# Statistical Inference and Prediction in Nonlinear Models using Additive Random Fields*


Christian M. Dahl†
Department of Economics
Purdue University
dahlc@mgmt.purdue.edu

Gloria González-Rivera
Department of Economics
University of California, Riverside
gloria.gonzalez@ucr.edu

Yu Qin
Countrywide Financial Corporation
yu_qin@countrywide.com


April 5, 2005


**Abstract**

We study nonlinear models within the context of the flexible parametric random field regression model proposed by Hamilton (2001). Though the model is parametric, it enjoys the flexibility of the nonparametric approach since it can approximate a large collection of nonlinear functions and it has the added advantage that there is no "curse of dimensionality."The fundamental premise of this paper is that the parametric random field model, though a good approximation to the true data generating process, is still a misspecified model. We study the asymptotic properties of the estimators of the parameters in the conditional mean and variance of a generalized additive random field regression under misspecification.


*The notation follows Abadir and Magnus (2002). All software developed for this paper can be obtained from the corresponding author.


†Corresponding author. Address: 403 West State Street, Purdue University, West Lafayette, IN 47907-2056, USA. E-mail: dahlc@mgmt.purdue.edu. Phone: +1 765-494-4503. Fax: +1 765-496-1778.

The additive specification approximates the contribution of each regressor to the model by an individual random field, such that the conditional mean is the sum of as many independent random fields as the number of regressors. This new specification can be viewed as a generalization of Hamilton's and therefore, our results will provide "tools" for classical statistical inference that also will apply to his model. We develop a test for additivity that is based on a measure of goodness of fit. The test has a limiting Gaussian distribution and is very easy to compute. Through extensive Monte Carlo simulations, we assess the out-of-sample predictive accuracy of the additive random field model and the finite sample properties of the test for additivity comparing its performance to existing tests. The results are very encouraging and render the parametric additive random field model as a good alternative specification relative to its nonparametric counterparts, particularly in small samples.

**Keywords:** Random Field Regression, Nonlocal Misspecification, Asymptotics, Test for Additivity.

**JEL classification:** C12; C15; C22.

# 1 Introduction

We study nonlinear regression models within the context of the parametric random field model proposed by Hamilton (2001). Stationary random fields have been long-standing tools for the analysis of spatial data primarily in the field of geostatistics, environmental sciences, agriculture, and computer design, see, e.g., Cressie (1993). Several applications can be found in Loh and Lam (2000), Abt and Welch (1998), Ying (1991, 1993) , and Mardia and Marshall (1984). The analysis of economic data with random fields is still in a preliminary state of development. In a regression framework, random fields were first introduced by Hamilton (2001) and further developed in Dahl (2002), Dahl and González-Rivera (2003), Dahl and Qin (2004), and Dahl and Hylleberg (2004).

Random fields are closely related to universal kriging and thin plate splines. Dahl (2002) points out that the Hamilton's estimator of the conditional mean function becomes identical to the cubic spline smoother when the conditional mean function is viewed as a realization of a Brownian motion process.[1] In addition, Dahl (2002) shows that the random field approach has superior predictive accuracy compared to popular nonparametric estimators, i.e. the spline smoother, when the data is generated from popular econometric models such as LSTAR/ESTAR and various bilinear specifications. Though the random field model is parametric, it enjoys the flexibility of the nonparametric approach since it can approximate a large collection of nonlinear functions but with the added advantage that there is no "curse of dimensionality" and there is no dependence on bandwidth selection or smoothing parameter.

In many of the aforementioned applications (those *not* related to econometrics and/or economics), the random field model has been treated as the true data generating process. The view presented in this paper is that the parametric random field model, though a good approximation to the true data generating mechanism, is still a misspecified model. However, despite of misspecification, Hamilton (2001) shows that it is possible to obtain a consistent estimator of the *overall conditional mean function* under very general conditions. The theoretical contribution of this paper focuses on the consequences of misspecification for the asymptotic properties of the *estimators of the parameters* found in the conditional mean and in the variance of the error term of the model. This is an important and additional contribution to the Hamilton's results, which only focused on the overall conditional mean function.

An additional innovation of this paper is the specification of additive random fields to model the conditional mean of the variable of interest. Within the nonparametric literature, additive models play a very predominant role because they mitigate the "curse of dimensionality" and their estimators have faster convergence rates than those of the nonadditive nonparametric models, see i.e. Stone (1985, 1986). Hastie and Tibshirani (1990) provide an important and thorough analysis of additive models. There is a vast literature on nonparametric estimation of additive models. Some recent contributions are: Sperlich, Tjostheim and Yang (2002), Yang (2002), and Carroll, Hardle and Mammen (2002). In this paper, we investigate how and to what extent imposing an additive

[1] Using the results of Kimeldorf and Wahba (1971) and Wahba (1978, 1990) we show how this result generalizes to $\boldsymbol{x}_t \in \mathbb{R}^k$.

structure on the random field regression model improves its statistical properties when the true data generating process is additive in one or more of its arguments (nonzero interactions terms are allowed). In particular, we propose (in the simplest case where there are no interaction terms,) to approximate the individual contribution of each of the $k$ regressors to the conditional mean by a random field, such that the nonlinear part is the sum of $k$ individual and independent random fields. Our approach differs from the Hamilton's model where one "comprehensive" random field approximates the joint contribution of the $k$ regressors. This new specification can be viewed as a generalized random field as it collapses to the Hamilton's model when no assumptions regarding additivity are imposed. This implies that all the asymptotic results that we derive in this paper, they will also apply to the Hamilton's (2001) model. We will assess the gains in the predictive accuracy of the model when the additive structure is incorporated.

First, we provide a complete characterization of the asymptotic theory associated with the estimators of the parameters of the generalized additive random field model. Our environment is similar to Hamilton's (2001). Our results facilitate classical statistical inference in random field regression models. Classical statistical inference is numerically far less computationally intensive than the Bayesian inference suggested by Hamilton (2001). Establishing the asymptotic properties of the estimators is non-trivial mainly because the random field model is viewed as an approximation to the true data generating process, and as such we need to deal with misspecification concerns.

Secondly, when imposing additivity, we need to evaluate the validity of such a restriction. There is a vast literature on specification tests for additivity in a nonparametric setting. For example, Barry (1993) developed a test for additivity within a nonparametric model with sampling in a discrete grid. Eubank, Hart, Simpson, and Stefanski (1995) showed the asymptotic performance of a Tukey–type additivity test based on Fourier series estimation of nonparametric models and they proposed a new test, which delivers a consistent estimation of the interaction terms. Chen, Liu, and Tsay (1995) relaxed the restriction of sampling in a grid and proposed an LM-type test for additivity in autoregressive processes. Hong and White (1995), and Wooldridge (1992) proposed several specification tests, which can also be used to test for additivity. Most of these studies on additivity testing rely on nonparametric models. In addition to being computationally demanding, these methods depend very heavily on the bandwidth selection and on the construction of the weight function, which typically are not easily obtained. Our contribution is a new test for additivity that does not depend on a nonparametric estimation procedure. It will be based on a measure of goodness of fit, which will permit a fast computation. The test has a limiting Gaussian distribution and a Monte Carlo study illustrates that it has very good size and power for a large class of additive and nonadditive data generating processes.

A potential drawback of additive models is the large number of parameters to be estimated. We introduce a more restrictive specification, called the "proportional additive random field model" where the weight ratio between any two random fields is kept fixed. This modification substantially improves the computational/numerical aspects of the estimation and testing procedures. Based on extensive Monte Carlo studies, we find that these improvements are achieved without sacrificing predictive efficiency in relation to the generalized additive model.

The organization of the paper is as follows. In Section 2 we present the additive random field model. In Section 3 we establish the asymptotic properties of the estimators of the parameters of the model. In Section 4, we develop a new test for additivity and characterize its asymptotic distribution. In Section 5 we conduct various Monte Carlo experiments to analyze the small sample properties of the predictive accuracy of the estimated additive random field model and the size and power properties of the proposed test for additivity. We conclude in Section 6. All proofs can be found in the mathematical appendix.

# 2 Preliminaries

## 2.1 The additive random field regression model

Let $y_t \in \mathbb{R}$, $\boldsymbol{x}_{t\cdot} \in \mathbb{R}^k$ and consider the model

$$y_t = \mu(\boldsymbol{x}_{t\cdot}) + \widetilde{\epsilon}_t, \tag{1}$$

where $\widetilde{\epsilon}_t$ is a sequence of independent and identically $N(0, \tilde{\sigma}^2)$ distributed random variables and $\mu(\cdot) : \mathbb{R}^k \to \mathbb{R}$ is a random function of a $k \times 1$ vector $\boldsymbol{x}_{t\cdot}$, which is assumed to be deterministic.[2] The vector of explanatory variables is partitioned as $(\boldsymbol{x}_{t1\cdot}, \ldots, \boldsymbol{x}_{tI\cdot})'$, where $\boldsymbol{x}_{ti\cdot} \in \mathbb{R}^{k_i}$ and $\sum_{i=1}^{I} k_i = k$.[3] Model (1) is called an $I$-dimensional *additive random field model* if the conditional mean of $y_t$, i.e. $\mu(\boldsymbol{x}_{t\cdot})$, has a linear component and a stochastic additive nonlinear component such as

$$\mu(\boldsymbol{x}_{t\cdot}) = \boldsymbol{x}_{t\cdot}\boldsymbol{\beta} + \sum_{i=1}^{I} \tilde{\lambda}_i m_i(\boldsymbol{g}_i \odot \boldsymbol{x}_{ti\cdot}), \tag{2}$$

for $i = 1, 2, ..., I$, where $\tilde{\lambda}_i \in \mathbb{R}_+$, $\boldsymbol{g}_i \in \mathbb{R}_+^{k_i}, I \leq k$, and for any choice of $\boldsymbol{z} \in \mathbb{R}^{k_i}$, where $m_i(\boldsymbol{z})$ is a realization of a random field. Model (2) is said to be fully additive when $I = k$; or partially additive when $I < k$. For example, suppose that the true functional form of the conditional mean is given by

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t1}^2 + \beta_3 \sin(x_{t2} x_{t3}) + \widetilde{\epsilon}_t.$$

This model is partially additive in $x_{t1}$. Individual random fields approximate the nonlinear components, i.e. $\tilde{\lambda}_1 m_1(g_1 \cdot x_{t1}) \approx \beta_2 x_{t1}^2$ and $\tilde{\lambda}_2 m_2(g_2 \cdot x_{t2}, g_3 \cdot x_{t3}) \approx \beta_3 \sin(x_{t2} x_{t3})$, for $I = 2, k_1 = 1$, and $k_2 = 2$. Each of the $I$ random fields is assumed to have the following distribution

$$m_i(\boldsymbol{z}) \sim N(0, 1), \tag{3}$$

$$\mathrm{E}(m_i(\boldsymbol{z}) m_j(\boldsymbol{w})) = \begin{cases} \boldsymbol{H}_i(h) & \text{if} \quad i = j \\ \mathbf{O}_T & \text{if} \quad i \neq j \end{cases}, \tag{4}$$

where $h$ is the Euclidean distance $h \equiv \frac{1}{2}[(\boldsymbol{z} - \boldsymbol{w})'(\boldsymbol{z} - \boldsymbol{w})]^{\frac{1}{2}}$.[4] The realization of $m_i(\cdot)$ for $i = 1, 2, ..., I$ is considered predetermined and independent of $\{\boldsymbol{x}_{1\cdot}, ..., \boldsymbol{x}_{T\cdot}, \widetilde{\epsilon}_1, ..., \widetilde{\epsilon}_T\}$. The $i$'th covariance matrix $\boldsymbol{H}_i(h)$ is defined as

$$\boldsymbol{H}_i(h) = \begin{cases} G_{k-1}(h, 1)/G_{k-1}(0, 1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases}, \tag{5}$$

[2]Without loss of generality we assume that all variables are demeaned.

[3]This partition of $\boldsymbol{x}_{t\cdot}$ is made to simplify the exposition. In general, regressors can enter one or more of the random fields in the additive random field model without altering any of the asymptotic results.

[4]$\boldsymbol{g}$ is a $k \times 1$ vector of parameters and $\odot$ denotes element-by-element multiplication i.e. $\boldsymbol{g} \odot \boldsymbol{x}_t$ is the Hadamard product. $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients.

where $G_k(h,r)$, $0 < h \leq r$ is[5]

$$G_k(h,r) = \int_h^r (r^2 - z^2)^{\frac{k}{2}} dz. \tag{6}$$

When the model is fully additive $\boldsymbol{x}_{ti\cdot}$ becomes a scalar for all $i$ and expression (5) reduces to $\boldsymbol{H}_i(h) = (1-h)\,1(h \leq 1)$.[6,7] Since $m_i(\boldsymbol{z})$ is not observable for any choice of $\boldsymbol{z}$, the functional form of $\mu(\boldsymbol{x}_{t\cdot})$ cannot be observed. Hence, inference about the unknown parameters $(\boldsymbol{\beta}, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I, \boldsymbol{g}_1, \ldots, \boldsymbol{g}_I, \tilde{\sigma}^2)$ must be based on the observed realizations of $y_t$ and $\boldsymbol{x}_{t\cdot}$ only. For this purpose we rewrite model (1) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\boldsymbol{y} \equiv \boldsymbol{y}_T$ is a $T \times 1$ vector with $t$-th element equal to $y_t$, $\boldsymbol{X} \equiv \boldsymbol{X}_T$ is a $T \times k$ matrix with $t$-th row equal to $\boldsymbol{x}_{t\cdot}$, $\boldsymbol{\varepsilon}$ is a $T \times 1$ random vector with $t$-th element equal to $\sum_{i=1}^{I} \tilde{\lambda}_i m_i(\boldsymbol{g}_i \odot \boldsymbol{x}_{ti\cdot}) + \widetilde{\epsilon}_t$, and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}_T, \boldsymbol{H}(\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T)$ where $\boldsymbol{H}(\boldsymbol{\lambda}) \equiv \sum_{i=1}^{I} \lambda_i \boldsymbol{H}_i$ and $(\lambda_i, \sigma) \equiv (\tilde{\lambda}_i^2, \tilde{\sigma}^2)$. To avoid identification issues, the shape parameters $\boldsymbol{g}_i$ in the random field model are assumed to be fixed as in Hamilton (2001). Furthermore, we assume that the components of the vector $(\lambda_1, \lambda_2, ..., \lambda_I, \sigma)'$ are strictly positive.

**Assumption i.** The parameter vector $\boldsymbol{g}_i \in \mathbb{R}_+^{k_i}$ for $i = 1, 2, ..., I$ in the additive random field model (1) is fixed with typical element equal to $g_{j_i} = \frac{1}{2\sqrt{k_i s_{j_i}^2}}$, for $i = 1, \ldots, I$, and $j_i \in [1; k]$, where $s_{j_i}^2 = \frac{1}{T}\sum_{t=1}^{T}(x_{j_i t} - \bar{x}_{j_i})^2$ and $\bar{x}_{j_i}$ is the sample mean of the $j_i$-th explanatory variable.

[5] From Hamilton (2001), $G_k(h,r)$ can be computed recursively from

$$\begin{aligned} G_0(h,r) &= r - h \\ G_1(h,r) &= (\pi/4)r^2 - 0.5h(r^2-h^2)^{1/2} - (r^2/2)\sin^{-1}(h/r) \\ G_k(h,r) &= -\frac{h}{1+k}(r^2-h^2)^{k/2} + \frac{kr^2}{1+k}G_{k-2}(h,r) \end{aligned}$$

for $k = 2, 3, \ldots$.

[6] The correlation between $m(\boldsymbol{z})$ and $m(\boldsymbol{w})$ is given by the volume of the intersection of a $k$ dimensional unit spheroid centered at $\boldsymbol{z}$ and a $k$ dimensional unit spheroid centered at $\boldsymbol{w}$ relative to the volume of a $k$ dimensional unit spheroid. Hence, the correlation between $m(\boldsymbol{z})$ and $m(\boldsymbol{w})$ is zero if the Euclidean distance between $\boldsymbol{z}$ and $\boldsymbol{w}$ is $h \geq 2$.

[7] The reader interested in a critical review on the choice of an appropriate covariance function is referred to Dahl and González-Rivera (2003).

**Assumption ii.** We define $\boldsymbol{\theta} \equiv (\lambda_1, \ldots, \lambda_I, \sigma)' \in \Theta \subseteq \mathbb{R}_+^{I+1}$, where $\Theta$ is a compact parameter space. There exist sufficiently small but positive real numbers $\underline{\boldsymbol{\theta}} = (\underline{\lambda}_1, \ldots, \underline{\lambda}_I, \underline{\sigma})'$ and sufficiently large positive real numbers $\overline{\boldsymbol{\theta}} = (\overline{\lambda}_1, \ldots, \overline{\lambda}_I, \overline{\sigma})'$, such that $\lambda_i \in [\underline{\lambda}_i, \overline{\lambda}_i]$ for $\forall i = 1, \ldots, I$ and $\sigma \in [\underline{\sigma}, \overline{\sigma}]$.

We focus on deriving the limiting properties of the maximum likelihood estimators of $\boldsymbol{\theta} = (\lambda_1, \lambda_2, ..., \lambda_I, \sigma)'$ (the parameters of the nonlinear part of the model). The vector $\boldsymbol{\beta}$ will be considered as a nuisance parameter vector, which can be estimated consistently by ordinary least squares by simply ignoring the possibly nonlinear part of the model (independently of whether this part is additive or not). Model (7) can be viewed as a generalized least squares representation where $\boldsymbol{\varepsilon}$ has a non-spherical covariance function $\boldsymbol{C} = \boldsymbol{H}(\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T$. We can write the average log-likelihood function of $\boldsymbol{\varepsilon}$ as (apart from a constant term)

$$Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{2T} \ln \det \boldsymbol{C} - \frac{1}{2T} (\boldsymbol{y} - \boldsymbol{X\beta})' \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}). \tag{8}$$

## 2.2 The data generating process

The random field model is an approximation to the true functional form of the conditional mean. We assume that there is a data generating mechanism that needs to be discovered. The important question is that of "representability" of the true functional form through some function of the covariance function of the random field. We follow similar arguments as in Hamilton (2001). We assume that $y_t$ is generated according to the process

$$y_t = \psi(\boldsymbol{x}_{t\cdot}) + e_t, \tag{9}$$

for $t = 1, 2, ..., T$, where $\psi : \mathbb{R}^k \to \mathbb{R}$ is given by the additive function

$$\psi(\boldsymbol{x}_{t\cdot}) = \boldsymbol{x}_{t\cdot}\boldsymbol{\alpha} + \sum_{i=1}^{I} l_i(\boldsymbol{x}_{ti\cdot}). \tag{10}$$

In addition, the following assumptions are imposed.

**Assumption 1: The sequence $\{\boldsymbol{x}_t\}$ is dense.** The deterministic sequence $\{\boldsymbol{x}_{ti\cdot}\}$, with $\boldsymbol{x}_{ti\cdot} \in A_i$ and $A_i = A_1 \times A_2 \times ... \times A_{k_i}$ a closed rectangular subset of $\mathbb{R}^{k_i}$ and $\lambda \in \Gamma_0$, is said to be *dense* for $A_i$ uniformly on the compact space $A_i \times \Gamma_0 \subset \mathbb{R}^{k_i} \times \mathbb{R}$ if there exists a continuous $f_i : A_i \to \mathbb{R}$ such that $f_i(\boldsymbol{x}_{ti\cdot}) > 0$ for all $i$ and $\boldsymbol{x}_{ti\cdot}$ and such that for any $\epsilon > 0$ and any continuous $\phi_i : A_i \times A_i \times \Gamma_0 \to \mathbb{R}$ there exists an $N$ such that for all $T \geq N$,

$$\sup_{A_i \times \Gamma_0} \left| \frac{1}{T} \sum_{s=1}^{T} \phi_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}_{si\cdot}; \lambda) - \int_{A_i} \phi_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}; \lambda) f_i(\boldsymbol{x}) d\boldsymbol{x} \right| < \epsilon$$

for all $i = 1, 2, ..., k$.

**Assumption 2: The function $l_i : \mathbb{R}^{k_i} \to \mathbb{R}$ is representable.** Let $A_i$, and $\Gamma_0$ be given as in Assumption 1 and let $l_i : A_i \times \Gamma_0 \to \mathbb{R}$ be an arbitrary continuous function. We say that $l_i(\cdot)$ is representable with respect to $\phi_i(\cdot)$ if there exists a continuous function $f_i : A_i \to \mathbb{R}$ such that

$$l_i(\boldsymbol{x}_{ti\cdot}; \lambda) = \int_{A_i} \phi_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}; \lambda) f_i(\boldsymbol{x}) d\boldsymbol{x}.$$

Assumption 1 is important because it implies that we can write

$$\begin{aligned} l_i(\boldsymbol{x}_{ti\cdot}; \lambda_i) &= \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \phi_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}_{si\cdot}; \lambda_i) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \lambda_i \boldsymbol{H}_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}_{si\cdot}) \tilde{\phi}_i(\boldsymbol{x}_{si\cdot}), \end{aligned}$$

where $\lambda_i$ is given as in Assumption *ii.*, $\boldsymbol{H}_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}_{si\cdot})$ denotes the $(t, s)$ entry in $\boldsymbol{H}_i$ given by (5), and $\tilde{\phi}_i : \mathbb{R}^{k_i} \to \mathbb{R}$ is an arbitrary continuous function. Note, that by varying $\tilde{\phi}_i(\cdot)$, Assumption 2 describes the general class of nonlinear/linear functions for which $l_i(\boldsymbol{x}_{ti\cdot}; \lambda_i)$ is representable in terms of the spherical covariance function. Importantly, Hamilton (2001) shows that Taylor and Fourier sine series expansions are representable under Assumption 2 . We can thus expect the random field model to have good approximation properties over a very broad class of functions. Defining the sample version of $l_i(\cdot)$ as

$$l_{Ti}(\boldsymbol{x}_{ti\cdot}; \lambda_i) = \frac{1}{T} \sum_{s=1}^{T} \lambda_i \boldsymbol{H}_i(\boldsymbol{x}_{ti\cdot}, \boldsymbol{x}_{si\cdot}) \tilde{\phi}_i(\boldsymbol{x}_{is\cdot}), \tag{11}$$

it follows that $\lim_{T\to\infty} l_{Ti}(\boldsymbol{x}_{ti\cdot}; \lambda_i) \to l_i(\boldsymbol{x}_{ti\cdot}; \lambda_i)$ uniformly on $A_i \times \Gamma_0$ for $\forall t$, hereby providing a necessary link between the approximating random field model and the true data generating process. Hamilton (2001) discusses pointwise convergence of $l_{Ti}(\cdot)$ to $l_i(\cdot)$ in $\boldsymbol{x}_{ti\cdot}$ for $\forall t$. Following similar arguments as in Dahl and Qin (2004), we generalize the convergence to uniform convergence.

**Assumption 3: Distribution of $e_t$.** The error term $e_t$ is assumed to be an i.i.d. Gaussian distributed random variable with zero mean and variance $\sigma_e^2$.

**Assumption 4: Limiting behavior of "second moments" of $\boldsymbol{\psi}(\boldsymbol{X})$ and $\boldsymbol{X}$.** Let $\boldsymbol{\psi}(\boldsymbol{X}) = (\psi(\boldsymbol{x}_{1\cdot}), ..., \psi(\boldsymbol{x}_{T\cdot}))'$ where $\psi(\cdot)$ is given by (10). Assume: *i.* $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$ converges to a finite nonsingular matrix. *ii.* $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{\psi}(\boldsymbol{X})'\boldsymbol{\psi}(\boldsymbol{X})$ converges to a finite scalar uniformly in $\boldsymbol{\alpha}$. *iii.* $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{\psi}(\boldsymbol{X})$ converges to a finite $k \times 1$ vector uniformly in $\boldsymbol{\alpha}$.

Assumptions 3 and 4 seem somewhat restrictive but are a consequence of working in Hamilton's (2001) environment and they serve primarily to shorten the proofs. We find it important to establish the asymptotic results under these basic assumptions first before extending Hamilton's (2001) fundamental assumptions. We conjecture that replacing Gaussianity with stationarity, ergodicity, a sufficient number of moment conditions in Assumption 3, and imposing similar conditions on $(y, \boldsymbol{x}')'$ in Assumption 4 would not alter the main results.

# 3 Asymptotics

In this section we establish three important asymptotic results; (1) consistency of the estimator of the conditional mean in model (2); (2) consistency of the maximum likelihood estimators of the parameters in the nonlinear component of the model; and (3) their asymptotic normality. The estimation procedure is a two-stage approach. In the first stage, we estimate the parameters $\boldsymbol{\beta}$ in the linear component of the model by OLS. In the second stage, the parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\beta})$ in the nonlinear part of the model are estimated by maximizing the objective function (8). We establish the asymptotic theory for the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ under the additivity assumption in (10). To this end, we need to derive a set of results on uniform convergence of deterministic functions. In the following subsections, we state the main theorems; their proofs can be found in the Mathematical Appendix.

### 3.1 Consistency of the conditional mean estimator

Let $\boldsymbol{\mu}_T = (\mu(\boldsymbol{x}_{1\cdot}), \mu(\boldsymbol{x}_{2\cdot}), ..., \mu(\boldsymbol{x}_{T\cdot}))'$ and $\boldsymbol{\xi}_T = (\xi_T(\boldsymbol{x}_{1\cdot}), \xi_T(\boldsymbol{x}_{2\cdot}), ..., \xi_T(\boldsymbol{x}_{T\cdot}))'$, where $\xi_T(\boldsymbol{x}_{t\cdot}) = \mathrm{E}\ (\mu(\boldsymbol{x}_{t\cdot})|y_T, \boldsymbol{x}_{T\cdot}, y_{T-1}, \boldsymbol{x}_{T-1\cdot}, ...)$. Under the assumption of joint normality of $\boldsymbol{y}_T$ and $\boldsymbol{\mu}_T$, the expectation of the conditional mean function, conditional on $y_T, \boldsymbol{x}_{T\cdot}, y_{T-1}, \boldsymbol{x}_{T-1\cdot}, ...$, is given by

$$\boldsymbol{\xi}_T = \boldsymbol{X\beta} + \boldsymbol{H}(\boldsymbol{\lambda}) \left(\boldsymbol{H}(\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T\right)^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}).$$

For the misspecified additive random field model (2), the following theorem states that $\boldsymbol{\xi}_T$ is a consistent estimator of $\boldsymbol{\mu}_T$.

**Theorem 1** Let assumptions 1, 2, and 3 hold and define

$$\boldsymbol{L}_T(\boldsymbol{\lambda}^0) = (l_T(\boldsymbol{x}_{1\cdot}; \boldsymbol{\lambda}), l_T(\boldsymbol{x}_{2\cdot}; \boldsymbol{\lambda}), ..., l_T(\boldsymbol{x}_{T\cdot}; \boldsymbol{\lambda}))',$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_I)'$, and

$$\boldsymbol{L}_T(\boldsymbol{\lambda}) = \frac{1}{T} \sum_{i=1}^{I} \lambda_i \boldsymbol{H}_i \widetilde{\boldsymbol{\phi}}_i,$$

for $\widetilde{\boldsymbol{\phi}}_i = \left(\widetilde{\phi}_i(\boldsymbol{x}_{1i\cdot}), \widetilde{\phi}_i(\boldsymbol{x}_{2i\cdot}), ..., \widetilde{\phi}_i(\boldsymbol{x}_{Ti\cdot})\right)'$. Then,

$$\lim_{T\to\infty} \sup_{\Theta\times A^2} \frac{1}{T} \mathrm{E}\left[(\boldsymbol{\xi}_T - \boldsymbol{X\beta} - \boldsymbol{L}_T(\boldsymbol{\lambda}))' (\boldsymbol{\xi}_T - \boldsymbol{X\beta} - \boldsymbol{L}_T(\boldsymbol{\lambda}))\right] \to 0. \tag{12}$$

Theorem 1 generalizes Theorem 4.7 in Hamilton (2001) in two directions: First, it establishes that the convergence is uniform on $\Theta \times A^2$, and secondly, it applies to a richer class of random field regression models. From Theorem 1, we can derive some additional results regarding uniform convergence of the following deterministic sequences. These results will play an important role later on, when we establish the asymptotic distribution of the estimators of the parameters.

**Corollary 1** Let assumptions i., ii., and 1 to 4 hold. Then,

$$\lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T}\left(\boldsymbol{\psi}(\boldsymbol{X})-\boldsymbol{X\beta}\right)'\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-2}\left(\boldsymbol{\psi}(\boldsymbol{X})-\boldsymbol{X\beta}\right) \rightarrow 0, \quad (13)$$

$$\lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T}\operatorname{tr}\left[\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\right] \rightarrow 0, \quad (14)$$

$$\lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T}\operatorname{tr}\left[\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\right]\rightarrow 0. \quad (15)$$

**Corollary 2** Let assumptions i., ii., and 1 to 4 hold. Then, for all $i,j=1,2,...,I$,

$$\lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T}\operatorname{tr}\left[\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}_i\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}_j\right]\rightarrow 0.$$

## 3.2 Consistency of the parameter estimators

Recall the average likelihood function (8) associated with model (7) is given as,

$$Q_T\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)=-\frac{1}{2T}\ln\det\boldsymbol{C}-\frac{1}{2T}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)'\boldsymbol{C}^{-1}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right).$$

In this section, we will establish consistency of the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}\equiv(\lambda_1,\ldots,\lambda_I,\sigma)'$. The consistency of $\widehat{\boldsymbol{\beta}}$, i.e., $\widehat{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}^*$, is shown by Dahl and Qin (2004). We proceed as follows: first, we prove that the expectation of $Q_T\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)$ uniformly converges to a limiting function $Q^*\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)$. In the second stage, we prove the main theorem that states $\hat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}^*$, where $\boldsymbol{\theta}^*$ maximizes $Q^*(\boldsymbol{\theta},\boldsymbol{\beta}^*)$. For these results to hold we need a set of auxiliary propositions on uniform convergence of deterministic sequences.

Specifically, let us define $R_T\equiv\frac{1}{T}\log\det\boldsymbol{C}$, and $U_T\equiv\frac{1}{T}\left(\boldsymbol{\psi}\left(\boldsymbol{x}\right)-\boldsymbol{X\beta}\right)'\boldsymbol{C}^{-1}\left(\boldsymbol{\psi}\left(\boldsymbol{x}\right)-\boldsymbol{X\beta}\right)$, with corresponding differentials $\mathrm{D}^l_{i_1,\ldots,i_l}R_T=\frac{\partial^{i_1+\ldots+i_l}R_T}{(\partial\lambda_1)^{i_1}\ldots(\partial\lambda_l)^{i_l}}$, and $\mathrm{D}^l_{i_1,\ldots,i_l}U_T=\frac{\partial^{i_1+\ldots+i_l}U_T}{(\partial\lambda_1)^{i_1}\ldots(\partial\lambda_l)^{i_l}}$, where $i_1,\ldots,i_l=0,1,\ldots,I$, $l=1,2,...$, $\mathrm{D}_0R_T=\frac{\partial R_T}{\partial\sigma}$, and $\mathrm{D}_0U_T=\frac{\partial U_T}{\partial\sigma}$. Furthermore, define the following limits $\mathrm{D}^l_{i_1,\ldots,i_l}R=\lim_{T\to\infty}\mathrm{D}^l_{i_1\ldots,i_l}R_T$ and $\mathrm{D}^l_{i_1,\ldots,i_l}U=\lim_{T\to\infty}\mathrm{D}_{i_1,\ldots,i_l}U_T$, and $\mathrm{D}_0R=\lim_{T\to\infty}\mathrm{D}_0R_T$, $\mathrm{D}_0U=\lim_{T\to\infty}\mathrm{D}_0U_T$. The following propositions establish that, under the assumptions stated above, the sequences of $R_T$ and $U_T$ as well as their differentials will converge uniformly.

**Proposition 1** Given assumptions i., ii., and 1 to 4, $\mathrm{D}_iR$, $\mathrm{D}^2_{i,j}R$ and $\mathrm{D}_0U$ are equal to zero uniformly on $\Theta\times B\times A$ for all $i,j=1,...,I$.

**Proposition 2** Given assumptions i., ii., and 1 to 4, all of the function sequences $\left\{ \mathrm{D}^l_{i_1 \ldots i_l} R_T \right\}_T$ for $l = 1, 2, ...$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$, are equicontinuous. Furthermore, each function sequence converges uniformly on $\Theta \times A$ as $T \to \infty$.

**Proposition 3** Given assumptions i., ii., and 1 to 4, all of the function sequences $\left\{ \mathrm{D}^l_{i_1 \ldots i_l} U_T \right\}_T$ for $l = 1, 2, ..$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$, are equicontinuous. Furthermore, each function sequence converges uniformly on $\Theta \times B \times A$ as $T \to \infty$.

Next, we define the following second order sample moment matrices

$$
\begin{aligned}
M_{x_{\cdot i} x_{\cdot j}} (\boldsymbol{\theta}) &\equiv \frac{1}{T} \boldsymbol{x}'_{\cdot i} \left( \boldsymbol{H} (\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T \right)^{-1} \boldsymbol{x}_{\cdot j}, \\
M_{\psi x_{\cdot i}} (\boldsymbol{\theta}) &\equiv \frac{1}{T} \boldsymbol{\psi}(\boldsymbol{X})' \left( \boldsymbol{H} (\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T \right)^{-1} \boldsymbol{x}_{\cdot i}, \\
M_{\psi \psi} (\boldsymbol{\theta}) &\equiv \frac{1}{T} \boldsymbol{\psi}(\boldsymbol{X})' \left( \boldsymbol{H} (\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T \right)^{-1} \boldsymbol{\psi}(\boldsymbol{X}),
\end{aligned}
$$

where $\boldsymbol{x}_{\cdot i} = (x_{1i}, \ldots, x_{Ti})'$ for $i, j = 1, 2, ..., k$. Then, we have the following result.

**Proposition 4** Given assumptions i., ii., and 1 to 4, all of the function sequences $\left\{ \mathrm{D}^l_{i_1 \ldots i_l} M_{x_{\cdot i} x_{\cdot j}} \right\}_T$, $\left\{ \mathrm{D}^l_{i_1 \ldots i_l} M_{\psi x_{\cdot i}} \right\}_T$, and $\left\{ \mathrm{D}^l_{i_1 \ldots i_l} M_{\psi \psi} \right\}_T$ for $l = 1, 2, ..$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$, are equicontinuous. Furthermore, each function sequence converges uniformly on $\Theta \times B \times A$ as $T \to \infty$, and

$$
\lim_{T \to \infty} \sup_{\Theta \times B \times A} \mathrm{D}_0 M_{x_{\cdot i} x_{\cdot j}} (\boldsymbol{\theta}) \to 0, \tag{16}
$$

$$
\lim_{T \to \infty} \sup_{\Theta \times B \times A} \mathrm{D}_0 M_{\psi x_{\cdot i}} (\boldsymbol{\theta}) \to 0, \tag{17}
$$

$$
\lim_{T \to \infty} \sup_{\Theta \times B \times A} \mathrm{D}_0 M_{\psi \psi} (\boldsymbol{\theta}) \to 0. \tag{18}
$$

Next, consider the objective function (8). Let $Q^*_T(\boldsymbol{\theta}, \boldsymbol{\beta}) = \mathrm{E} \left( Q_T (\boldsymbol{\theta}, \boldsymbol{\beta}) \right)$. Substituting $\boldsymbol{y} = \boldsymbol{\psi} (\boldsymbol{X}) + \boldsymbol{e}$ in (8) and taking expectations, we have

$$
\begin{aligned}
Q^*_T(\boldsymbol{\theta}; \boldsymbol{\beta}) \quad = \quad & -\frac{1}{2T} \log \det \boldsymbol{C} + \\
& -\frac{1}{2T} \left( \boldsymbol{\psi} (\boldsymbol{X}) - \boldsymbol{X} \boldsymbol{\beta} \right)' \boldsymbol{C}^{-1} \left( \boldsymbol{\psi} (\boldsymbol{X}) - \boldsymbol{X} \boldsymbol{\beta} \right) - \frac{1}{2T} \sigma^2_e \, \mathrm{tr} \left( \boldsymbol{C}^{-1} \right),
\end{aligned} \tag{19}
$$

which can also be written according to the notation established previously as

$$
Q^*_T (\boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{2} R_T - \frac{1}{2} U_T - \frac{\sigma^2_e}{2} \mathrm{D}_0 R_T.
$$

**Theorem 2** Let assumptions i., ii., and 1 to 4 hold. Then

$$\lim_{T\to\infty} \sup_{\Theta\times B\times A} |Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) - Q_T^*(\boldsymbol{\theta},\boldsymbol{\beta})| \xrightarrow{p} 0, \tag{20}$$

and

$$\lim_{T\to\infty} \sup_{\Theta\times B\times A} |Q_T^*(\boldsymbol{\theta},\boldsymbol{\beta}) - Q^*(\boldsymbol{\theta},\boldsymbol{\beta})| \to 0, \tag{21}$$

where $Q^*(\boldsymbol{\theta},\boldsymbol{\beta}) = -\frac{1}{2}R - \frac{1}{2}U - \frac{\sigma_e^2}{2}\mathrm{D}_0 R$.

By Theorem 2, it can be seen that $Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) \xrightarrow{p} Q^*(\boldsymbol{\theta},\boldsymbol{\beta})$. The unique existence of $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ that maximizes $Q^*(\boldsymbol{\theta},\boldsymbol{\beta})$ for a given $\boldsymbol{\beta}$ is guaranteed in the following theorem.

**Theorem 3** Let $U(\boldsymbol{\theta})$ be a convex function of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\operatorname{tr} C^{-1}C^{-1} < 2\sigma_e^2 \operatorname{tr} C^{-1}C^{-1}C^{-1}$. Then $Q^*(\boldsymbol{\theta};\boldsymbol{\beta})$ is a concave function of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and has a unique maximizer $\boldsymbol{\theta}^*$ in $\boldsymbol{\Theta}$. A necessary condition for concavity of $Q^*(\boldsymbol{\theta};\boldsymbol{\beta})$ in $\sigma$ is given by the condition $\sigma \leq 2\sigma_e^2$.

In the following sections, we will derive a consistent estimator of $\sigma_e^2$ that will permit actual empirical verification of the necessary and sufficient conditions for identification. Now, we can establish consistency of the estimators of the parameters.

**Theorem 4** Let assumptions i., ii., and 1 to 4 hold, and let $\hat{\boldsymbol{\beta}}$ be the first-stage OLS estimator of the linear parameter $\boldsymbol{\beta}$ such that $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^*$. Then

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*,$$

where $\hat{\boldsymbol{\theta}}$ is the two-stage estimator and $\boldsymbol{\theta}^*$ maximizes $Q^*(\boldsymbol{\theta},\boldsymbol{\beta}^*)$.

### 3.3 Asymptotic normality

In this section we establish the asymptotic distribution of $\hat{\boldsymbol{\theta}}$. In the first stage estimation, we compute the OLS estimator $\hat{\boldsymbol{\beta}}$, which is taken as a nuisance parameter in the second stage. In the second stage, we estimate $\boldsymbol{\theta}$. The estimation of $\boldsymbol{\beta}$ will introduce further variability in the estimation of $\boldsymbol{\theta}$ that will be reflected in the moments of its asymptotic distribution. We begin with the derivation of the asymptotic distribution of the score function. The objective function associated with the first stage OLS estimation is given as

$$m_T(\boldsymbol{\beta}) = -\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \boldsymbol{x}_{t\cdot}\boldsymbol{\beta}\right)^2. \tag{22}$$

We stack the gradient vector of (22) and the score vector associated with (8) in a $(I+1+k)\times 1$ vector as

$$\boldsymbol{g}_T(\boldsymbol{\theta},\boldsymbol{\beta}) = \left(\mathrm{D}_{\boldsymbol{\theta}}Q_T(\boldsymbol{\theta},\boldsymbol{\beta})', \mathrm{D}_{\boldsymbol{\beta}}m_T(\boldsymbol{\beta})'\right)', \tag{23}$$

where

$$\begin{aligned}\mathrm{D}_{\boldsymbol{\theta}}Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) &= \begin{pmatrix} \mathrm{D}_1 Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) \\ \vdots \\ \mathrm{D}_I Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) \\ \mathrm{D}_0 Q_T(\boldsymbol{\theta},\boldsymbol{\beta}) \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_1\right) + \frac{1}{2T}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}\boldsymbol{H}_1\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \\ \vdots \\ -\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_I\right) + \frac{1}{2T}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}\boldsymbol{H}_I\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \\ -\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\right) + \frac{1}{2T}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \end{pmatrix}\end{aligned} \tag{24}$$

and

$$\mathrm{D}_{\boldsymbol{\beta}}m_T(\boldsymbol{\beta}) = \frac{2}{T}\left(\sum_{t=1}^{T}x_{t1}(y_t - \boldsymbol{x}_{t\cdot}\boldsymbol{\beta}), \ldots, \sum_{t=1}^{T}x_{tk}(y_t - \boldsymbol{x}_{t\cdot}\boldsymbol{\beta})\right)'.$$

The following proposition provides us with the exact variance of $\sqrt{T}\boldsymbol{g}_T$.

**Proposition 5** Let assumptions i., ii., and 1 to 4 hold. Then

$$\begin{aligned}\operatorname{cov}\left(\sqrt{T}\mathrm{D}_{i_1}Q_T(\boldsymbol{\theta},\boldsymbol{\beta}), \sqrt{T}\mathrm{D}_{i_2}Q_T(\boldsymbol{\theta},\boldsymbol{\beta})\right) &= \frac{\sigma_e^4}{2}\frac{1}{T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\right) + \\ &\quad \sigma_e^2\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{c}, \quad (25) \\ \operatorname{cov}\left(\sqrt{T}\mathrm{D}_{i_1}Q_T(\boldsymbol{\theta},\boldsymbol{\beta}), \sqrt{T}\mathrm{D}_{j_1}m_T(\boldsymbol{\beta})\right) &= -2\sigma_e^2\mathrm{D}_{i_1}\left(M_{x_{j_1}\psi} - \sum_{j=1}^{k}M_{x_{j_1}x_j}\beta_j\right), \quad (26) \\ \operatorname{cov}\left(\sqrt{T}\mathrm{D}_{j_1}m_T(\boldsymbol{\beta}), \sqrt{T}\mathrm{D}_{j_2}m_T(\boldsymbol{\beta})\right) &= \frac{4\sigma_e^2}{T}\boldsymbol{x}_{\cdot j_1}'\boldsymbol{x}_{\cdot j_2}, \quad (27)\end{aligned}$$

for all $i_1, i_2 = 0, 1, 2, ..., I$ and $j_1, j_2 = 1, 2, ..., k$. Furthermore, (25) (26) and (27) converge uniformly on $\Theta \times B \times A$ as $T \to \infty$.

The asymptotic distribution of $\sqrt{T}\boldsymbol{g}_T$ is a result of the following theorem.

**Theorem 5** Let $\boldsymbol{g}_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ be given by (23) and let assumptions i., ii., and 1 to 4 hold. Then, as $T \to \infty$,

$$\sqrt{T}\boldsymbol{g}_T\left(\boldsymbol{\theta}^*, \boldsymbol{\beta}^*\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}\left(\boldsymbol{\theta}^*, \boldsymbol{\beta}^*\right) = [\boldsymbol{\Sigma}^*_{11} \quad \boldsymbol{\Sigma}^*_{12} : \boldsymbol{\Sigma}^{*\prime}_{12} \quad \boldsymbol{\Sigma}^*_{22}]$ and

$$\begin{aligned}
\boldsymbol{\Sigma}^*_{11} &= \frac{\sigma_e^4}{2} \lim_{T\to\infty} \frac{1}{T} \operatorname{tr}(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}), \\
\boldsymbol{\Sigma}^*_{12} &= -2\sigma_e^2 \lim_{T\to\infty} \mathrm{D}_{i_1} M_{x_{j_1}\psi}, \\
\boldsymbol{\Sigma}^*_{22} &= 4\sigma_e^2 \lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{X}.
\end{aligned}$$

Then, the asymptotic normality of $\hat{\boldsymbol{\theta}}$ can be established in the following Theorem 6.

**Theorem 6** Let assumptions i., ii., and 1 to 4 hold. Define $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\beta}')'$, let $Q_T(\boldsymbol{\zeta})$ and $m_T(\boldsymbol{\beta})$ be given by (8) and (22) respectively, and let $\boldsymbol{\Sigma}^*$ be defined as in Theorem 5. Then

$$\sqrt{T}\left(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}^*\right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{M}^*), \tag{28}$$

where $\boldsymbol{M}^* = \boldsymbol{G}^{*-1}\boldsymbol{\Sigma}^*\boldsymbol{G}^{*-1\prime}$, for $\boldsymbol{G}^* = \lim_{T\to\infty} \boldsymbol{G}_T(\boldsymbol{\zeta}^*)$, and

$$\begin{aligned}
\boldsymbol{G}_T(\boldsymbol{\zeta}) &= \mathrm{D}_{\boldsymbol{\zeta}}\boldsymbol{g}_T(\boldsymbol{\zeta}) \\
&= \begin{pmatrix} \mathrm{D}^2_{\boldsymbol{\theta\theta}} Q_T(\boldsymbol{\zeta}) & \mathrm{D}^2_{\boldsymbol{\theta\beta}} Q_T(\boldsymbol{\zeta}) \\ \mathbf{0}_{k\times I} & \mathrm{D}^2_{\boldsymbol{\beta\beta}} m_T(\boldsymbol{\beta}) \end{pmatrix}.
\end{aligned}$$

For the parameter of interest $\boldsymbol{\theta}$, notice that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^*) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{M}^*_{11}),$$

where $\boldsymbol{M}_{11}$ is the upper left corner of the matrix $\boldsymbol{M}$, which is equal to

$$\begin{aligned}
\boldsymbol{M}^*_{11} &= (\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q^*(\boldsymbol{\zeta}^*))^{-1}\boldsymbol{\Sigma}^*_{11}(\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q^*(\boldsymbol{\zeta}^*))^{-1} + \\
&\quad \sigma_e^2(\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q^*(\boldsymbol{\zeta}^*))^{-1}\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q^*(\boldsymbol{\zeta}^*)(\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{X})^{-1}(\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q^*(\boldsymbol{\zeta}^*))'(\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q^*(\boldsymbol{\zeta}^*))^{-1}.
\end{aligned} \tag{29}$$

A consistent estimator of the variance $\boldsymbol{M}_{11}^{*}$ is obtained by substituting $\sigma_e^2$ and $\boldsymbol{\psi}(\boldsymbol{X})$ with their respective consistent estimators $\hat{\sigma}_e^2$ and $\hat{\mu}$. Establishing the asymptotic normality and consistency of the estimators of the parameters of the additive random field model in Theorem 6 is useful for a number of reasons. These asymptotic results can be used to construct confidence bands for the estimated conditional mean function without the use of the computer intensive Bayesian methods suggested by Hamilton (2001). Such bands are extremely useful in testing hypothesis about the functional form of the data generating process. Another important application of the asymptotic distribution is that it can be used to establish the asymptotic distribution of a very simple parametric test for additivity, which is discussed in the following section.

# 4 Testing for additivity

There are several additivity tests in the nonparametric literature, for example Chen, Liu, and Tsay (1995), and in the parametric literature, for example, Barry (1993), and Eubank, Hart, Simpson, and Stefanski (1995). The last two tests suffer from a restrictive sampling scheme because the explanatory variables are sampled on a grid, and the first two rely on the selection of a data-dependent bandwidth. We propose an additive test within the framework of a parametric random field model that does not depend on either a sampling scheme or bandwidth selection. As a by-product, we also propose a "new" consistent estimator of $\sigma_e^2$.

The null hypothesis is that the data generating process is given by (9) and (10) described in section 2.2. The test for additivity will assess the goodness of fit of the misspecified additive random field model $\mu(\boldsymbol{x}_{t\cdot}) = \boldsymbol{x}_{t\cdot}\boldsymbol{\beta} + \sum_{i=1}^{I} \tilde{\lambda}_i m_i(\boldsymbol{g}_i \odot \boldsymbol{x}_{ti\cdot})$. A direct measure of the goodness of fit of the additive random field model is the estimation error. By Theorem 1, the mean squared error in the estimation of the conditional mean converges to 0 uniformly, if the data generating process is additive in the regressors. Using this result, the test is based on the estimation error of the observed response $y_t$, which is the sum of the true conditional mean $\psi(\boldsymbol{x}_{t\cdot})$ and the error term $e_t$.

**Theorem 7** Let assumptions i., ii., and 1 to 4 hold. Define the estimation error as $\hat{\varepsilon} = \boldsymbol{y} - \left(\boldsymbol{X\beta} + \boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right)$, where $\boldsymbol{C} = (\boldsymbol{H}(\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T)$. Then

$$\frac{1}{\sqrt{T}}\left(\hat{\varepsilon}'\hat{\varepsilon} + \sigma^2 T \mathrm{D}_0 U_T + \sigma^2\sigma_e^2 T \mathrm{D}_{00}^2 R_T\right) \overset{a}{\sim} \mathrm{N}\left(0, -\frac{1}{3}\sigma^4\sigma_e^4 \mathrm{D}_{0000}^4 R\right), \quad (30)$$

for any $(\boldsymbol{\theta}', \boldsymbol{\beta}')' \in \Theta \times B$.

Note that the quantities $\mathrm{D}_0 U_T(\boldsymbol{\theta}), \mathrm{D}_{00}^2 R_T(\boldsymbol{\theta})$, and $\mathrm{D}_{0000}^4 R(\theta)$ are all nonpositive. It follows from Theorem 7 that

$$\lim_{T\to\infty} \sup_{\Theta\times B} \left| \frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon} + \sigma^2\sigma_e^2 \mathrm{D}_{00}^2 R_T \right| = 0 \quad (31)$$

because the deterministic term $\sigma^2 \mathrm{D}_0 U_T$ converges to zero by Proposition 1. Furthermore, it follows, from (31), that a consistent estimator of $\sigma_e^2$ can be constructed as described in the following Corollary.

**Corollary 3** Let assumptions i., ii., and 1 to 4 hold. Let $\hat{\boldsymbol{\varepsilon}}$ be defined as in Theorem 7. Then, as $T \to \infty$,

$$\frac{\frac{1}{T}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{-\hat{\sigma}^2 \mathrm{D}_{00}^2 R_T} \xrightarrow{p} \sigma_e^2, \quad (32)$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\lambda}}', \hat{\sigma})'$, and $\widehat{\boldsymbol{\beta}}$ are the consistent two-stage estimators.

Note that a careful examination of the proof of Theorem 7 indicates that the results in (31) and (32) do not specifically rely on the additive random field framework. They hold for any general random field model. Therefore, the consistent estimator of $\sigma_e^2$ can also be obtained by fitting a random field model to a nonadditive or additive model. One only needs to replace the corresponding quantities, for example $\hat{\varepsilon}$ and $\mathrm{D}_{00}^2 R_T$, with their sample counterparts in the random field model under consideration. Based on Theorem 7, we propose the following test statistic for additivity.

**Corollary 4** Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\lambda}}', \hat{\sigma})', \hat{\boldsymbol{\beta}}$, and $\hat{\sigma}_e^2$ be consistent estimators. Under the null hypothesis of an additive data generating process, the statistic $ResT_{DGQ}$ is asymptotically standard normally distributed, i.e.,

$$ResT_{DGQ} \equiv \frac{\frac{1}{T}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + \hat{\sigma}^2\hat{\sigma}_e^2 \mathrm{D}_{00}^2 R_T}{\sqrt{-\frac{2}{3T}\hat{\sigma}^4\hat{\sigma}_e^2 \mathrm{D}_{000}^3 U_T - \frac{1}{3T}\hat{\sigma}^4\hat{\sigma}_e^4 \mathrm{D}_{0000}^4 R_T}} \xrightarrow{d} \mathrm{N}(0,1). \quad (33)$$

We will refer to $ResT_{DGQ}$ as the *residual based test statistic*. It should be noticed that we have added the asymptotically vanishing term $\mathrm{D}^3_{000}U_T$ to the variance and deleted the asymptotic vanishing term $\mathrm{D}_0U_T$ from the mean to obtain better finite sample properties. In particular, the additional term in the variance corrects the tendency of the test to over-reject a true null hypothesis in finite samples. Note that, by Proposition 1, $\lim_{T\to\infty}\mathrm{D}^2_{ij}R_T \to 0$ and $\lim_{T\to\infty}\mathrm{D}_0U_T \to 0$ uniformly. While the first term relies solely on the covariance matrix of the random field, it is the second term that actually reflects the approximation of the true data generating process (hereafter, DGP) by the additive random field model. In other words, when the true DGP is not additive, the additive random field model will not be a good approximation, that is, $\lim_{T\to\infty}\mathrm{D}_0U_T \nrightarrow 0$. However, $\mathrm{D}_0U_T$ is bounded as

$$\mathrm{D}_0U_T \leq \frac{1}{T\sigma^2}(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X\beta})'(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X\beta}).$$

We can assume that when the true DGP is not additive then $\mathrm{D}_0U_T = O(1)$ for every $(\boldsymbol{\theta}', \boldsymbol{\beta}')' \in \Theta \times B$, thus under the alternative hypothesis, $ResT_{DGQ} = O_p(\sqrt{T})$.[8] This property agrees with most of the specification tests in the literature, see, e.g., Wooldridge (1992), providing an argument for the consistency of the test statistic $ResT_{DGQ}$. Generally, under a true alternative, $ResT_{DGQ}$ gives a large value. The test statistic $ResT_{DGQ}$ is not yet directly computable due to the unknown but asymptotically vanishing term $\mathrm{D}^3_{000}U_T$. Since the estimator of the conditional mean $\boldsymbol{\psi}(\boldsymbol{X})$ is given by $\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{H}(\hat{\boldsymbol{\lambda}})\boldsymbol{C}^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ and

$$\lim_{T\to\infty} \sup_{\Theta\times B} |\mathrm{D}^3_{000}\hat{U}_T| = 0, \tag{34}$$

where $\hat{U}_T = \frac{1}{T}(\hat{\boldsymbol{\mu}} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{C}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\mu}} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, we suggest replacing $\mathrm{D}^3_{000}U_T$ by $\mathrm{D}^3_{000}\hat{U}_T$ when computing the test statistic $ResT_{DGQ}$. This will not result in a loss of asymptotic power but will improve the small sample properties of the test. Summarizing, the test statistic $ResT_{DGQ}$ can be computed by the following simple 3-step procedure:

[8] This result can be seen by multiplying (57) in the Mathematical Appendix by $\sqrt{T}$ and noticing that the resulting two last terms on the right hand side will be bounded in probability by a central limit theorem. After having multiplied by $\sqrt{T}$, the first term on the right hand side of (57), however, will be $O(\sqrt{T})$ when $\mathrm{D}_0U_T = O(1)$.

Step 1 Fit a random field model with only one comprehensive random field as in Hamilton (2001), Dahl and González-Rivera (2002), or Dahl and Qin (2004). Compute the consistent estimator $\hat{\sigma}_e^2$ as in (32) based on this random field model.

Step 2 Fit the additive random field model (2), and then use the $\hat{\sigma}_e^2$ obtained in Step 1 to calculate the test statistic

$$\widehat{ResT}_{DGQ} = \frac{\frac{1}{T}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + \hat{\sigma}^2\hat{\sigma}_e^2 \mathrm{D}_{00}^2 R_T}{\sqrt{-\frac{2}{3T}\hat{\sigma}^4\hat{\sigma}_e^2\mathrm{D}_{000}^3\hat{U}_T - \frac{1}{3T}\hat{\sigma}^4\hat{\sigma}_e^4\mathrm{D}_{0000}^4 R_T}},$$

by plugging in the two-stage estimates $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\theta}}$.

Step 3 Reject the null if $\left|\widehat{ResT}_{DGQ}\right| > \Phi^{-1}\left(1-\frac{1}{2}\alpha\right)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution and $\alpha$ denotes the nominal level of the test.

The auxiliary random field model in Step 1 is only used to estimate the variance $\sigma_e^2$ of the true disturbance term. This objective can also be achieved by other consistent estimates of $\sigma_e^2$, such as the nonparametric estimator suggested by Hall, Kay, and Titterington (1990) that enjoys the parametric rate of convergence. We notice that a too high estimate of $\sigma_e^2$ usually results in a less powerful but more conservative test. In the Monte Carlo experiment section we will discuss the degree of robustness of the residual based test to "overestimates" of $\sigma_e^2$ within a large class of nonadditive data generating processes.

We conclude this section by discussing why it is not appropriate to construct a nested additivity test based on the likelihood function, such as an LM-type test as proposed by Hamilton (2001) and Dahl and González-Rivera (2003) to detect neglected nonlinearity of a more general form. To perform a nested additivity test, we need under the alternative hypothesis a general model that includes a comprehensive random field, i.e., $y_t = \boldsymbol{x}_{t\cdot}\boldsymbol{\beta} + \sum_{i=1}^{I}\tilde{\lambda}_i m_i(\boldsymbol{x}_{ti\cdot}) + \tilde{\lambda}_h m_h(\boldsymbol{x}_{t\cdot}) + \epsilon_t$, where $m_h(\boldsymbol{x}_{t\cdot})$ denotes the comprehensive random field defined on the compact set $A^k \in \mathbb{R}^k$. Then, an LM-type test for additivity will have a null hypothesis defined as

$$H_0 : \lambda_h \equiv \tilde{\lambda}_h^2 = 0. \tag{35}$$

In this setting, one has to allow the domain of the parameter $\lambda_h$ to be a compact set $A_h \in \mathbb{R}$ containing the origin. On theoretical grounds, the inclusion of the origin in the parameter space invalidates assumption ii. with critical consequences for the validity of the asymptotic results presented. On numerical grounds, we find that, under the null hypothesis (35), the elements of the Hessian matrix $\mathrm{D}^2_{hh}Q_T(\boldsymbol{\theta};\boldsymbol{\beta})$, i.e., the expression $\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_h\boldsymbol{C}^{-1}\boldsymbol{H}_h\boldsymbol{C}^{-1}\boldsymbol{c}$, do not converge. In most cases, this quantity actually explodes!

# 5 Simulation experiments

We perform several Monte Carlo experiments with a wide range of additive and nonadditive data generating processes to evaluate the predictive power of the additive random field model, and the performance of the residual based additivity test. In Table 1, we present sixteen data generating processes; eight with an additive structure (A1 to A8) and eight with a non-additive structure (N1 to N8). The nonlinear specifications are diverse: polynomials, logarithmic, exponential, thresholds, and sine functions, which cover many of the most popular econometric models used in applied work. The explanatory variables $(x_{t1}, x_{t2})'$ are sampled independently from a uniform distribution, i.e., $x_{ti} \sim U(-5, 5)$ for $i = 1, 2$. Models A5-A8 and N5-N8 are adapted from Chen, Liu, and Tsay (1995), with certain modifications of the coefficients to accommodate the uniformly designed sampling domain. We conduct 1000 Monte Carlo replications with a sample size of 100 observations for estimation and 100 out-of-sample points for the one-step ahead prediction.

In this section, we introduce an estimator/algorithm based on a simplified version of the additive random field model, which reduces the computational burden significantly but maintains almost the same predictive accuracy as the additive model described in the previous sections. We call the alternative specification the "proportional additive random field model" and it is given as

$$y_t = \boldsymbol{x}_{t\cdot}\boldsymbol{\beta} + \tilde{\lambda}\sum_{i=1}^{I} c_i m_i(\boldsymbol{g}_i \odot \boldsymbol{x}_{ti\cdot}) + \tilde{\epsilon}_t, \tag{36}$$

where $c_i$ is a predetermined proportional weight of the $i$-th random field, $\tilde{\lambda}$ is the total weight of the random field component, and $\tilde{\epsilon}_t \sim \mathrm{IN}(0, \tilde{\sigma}^2)$. Studies of proportional additivity in nonparametric models can be found in Yang (2002) and Carroll, Hardle,

Table 1: Additive and nonadditive data generating processes. It is assumed that $x_{it} \sim i.i.d.U(-5,5)$ for $i = 1,2$ and $e_t \sim \mathrm{N}(0, \sigma_e^2)$.

| Model | True DGP |
|---|---|
| A1 | $y_t = 1 + 0.1x_{t1}^2 + 0.2x_{t2} + e_t$ |
| A2 | $y_t = 0.1x_{t1}^2 + 0.2\ln(x_{t2} + 6) + e_t$ |
| A3 | $y_t = \frac{0.5x_{t1}-1}{x_{t1}+6} - 0.5\exp(0.5x_{t2}) + e_t$ |
| A4 | $y_t = 1.5\sin(x_{t1}) + 2\sin(x_{t2}) + e_t$ |
| A5 | $y_t = 0.5x_{t1} + \sin x_{t2} + e_t$ |
| A6 | $y_t = 0.8x_{t1} - 0.3x_{t2} + e_t$ |
| A7 | $y_t = \exp(-0.5x_{t1}^2)x_{t1} - \exp(-0.5x_{t2}^2)x_{t2} + e_t$ |
| A8 | $y_t = -2x_{t1}1_{\{x_{t1}<=0\}} + 0.4x_{t2}1_{\{x_{t2}>0\}} + e_t$ |
| N1 | $y_t = 0.3x_{t1}^2 x_{t2} + 0.2x_{t2}^2 + e_t$ |
| N2 | $y_t = \sin(x_{t1} + x_{t2}) + e_t$ |
| N3 | $y_t = 1.5\sin(x_{t1} + 0.2x_{t2}) + 2\sin(0.5x_{t1} + x_{t2}) + e_t$ |
| N4 | $y_t = 2 \times 1_{\{x_{t1}+x_{t2}<-8\}} + 1_{\{-8\leq x_{t1}+x_{t2}<-6\}} - 1_{\{-6\leq x_{t1}+x_{t2}<-4\}} +$ $2 \times 1_{\{-4\leq x_{t1}+x_{t2}<-2\}} + 1_{\{-2\leq x_{t1}+x_{t2}<0\}} - 1_{\{0\leq x_{t1}+x_{t2}<2\}} + 2 \times 1_{\{2\leq x_{t1}+x_{t2}<4\}} +$ $1_{\{4\leq x_{t1}+x_{t2}<6\}} - 1_{\{6\leq x_{t1}+x_{t2}<8\}} + 2 \times 1_{\{x_{t1}+x_{t2}\geq 8\}} + e_t$ |
| N5 | $y_t = x_{t1}\sin(x_{t2}) + e_t$ |
| N6 | $y_t = 2\exp(-0.5x_{t1}^2)x_{t1} - \exp(-0.5x_{t1}^2)x_{t2} + e_t$ |
| N7 | $y_t = (0.5x_{t1} - 0.4x_{t2})1_{\{x_{t1}<0\}} + (0.5x_{t1} + 0.3x_{t2})1_{\{x_{t1}>=0\}} + e_t$ |
| N8 | $y_t = x_{t1}(x_{t2} - 0.5)^+ + 0.8x_{t1}(0.5 - x_{t2})^+ + e_t$ |

Table 2: Prediction mean squared errors (MSE) and simulated standard errors (in parenthesis) for the additive random field model (MSE$_{ARF}$), the proportional additive random field model (MSE$_{PARF}$) and the Hamilton's random field model (MSE$_{RF}$). The data generating processes A1-A8 are specified in Table 1. The sample size and Monte Carlo replications are equal to 100 and 1000 respectively. The MSE's are based on 100 out-of-sample one-step ahead predictions.

| Model | $\sigma_e^2$ | $MSE_{ARF}$ | $MSE_{PARF}$ | $MSE_{RF}$ |
|---|---|---|---|---|
| | 0.01 | 0.0126 (0.0030) | 0.0149 (0.0034) | 0.0264 (0.0028) |
| A1 | 0.04 | 0.0397 (0.0080) | 0.0455 (0.0159) | 0.0690 (0.0076) |
| | 0.25 | 0.2517 (0.0270) | 0.2605 (0.0283) | 0.3154 (0.0296) |
| | 0.01 | 0.0127 (0.0033) | 0.0161 (0.0034) | 0.0265 (0.0027) |
| A2 | 0.04 | 0.0402 (0.0080) | 0.0459 (0.0158) | 0.0695 (0.0076) |
| | 0.25 | 0.2541 (0.0271) | 0.2616 (0.0284) | 0.3176 (0.0298) |
| | 0.01 | 0.0204 (0.0034) | 0.0217 (0.0040) | 0.0746 (0.0055) |
| A3 | 0.04 | 0.0690 (0.0096) | 0.0730 (0.0142) | 0.1293 (0.0120) |
| | 0.25 | 0.3185 (0.0369) | 0.3099 (0.0382) | 0.4524 (0.0414) |
| | 0.01 | 0.0176 (0.0031) | 0.0192 (0.0037) | 0.0480 (0.0045) |
| A4 | 0.04 | 0.0628 (0.0123) | 0.0708 (0.0143) | 0.0996 (0.0107) |
| | 0.25 | 0.3189 (0.0455) | 0.3128 (0.0677) | 0.4274 (0.0458) |
| | 0.01 | 0.0166 (0.0028) | 0.0175 (0.0034) | 0.0177 (0.0020) |
| A5 | 0.04 | 0.0602 (0.0076) | 0.0499 (0.0136) | 0.0607 (0.0066) |
| | 0.25 | 0.2746 (0.0228) | 0.2797 (0.0296) | 0.3148 (0.0304) |
| | 0.01 | 0.0106 (0.0017) | 0.0099 (0.0015) | 0.0098 (0.0007) |
| A6 | 0.04 | 0.0548 (0.0085) | 0.0395 (0.0071) | 0.0392 (0.0038) |
| | 0.25 | 0.2473 (0.0156) | 0.2446 (0.0149) | 0.2443 (0.0150) |
| | 0.01 | 0.0141 (0.0030) | 0.0129 (0.0051) | 0.0279 (0.0044) |
| A7 | 0.04 | 0.0519 (0.0103) | 0.0504 (0.0119) | 0.0779 (0.0092) |
| | 0.25 | 0.3188 (0.0203) | 0.3192 (0.0203) | 0.3192 (0.0196) |
| | 0.01 | 0.0146 (0.0031) | 0.0144 (0.0034) | 0.0295 (0.0030) |
| A8 | 0.04 | 0.0639 (0.0099) | 0.0446 (0.0130) | 0.0704 (0.0077) |
| | 0.25 | 0.2523 (0.0270) | 0.2601 (0.0269) | 0.3106 (0.0283) |

and Mammen (2002). As already indicated, the estimation of (36) is less computationally demanding than the more flexible additive random field model (2). In Table 2 we compare the mean squared error of the one-step ahead prediction for the eight additive models A1-A8 using three different specifications of the random field: the additive, the proportional additive, and the comprehensive random field. In each table we consider the same models but with different error variances: $\sigma_e^2 = 0.01, 0.04$, and 0.25. Across Table 2, the additive and the proportional additive random field models exhibit smaller mean squared errors than the model with one comprehensive random field, with the exception of specification A6, which has a very simple linear structure. Though the proportional additive random field has a larger mean squared error relative to the additive model, their differences are almost insignificant for most of the data generating processes. As expected, the prediction MSEs increase significantly when $\sigma_e^2$ increases. As indicated in Table 2, when $\sigma_e^2 = 0.25$ the differences in the MSEs are less noticeable though the additive and proportional additive specifications are still superior to that of the comprehensive random field.

In Tables 3 and 4 we report the rejection frequencies of the residual-based test statistic for the sixteen additive and nonadditive data generating processes of Table 1. We compare the performance of our test with that of the Lagrange Multiplier statistic proposed by Chen, Liu, and Tsay (1995), which has been documented to have extremely good size and power properties in finite samples. We denote this test statistic $\text{LM}_{CLT}$. The nominal size is chosen to be $\alpha = 0.05$. In Table 3 we present the rejection frequencies of the two tests for the eight additive models A1-A8. Both tests have an acceptable empirical size, though for the simple linear structure A6, the $\text{LM}_{CLT}$ test has a perhaps marginal better size than the $\widehat{ResT}_{DGQ}$ test, which is slightly oversized. Recall that the residual-based test is based on the estimation of the parameters of a random field designed to detect nonlinearities. Thus, it is not surprising that its performance is inferior when the process is truly linear. In addition, across all specifications, the estimator of the error variance proposed in Corollary 3 performs very well by providing an accurate estimate of the population value.

In Table 4, we report the power of the tests for the eight nonadditive models N1-N8. The residual-based test has perfect power across all eight specifications. Both tests have perfect power for functional forms like N1 and N6-N8, which can be approximated

Table 3: Rejection frequencies (power) for the additive models in Table 1 with nominal size $\alpha = 0.05$. Consistent estimates of $\sigma_e^2$ obtained according to (32) are reported in parenthesis. Sample size and Monte Carlo replications are equal to 100 and 1000 respectively.

| Model | $\sigma_e^2(\hat{\sigma}_e^2)$ | $\widehat{ResT}_{DGQ}$ | $\mathrm{LM}_{CLT}$ |
|---|---|---|---|
| | 0.01(0.0125) | 0.062 | 0.058 |
| A1 | 0.04(0.0389) | 0.049 | 0.048 |
| | 0.25(0.2699) | 0.051 | 0.058 |
| | 0.01(0.0130) | 0.093 | 0.054 |
| A2 | 0.04(0.0419) | 0.035 | 0.046 |
| | 0.25(0.2692) | 0.048 | 0.058 |
| | 0.01(0.0198) | 0.063 | 0.084 |
| A3 | 0.04(0.0443) | 0.000 | 0.064 |
| | 0.25(0.2738) | 0.038 | 0.050 |
| | 0.01(0.0179) | 0.000 | 0.104 |
| A4 | 0.04(0.0446) | 0.005 | 0.030 |
| | 0.25(0.2512) | 0.041 | 0.036 |
| | 0.01(0.0103) | 0.033 | 0.034 |
| A5 | 0.04(0.0391) | 0.048 | 0.042 |
| | 0.25(0.2697) | 0.049 | 0.044 |
| | 0.01(0.0097) | 0.075 | 0.058 |
| A6 | 0.04(0.0387) | 0.077 | 0.058 |
| | 0.25(0.2419) | 0.088 | 0.058 |
| | 0.01(0.0106) | 0.016 | 0.042 |
| A7 | 0.04(0.0487) | 0.050 | 0.040 |
| | 0.25(0.2635) | 0.072 | 0.046 |
| | 0.01(0.0126) | 0.037 | 0.044 |
| A8 | 0.04(0.0420) | 0.019 | 0.040 |
| | 0.25(0.2628) | 0.048 | 0.040 |

Table 4: Rejection frequencies (power) for the non-additive models in Table 1 with nominal size $\alpha = 0.05$. Consistent estimates of $\sigma_e^2$ obtained according to (32) are reported in parenthesis. Sample size and Monte Carlo replications are equal to 100 and 1000 respectively

| Model | $\sigma_e(\hat{\sigma}_e^2)$ | $\widehat{ResT}_{DGQ}$ | LM test |
|---|---|---|---|
| N1 | 0.01(0.4310) | 1.000 | 1.000 |
| | 0.04(0.4558) | 1.000 | 1.000 |
| | 0.25(0.6342) | 1.000 | 1.000 |
| N2 | 0.01(0.0238) | 1.000 | 0.000 |
| | 0.04(0.0499) | 1.000 | 0.000 |
| | 0.25(0.3102) | 1.000 | 0.002 |
| N3 | 0.01(0.0300) | 1.000 | 0.000 |
| | 0.04(0.0576) | 1.000 | 0.000 |
| | 0.25(0.2591) | 1.000 | 0.000 |
| N4 | 0.01(0.2108) | 1.000 | 0.972 |
| | 0.04(0.2444) | 1.000 | 0.766 |
| | 0.25(0.4749) | 0.994 | 0.386 |
| N5 | 0.01(0.0389) | 1.000 | 1.000 |
| | 0.04(0.0613) | 1.000 | 0.986 |
| | 0.25(0.2324) | 1.000 | 0.736 |
| N6 | 0.01(0.0311) | 1.000 | 1.000 |
| | 0.04(0.0540) | 1.000 | 1.000 |
| | 0.25(0.2816) | 1.000 | 1.000 |
| N7 | 0.01(0.0453) | 1.000 | 1.000 |
| | 0.04(0.0716) | 1.000 | 1.000 |
| | 0.25(0.3613) | 1.000 | 1.000 |
| N8 | 0.01(0.0963) | 1.000 | 1.000 |
| | 0.04(0.1234) | 1.000 | 1.000 |
| | 0.25(0.3096) | 1.000 | 1.000 |

reasonably well by a Taylor's series expansion[9]. However, the $LM_{CLT}$ test has no power for specifications that are Fourier series expansions, like N2 and N3. These models can not be approximated well by low order polynomials, nevertheless, the residual-based test $\widehat{ResT}_{DGQ}$ has power equal to one. Note that although the error variance is overestimated across the eight non-additive specifications, the residual-based test is in this case fairly robust and there is no loss of power.

## 6 Conclusion

In this paper we have proposed and analyzed the large sample behavior of the non-locally misspecified additive random field models, which can be viewed as a generalization of the parametric random field proposed by Hamilton (2001). We establish the consistency and asymptotic normality of the estimators of the parameters entering the model under a set of similar assumptions as those in Hamilton (2001). As a by-product, we have proposed a test for additivity that is fully parametric, very easy to compute, and - based on simulation studies - appears to have very good small sample size and power properties when applied to a large class of additive and non-additive specifications. Finally, we provide simulation evidence showing that the additive random field model substantially improve the accuracy of out-of-sample predictions relative to Hamilton's (2001) comprehensive random field model when the data generating mechanism is additive.

[9]Hamilton (2001, Lemma 4.9) proved that the random field regression model estimates very well specifications that can be approximated by a Taylor expansion. The LM test proposed by Chen, Liu, and Tsay (1995) is basically designed on a Volterra expansion.

# References


Abadir, K. M. and J. R. Magnus (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal 5*, 76–90.

Abt, M. and W. J. Welch (1998). Fisher information and maximum likelihood estimation of covariance parameters in gaussian stochastic processes. *The Canadian Journal of Statistics 26(1)*, 127–137.

Barry, D. (1993). Testing for additivity of a regression function. *The Annals of Statistics 21(1)*, 235–254.

Carroll, R. J., W. Hardle, and E. Mammen (2002). Estimation in an additive model when the components are linked parametrically. *Econometric Theory 18*, 886–912.

Chen, R., J. S. Liu, and R. S. Tsay (1995). Additivity tests for nonlinear autoregression. *Biometrika 82(2)*, 369–383.

Cressie, N. A. (1993). *Statistics for Spatial Data*. Wiley, NY.

Dahl, C. M. (2002). An investigation of tests for linearity and the accuracy of likelihood based inference using random fields. *Econometrics Journal 5*, 263–284.

Dahl, C. M. and G. Gonzalez-Rivera (2003). Testing for neglected nonlinearity in regression models based on the theory of random fields. *Journal of Econometrics 114(1)*, 141–164.

Dahl, C. M. and S. Hylleberg (2004). Flexible regression models and relative forecast performance. *International Journal of Forecasting 20*, 201–217.

Dahl, Christian, M. and Y. Qin (2004). The limiting behavior of the estimated parameters in a misspecified random field regression model. *Unpublished Manuscript, Purdue University*.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.

Eubank, R., J. Hart, D. Simpson, and L. Stefanski (1995). Testing for additivity in nonparametric regression. *The Annals of Statistics 23(6)*, 1896–1920.

Hall, P., J. W. Kay, and D. M. Titterington (1990). Asymptotically optimal difference based estimation of variance in nonparametric regression. *Biometrika 77*, 521–528.

Hamilton, J. D. (2001). A parametric approach to flexible nonlinear inference. *Econometrica 69(3)*, 537–573.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*. Chapman and Hall, London.

Hong, Y. and H. White (1995). Consistent specification testing via nonparametric series regression. *Econometrica 63(5)*, 1133–1159.

Loh, W.-L. and T.-K. Lam (2000). Estimating structured correlation matrices in smooth gaussian random field models. *The Annals of Statistics 28(3)*, 880–904.

Lutkepohl, H. (1996). *Handbook of Matrices*. John Wiley and Sons, New York, USA.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley.

Mardia, K. and R. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika 71(1)*, 135–146.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.

Sperlich, S., D. Tjostheim, and L. Yang (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory 18*, 197–251.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics 13(2)*, 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics 14(2)*, 590–606.

Wooldridge, J. M. (1992). A test of functional form against nonparametric alternatives. *Econometric Theory 8*, 452–475.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV, Chapter 45. Amsterdam, North Holland.

Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica 12*, 801–821.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a gaussian process. *Journal of Multivariate Analysis 36*, 280–296.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics 21(3)*, 1567–1590.

# Mathematical Appendix

We need some auxiliary results on uniform convergence of deterministic sequences before proceeding with the proofs of consistency and asymptotic normality theorems.

**Lemma A.1** Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_I)$ and $\sigma$ be the nonlinear parameters in the additive random field model, and $\boldsymbol{C} \equiv \boldsymbol{H}(\boldsymbol{\lambda}) + \sigma \boldsymbol{I}_T$. For any permissible covariance matrix $\boldsymbol{H}(\boldsymbol{\lambda})$,

$$\boldsymbol{C}^{-1}\boldsymbol{H}(\boldsymbol{\lambda}) = \boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1} = \boldsymbol{I}_T - \sigma\boldsymbol{C}^{-1}.$$

**Proof of Lemma A.1** Consider

$$\boldsymbol{C}\boldsymbol{C}^{-1} = (\boldsymbol{H}(\boldsymbol{\lambda}) + \sigma\boldsymbol{I}_T)(\boldsymbol{H}(\boldsymbol{\lambda}) + \sigma\boldsymbol{I}_T)^{-1} = \boldsymbol{I},$$

then

$$\boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1} + \sigma\boldsymbol{C}^{-1} = \boldsymbol{I},$$

hence

$$\boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1} = \boldsymbol{I} - \sigma\boldsymbol{C}^{-1}.$$

Similarly, considering $\boldsymbol{C}^{-1}\boldsymbol{C} = \boldsymbol{I}$, the first equality of the lemma follows. By Lemma A.1, the covariance matrix $\boldsymbol{P}_T \equiv \mathrm{E}\left[(\boldsymbol{\xi}_T - \boldsymbol{\mu}_T)(\boldsymbol{\xi}_T - \boldsymbol{\mu}_T)'\right]$ can be written as,

$$\boldsymbol{P}_T = \boldsymbol{H}(\boldsymbol{\lambda}) - \boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1}\boldsymbol{H}(\boldsymbol{\lambda}) = \sigma\boldsymbol{C}^{-1}\boldsymbol{H}(\boldsymbol{\lambda}).$$

**Lemma A.2** Let Assumptions *ii.*,1,2, and 3 hold. Then

$$\lim_{T\to\infty} \sup_{\Theta\times A^2} \boldsymbol{P}_T \to \boldsymbol{O}_T.$$

For each individual random field, let $\boldsymbol{P}_{Ti} \equiv \sigma(\lambda_i\boldsymbol{H}_i + \sigma\boldsymbol{I}_T)^{-1}\lambda_i\boldsymbol{H}_i$, then

$$\lim_{T\to\infty} \sup_{\Theta\times A_i^2} \boldsymbol{P}_{Ti} \to \boldsymbol{O}_T.$$

**Proof of Lemma A.2** It follows directly from Theorem 2 in Dahl and Qin (2004).

**Corollary A.1** Let $\tau_i \equiv \frac{1}{T^3}\widetilde{\boldsymbol{\phi}}_i' \boldsymbol{P}_T' \boldsymbol{P}_T \widetilde{\boldsymbol{\phi}}_i$ where $\widetilde{\phi}_i : A_i \to \mathbb{R}$ is a continuous function for $i = 1, 2, ..., I$ and $\widetilde{\boldsymbol{\phi}}_i = \left(\widetilde{\phi}_i(\boldsymbol{x}_{1i\cdot}), \widetilde{\phi}_i(\boldsymbol{x}_{2i\cdot}), ..., \widetilde{\phi}_i(\boldsymbol{x}_{Ti\cdot})\right)'$. Then

$$\lim_{T\to\infty} \sup_{\Theta \times A_i^2} \tau_i \to 0, \tag{37}$$

for all $i = 1, 2, ..., I$. Furthermore, for each individual random field, let $\tau_{ii} \equiv \frac{1}{T^3}\widetilde{\boldsymbol{\phi}}_i' \boldsymbol{P}_{Ti}' \boldsymbol{P}_{Ti} \widetilde{\boldsymbol{\phi}}_i$. Then

$$\lim_{T\to\infty} \sup_{\Theta \times A_i^2} \tau_{ii} \to 0, \tag{38}$$

for all $i = 1, 2, ..., I$.

**Proof of Corollary A.1** It follows directly from Theorem 3 in Dahl and Qin (2004) and from Theorem 4.7 in Hamilton (2001).

**Corollary A.2** Let $\widetilde{\boldsymbol{\phi}}_i$ be defined as in Corollary A.1, and let $\boldsymbol{P}_T^i \equiv \sigma \boldsymbol{C}^{-1} \lambda_i \boldsymbol{H}_i$. Define $\tau_i^i \equiv \frac{1}{T^3}\widetilde{\boldsymbol{\phi}}_i' (\boldsymbol{P}_T^i)' \boldsymbol{P}_T^i \widetilde{\boldsymbol{\phi}}_i$. Then

$$\lim_{T\to\infty} \sup_{\Theta \times A^2} \tau_i^i \to 0, \tag{39}$$

for all $i = 1, 2, ..., I$.

**Proof of Corollary A.2** For any two $n \times n$ positive definite matrices $\boldsymbol{A}, \boldsymbol{B}$, we say that $\boldsymbol{A} > \boldsymbol{B}$ if $\boldsymbol{A} - \boldsymbol{B}$ is also a positive definite matrix. Since $\boldsymbol{P}_T^i \leq \boldsymbol{P}_T$ for $\forall i, T$, it follows that $\tau_i^i \leq \tau_i$ for all $\forall i, T$ . Since $\tau_i^i \geq 0$, Corollary A.2 immediately follows.

**Proof of Theorem 1** Let $\boldsymbol{e} = (e_1, e_2, ..., e_T)'$ be the vector of disturbances in the data generating process (Assumption 3). We can write

$$\begin{aligned}
\boldsymbol{\xi}_T - \boldsymbol{X\beta} - \boldsymbol{L}_T(\boldsymbol{\lambda}) &= \boldsymbol{H}(\boldsymbol{\lambda}) \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) - \boldsymbol{L}_T(\boldsymbol{\lambda}) \\
&= \boldsymbol{H}(\boldsymbol{\lambda}) \boldsymbol{C}^{-1} (\boldsymbol{L}_T(\boldsymbol{\lambda}) + \boldsymbol{e}) - \boldsymbol{L}_T(\boldsymbol{\lambda}) \\
&= \frac{1}{T}\left(\boldsymbol{H}(\boldsymbol{\lambda}) \boldsymbol{C}^{-1} \sum_{i=1}^{I} \lambda_i \boldsymbol{H}_i \widetilde{\boldsymbol{\phi}}_i - \sum_{i=1}^{I} \lambda_i \boldsymbol{H}_i \widetilde{\boldsymbol{\phi}}_i\right) \\
&\quad + \boldsymbol{H}(\boldsymbol{\lambda}) \boldsymbol{C}^{-1} \boldsymbol{e} \\
&= \frac{1}{T}\sum_{i=1}^{I} \boldsymbol{P}_T^i \widetilde{\boldsymbol{\phi}}_i + \boldsymbol{H}(\boldsymbol{\lambda}) \boldsymbol{C}^{-1} \boldsymbol{e},
\end{aligned}$$

where $\boldsymbol{P}_T^i$ is defined in Corollary A.2, and the last equality follows from Lemma A.1. In order to show that

$$\lim_{T\to\infty}\sup_{\Theta\times A^2}\frac{1}{T}\,\mathrm{E}\left[\left(\boldsymbol{\xi}_T-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{L}_T\left(\boldsymbol{\lambda}\right)\right)'\left(\boldsymbol{\xi}_T-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{L}_T\left(\boldsymbol{\lambda}\right)\right)\right]\to 0, \tag{40}$$

we need to show that

$$\lim_{T\to\infty}\sup_{\Theta\times A^2}\frac{1}{T^3}\left(\sum_{i=1}^{I}\boldsymbol{P}_T^i\widetilde{\boldsymbol{\phi}}_i\right)'\left(\sum_{i=1}^{I}\boldsymbol{P}_T^i\widetilde{\boldsymbol{\phi}}_i\right)=\lim_{T\to\infty}\sup_{\Theta\times A^2}\left(\sum_{i=1}^{I}\sqrt{\tau_i^i}\right)\left(\sum_{j=1}^{I}\sqrt{\tau_j^j}\right)\to 0, \tag{41}$$

where $\tau_i^i$ is given by (39) in Corollary A.2. The convergence of the remaining terms in (40) is shown in Dahl and Qin (2004). Let $a_{ij}=\sqrt{\tau_i^i}\sqrt{\tau_j^j}$. By the Cauchy-Schwartz inequality, we have $a_{ij}\leq\sqrt{a_{ii}a_{jj}}$. Hence,

$$\begin{aligned}
\lim_{T\to\infty}\sup_{\Theta\times A^2}\left(\sum_{i=1}^{I}\sqrt{\tau_i^i}\right)\left(\sum_{j=1}^{I}\sqrt{\tau_j^j}\right) &= \lim_{T\to\infty}\sup_{\Theta\times A^2}\sum_{i=1}^{I}\sum_{j=1}^{I}a_{ij}\\
&\leq \lim_{T\to\infty}\sup_{\Theta\times A^2}\left(\sum_{i=1}^{I}\sqrt{a_{ii}}\right)^2\\
&\leq \lim_{T\to\infty}\sup_{\Theta\times A^2}I\sum_{i=1}^{I}a_{ii}\\
&= \lim_{T\to\infty}\sup_{\Theta\times A^2}I\sum_{i=1}^{I}\tau_i^i,\to 0,
\end{aligned}$$

by Corollary A.2.

**Proof of Corollary 1** Define

$$\boldsymbol{S}\equiv\boldsymbol{\psi}\left(\boldsymbol{X}\right)-\Big(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)\Big).$$

By Theorem 1 we have that $\lim_{T\to\infty}\frac{1}{T}\,\mathrm{E}\left(\boldsymbol{S}'\boldsymbol{S}\right)\to 0$ uniformly in $\boldsymbol{\theta},\boldsymbol{\beta}$, and $\boldsymbol{X}$.

$$\begin{aligned}
\frac{1}{T}\,\mathrm{E}\left(\boldsymbol{S}'\boldsymbol{S}\right) &= \frac{\sigma^2}{T}\left(\boldsymbol{\psi}\left(\boldsymbol{X}\right)-\boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-2}\left(\boldsymbol{\psi}\left(\boldsymbol{X}\right)-\boldsymbol{X}\boldsymbol{\beta}\right)\\
&\quad+\frac{\sigma_e^2}{T}\,\mathrm{tr}\left[\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\boldsymbol{H}\left(\boldsymbol{\lambda}\right)\left(\boldsymbol{H}\left(\boldsymbol{\lambda}\right)+\sigma\boldsymbol{I}_T\right)^{-1}\right].
\end{aligned}$$

Since both terms on the right hand side are positive, results (13) and (14) of Corollary 1 follow. Result (15) is a direct consequence of Lemma A.1 and (14).

**Lemma A.3** Suppose $\boldsymbol{A}, \boldsymbol{B}$ are two $(n, n)$ positive definite matrices. Then,

$$\operatorname{tr}(\boldsymbol{AB}) > 0. \tag{42}$$

**Proof of Lemma A.3** From Magnus and Neudecker (1999) Theorem 13, p. 16, there exists an orthogonal matrix $\boldsymbol{P}$, such that $\boldsymbol{A} = \boldsymbol{P}'\boldsymbol{V}\boldsymbol{P}$, where $\boldsymbol{V} = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$ and $\gamma_i$ for $i = 1, \ldots, n$ are eigenvalues of $\boldsymbol{A}$. Consequently,

$$\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{P}'\boldsymbol{V}\boldsymbol{P}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{V}(\boldsymbol{P}\boldsymbol{B}\boldsymbol{P}')).$$

Since $\boldsymbol{B}$ is a positive definite matrix, $\boldsymbol{PBP}'$ is a positive definite matrix. Therefore, all diagonal elements of $\boldsymbol{PBP}'$, denoted as $\alpha_1, \ldots, \alpha_n$ must be positive. Hence,

$$\operatorname{tr}(\boldsymbol{V}(\boldsymbol{PBP}')) = \sum_{i=1}^{n} \gamma_i \alpha_i > 0,$$

which completes the proof.

**Proof of Corollary 2** We can write

$$\begin{aligned}
\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{C}^{-1}\boldsymbol{H}(\boldsymbol{\lambda})\right) &= \operatorname{tr}\left[\boldsymbol{C}^{-1}\left(\sum_{i=1}^{I}\lambda_i\boldsymbol{H}_i\right)\boldsymbol{C}^{-1}\left(\sum_{i=1}^{I}\lambda_i\boldsymbol{H}_i\right)\right] \\
&= \operatorname{tr}\left[\left(\sum_{i=1}^{I}\lambda_i\boldsymbol{C}^{-1}\boldsymbol{H}_i\right)\left(\sum_{i=1}^{I}\lambda_i\boldsymbol{C}^{-1}\boldsymbol{H}_i\right)\right] \\
&= \sum_{i=1}^{I}\lambda_i^2\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{H}_i\right) \\
&\quad +2\sum_{i=1}^{I}\sum_{j>i}^{I}\lambda_i\lambda_j\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{H}_j\right).
\end{aligned}$$

By combining (15) of Corollary 1 and Lemma A.3 (taking $\boldsymbol{A} = \boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}$ and $\boldsymbol{B} = \boldsymbol{H}_j$) it is obvious that each term on the right hand side is positive and converges uniformly to zero on $\Theta \times B \times A$ as $T \to \infty$, which concludes the proof.

**Proof of Proposition 1** From Magnus and Neudecker (1999, chapter 8)

$$\mathrm{D}_i R_T = \frac{1}{T}\mathrm{D}_i \log\det\boldsymbol{C} = \frac{1}{T}\operatorname{tr}\left(\boldsymbol{C}^{-1}(\mathrm{D}_i\boldsymbol{C})\right) = \frac{1}{T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\right),$$

and

$$\mathrm{D}^2_{i,j} R_T = \frac{1}{T}\,\mathrm{tr}\left(-\boldsymbol{C}^{-1}\left(\mathrm{D}_i\boldsymbol{C}\right)\boldsymbol{C}^{-1}\left(\mathrm{D}_j\boldsymbol{C}\right)\right) = \frac{1}{T}\,\mathrm{tr}\left(-\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{H}_j\right).$$

From Corollary 2, it follows that $\mathrm{D}^2_{i,j} R = \lim_{T\to\infty} \sup_{\Theta\times A} \mathrm{D}^2_{i,j} R_T = 0$ for all $i, j = 1, ..., I$. From Corollary 1, it follows that $\mathrm{D}_0 U = 0$. Finally, let the eigenvalues of $\boldsymbol{C}^{-1}\boldsymbol{H}_i$ be $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_T$. Then

$$\mathrm{D}_i R_T = \frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\right) = \frac{1}{T}\sum_{t=1}^{T}\gamma_t \leq \sqrt{\frac{\sum_{t=1}^{T}\gamma_t^2}{T}} = \sqrt{\mathrm{D}^2_{i,i} R_T},$$

and $\mathrm{D}_i R = 0$ immediately follows, for all $i = 1, 2, ..., I$.

**Proof of Proposition 2** The sequences of differentials of $R_T$ for $l = 3, 4$ are given by

$$\begin{aligned}\mathrm{D}^3_{i_1,i_2,i_3} R_T &= \frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad + \frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right),\end{aligned} \tag{43}$$

and

$$\begin{aligned}\mathrm{D}^4_{i_1,i_2,i_3,i_4} R_T &= -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) \\ &\quad -\frac{1}{T}\,\mathrm{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_4}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right).\end{aligned} \tag{44}$$

Let us focus on the first term of (43). Recall the following property of the trace operator,

$$\left|\mathrm{tr}\left(\boldsymbol{A}'\boldsymbol{B}\right)\right| \leq \sqrt{\mathrm{tr}\left(\boldsymbol{A}'\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{B}'\boldsymbol{B}\right)}. \tag{45}$$

Then,

$$
\begin{aligned}
& \left|\frac{1}{T}\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)\right| \qquad (46)\\
\leq\ & \frac{1}{T}\sqrt{\operatorname{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\right) tr\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)}\\
\leq\ & \frac{1}{T}\sqrt{\frac{T}{\lambda_{i_3}^2}}\sqrt{\operatorname{tr}\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)}\\
\leq\ & \frac{1}{T}\sqrt{\frac{T}{\lambda_{i_3}^2}}\sqrt{\frac{T}{\lambda_{i_2}^2\lambda_{i_1}^2}}\\
=\ & \frac{1}{\lambda_{i_3}\lambda_{i_2}\lambda_{i_1}}.
\end{aligned}
$$

The second inequality in (46) follows from $\left(\boldsymbol{C}^{-1}\right)^2 \leq \left(\left(\lambda_{i_3}\boldsymbol{H}_{i_3}\right)^{-1}\right)^2$ for all $i_3$. Thus,

$$
tr\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_3}\boldsymbol{H}_{i_3}\boldsymbol{C}^{-1}\right) = tr\left(\left(\boldsymbol{C}^{-1}\right)^2\left(\boldsymbol{H}_{i_3}\right)^2\right) = \frac{1}{\lambda_{i_3}^2} tr\left(\left(\boldsymbol{C}^{-1}\right)^2\left(\lambda_{i_3}\boldsymbol{H}_{i_3}\right)^2\right) \leq \frac{T}{\lambda_{i_3}^2}.
$$

Similarly, the third inequality in (46) follows from

$$
\begin{aligned}
\operatorname{tr}\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right) &= \operatorname{tr}\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)\\
&\leq \operatorname{tr}\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\left(\left(\lambda_{i_2}\boldsymbol{H}_{i_2}\right)^{-1}\right)^2\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)\\
&= \frac{1}{\lambda_{i_2}^2}\operatorname{tr}\left(\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\right)\\
&\leq \frac{1}{\lambda_{i_2}^2}\operatorname{tr}\left(\boldsymbol{H}_{i_1}\left(\left(\lambda_{i_1}\boldsymbol{H}_{i_1}\right)^{-1}\right)^2\boldsymbol{H}_{i_1}\right)\\
&= \frac{T}{\lambda_{i_2}^2\lambda_{i_1}^2}.
\end{aligned}
$$

The same argument can be applied to the second term of (43), which subsequently can be shown to be bounded by $\frac{1}{\lambda_{i_3}\lambda_{i_2}\lambda_{i_1}}$. It follows that $\sup\limits_{\Theta\times A}\mathrm{D}^3_{i_1,i_2,i_3}R_T$ is bounded for all $T$, and all $i_1, i_2, i_3 = 0, 1, 2, ..., I$. Using similar arguments, we can show that $\sup\limits_{\Theta\times A}\mathrm{D}_{i_1}R_T$, $\sup\limits_{\Theta\times A}\mathrm{D}^2_{i_1,i_2}R_T$, and $\sup\limits_{\Theta\times A}\mathrm{D}^4_{i_1,i_2,i_3,i_4}R_T$ are all bounded for all $T$ and for all $i_1, i_2, i_3, i_4 = 0, 1, 2, ..., I$. Since $\left\{\mathrm{D}^l_{i_1\ldots i_l}R_T\right\}_T$ for $l = 1, 2, ..., 4$, $i_1, \ldots, i_l = 0, 1, \ldots, I$, and all $T$ are bounded, the sequences $\left\{\mathrm{D}^l_{i_1\ldots i_l}R_T\right\}_T$ for $l = 1, 2, 3$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$ are equicontinuous. Since at least one of these equicontinuous sequences converges − for example $\left\{\mathrm{D}^2_{i_1,i_2}R_T\right\}_T$ in Proposition 1 − Theorem 5 in Dahl and Yu (2004) and Theorem 7.16 in Rudin (1976) imply that all of the function sequences $\left\{\mathrm{D}^l_{i_1\ldots i_l}R_T\right\}_T$ for $l = 1, 2, 3$

and $i_1, \ldots, i_l = 0, 1, \ldots, I$ are uniformly convergent on $\Theta \times A$ as $T \to \infty$. This completes the proof.

**Proof of Proposition 3** Define $\boldsymbol{c} \equiv (\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta})$. The sequences of differentials of $U_T$ for $l = 1, 2$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$ are given by

$$\mathrm{D}_{i_1} U_T = -\frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right),$$

and

$$\mathrm{D}^2_{i_1,i_2} U_T = \frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right) + \frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{c}\right).$$

Notice that

$$\begin{aligned}
\frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right) &= \frac{1}{T\lambda_{i_1}} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\lambda_{i_1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right) \\
&\leq \frac{1}{T\lambda_{i_1}} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{c}\right) \\
&\leq \frac{1}{\sigma\lambda_{i_1}} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{c}\right),
\end{aligned}$$

where, by Assumption 4, $\frac{1}{T} \operatorname{tr}(\boldsymbol{c}'\boldsymbol{c})$ is bounded for all $T$. Note that the last inequality follows as $\boldsymbol{C}^{-1} < \sigma^{-1}\boldsymbol{I}$ while Assumption 2 guarantees the existence of $\frac{1}{\sigma\lambda_{i_1}}$. This implies that $\sup\limits_{\Theta\times B\times A} \mathrm{D}_{i_1} U_T$ is bounded for all $i_1 = 1, 2, ..., I$ and for all $T$. Following a similar argument as in Proposition 2, the first term of $\mathrm{D}^2_{i_1,i_2} U_T$ is bounded,

$$\begin{aligned}
\left|\frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right)\right| &= \left|\frac{1}{\lambda_{i_1}\lambda_{i_2}} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\lambda_{i_2}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\lambda_{i_1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}\right)\right| \\
&\leq \frac{1}{\sigma\lambda_{i_1}\lambda_{i_2}} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{c}\right).
\end{aligned}$$

The same bound is obtained for the second term of $\mathrm{D}^2_{i_1,i_2} U_T$. Consequently, $\sup\limits_{\Theta\times B\times A} \mathrm{D}^2_{i_1,i_2} U_T$ is bounded for all $i_1, i_2 = 0, 1, ..., I$ and for all $T$. It is not difficult to show that all terms of $\mathrm{D}^l_{i_1\ldots i_l} U_T$ will be bounded by $\left(\sigma^l_{i=1}\lambda_i\right)^{-1} \frac{1}{T} \operatorname{tr}(\boldsymbol{c}'\boldsymbol{c})$ implying that $\sup\limits_{\Theta\times B\times A} \mathrm{D}^l_{i_1\ldots i_l} U_T$ for $l = 1, 2, ..$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$ is bounded for all $T$. Since $\left\{\mathrm{D}^l_{i_1\ldots i_l} U_T\right\}_T$ for $l = 1, 2, ...,$ $i_1, \ldots, i_l = 0, 1, \ldots, I$, and all $T$ are bounded, the sequences $\left\{\mathrm{D}^l_{i_1\ldots i_l} U_T\right\}_T$ for $l = 1, 2, ..$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$ are equicontinuous. Since at least one of these equicontinuous function sequences converges – for example $\{\mathrm{D}_0 U_T\}_T$ in Proposition 1

– Theorem 5 in Dahl and Qin (2004) and Theorem 7.16 in Rudin (1976) imply that all of the function sequences $\left\{\mathrm{D}^l_{i_1\ldots i_l} U_T\right\}_T$ for $l = 1, 2, ..$ and $i_1, \ldots, i_l = 0, 1, \ldots, I$ are uniformly convergent on $\Theta \times B \times A$ as $T \to \infty$. This completes the proof.

**Proof of Proposition 4** The proof of equicontinuity proceeds in a similar fashion as in Propositions 2 and 3. The uniform convergence results of (16),(17), and (18) follow directly from Corollary 1. Inserting $\boldsymbol{\psi}(\boldsymbol{X}) = \boldsymbol{x}_{\cdot i}$ and $\boldsymbol{\beta} = \mathbf{0}$ in (13), we have

$$\lim_{T\to\infty} \sup_{\Theta\times A} \frac{1}{T} \boldsymbol{x}'_{\cdot i} \left( \left( \boldsymbol{H}(\boldsymbol{\lambda}) + \sigma_1 \boldsymbol{I}_T \right)^{-1} \right)^2 \boldsymbol{x}_{\cdot i} \to 0.$$

Inserting $\boldsymbol{\psi}(\boldsymbol{X}) = \boldsymbol{x}_{\cdot i} + \boldsymbol{x}_{\cdot i}$ and $\boldsymbol{\beta} = \mathbf{0}$ in (13), we can write

$$\begin{aligned} \lim_{T\to\infty} \sup_{\Theta\times B\times A} \tfrac{1}{T} \boldsymbol{c}' \left(\boldsymbol{C}^{-1}\right)^2 \boldsymbol{c} &= \lim_{T\to\infty} \sup_{\Theta\times B\times A} \frac{1}{T} \boldsymbol{x}'_{\cdot i} \left(\boldsymbol{C}^{-1}\right)^2 \boldsymbol{x}_{\cdot i} \\ &+ \lim_{T\to\infty} \sup_{\Theta\times B\times A} \frac{1}{T} \boldsymbol{x}'_{\cdot j} \left(\boldsymbol{C}^{-1}\right)^2 \boldsymbol{x}_{\cdot j} + \lim_{T\to\infty} \sup_{\Theta\times B\times A} \frac{2}{T} \boldsymbol{x}'_{\cdot i} \left(\boldsymbol{C}^{-1}\right)^2 \boldsymbol{x}_{\cdot j}. \end{aligned}$$

Since the first two terms converge to zero, the last term must also converge to zero for (13) to hold. Result (17) follows from (13) when $\boldsymbol{\beta} = \mathbf{0}$. Result (18) follows from (13) when $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{x}_{\cdot i}$.

**Proof of Theorem 2** Define $f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) \equiv Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) - Q^*_T(\boldsymbol{\theta}, \boldsymbol{\beta})$. We write

$$f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{T} \left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)' \boldsymbol{C}^{-1} \boldsymbol{e} - \frac{1}{2T} \boldsymbol{e}' \boldsymbol{C}^{-1} \boldsymbol{e} + \frac{\sigma_e^2}{2T} \operatorname{tr}\left(\boldsymbol{C}^{-1}\right).$$

We wish to show that

$$\lim_{T\to\infty} \sup_{\Theta\times B\times A} \left| f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) \right| \xrightarrow{p} 0,$$

which (according to Theorem 21.9 in Davidson (1994)) will be satisfied if and only if a) $\lim f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{p} 0$ for each $(\boldsymbol{\theta}, \boldsymbol{\beta}) \in \Theta \times B$ and b) $f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta})$ is stochastically equicontinuous. Notice first that $\mathrm{E}\left[f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta})\right] = 0$. By Chebyshev's inequality, condition a) will the be satisfied if

$$\lim_{T\to\infty} \sup_{\Theta\times B\times A} \mathrm{E}\left[f_T(\boldsymbol{e}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\beta})^2\right] \xrightarrow{p} 0.$$

Now, let $0 < \gamma_1 \leq \ldots \leq \gamma_T$ be the eigenvalues of $\boldsymbol{H}(\boldsymbol{\lambda})$, and $\boldsymbol{\nu}_t$ for $t = 1, \ldots, T$ be the corresponding eigenvectors. Let $z_t \equiv \frac{1}{\sigma_e} \boldsymbol{\nu}'_t \boldsymbol{e}$ and $a_t \equiv \boldsymbol{\nu}'_t(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta})$ for $t = 1, \ldots, T$.

Then $z_t \sim \mathrm{IN}(0,1)$ and we can write

$$\begin{aligned} f_T(\boldsymbol{e},\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\beta}) &= \frac{1}{T}\sum_{t=1}^{T}[-\frac{\sigma_e a_t z_t}{\gamma_t+\sigma} - \frac{\sigma_e^2 z_t^2}{2(\gamma_t+\sigma)} + \frac{\sigma_e^2}{2(\gamma_t+\sigma)}] \\ &= -\frac{1}{T}\sum_{t=1}^{T}\frac{2\sigma_e a_t z_t + \sigma_e^2 z_t^2 - \sigma_e^2}{2(\gamma_t+\sigma)}. \end{aligned} \tag{47}$$

Therefore,

$$\mathrm{E}\left[f_T\left(\boldsymbol{e},\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\beta}\right)^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}\frac{4\sigma_e^2 a_t^2\,\mathrm{E}(z_t^2) + \sigma_e^4\,\mathrm{E}(z_t^4) + \sigma_e^4 + 4a_t\sigma_e^3\,\mathrm{E}(z_t^3) - 4a_t\sigma_e^3\,\mathrm{E}(z_t) - 2\sigma_e^4\,\mathrm{E}(z_t^2)}{4(\gamma_t+\sigma)^2},$$

and consequently

$$\begin{aligned} \lim_{T\to\infty}\sup_{\Theta\times B\times A}\mathrm{E}\left[f_T\left(\boldsymbol{e},\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\beta}\right)^2\right] &= \lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T^2}\sum_{t=1}^{T}\frac{4\sigma_e^2 a_t^2 + 3\sigma_e^4 + \sigma_e^4 - 2\sigma_e^4}{4(\gamma_t+\sigma)^2} \\ &= \lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{T^2}\sum_{t=1}^{T}\frac{2\sigma_e^2 a_t^2 + \sigma_e^4}{2(\gamma_t+\sigma)^2} \\ &= \lim_{T\to\infty}\sup_{\Theta\times B\times A}\left(-\frac{1}{T}\sigma_e^2\mathrm{D}_0 U_T + \frac{1}{T^2}\sum_{t=1}^{T}\frac{\sigma_e^4}{2(\gamma_t+\sigma)^2}\right) \\ &\leq -\sigma_e^2\mathrm{D}_0 U + \lim_{T\to\infty}\sup_{\Theta\times B\times A}\frac{1}{2T}\frac{\sigma_e^4}{\sigma^2} \\ &\to 0, \end{aligned}$$

where the last inequality follows from assumption ii. and Proposition 1. This completes the proof of condition a). To verify condition b) define

$$\widetilde{f}_T = f_T\left(\boldsymbol{e},\boldsymbol{X},\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\beta}}\right),$$

and note that

$$\begin{aligned} \left|f_T - \widetilde{f}_T\right| &\leq \frac{1}{T}\left|\mathrm{tr}\left(\left(\left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{C}^{-1} - \left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)'\widetilde{\boldsymbol{C}}^{-1}\right)\boldsymbol{e}\right)\right| \\ &\quad + \frac{1}{2T}\left|\mathrm{tr}\left(\boldsymbol{e}'\left(\boldsymbol{C}^{-1} - \widetilde{\boldsymbol{C}}^{-1}\right)\boldsymbol{e}\right)\right| + \frac{\sigma_e^2}{2T}\left|\mathrm{tr}\left(\boldsymbol{C}^{-1} - \widetilde{\boldsymbol{C}}^{-1}\right)\right| \\ &\leq \sqrt{\left(\frac{1}{T}\sum e_t^2\right)}\left\|\left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{C}^{-1} - \left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)'\widetilde{\boldsymbol{C}}^{-1}\right\| \\ &\quad + \left(\left(\frac{1}{T}\sum e_t^2\right) + \frac{\sigma_e^2}{2T}\right)\left\|\boldsymbol{C}^{-1} - \widetilde{\boldsymbol{C}}^{-1}\right\|. \end{aligned}$$

It follows immediately (as $\boldsymbol{X}$ is non-stochastic) that $\left\| (\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X\beta})' \boldsymbol{C}^{-1} - \left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)' \widetilde{\boldsymbol{C}}^{-1} \right\| \downarrow 0$ and $\left\| \boldsymbol{C}^{-1} - \widetilde{\boldsymbol{C}}^{-1} \right\| \downarrow 0$ when $(\boldsymbol{\theta}', \boldsymbol{\beta}')' \to \left(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\beta}}'\right)'$. Furthermore, since $\frac{1}{T}\sum e_t^2 = O_p(1)$ (by LLN) and $\frac{\sigma_e^2}{2T} = O(T^{-1})$ we can conclude that condition b) holds according to, e.g., Theorem 21.10 in Davidson (1994). This completes the proof of (20). Condition (21) follows directly from Propositions 2 and 3.

**Proof of Theorem 3** Calculate the first and second order conditions of the function $Q^*(\boldsymbol{\theta};\boldsymbol{\beta})$. From Propositions 2 and 3, and Theorem 2, we can write the Hessian matrix as

$$\boldsymbol{\mathcal{H}}(\boldsymbol{\theta};\boldsymbol{\beta}) = \begin{pmatrix} -\frac{1}{2}\mathrm{D}_{11}^2 U(\boldsymbol{\theta}) & \cdots & -\frac{1}{2}\mathrm{D}_{1I}^2 U(\boldsymbol{\theta}) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{2}\mathrm{D}_{I1}^2 U(\boldsymbol{\theta}) & \cdots & -\frac{1}{2}\mathrm{D}_{II}^2 U(\boldsymbol{\theta}) & 0 \\ 0 & \cdots & 0 & -\frac{1}{2}(\mathrm{D}_{00}^2 R(\boldsymbol{\theta}) + \sigma_e^2 \mathrm{D}_{000}^3 R(\boldsymbol{\theta})) \end{pmatrix}. \quad (48)$$

The assumption of convexity of $U(\boldsymbol{\theta})$ guarantees that the upper block of the Hessian matrix (48) is negative definite and the function $Q^*(\boldsymbol{\theta};\boldsymbol{\beta})$ is concave in $(\lambda_1, \lambda_2, ...\lambda_I)$. The right lower element of the Hessian matrix is

$$-\frac{1}{2}(\mathrm{D}_{00}^2 R(\boldsymbol{\theta}) + \sigma_e^2 \mathrm{D}_{000}^3 R(\boldsymbol{\theta})) = \frac{1}{2}\operatorname{tr} C^{-1}C^{-1} - \sigma_e^2 \operatorname{tr} C^{-1}C^{-1}C^{-1}.$$

For this term to be negative, it is necessary and sufficient that $\operatorname{tr} C^{-1}C^{-1} < 2\sigma_e^2 \operatorname{tr} C^{-1}C^{-1}C^{-1}$. A necessary condition for concavity of $Q^*(\boldsymbol{\theta};\boldsymbol{\beta})$ in $\sigma$ is given by $\sigma \leq 2\sigma_e^2$. This condition comes from considering the eigenvalues of the matrix $C^{-1}\sigma$, which are less than one (Magnus and Neudecker, 1999, p. 25). Then, $\operatorname{tr} C^{-1}C^{-1}C^{-1}\sigma^3 \leq \operatorname{tr} C^{-1}C^{-1}\sigma^2$ and $\sigma \leq \frac{\operatorname{tr} C^{-1}C^{-1}}{\operatorname{tr} C^{-1}C^{-1}C^{-1}} < 2\sigma_e^2$.

**Proof of Theorem 4** We follow the five conditions for consistency in Dahl and Qin (2004, Theorem 1). Conditions *i.* requires that $\Theta$ and $B$ are compact parameter spaces, which is also required in our assumption *ii*. Condition *ii*. $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* \in B$ can be verified by a similar argument as in Dahl and Qin (2004). Condition *iii.* requires that $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ is a continuous measurable function for all $T$ and it is satisfied trivially. Condition *iv.* requires that $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{p} Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ uniformly in $\Theta \times B$ and it is satisfied by Theorem

2. Condition $v$. requires the existence of a unique maximizer $\boldsymbol{\theta}^* \in \Theta$ of $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ and it is satisfied by Theorem 3. This completes the proof of consistency of $\hat{\boldsymbol{\theta}}$.

**Proof of Proposition 5** Define $\boldsymbol{v} \equiv (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ and $\boldsymbol{B}_{i_1} \equiv \frac{1}{\sqrt{T}}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}$ and notice that $\boldsymbol{v} \sim \mathrm{N}_T(\boldsymbol{c}, \sigma_e^2 \boldsymbol{I}_T)$. First, we focus on proving equation (25) and for that purpose we use the moment generating function $M(s_{i_1}, s_{i_2})$ defined as

$$M(s_{i_1}, s_{i_2}) = \mathrm{E}\left(\exp(s_{i_1}\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v} + s_{i_2}\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v})\right),$$

where $s_{i_1}, s_{i_2} \in \mathbb{R}$. Now, let

$$\begin{aligned}
\varkappa &\equiv -\frac{1}{2\sigma_e^2}(\boldsymbol{v} - \boldsymbol{c})'(\boldsymbol{v} - \boldsymbol{c}) + \boldsymbol{v}'(s_{i_1}\boldsymbol{B}_{i_1} + s_{i_2}\boldsymbol{B}_{i_2})\boldsymbol{v} \\
&= -\frac{1}{2\sigma_e^2}\boldsymbol{v}'\boldsymbol{v} + \frac{1}{\sigma_e^2}\boldsymbol{c}'\boldsymbol{v} \\
&\quad -\frac{1}{2\sigma_e^2}\boldsymbol{c}'\boldsymbol{c} + \boldsymbol{v}'(s_{i_1}\boldsymbol{B}_{i_1} + s_{i_2}\boldsymbol{B}_{i_2})\boldsymbol{v} \\
&= -\frac{1}{2\sigma_e^2}\left(\boldsymbol{v}'\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)\boldsymbol{v} - \boldsymbol{c}'\boldsymbol{v} - \boldsymbol{v}'\boldsymbol{c} + \boldsymbol{c}'\boldsymbol{c}\right) \\
&= -\frac{1}{2\sigma_e^2}\Big[\left(\boldsymbol{v} - \left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1}\boldsymbol{c}\right)'\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right) \\
&\quad \times \left(\boldsymbol{v} - \left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1}\boldsymbol{c}\right) \\
&\quad - \boldsymbol{c}'\left(\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1} - \boldsymbol{I}_T\right)\boldsymbol{c}\Big] \\
&= -\frac{1}{2\sigma_e^2}(\boldsymbol{v} - \tilde{\boldsymbol{c}})'\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)(\boldsymbol{v} - \tilde{\boldsymbol{c}}) \\
&\quad + \boldsymbol{c}'(s_{i_1}\boldsymbol{B}_{i_1} + s_{i_2}\boldsymbol{B}_{i_2})\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1}\boldsymbol{c},
\end{aligned}$$

where $\tilde{\boldsymbol{c}} = \left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1}\boldsymbol{c}$. By using the above formulation we can write

$$\begin{aligned}
M(s_{i_1}, s_{i_2}) &= \int \frac{1}{(2\pi\sigma_e^2)^{\frac{T}{2}}}\exp(\varkappa)\,d\boldsymbol{v} \\
&= \int \frac{\left|\boldsymbol{I} - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right|^{-\frac{1}{2}}}{(2\pi\sigma_e^2)^{\frac{T}{2}}\left|\boldsymbol{I} - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right|^{-\frac{1}{2}}}\exp(\varkappa)\,d\boldsymbol{v} \\
&= \left|\boldsymbol{I} - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right|^{-\frac{1}{2}} \\
&\quad \times \exp\left(\boldsymbol{c}'(s_{i_1}\boldsymbol{B}_{i_1} + s_{i_2}\boldsymbol{B}_{i_2})\left(\boldsymbol{I}_T - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1} - 2s_{i_2}\sigma_e^2\boldsymbol{B}_{i_2}\right)^{-1}\boldsymbol{c}\right).
\end{aligned}$$

Since $\text{cov}\,(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v},\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v}) = \text{E}[(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v})(\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v})] - \text{E}\,(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v})\,\text{E}\,(\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v})$ and $\text{E}[(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v})(\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v})] = \text{D}^2_{s_{i_1}s_{i_2}}M(s_{i_1},s_{i_2})|_{s_{i_1}=0,s_{i_2}=0}$, we can write

$$\begin{aligned}
\text{cov}\,(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v},\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v}) &= 2\sigma_e^4\,\text{tr}\,(\boldsymbol{B}_{i_1}\boldsymbol{B}_{i_2}) + 4\sigma_e^2\boldsymbol{c}'\boldsymbol{B}_{i_1}\boldsymbol{B}_{i_2}\boldsymbol{c} \\
&= 2\sigma_e^4\frac{1}{T}\,\text{tr}(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}) \\
&\quad +4\sigma_e^2\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{c}.
\end{aligned}$$

Noticing that

$$\begin{aligned}
\text{cov}\left(\sqrt{T}\text{D}_{i_1}Q_T\,(\boldsymbol{\theta},\boldsymbol{\beta})\,,\sqrt{T}\text{D}_{i_2}Q_T\,(\boldsymbol{\theta},\boldsymbol{\beta})\right) &= \text{cov}\left(\frac{1}{2}\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v},\frac{1}{2}\boldsymbol{v}'\boldsymbol{B}_{i_2}\boldsymbol{v}\right) \\
&= \frac{\sigma_e^4}{2}\frac{1}{T}\,\text{tr}\left(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\right) \\
&\quad +\sigma_e^2\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_2}\boldsymbol{C}^{-1}\boldsymbol{c},
\end{aligned}$$

completes the proof of equation (25). To show that equation (26) holds we proceed exactly as above. Let $\tilde{\boldsymbol{x}}_{\cdot j_1} = \frac{1}{\sqrt{T}}\boldsymbol{x}_{\cdot j_1}$. We write,

$$\begin{aligned}
\varkappa &= -\frac{1}{2\sigma_e^2}(\boldsymbol{v}-\boldsymbol{c})'(\boldsymbol{v}-\boldsymbol{c}) + s_{i_1}\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v} + s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1}\boldsymbol{v} \\
&= -\frac{1}{2\sigma_e^2}[\boldsymbol{v} - (\boldsymbol{I}_T - 2\sigma_e^2 s_{i_1}\boldsymbol{B}_{i_1})^{-1}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1})]'(\boldsymbol{I}_T - 2\sigma_e^2 s_{i_1}B_{i_1}) \\
&\quad [\boldsymbol{v} - (\boldsymbol{I}_T - 2\sigma_e^2 s_{i_1}\boldsymbol{B}_{i_1})^{-1}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1})] + \\
&\quad \frac{1}{2\sigma_e^2}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1})'(\boldsymbol{I}_T - 2\sigma_e^2 s_{i_1}\boldsymbol{B}_{i_1})^{-1}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1}) - \frac{1}{2\sigma_e^2}\boldsymbol{c}'\boldsymbol{c}.
\end{aligned}$$

Then, the corresponding moment generating function can be written as

$$\begin{aligned}
M(s_{i_1},s_{j_1}) &= \text{E}\exp(s_{i_1}\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v} + s_{j_1}\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{v}) \\
&= \left|\boldsymbol{I} - 2s_{i_1}\sigma_e^2\boldsymbol{B}_{i_1}\right|^{-\frac{1}{2}} \\
&\quad \times\exp(\frac{1}{2\sigma_e^2}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1})'(\boldsymbol{I}_T - 2\sigma_e^2 s_{i_1}\boldsymbol{B}_{i_1})^{-1}(\boldsymbol{c}+\sigma_e^2 s_{j_1}\tilde{\boldsymbol{x}}_{\cdot j_1}) \\
&\quad -\frac{1}{2\sigma_e^2}\boldsymbol{c}'\boldsymbol{c}).
\end{aligned}$$

Immediately, using the matrix differentials, $\text{D}\det(\boldsymbol{B}) = \det(\boldsymbol{B})\,\text{tr}\left(\boldsymbol{B}^{-1}\text{D}\boldsymbol{B}\right)$ and $\text{D}\boldsymbol{B}^{-1} = -\boldsymbol{B}^{-1}\text{D}\boldsymbol{B}\boldsymbol{B}^{-1}$, we get

$$\begin{aligned}
\text{E}\left((\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v})(\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{v})\right) &= \text{D}^2_{s_{i_1}s_{j_1}}M(s_{i_1},s_{j_1})|_{s_{i_1}=0,s_{j_1}=0} \\
&= \sigma_e^2\,\text{tr}\,(\boldsymbol{B}_{i_1})\left(\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{c}\right) + \left(\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{c}\right)(\boldsymbol{c}'\boldsymbol{B}_{i_1}\boldsymbol{c}) + 2\sigma_e^2\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{B}_{i_1}\boldsymbol{c}.
\end{aligned}$$

Therefore, we have $\operatorname{cov}\left(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v}, \tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{v}\right) = \mathrm{E}[(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v})(\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{v})] - \mathrm{E}\left(\boldsymbol{v}'\boldsymbol{B}_{i_1}\boldsymbol{v}\right)\mathrm{E}\left(\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{v}\right) = 2\sigma_e^2\tilde{\boldsymbol{x}}'_{\cdot j_1}\boldsymbol{B}_{i_1}\boldsymbol{c}$. Equation (27) follows trivially. Finally, we note that, in expression (25), $\frac{1}{T}\operatorname{tr}(\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1})$ is a component in $\mathrm{D}_{0i_10i_2}R_T$ and $\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_{i_1}\boldsymbol{C}^{-1}\boldsymbol{c}$ is a component in $\mathrm{D}_{i_10i_2}U_T$. Therefore, (25) converges uniformly following Propositions 2 and 3. Uniform convergence of (26) and (27) results from Proposition 4 and Assumption 4, respectively.

**Theorem A.1 (adapted from Davidson (1994))** Let $\{\boldsymbol{g}_{t,T}\}_{t=1}^{T}$ be a triangular array of $p$ dimensional random vectors and let $\frac{1}{T}\operatorname{var}\left(\sum_{t=1}^{T}\boldsymbol{g}_{t,T}\right)$ converge to a semi-positive definite matrix $\boldsymbol{\Sigma}$. If for $\forall\boldsymbol{\alpha}\in\mathbb{R}^p$ satisfying $\boldsymbol{\alpha}'\boldsymbol{\alpha}=1$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\alpha}'\boldsymbol{g}_{t,T}$ converges to a normal random variable in distribution as $T\to\infty$ and $\mathrm{E}\left(\sum_{t=1}^{T}\boldsymbol{g}_{t,T}\right)=\boldsymbol{0}_p$. Then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{g}_{t,T}\xrightarrow{d}\mathrm{N}\left(\boldsymbol{0}_p,\boldsymbol{\Sigma}\right)$.

**Proof of Theorem 5** Building on Theorem A.1, let $p=k+I+1$ and notice that $\boldsymbol{g}_T\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}\right)=\boldsymbol{0}_p$. In Theorem 4, we have proven the consistency of $\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}\right)$ with respect to $(\boldsymbol{\theta}^*,\boldsymbol{\beta}^*)$. To prove the asymptotic normality of $\boldsymbol{g}_T\left(\boldsymbol{\theta}^*,\boldsymbol{\beta}^*\right)$, we need to show that $\boldsymbol{g}_T\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}\right)$ is normally distributed and, given the equicontinuity of the function $\boldsymbol{g}_T\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)$ together with the consistency property, the desired result will follow. Note that $\mathrm{E}\left(\boldsymbol{g}_T\left(\boldsymbol{\theta}^*,\boldsymbol{\beta}^*\right)\right)=\boldsymbol{0}_p$ and, by Proposition 5, $\operatorname{var}(\boldsymbol{g}_T(\boldsymbol{\theta}^*,\boldsymbol{\beta}^*))$ uniformly converges to a semi-positive definite matrix $\boldsymbol{\Sigma}^*$ as $T\to\infty$. The normality of the last $k$ rows of $\boldsymbol{g}_T(\boldsymbol{\theta},\boldsymbol{\beta})$, given by $\mathrm{D}_{\boldsymbol{\beta}}m_T(\boldsymbol{\beta})$, is a direct consequence of the normality of $\boldsymbol{y}$. We need to show the asymptotic normality of the first $I+1$ rows of $\boldsymbol{g}_T(\boldsymbol{\theta},\boldsymbol{\beta})$ given by $\mathrm{D}_{\boldsymbol{\theta}}Q_T(\boldsymbol{\theta};\boldsymbol{\beta})$. Using Theorem A.1, we define $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{t,T}=\mathrm{D}_{\boldsymbol{\theta}}Q_T\left(\boldsymbol{\theta},\boldsymbol{\beta}\right)$. Let $C_0$ denote a non-stochastic term that may depend on $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. For any $\boldsymbol{\alpha}\in\mathbb{R}^{I+1}$ we have

$$
\begin{aligned}
\sqrt{T}\sum_{i=0}^{I}\alpha_i\mathrm{D}_iQ_T\left(\boldsymbol{\theta},\boldsymbol{\beta}\right) &= \frac{1}{\sqrt{T}}\left(\boldsymbol{\psi}(\boldsymbol{X})-\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{C}^{-1}\sum_{i=0}^{I}\alpha_i\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{e} \\
&\quad +\frac{1}{2\sqrt{T}}\boldsymbol{e}'\boldsymbol{C}^{-1}\sum_{i=0}^{I}\alpha_i\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{e}+C_0 \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\gamma_t\sigma_e\delta_tz_t+\frac{1}{2}\gamma_t\sigma_e^2z_t^2)+C_0,
\end{aligned}
$$

where $\gamma_t$ for $t = 1, \ldots, T$ are the eigenvalues of $\boldsymbol{C}^{-1}\left\{\sum_{i=0}^{I}\alpha_i \boldsymbol{H}_i\right\}\boldsymbol{C}^{-1}$, $\boldsymbol{\nu}_t$ are the corresponding eigenvectors such that $\boldsymbol{\nu}_t'\boldsymbol{\nu}_t = 1$, $z_t \equiv \boldsymbol{\nu}_t'\boldsymbol{e}/\sigma_e$ with $z_t \sim \mathrm{N}(0,1)$, and $\delta_t \equiv \boldsymbol{\nu}_t'\left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)$. Notice that by the Cauchy-Schwarz inequality

$$\frac{1}{\sqrt{T}}|\delta_t| \leq \sqrt{\frac{1}{T}\left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{\psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta}\right)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\psi(\boldsymbol{x}_{t\cdot}) - \boldsymbol{x}_{t\cdot}\boldsymbol{\beta})^2}. \tag{49}$$

By Assumption 4 this implies that there exist a constant $C_1 < \infty$ such that

$$\frac{1}{\sqrt{T}}|\delta_t| \leq C_1,$$

for any $T$ and $t = 1, \ldots, T$ on $B$. Next we will show that the condition

$$|\gamma_t| \leq C_2 < \infty, \tag{50}$$

is satisfied for all $t$. Let $\kappa_t(\boldsymbol{A})$ denote the $t$'th eigenvalue of the $T \times T$ matrix $\boldsymbol{A}$ where $\kappa_1(\boldsymbol{A}) \leq \kappa_2(\boldsymbol{A}) \leq ... \leq \kappa_T(\boldsymbol{A})$. From Lutkepohl (1996, page 66, expression 13.a), we can write

$$\kappa_t\left(\boldsymbol{C}^{-1}\alpha_i\boldsymbol{H}_i\boldsymbol{C}^{-1}\right) = \alpha_i\kappa_t\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right), \tag{51}$$

and, from Weyl's theorem (Lutkepohl, 1996, page 162), we have

$$\begin{aligned}\gamma_T &\equiv \kappa_T\left(\boldsymbol{C}^{-1}\left\{\sum_{i=0}^{I}\alpha_i\boldsymbol{H}_i\right\}\boldsymbol{C}^{-1}\right)\\ &\leq \sum_{i=0}^{I}\kappa_T\left(\boldsymbol{C}^{-1}\alpha_i\boldsymbol{H}_i\boldsymbol{C}^{-1}\right)\\ &= \sum_{i=0}^{I}\alpha_i\kappa_T\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right),\end{aligned}$$

where the last equality follows from (51). By the triangular inequality we can write

$$|\gamma_T| \leq \sum_{i=0}^{I}|\alpha_i|\left|\kappa_T\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right)\right|.$$

Since $\kappa_t\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right) \geq 0$ for all $i = 0, 1, ..., I$, and all $t = 1, 2, ..., T$ , condition (50) will hold if $\kappa_T\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right)$ is bounded from above for all values of $i$. Since $\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}$ and $\frac{1}{\lambda_0\lambda_i}\boldsymbol{I} - \boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}$ are both positive definite (real symmetric) matrices for all $i$ (and $t$), we have that (Lutkepohl, 1996, page 162)

$$\kappa_T\left(\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\right) \leq \frac{1}{\lambda_0\lambda_i} < \infty,$$

for all $i = 1, 2, ..., I$, and we conclude that (50) holds (the last inequality follows from Assumption ii. In order to show asymptotic normality of $\sqrt{T}\sum_{i=0}^{I} \alpha_i \mathrm{D}_i Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ we need to check the moment condition in Liapunov's Theorem (Theorem 23.11 in Davidson (1994)), i.e. verify that the following condition is satisfied

$$\lim_{T\to\infty} \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \mathrm{E}\left|\gamma_t \sigma_e \delta_t z_t + \frac{1}{2}\gamma_t \sigma_e^2 z_t^2\right|^3 = 0. \tag{52}$$

Define $C_3 \equiv \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \mathrm{E}\left|\gamma_t \sigma_e \delta_t z_t + \frac{1}{2}\gamma_t \sigma_e^2 z_t^2\right|^3$. We have

$$\begin{aligned}
C_3 &\leq \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \mathrm{E}\left(|\gamma_t \sigma_e \delta_t z_t| + \left|\frac{1}{2}\gamma_t \sigma_e^2 z_t^2\right|\right)^3 \\
&= \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \left(\mathrm{E}|\gamma_t \sigma_e \delta_t z_t|^3 + \mathrm{E}\left|\frac{1}{2}\gamma_t \sigma_e^2 z_t^2\right|^3 + \frac{3}{2}\mathrm{E}|\gamma_t \sigma_e \delta_t z_t|^2 \left|\gamma_t \sigma_e^2 z_t^2\right| + \frac{3}{4}\mathrm{E}|\gamma_t \sigma_e \delta_t z_t| \left|\gamma_t \sigma_e^2 z_t^2\right|^2\right) \\
&= \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \left(|\gamma_t|^3 \sigma_e^3 |\delta_t|^3 \mathrm{E}|z_t|^3 + \frac{1}{8}|\gamma_t|^3 \sigma_e^6 \mathrm{E}|z_t|^6 + \frac{3}{2}|\gamma_t|^3 |\delta_t|^2 \sigma_e^4 \mathrm{E}|z_t|^4 + \frac{3}{4}|\gamma_t|^3 |\delta_t| \sigma_e^5 \mathrm{E}|z_t|^5\right) \\
&\leq C_2 C_1 \sigma_e^3 \mathrm{E}|z_t|^3 \frac{1}{T}\sum_{t=1}^{T} \gamma_t^2 \delta_t^2 + \frac{C_2^3 \sigma_e^6 \mathrm{E}|z_t|^6}{8\sqrt{T}} + \frac{3}{2} C_2^2 C_1 \sigma_e^4 \mathrm{E}|z_t|^4 \frac{1}{T}\sum_{t=1}^{T} |\gamma_t| |\delta_t| \\
&\quad + \frac{\frac{3}{4} C_2^2 \sigma_e^5 \mathrm{E}|z_t|^5}{\sqrt{T}} \frac{1}{T}\sum_{t=1}^{T} |\gamma_t| |\delta_t|.
\end{aligned}$$

Since $z_t$ is a Gaussian random variable, $\mathrm{E}|z_t|^j$ is bounded for all $j \in \mathbb{N}$. For (52) to be satisfied, it is sufficient to show that the following two conditions hold:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \gamma_t^2 \delta_t^2 = 0 \tag{53}$$

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} |\gamma_t| |\delta_t| = 0. \tag{54}$$

To verify (53) notice that we can write the condition as

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\gamma_t^2\delta_t^2 &= \frac{1}{T}\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\left\{\sum_{i=0}^{I}\alpha_i\boldsymbol{H}_i\right\}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\left\{\sum_{i=0}^{I}\alpha_i\boldsymbol{H}_i\right\}\boldsymbol{C}^{-1}\boldsymbol{c}\right) \\
&= \frac{1}{T}\sum_{i=0}^{I}\alpha_i^2\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{c}\right)+ \\
&\quad \frac{1}{T}\sum_{i=0}^{I}\sum_{j>i}^{I}\alpha_i\alpha_j\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_j\boldsymbol{C}^{-1}\boldsymbol{c}\right) \\
&\quad \frac{1}{T}\sum_{j=0}^{I}\sum_{i>j}^{I}\alpha_i\alpha_j\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_j\boldsymbol{C}^{-1}\boldsymbol{c}\right).
\end{aligned}
$$

Define

$$
\Upsilon_{ijT}\equiv\frac{1}{T}\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{H}_i\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{H}_j\boldsymbol{C}^{-1}\boldsymbol{c}\right),
$$

and notice that by Cauchy-Schwartz (Lutkepohl, 1996, page 43)

$$
|\Upsilon_{ijT}|\leq\sqrt{\Upsilon_{iiT}}\sqrt{\Upsilon_{jjT}}.
$$

From this inequality, it is easy to verify that $\lim_{T\to\infty}|\Upsilon_{ijT}|=0$, since

$$
\begin{aligned}
\lim_{T\to\infty}\Upsilon_{iiT} &\leq \lim_{T\to\infty}\frac{1}{T}\frac{1}{\lambda_i^2}\operatorname{tr}\left(\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{c}\right) \\
&= -\frac{1}{\lambda_i^2}\lim_{T\to\infty}\mathrm{D}_0U_T \\
&= 0,
\end{aligned}
$$

for all $i=0,1,2,...,I$, and from Proposition 1, we have that $\lim_{T\to\infty}\mathrm{D}_0U_T=0$, and $\lambda_i$ for $i=0,1,2,...,I$ is bounded away from zero by Assumption ii. Furthermore, since $\sum_{i=0}^{I}\alpha_i^2=1$ (see Theorem A.1), it follows that $|\alpha_i|<1$ and $|\alpha_i\alpha_j|<1$ for all $i,j=0,1,...,I$. Consequently, we can write

$$
\begin{aligned}
\lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T}\gamma_t^2\delta_t^2\right| &\leq \sum_{i=0}^{I}\left|\alpha_i^2\right|\lim_{T\to\infty}|\Upsilon_{iiT}|+\sum_{i=0}^{I}\sum_{j>i}^{I}|\alpha_i\alpha_j|\lim_{T\to\infty}|\Upsilon_{ijT}|+\sum_{i=0}^{I}\sum_{i>j}^{I}|\alpha_i\alpha_j|\lim_{T\to\infty}|\Upsilon_{ijT}| \\
&= 0,
\end{aligned}
$$

and condition (53) is verified. To verify condition (54), we use Cauchy's inequality and obtain

$$
\frac{1}{T}\sum_{t=1}^{T}|\gamma_t|\,|\delta_t|\leq\sqrt{\frac{1}{T}\sum\gamma_t^2\delta_t^2},
$$

and the desired result follows. Finally, the asymptotic variance $\boldsymbol{\Sigma}^*$ is obtained once we show that

$$\lim_{T\to\infty} \mathrm{D}_{i_1} M_{x_{\cdot j_1} x_{\cdot j_1}}(\boldsymbol{\theta}) \to 0, \tag{55}$$

uniformly on $\Theta \times B \times A$ for all $j_1 = 1, 2, ...., k$. To show condition (55), notice that

$$\mathrm{D}_i M_{x_{\cdot j} x_{\cdot j}}(\boldsymbol{\theta}) = -\frac{1}{T} \operatorname{tr}\left(\boldsymbol{x}'_{\cdot j} \boldsymbol{C}^{-1} \boldsymbol{H}_i \boldsymbol{C}^{-1} \boldsymbol{x}_{\cdot j}\right).$$

Then

$$\begin{aligned}
\left|\lim_{T\to\infty} \mathrm{D}_i M_{x_{\cdot j} x_{\cdot j}}(\boldsymbol{\theta})\right| &\leq \sqrt{\lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{x}'_{\cdot j} \boldsymbol{C}^{-1} \boldsymbol{H}_i \boldsymbol{C}^{-1} \boldsymbol{C}^{-1} \boldsymbol{H}_i \boldsymbol{C}^{-1} \boldsymbol{x}_{\cdot j}\right)} \sqrt{\lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{x}'_{\cdot j} \boldsymbol{x}_{\cdot j}\right)} \\
&\leq \sqrt{\frac{1}{\lambda_i^2} \lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\boldsymbol{x}'_{\cdot j} \boldsymbol{C}^{-1} \boldsymbol{C}^{-1} \boldsymbol{x}_{\cdot j}\right)} \sqrt{\lim_{T\to\infty} \frac{1}{T} \boldsymbol{x}'_{\cdot j} \boldsymbol{x}_{\cdot j}} \\
&= 0,
\end{aligned}$$

for all $i = 0, 1, ..., I$ and $j = 1, 2, ..., K$, where the last equality follows from Proposition 4 (first term converges to zero) and Assumption 4 (last term converges to a finite constant). This completes the proof.

**Proof of Theorem 6** First, we need to show the convergence of the matrices $\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$, $\mathrm{D}^2_{\boldsymbol{\beta\beta}} m_T(\boldsymbol{\beta})$, and $\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$. The convergence of $\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ is established in Theorem 3. The convergence of $\mathrm{D}^2_{\boldsymbol{\beta\beta}} m_T(\boldsymbol{\beta})$ follows trivially from Assumption 4. We need to prove the convergence of $\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$. Consider a typical element of $\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ given by $\mathrm{D}^2_{i_1 j_1} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ for $i_1 = 0, 1, ..., I$ and $j_1 = 1, 2, ..., k$. We can write

$$\begin{aligned}
\mathrm{D}^2_{i_1 j_1} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) &= -\frac{1}{T} \boldsymbol{x}'_{\cdot j_1} \boldsymbol{C}^{-1} \boldsymbol{H}_{i_i} \boldsymbol{C}^{-1} \boldsymbol{c} - \frac{1}{T} \boldsymbol{x}'_{\cdot j_1} \boldsymbol{C}^{-1} \boldsymbol{H}_{i_i} \boldsymbol{C}^{-1} \boldsymbol{e} \\
&= -\mathrm{D}_{i_1} M_{\boldsymbol{\psi} \boldsymbol{x}_{\cdot j_1}} - \sum_{j=1}^{k} \beta_j \mathrm{D}_{i_1} M_{\boldsymbol{x}_{\cdot j_1} \boldsymbol{x}_{\cdot j}} - \frac{1}{T} \boldsymbol{x}'_{\cdot j_1} \boldsymbol{C}^{-1} \boldsymbol{H}_{i_i} \boldsymbol{C}^{-1} \boldsymbol{e},
\end{aligned}$$

where the first term converges as $T \to \infty$ according to Proposition 4 for all $i_1, j_1$, and the last two terms converge to zero by Proposition 5 and Assumption 3, respectively. Now, define $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\beta}')'$ and let $Q_T(\boldsymbol{\zeta})$ and $m_T(\boldsymbol{\beta})$ be given by (8) and (22), respectively. Under Assumptions 1 - 4, the following conditions are satisfied: *i.* $\widehat{\boldsymbol{\zeta}}_T \xrightarrow{p} \boldsymbol{\zeta}^*$ (by Theorem 4). *ii.* $Q_T(\boldsymbol{\zeta})$ and $m_T(\boldsymbol{\beta})$ are twice continuously differentiable. *iii.* $\sqrt{T}\boldsymbol{g}_T(\boldsymbol{\zeta}^*)$=$(\sqrt{T}\mathrm{D}_{\boldsymbol{\theta}} Q_T(\boldsymbol{\zeta}^*)', \sqrt{T}\mathrm{D}_{\boldsymbol{\beta}} m_T(\boldsymbol{\beta}^*)')'$ converges to a normal random variable

$\mathrm{N}(0, \boldsymbol{\Sigma}^{*})$ in distribution (by Theorem 5). *iv.* $\mathrm{D}^2_{\boldsymbol{\theta\theta}} Q_T(\boldsymbol{\zeta})$, $\mathrm{D}^2_{\boldsymbol{\beta\beta}} m_T(\boldsymbol{\beta})$ and $\mathrm{D}^2_{\boldsymbol{\theta\beta}} Q_T(\boldsymbol{\zeta})$ converge to nonsingular matrices for any $\boldsymbol{\zeta}$ in a neighborhood of $\boldsymbol{\zeta}^{*}$. Conditions *i.*-*iv.* are sufficient conditions to obtain the desired result (Dahl and Qin, 2004; Theorem 9).

**Proof of Theorem 7** Derive the first and second moments of $\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sqrt{T}}$:

$$\begin{aligned}\frac{1}{T}\,\mathrm{E}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) &= \sigma^2\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{c} + \sigma^2\frac{1}{T}\,\mathrm{E}\left(\boldsymbol{e}'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{e}\right)\\ &= -\sigma^2\mathrm{D}_0 U_T - \sigma^2\sigma_e^2\mathrm{D}^2_{00}R_T,\end{aligned}$$

then

$$\mathrm{E}\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sqrt{T}}\right) = -\sigma^2\sqrt{T}\mathrm{D}_0 U_T - \sigma^2\sigma_e^2\sqrt{T}\mathrm{D}^2_{00}R_T.$$

Notice that

$$\frac{1}{T}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \sigma^2\mathrm{D}_0U_T + 2\sigma^2\frac{1}{T}\boldsymbol{c}'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{e} + \sigma^2\frac{1}{T}\boldsymbol{e}'\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\boldsymbol{e}. \qquad (56)$$

Now, let the eigenvalues and eigenvectors of $\boldsymbol{C}^{-1}$ be $\gamma_t$ and $\boldsymbol{\nu}_t$, for $t = 1, \ldots, T$, respectively. Let $z_t = \frac{\boldsymbol{\nu}'\boldsymbol{e}}{\sigma_e} \sim \mathrm{N}(0,1)$ and $a_t = \boldsymbol{\nu}'\boldsymbol{c}$, $t = 1, \ldots, T$. Then (56) can be written as

$$\frac{1}{T}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \sigma^2\mathrm{D}_0U_T + 2\sigma^2\sigma_e\frac{1}{T}\sum_{t=1}^{T}a_t\gamma_t^2 z_t + \sigma^2\sigma_e^2\frac{1}{T}\sum_{t=1}^{T}\gamma_t^2 z_t^2. \qquad (57)$$

Using that $\mathrm{cov}(z_t, z_t^2) = 0$ and $\mathrm{var}(z_t^2) = 2$, we write

$$\begin{aligned}\mathrm{var}\left(\frac{1}{\sqrt{T}}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\right) &= \frac{4\sigma^4\sigma_e^2}{T}\sum_{t=1}^{T}a_t^2\gamma_t^4\,\mathrm{var}(z_t) + \\ &\quad + \frac{4\sigma^4\sigma_e^3}{T}\sum_{t=1}^{T}a_t\gamma_t^4\,\mathrm{cov}(z_t, z_t^2) + \frac{\sigma^4\sigma_e^4}{T}\sum_{t=1}^{T}\gamma_t^4\,\mathrm{var}(z_t^2) = \\ &= -\frac{2}{3}\sigma^4\sigma_e^2\mathrm{D}^3_{000}U_T - \frac{1}{3}\sigma^4\sigma_e^4\mathrm{D}^4_{0000}R_T. \qquad (58)\end{aligned}$$

Asymptotically

$$\lim_{T\to\infty}\mathrm{var}\left(\frac{1}{\sqrt{T}}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\right) \to -\frac{1}{3}\sigma^4\sigma_e^4\mathrm{D}^4_{0000}R,$$

by Propositions 1, 2, and 3. The proof of asymptotic normality of $\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sqrt{T}}$ follows in a similar fashion to that of Theorem 5. The asymptotic normality of the last term of (56), which is a multiple of $\mathrm{D}_0Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$, follows immediately from Theorem 5. The second term of (56) is already a normal random variable by Assumption 3.

**Proof of Corollary 4** We standardize the asymptotic normal random variable on the left hand side of (30) by the actual standard deviation of $\frac{1}{\sqrt{T}}\hat{\varepsilon}'\hat{\varepsilon}$ in (58) and get

$$\frac{\frac{1}{\sqrt{T}}\left(\hat{\varepsilon}'\hat{\varepsilon}+\sigma^2 T\mathrm{D}_0 U_T+\sigma^2\sigma_e^2 T\mathrm{D}_{00}^2 R_T\right)}{\sqrt{-\frac{2}{3}\sigma^4\sigma_e^2\mathrm{D}_{000}^3 U_T-\frac{1}{3}\sigma^4\sigma_e^4\mathrm{D}_{0000}^4 R_T}}\overset{a}{\sim}\mathrm{N}(0,1).$$

We replace unknown parameters in the above expression by their consistent estimates. Then we multiply the numerator and the denominator of the left hand side of the above expression by $\frac{1}{\sqrt{T}}$. After removing the term $\sigma^2\mathrm{D}_0 U_T$, which converges to 0, we get the test $ResT_{DGQ}$.