma^4_{\epsilon,0}) \\
=\ & \mathrm{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}^{-2}((\sigma^2_\eta, \kappa_\eta, \sigma^2_{\epsilon,0})')) - 2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}((\sigma^2_{\eta,\alpha}, \kappa_{\eta,\alpha}, \sigma^2_{\epsilon,0})')) \\
=\ & \frac{n}{\sigma^4_\epsilon} + \frac{n}{\sigma^4_{\epsilon,0}} - \frac{2n}{\sigma^2_\epsilon\sigma^2_{\epsilon,0}} + O(n^{1/2}) \\
=\ & o(n).
\end{aligned}$$

Thus, (9.95) and hence (5.91) are obtained. The proof is then complete. □

Proof of Corollary 5

Let $\sigma^2_{\eta,\alpha} \equiv \sum_{j\in\alpha^c\setminus\alpha} \beta^2_j\sigma^2_j + \sigma^2_{\eta,0}$ and $\kappa_{\eta,\alpha} \equiv \theta_{\eta,\alpha}/\sigma^2_{\eta,\alpha}$ for $\alpha \in \mathcal{A}$, where $\theta_{\eta,\alpha}$ is defined in (5.85).

First, we prove (5.119). By (5.108) and (5.109)-(5.111), for $|\sigma_\epsilon - \sigma^2_{\epsilon,0}| = o(n^{-(1-\delta)/2})$ and $|\kappa_\eta\sigma^2_\eta - \theta_{\eta,\alpha}| = o(n^{-(1-\delta)/4})$, we have

$$\sup_{|\kappa_\eta - \kappa_{\eta,\alpha}| = o(1)} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) \quad = \quad o(n^\delta).$$

It then gives (5.119).

Second, we prove (5.120). By (5.115), for $|\sigma_\epsilon - \sigma^2_{\epsilon,0}| = O(n^{-1/2})$, we have

$$\sup_{|\kappa_\eta \sigma^2_\eta - \theta_{\eta,\alpha}| = O(n^{-1/4})} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) \;=\; O_p(1).$$

It then gives (5.120).

Finally, we prove (5.121). First, for $\delta \in (0,1)$, by (9.82), we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))\\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta})\\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0)\\
&\quad +2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon})\\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0)\\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta), \qquad (9.96)
\end{aligned}$$

where $\xi(\boldsymbol{\theta})$ and $\xi_2(\boldsymbol{\theta};\alpha)$ are defined in (5.53) and (5.98), respectively, the second equality follows from (3.3) and

$$\begin{aligned}
(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})) &= \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu}\\
&\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}),
\end{aligned}$$

and the last equality follows from (9.93). Then, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = o(n^{-(1-\delta)/2})$, $|\sigma^2_\eta\kappa_\eta - \theta_{\eta,\alpha}| = o(n^{-(1-\delta)/4})$ and $|\kappa_\eta - \kappa_{\eta,\alpha}| = o(1)$, we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}^{(2)}_\alpha) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}^{(2)}_\alpha)\\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) - \xi_2(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) + o_p(n^\delta)\\
&= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}^{(2)}_\alpha) + o_p(n^\delta)\\
&= o_p(n^\delta),
\end{aligned}$$

where the second equality follows from (5.110) and (5.111) and the last equality follows from (5.78). In addition, by (5.122) and (9.93), we have

$$\begin{aligned}
L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha)\boldsymbol{\mu} + O_p(1)\\
&= \frac{1}{2}\sum_{j\in\alpha^c\setminus\alpha}\frac{\beta^2_j\sigma^2_j\kappa_j}{(2\kappa_\eta\sigma^2_\eta\sigma^2_\epsilon)^{1/2}}n^{(1+\delta)/2} + o_p(n^{(1+\delta)/2}) \;>\; 0,
\end{aligned}$$

as $n\to\infty$ with probability tending to 1. Then, we have

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})/L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}^{(2)}_\alpha) = 1,$$

which gives (5.121) for $\delta \in (0,1)$. Second, for $\delta = 0$, similar to (9.96),

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0)\\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) + O_p(1).
\end{aligned}$$

Hence, for $\alpha \in \mathcal{A}$, $|\sigma^2_\epsilon - \sigma^2_{\epsilon,0}| = O(n^{1/2})$ and $|\kappa_\eta\sigma^2_\eta - \theta_{\eta,\alpha}| = O(n^{-1/4})$,

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}^{(2)}_\alpha) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}^{(2)}_\alpha) + O_p(1)\\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) - \xi_2(\boldsymbol{\theta}^{(2)}_\alpha;\alpha) + O_p(1)\\
&= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}^{(2)}_\alpha) + O_p(1)\\
&= O_p(1),
\end{aligned}$$

where the second equality follows from (5.116) and (5.117), and the last equality follows from (5.120). Then

$$L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(2)}) = O_p(1).$$

In addition, by (5.122) and (9.93), we have

$$\begin{aligned} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha)\boldsymbol{\mu} + O_p(1) \\ &= \frac{1}{2}\sum_{j\in\alpha^c\setminus\alpha}\frac{\beta_j^2\sigma_j^2\kappa_j}{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^2)^{1/2}}n^{1/2} + o_p(n^{1/2}) \;>\; 0, \end{aligned}$$

as $n\to\infty$ with probability tending to 1. Then, we have

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})/L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(2)}) = 1,$$

which gives (5.121) for $\delta = 0$. This completes the proof. □

Proof of Proposition 6

Let $\sigma_{\epsilon,\alpha}^2 \equiv \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2$ for $\alpha\in\mathcal{A}$.

First, we prove (5.124). By (4.9), we have

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{Z}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{Z} \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\ &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \sum_{j\in\alpha^c\setminus\alpha}\beta_j^2\sigma_j^2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\ &\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(3)}(\alpha;\boldsymbol{\theta}), \end{aligned} \tag{9.97}$$

where the last equality follows from (5.125). In addition, by (5.29),

$$\begin{aligned} \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) &= n\log\sigma_\epsilon^2 - \frac{1-\delta}{2}\log n + (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}n^{(1+\delta)/2} - (2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})n^\delta \\ &\quad +(\sigma_\eta^2-\sigma_\epsilon^2)\sigma_\epsilon^{-2}\kappa_\eta n^\delta + o(n^\delta) + O(1), \end{aligned}$$

by (5.32),

$$\begin{aligned} \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) &= \frac{\sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}n + \left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon}\right)^{1/2}\left(-\frac{\sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\sigma_{\eta,0}^2\kappa_{\eta,0}}{2\kappa_\eta\sigma_\eta^2}\right)n^{(1+\delta)/2} \\ &\quad + \frac{\sigma_{\eta,0}^2\kappa_{\eta,0}(\kappa_\eta-\kappa_{\eta,0})}{\kappa_\eta\sigma_\eta^2}n^\delta + \frac{\sigma_{\eta,0}^2(\kappa_\eta-\kappa_{\eta,0})^2}{2\kappa_\eta\sigma_\eta^2}n^\delta + o(n^\delta) + O(1), \end{aligned}$$

and by (5.31),

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) = \frac{n}{\sigma_\epsilon^2} - \frac{(2\kappa_\eta\sigma_\eta^2\sigma_\epsilon^{-2})^{1/2}}{2\sigma_\epsilon^2}n^{(1+\delta)/2} + o(n^\delta) + O(1),$$

It then follows together with and (9.97), for $\delta \in (0,1)$,

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \bigg(\log\sigma_\epsilon^2 + \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\bigg)n \\
&\quad + \bigg(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\bigg)^{1/2}\bigg(1 - \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\bigg)n^{(1+\delta)/2} \\
&\quad + \bigg(-\frac{\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2} + \kappa_\eta\bigg(\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{\kappa_\eta\sigma_\eta^2} - 1\bigg) - \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\bigg)n^\delta + \frac{\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}(\kappa_\eta - \kappa_{\eta,0})^2 n^\delta \\
&\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(3)}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta) + O(1). \qquad (9.98)
\end{aligned}$$

Therefore, for (5.124) to hold, it remains to show that for $\delta \in [0,1)$,

$$(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) = O_p(1), \qquad (9.99)$$

which follows in a similar way as (9.84). This completes the proof of (5.124).

Finally, we prove (5.88). For $\delta = 0$, it can be followed easily from (9.98) that

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) - \frac{1-\delta}{2}\log n + \bigg(\log\sigma_\epsilon^2 + \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\bigg)n \\
&\quad + \bigg(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\bigg)^{1/2}\bigg(1 - \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\bigg)n^{1/2} \\
&\quad - (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi^{(3)}(\alpha;\boldsymbol{\theta}) + O_p(1) \\
&= n\log(2\pi) - \frac{1-\delta}{2}\log n + \bigg(\log\sigma_\epsilon^2 + \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{\sigma_\epsilon^2}\bigg)n \\
&\quad + \bigg(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\bigg)^{1/2}\bigg(1 - \frac{\sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\bigg)n^{1/2} + \xi^{(3)}(\alpha;\boldsymbol{\theta}) + O_p(1),
\end{aligned}$$

where the last equality follows from (9.99). This completes the proof. □

Proof of Lemma 13

Let $\sigma_{\epsilon,\alpha}^2 \equiv \sum_{j\in\alpha^c\backslash\alpha}\beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2$.

First, we prove (5.128). It suffices to show that for $\delta \in [0,1)$ and any $\varepsilon > 0$,

$$\inf_{|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| \geq \varepsilon}\big(-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha)\big) > 0, \qquad (9.100)$$

as $n \to \infty$ with probability tending to 1. By (5.140) and (5.149), we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha) + 2\ell((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) \\
&= \bigg(\log\sigma_\epsilon^2 + \frac{\sigma_{\epsilon,\alpha}^2}{\sigma_\epsilon^2} - \log\sigma_{\epsilon,\alpha}^2 - 1\bigg)n \\
&\quad + \bigg\{\bigg(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_\epsilon^2}\bigg)^{1/2}\bigg(1 - \frac{\sigma_{\epsilon,\alpha}^2}{2\sigma_\epsilon^2} + \frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\bigg) - \bigg(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,\alpha}^2}\bigg)^{1/2}\bigg\}n^{(1+\delta)/2} \\
&\quad + \xi^{(3)}(\alpha;\boldsymbol{\theta}) - \xi^{(3)}(\alpha;(\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') + o_p(n^{(1+\delta)/2}) \\
&= \frac{(\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2)^2 n}{2\sigma_{\epsilon,\alpha}^4} + \xi^{(3)}(\alpha;\boldsymbol{\theta}) - \xi^{(3)}(\alpha;(\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') + o_p(n^{(1+\delta)/2}),
\end{aligned}$$

where the last equality follows from (5.39) and (5.40). Therefore, for (9.100) to hold, it remains to show that

$$\xi^{(3)}(\alpha;\boldsymbol{\theta})-\xi^{(3)}(\alpha;(\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') \;=\; o_p(n).$$

By (5.138), it is enough to show that

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(n),\\
\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(n),\\
\xi(\boldsymbol{\theta})-\xi((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)') &= o_p(n),
\end{aligned}$$

which can be obtained from Chebyshev's inequality and using the following moment conditions given by (5.34) and (5.35):

$$\begin{aligned}
\mathrm{var}(\boldsymbol{X}_j'((\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)')))\boldsymbol{X}_{j'}) &= O(n),\\
\mathrm{var}(\boldsymbol{X}_j'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)'))\boldsymbol{\eta}) &= O(n),\\
\mathrm{var}(\boldsymbol{X}_j'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma_\eta^2,\kappa_\eta,\sigma_{\epsilon,\alpha}^2)'))\boldsymbol{\epsilon}) &= O(n).
\end{aligned}$$

Thus, (9.100) and hence (5.128) are obtained.

Second, we prove (5.129). By (5.128), it suffices to show that for $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(1)$, $\delta\in[0,1)$ and any $\varepsilon>0$

$$\inf_{|\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2|\geq\varepsilon}\big(-2\ell(\hat{\boldsymbol{\theta}}(\alpha);\alpha)+2\ell((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\big) \;>\; 0, \qquad (9.101)$$

as $n\to\infty$ with probability tending to 1. By (5.140) and (5.149), for $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(1)$ and any $\varepsilon>0$, we have

$$\begin{aligned}
&-2\ell(\boldsymbol{\theta};\alpha)+2\ell((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\
&\geq \left\{\left(\frac{2\kappa_\eta\sigma_\eta^2}{\sigma_{\epsilon,\alpha}^2}\right)^{1/2}\left(\frac{1}{2}+\frac{\kappa_{\eta,0}\sigma_{\eta,0}^2}{2\kappa_\eta\sigma_\eta^2}\right)-\left(\frac{2\kappa_{\eta,0}\sigma_{\eta,0}^2}{\sigma_{\epsilon,\alpha}^2}\right)^{1/2}\right\}n^{(1+\delta)/2}+\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\
&\quad+\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)+\xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)')+o_p(n^{(1+\delta)/2})\\
&= \frac{(\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2)^2n^{(1+\delta)/2}}{2^{5/2}\sigma_{\epsilon,\alpha}(\kappa_{\eta,0}\sigma_{\eta,0}^2)^{3/2}}+\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)\\
&\quad+\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha)+\xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)')+o_p(n^{(1+\delta)/2}),
\end{aligned}$$

where the first inequality follows from $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=o(1)$ and the lefthand side of (5.39) is positive, the first equality follows from (5.41). Therefore, for (9.101) to hold, it remains to show that

$$\begin{aligned}
\xi_1(\boldsymbol{\theta};\alpha)-\xi_1((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(n^{(1+\delta)/2}),\\
\xi_2(\boldsymbol{\theta};\alpha)-\xi_2((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)';\alpha) &= o_p(n^{(1+\delta)/2}),\\
\xi(\boldsymbol{\theta})-\xi((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)') &= o_p(n^{(1+\delta)/2}),
\end{aligned}$$

which follows similarly from (5.141)-(5.142), by using the following moment conditions given by (5.33) and (5.34):

$$\begin{aligned}
\mathrm{var}(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)'))\boldsymbol{\eta}) &= o(n),\\
\mathrm{var}(\boldsymbol{\eta}'(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\boldsymbol{\Sigma}^{-1}((\sigma_{\eta,0}^2,\kappa_{\eta,0},\sigma_{\epsilon,\alpha}^2)'))\boldsymbol{\epsilon}) &= o(n),
\end{aligned}$$

and for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(1)$,

$$\begin{aligned}
&\mathrm{var}\big(\boldsymbol{\epsilon}'\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)')\big)\boldsymbol{\epsilon}\big)/(2\sigma_{\epsilon,0}^4) \\
&= \mathrm{tr}(\boldsymbol{\Sigma}^{-2}(\boldsymbol{\theta})) + \mathrm{tr}(\boldsymbol{\Sigma}^{-2}((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)')) - 2\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}((\sigma_{\eta,0}^2, \kappa_{\eta,0}, \sigma_{\epsilon,\alpha}^2)')) \\
&= \frac{n}{\sigma_\epsilon^4} + \frac{n}{\sigma_{\epsilon,\alpha}^4} - \frac{2n}{\sigma_\epsilon^2\sigma_{\epsilon,\alpha}^2} + O(n^{1/2}) \\
&= o(n).
\end{aligned}$$

Thus, (9.101) and hence (5.129) are obtained. The proof is then complete. □

Proof of Corollary 6

Let $\sigma_{\epsilon,\alpha}^2 \equiv \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2\sigma_j^2 + \sigma_{\epsilon,0}^2$ for $\alpha \in \mathcal{A}$.

First, we prove (5.155). By (5.144) and (5.145)-(5.147), for $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(n^{-(1-\delta)/2})$ and $|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(n^{-(1-\delta)/4})$, we have

$$\sup_{|\kappa_\eta - \kappa_{\eta,0}| = o(1)} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha) \;=\; o(n^\delta).$$

It then gives (5.155).

Second, we prove (5.156). By (5.151), for $|\sigma_\epsilon - \sigma_{\epsilon,\alpha}^2| = O(n^{-1/2})$, we have

$$\sup_{|\kappa_\eta\sigma_\eta^2 - \kappa_{\eta,0}\sigma_{\eta,0}^2| = O(n^{-1/4})} -2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha) \;=\; O_p(1).$$

It then gives (5.156).

Finally, we prove (5.157). First, for $\delta \in (0,1)$, by (9.97), we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) + \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \\
&\quad -(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0) \\
&\quad +2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) - 2(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}) \\
&= n\log(2\pi) + \log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) + (\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_0) \\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) + 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) + o_p(n^\delta), \qquad (9.102)
\end{aligned}$$

where $\xi(\boldsymbol{\theta})$ and $\xi_2(\boldsymbol{\theta};\alpha)$ are defined in (5.53) and (5.137), respectively, the second equality follows from (3.3) and

$$\begin{aligned}
(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\alpha;\boldsymbol{\theta})) &= \boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha;\boldsymbol{\theta})\boldsymbol{\mu} \\
&\quad +(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{M}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{M}(\alpha;\boldsymbol{\theta})(\boldsymbol{\eta}+\boldsymbol{\epsilon}),
\end{aligned}$$

and the last equality follows from (9.99). Then, for $\alpha \in \mathcal{A}\setminus\mathcal{A}^c$, $|\sigma_\epsilon^2 - \sigma_{\epsilon,\alpha}^2| = o(n^{-(1-\delta)/2})$, $|\sigma_\eta^2\kappa_\eta - \kappa_{\eta,0}\sigma_{\eta,0}^2| = o(n^{-(1-\delta)/4})$ and $|\kappa_\eta - \kappa_{\eta,0}| = o(1)$, we have

$$\begin{aligned}
-2\ell(\boldsymbol{\theta};\alpha) + 2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)}) + \xi(\boldsymbol{\theta}) - \xi(\boldsymbol{\theta}_\alpha^{(3)}) \\
&\quad +\xi_2(\boldsymbol{\theta};\alpha) - \xi_2(\boldsymbol{\theta}_\alpha^{(3)};\alpha) + o_p(n^\delta) \\
&= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) - 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)}) + o_p(n^\delta) \\
&= o_p(n^\delta),
\end{aligned}$$

where the second equality follows from (5.146) and (5.147) and the last equality follows from (5.78). In addition, by (5.159) and (9.99), we have

$$\begin{aligned} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha)\boldsymbol{\mu}+O_p(1) \\ &= \frac{1}{2}\sum_{j\in\alpha^c\setminus\alpha}\frac{\beta_j^2\sigma_j^2}{\sigma_\epsilon^2}n+o_p(n) \;>\; 0, \end{aligned}$$

as $n\to\infty$ with probability tending to 1. Then, we have

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})/L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)})=1,$$

which gives (5.157) for $\delta\in(0,1)$. Second, for $\delta=0$, similar to (9.102),

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha) &= n\log(2\pi)+\log\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))+(\boldsymbol{\eta}+\boldsymbol{\epsilon})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{\eta}+\boldsymbol{\epsilon})+\xi(\boldsymbol{\theta})-\xi(\boldsymbol{\theta}_0) \\ &\quad +\xi_2(\boldsymbol{\theta};\alpha)+2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})+O_p(1). \end{aligned}$$

Hence, for $\alpha\in\mathcal{A}$, $|\sigma_\epsilon^2-\sigma_{\epsilon,\alpha}^2|=O(n^{1/2})$ and $|\kappa_\eta\sigma_\eta^2-\kappa_{\eta,0}\sigma_{\eta,0}^2|=O(n^{-1/4})$,

$$\begin{aligned} -2\ell(\boldsymbol{\theta};\alpha)+2\ell(\boldsymbol{\theta}_\alpha^{(3)};\alpha) &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})-2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)})+\xi(\boldsymbol{\theta})-\xi(\boldsymbol{\theta}_\alpha^{(3)})+O_p(1) \\ &\quad +\xi_2(\boldsymbol{\theta};\alpha)-\xi_2(\boldsymbol{\theta}_\alpha^{(3)};\alpha)+O_p(1) \\ &= 2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})-2L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)})+O_p(1) \\ &= O_p(1), \end{aligned}$$

where the second equality follows from (5.152) and (5.153), and the last equality follows from (5.156). Then

$$L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})-L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)})=O_p(1).$$

In addition, by (5.122) and (9.93), we have

$$\begin{aligned} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}(\alpha;\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{A}(\alpha)\boldsymbol{\mu}+O_p(1) \\ &= \frac{1}{2}\sum_{j\in\alpha^c\setminus\alpha}\frac{\beta_j^2\sigma_j^2}{\sigma_\epsilon^2}n+o_p(n) \;>\; 0, \end{aligned}$$

as $n\to\infty$ with probability tending to 1. Then, we have

$$\operatorname*{plim}_{n\to\infty} L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta})/L^{\mathrm{KL}}(\alpha;\boldsymbol{\theta}_\alpha^{(3)})=1,$$

which gives (5.157) for $\delta=0$. This completes the proof. □

Proof of Corollary 7

It suffices to show that (6.3) is satisfied all the time in this case. By (3.17),

$$\mathrm{E}(L(\alpha)) \;=\; R_1(\alpha)+R_2(\alpha)+\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1})\geq\sigma_\epsilon^2\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}^{-1}),$$

where the inequality follows from $R_1(\alpha) \geq 0$ and $R_2(\alpha) \geq 0$, for all $\alpha \in \mathcal{A}$. Also, from (5.30),

$$\lim_{n\to\infty} \frac{\sigma_\epsilon^2 \mathrm{tr}(\mathbf{\Sigma}_\eta \mathbf{\Sigma}^{-1})}{n^{(1+\delta)/2}} = 2^{-1/2}(\sigma_\eta^2 \kappa_\eta \sigma_\epsilon^2)^{1/2} > 0. \tag{9.103}$$

Hence, (6.3) is then satisfied for fixed p.

For $\mathcal{A} \neq \emptyset$ and (6.3) is satisfied from above, it is shown that for fixed p and $\alpha \in \mathcal{A}$,

$$\operatorname*{plim}_{n\to\infty} \left| \frac{L(\alpha)}{\mathrm{E}(L(\alpha))} - 1 \right| \;=\; 0, \tag{9.104}$$

by (6.9). Also, from (3.17), we have for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$,

$$\lim_{n\to\infty} \frac{\mathrm{E}(L(\alpha))}{\mathrm{E}(L(\alpha^c))} \;\geq\; 1,$$

by $\sigma_\epsilon^2 \mathrm{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{M}(\alpha)) \leq p(\alpha) < \infty$ and (9.103), where the equality holds for

$$\lim_{n\to\infty} \frac{\sigma_\epsilon^4 \boldsymbol{\mu}' \boldsymbol{A}(\alpha)' \mathbf{\Sigma}^{-2} \boldsymbol{A}(\alpha)\boldsymbol{\mu}}{\mathrm{E}(L(\alpha^c))} \;=\; 0. \tag{9.105}$$

Let $\alpha^L = \arg\min_{\alpha\in\mathcal{A}} L(\alpha)$. If (9.105) is not satisfied, by (9.104), we have

$$\operatorname*{plim}_{n\to\infty} \frac{L(\alpha)}{L(\alpha^c)} \;=\; c,$$

for some $c > 1$ and $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$, which follows that $P\big(\alpha^L \in \mathcal{A}^c\big) = 1$. In addition, by (3.17) and (9.103), we have for any $\alpha \in \mathcal{A}^c$,

$$\lim_{n\to\infty} \frac{\mathrm{E}(L(\alpha^L))}{\mathrm{E}(L(\alpha))} \;=\; 1,$$

Then by (9.104), for any $\alpha \in \mathcal{A}^c$, we have

$$\operatorname*{plim}_{n\to\infty} \frac{L(\alpha)}{L(\alpha^L)} \;=\; 1.$$

It then gives $\operatorname*{plim}_{n\to\infty} L(\hat\alpha)/\inf_{\alpha\in\mathcal{A}} L(\alpha) = 1$, if $\lim_{n\to\infty} P\big(\hat\alpha \in \mathcal{A}^c\big) = 1$. If (9.105) is satisfied, we then have for any $\alpha \in \mathcal{A}^c$,

$$\lim_{n\to\infty} \frac{\mathrm{E}(L(\alpha^L))}{\mathrm{E}(L(\alpha))} \;=\; 1,$$

which gives that for any $\alpha \in \mathcal{A}$,

$$\operatorname*{plim}_{n\to\infty} \frac{L(\alpha)}{L(\alpha^L)} \;=\; 1, \tag{9.106}$$

by (9.104), which also gives $\operatorname*{plim}_{n\to\infty} L(\hat\alpha)/\inf_{\alpha\in\mathcal{A}} L(\alpha) = 1$, for $\lim_{n\to\infty} P\big(\hat\alpha \in \mathcal{A}^c\big) = 1$. Then, it completes the proof. □

Proof of Corollary 8

For fixed p, (6.3) is shown by (9.103). Hence, by Theorem 14, it gives

$$\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}_{\text{CAIC}}) \Big/ \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1.$$

For any $\delta \in [0, 1)$, by (5.38), we have

$$\sigma_\epsilon^4 \boldsymbol{\mu}' \boldsymbol{A}(\alpha) \boldsymbol{\Sigma}^{-2} \boldsymbol{A}(\alpha) \boldsymbol{\mu} = O(n^\delta),$$

which gives (9.105) by (9.103). It then follows (9.106), which gives $\operatorname*{plim}_{n\to\infty} L(\hat{\alpha}) \Big/ \inf_{\alpha\in\mathcal{A}} L(\alpha) = 1$, for any arbitrary selected model $\hat{\alpha}$. It completes the proof. □

Proof of Corollary 9

For $\mathcal{A}^c = \{\alpha^c\}$, it suffices to show that $\lim_{n\to\infty} \mathrm{E}(L(\alpha)) - \sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}^{-1}) = \infty$ for $\alpha \in \mathcal{A} \setminus \mathcal{A}^c$ and finite fixed p by

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{\mu}' \boldsymbol{A}(\alpha)' \boldsymbol{\Sigma}^{-1} \boldsymbol{A}(\alpha) \boldsymbol{\mu} = \infty.$$

First, for $i = 1, \ldots, p$, let $\boldsymbol{X}_j$ be the jth column of $\boldsymbol{X}$. Then, for $j = 1, \ldots, p$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-2})} \boldsymbol{X}_j' \boldsymbol{\Sigma}^{-2} \boldsymbol{X}_j = \sigma_j^2,$$

by its expectation equals to σ_j^2 and its variance

$$\begin{aligned}
\lim_{n\to\infty} \sigma_j^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-4}) \Big/ \Big(\mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big)^2 &\leq \lim_{n\to\infty} \sigma_j^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) \Big/ \Big(\sigma_\epsilon^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big)^2 \\
&= \sigma_j^4 \Big/ \lim_{n\to\infty} \Big(\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big) = 0,
\end{aligned}$$

by $\sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$ and $\lim_{n\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) = \infty$, and for $j, k = 1, \ldots, p$ and $j \neq k$,

$$\lim_{n\to\infty} \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-2})} \boldsymbol{X}_j' \boldsymbol{\Sigma}^{-2} \boldsymbol{X}_k = 0,$$

by its expectation equals to 0 and its variance

$$\begin{aligned}
\lim_{n\to\infty} \sigma_j^2 \sigma_k^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-4}) \Big/ \Big(\mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big)^2 &\leq \lim_{n\to\infty} \sigma_j^2 \sigma_k^2 \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) \Big/ \Big(\mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big)^2 \\
&= \sigma_j^2 \sigma_k^2 \Big/ \lim_{n\to\infty} \Big(\sigma_\epsilon^4 \mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big) = 0.
\end{aligned}$$

It then follows that for $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_j)$,

$$\lim_{n\to\infty} \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-2})} \boldsymbol{\mu}' \boldsymbol{A}(\alpha)' \boldsymbol{\Sigma}^{-2} \boldsymbol{A}(\alpha) \boldsymbol{\mu} = \sum_{j\in\alpha^c\setminus\alpha} \beta_j^2 \sigma_j^2 > 0, \tag{9.107}$$

for $\alpha \in \mathcal{A}^c$. Hence, (6.26) holds by $\lim_{n\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) = \infty$, we then complete the proof. □

Proof of Corollary 10

For $\mathcal{A}^c \neq \emptyset$ and $\lambda \to \infty$, by (9.107) and $\lim_{n\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}^{-2})/\lambda = \infty$, (6.46) is then satisfied. In addition, we have

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M}(\alpha)) &= \operatorname*{plim}_{n\to\infty} \mathrm{tr}\big((\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{\Sigma}^{-2}\boldsymbol{X}(\alpha)\big) \\ &= \lim_{n\to\infty} \sum_{j\in\alpha} \sigma_j^2 \frac{\mathrm{tr}(\boldsymbol{\Sigma}^{-2})}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} > 0, \end{aligned}$$

which gives (6.47) for fixed p. It then follows that $\lim_{n\to\infty} P\big(\hat{\alpha}_{\mathrm{GIC}(\lambda)} = \alpha^c\big) = 1$, which completes the proof. □

References

Akaike, H., "Information theory and the maximum likelihood principle." In Second International Symposium on Information Theory (V. Petrov and F. Csáki eds.), Akademiai Kiádo, Budapest, 267-281, 1973.

Chen, H.-S., Simpson, D. G. Ying, Z., "Infill asymptotics for a stochastic process model with measurement error," Statistica Sinica, **10**, 141-156, 2000.

Cressie, N., Statistics for Spatial Data (revised edition), Wiley: New York, 1993.

Hoeting, J. A., Davis, R. A., Merton, A. A., and Thompson, S. E,, "Mode selection for geostatistical models," Ecological Applications, **16**, 87-98, 2006.

Huang, H.-C. and Chen, C.-S., "Optimal geostatistical model selection." Journal of the American Statistical Association, **102**, 1009-1024, 2007.

Journel, A. G., "Nonparametric estimation of spatial distributions," Mathematical Geology, **15**(3), 445-468, 1983.

Krige, D. G., "A statistical approach to some basic mine valuation problems on the Witwatersrand," Journal of the Chemical, Metallurgical and Mining Society of South Africa, **52**, 119-139, 1951.

Mardia, K. V., and Marshall, R. J., "Maximum likelihood estimation of models for residual covariance in spatial regression," Biometrika, **71**, 135-146, 1984.

Matérn, B., Spatial Variation (second edition). Lecture Notes in Statistics, Springer: New York, 1986.

Matheron, G., The Theory of Regionalized Variables and its Applications. Fontainebleau: Ecole des Mines, 1971.

Matheron, G., Estimating and Choosing: An Essay on Probability in Practice. Translated by A. M. Hasofer, Berlin: Springer-Verlag, 1989.

McQuarrie, A. D. R., and Tsai, C.-L., Regression and Time Series Model Selection, World Scientific: Singapore, 1989.

Nishii, R., "Asymptotic properties of criteria for selection of variables in multiple regression," *The Annals of Statistics*, **12**, 758-765, 1984.

Pu, W., and Niu, X. F., "Selecting mixed-effects models based on a generalized information criterion," Journal of Multivariate Analysis, **97**, 733-758, 2006.

Schabenberger, O. L., and Gotway, C. A., Statistical Methods for Spatial Data Analysis, Chapman & Hall/CRC: Boca Raton, 2005.

Schwartz, G., "Estimating the dimensions of a model," The Annals of Statistics, **6**, 461-464, 1978.

Shao, J., "An asymptotic theory for linear model selection," The Annals of Statistics, **7**, 221-242, 1997.

Shibata, R., "An Optimal Selection of Regression Variables," Biometrika, **68**, 45-54, 1981.

Stein, M., Interpolation of Spatial Data: Some Theory for Kriging, Springer: New York, 1999.

Uhlenbeck, G. E. and Ornstein, L. S., "On the theory of Brownian motion," Physical Review, **36**, 823-841, 1930.

Vaida, F., and Blanchard, S., "Conditional Akaike information for mixed-effects models," Biometrika, **92**, 351-370, 2005.

Whittle, P., "Bounds for the moments of linear and quadratic forms in independent variables," Theory of Probability and Its Applications, **5**, 302-305, 1960.

Ying, Z. (1991), "Asymptotic Properties of a Maximum Likelihood Estimator With Data From a Gaussian Process," Journal of Multivariate Analysis, **36**, 280-296, 1991.

Zhang, H., "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," Journal of the American Statistical Association, **99**, 250-261, 2004.

Zhang, H. and Zimmerman, D. L., "Towards reconciling two asymptotic frameworks in spatial statistics," Biometrika, **92**, 921-936, 2005.