

The limiting behavior of the estimated parameters in a misspecified random field regression model*

Christian M. Dahl[†]

Yu Qin

Department of Economics
Purdue University

Department of Statistics
Purdue University

August 15, 2005

Abstract

This paper examines the limiting properties of the estimated parameters in the random field regression model recently proposed by Hamilton (Econometrica, 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.” Contrary to existing literature on the asymptotic properties of the estimated parameters in random field models our results do not require that the explanatory variables are sampled on a grid. However, as a consequence the random field model specification introduces non-stationarity and non-ergodicity in the misspecified model and it becomes non-trivial, relative to the existing literature, to establish the limiting behavior of the estimated parameters. The asymptotic results are obtained by applying some convenient new uniform convergence results that we propose. This theory may have applications beyond those presented here. Our results indicate that classical statistical inference techniques, in general, works very well for random field

*We would like to thank Jim Hamilton, Gloria González-Rivera and seminar participants at Cornell, Ohio State and Purdue for very helpful comments. Notation follows Abadir and Magnus (2002).

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

regression models in finite samples and that these models successfully can fit and uncover many types of nonlinear structures in data.

1 Introduction

Hamilton (2001) proposes a novel parametric regression model approach based on random fields for estimating an unknown conditional mean function. The strength of this approach is that it does not rely on any functional form being specified prior to estimation. Random field models have wide applications in geostatistics, environmental, ecological and agricultural sciences and have been used to model images, temperature, wind, epidemiology, pollution etc. However, in all of these applications, the random field model has been treated as the true data generating process. In most economic applications it would be safe to assume that the parametric random field almost surely would be misspecified in the sense that the true data generating process is not a random field process. However, despite of misspecification, Hamilton (2001, p. 550-551) shows that it is possible to obtain a consistent estimator of the conditional mean function under very general conditions. Furthermore, empirical evidence and evidence based on simulations, see, e.g., Dahl (2002), Hamilton (2001,2003), Dahl and González-Rivera (2003a,b) and Dahl and Hylleberg (2004) show that the random field approach is promising in terms of having relatively better small sample fitting and forecasting abilities compared to a wide range of parametric and nonparametric alternatives based on samples of macroeconomic variables.

Allowing for misspecification in parametric regression analysis has become popular in econometrics, since typically very limited information is available regarding the actual data generating process. There are many approaches to analyzing, estimating and viewing misspecified parametric models. Distributional misspecifications or quasi maximum likelihood theory is perhaps the most common form of misspecification and originated in econometrics from White (1982, 1994) and Gouriéroux, Monfort and Trognon (1984). In addition there is an extensive literature on so-called local misspecification of population moment conditions, where misspecification is present in finite samples but not asymptotically, see, e.g., Gallant and White (1988), Hansen and Sargent(1993), Watson

(1993), Diebold et al. (1998), Hansen and Jagannathan (1997), and Kitamura (1997). In this paper we restrict attention to non-local misspecification of the first conditional population moment of the variable of interest, such that the nature of the misspecification is considered to remain constant throughout the sample. White (1980b,1981,1982) and Domowitz and White (1982) considered this type of misspecification and elegantly derived large sample behavior of the least squares estimator. More recent work includes Maasoumi and Phillips (1982), White (1994) and Hall and Inoue (2003) who derive the limiting properties of the non-locally misspecified IV, QMLE and GMM estimators respectively under standard regularity conditions.

In particular, the main purpose of this paper is to establish the large sample behavior of the likelihood based parameter estimates in a possible functional form misspecified random field regression model. An important common feature in most of the existing work on the asymptotic theory is that the objective function (either the log likelihood or the moment conditions) can be written in sums of a double array of stationary and ergodic random variables, or at least random variables with bounded moment conditions, see for example White (1994) or Davidson (1994). It then becomes a relatively straightforward task to establish the limiting behavior of the estimated parameters using a battery of law of large numbers and central limit theorems. Contrary to most (if not all) of the existing literature on the asymptotic properties of the estimated parameters in (correctly specified) random field models, our results are not based on an assumption that the explanatory variables are sampled on a grid, which would be too restrictive for the method to claim any empirical relevance in economics. However, under the general settings of the random field model of Hamilton (2001), although the objective function, which is the log likelihood function, can be written in sums of a double array of random variables (as shown in subsection 2.3), these will not be stationary nor ergodic and do not have bounded moment conditions. Consequently, a new approach is proposed to establish the required convergent results.

Our results not only provide theoretical/asymptotic descriptions of the properties of the estimated parameters in Hamilton's random field regression model, but many of the results may also be useful in spatial statistical analysis where a wide range of new and alternative parameterized covariance functions are used but where the limiting

properties of the estimated models are still unknown. It should be emphasized that although convergence results are provided, the limiting functions typically cannot be given any closed form solutions. As a result one should be careful interpreting, say, the estimated parameters in any strict sense. Finally, it is worth mentioning that as a by-product of the theoretical analysis, a new simple two-stage estimator is proposed, which based on a limited set of simulation experiments seem to be as accurate, but much faster, relative to Hamilton's (2001) original proposed maximum likelihood estimator.

The paper is organized as follows: In Section 2 the random field regression model is introduced and assumptions about the data generating process are made. Section 3 establishes consistency and asymptotic normality of the estimated parameters of the possible misspecified random field regression. In Section 4 various finite sample properties of the estimated parameters are illustrated based on a small simulation study. Finally, Section 5 concludes.

2 Preliminaries

In this section, the random field regression model introduced by Hamilton (2001) is presented. Generally, Hamilton's notation is adapted with only minor modifications. The required assumptions will also be discussed and motivated in this section.

2.1 The random field regression model

The basic idea underlying the flexible regression model approach suggested by Hamilton (2001) is to view not only the endogenous variable as a realization of a stochastic process but the functional form of the conditional mean function itself as the outcome of a random process. Let $y_t \in \mathbb{R}$, $\mathbf{x}_{t.} \in \mathbb{R}^k$ and consider the model

$$y_t = \mu(\mathbf{x}_{t.}) + \tilde{\epsilon}_t, \tag{1}$$

where $\tilde{\epsilon}_t$ is a sequence of independent $N(0, \sigma^2)$ distributed error terms and $\mu(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$ is a random function of a $(k, 1)$ vector $\mathbf{x}_{t.}$, which is assumed to be deterministic.¹ Let

¹Without loss of generality it is assumed that all variables are demeaned.

the mean of the conditional distribution, i.e. $\mu(\mathbf{x}_{t.})$, be represented as having a linear part and a stochastic nonlinear part, i.e., as

$$\mu(\mathbf{x}_{t.}) = \mathbf{x}_{t.}\boldsymbol{\beta} + \lambda m(\mathbf{g} \odot \mathbf{x}_{t.}), \quad (2)$$

where for any choice of \mathbf{z} , $m(\mathbf{z})$ is a realization from a Gaussian random field satisfying

$$m(\mathbf{z}) \sim N(0, 1), \quad (3)$$

$$\mathbb{E}(m(\mathbf{z})'m(\mathbf{w})) = \mathbf{H}(h), \quad (4)$$

and where h is defined as $h \equiv \frac{1}{2}[(\mathbf{z} - \mathbf{w})'(\mathbf{z} - \mathbf{w})]^{1/2}$.² The realization of $m(\cdot)$ is viewed as being predetermined and independent of $\{\mathbf{x}_{1.}, \dots, \mathbf{x}_{T.}, \tilde{\epsilon}_1, \dots, \tilde{\epsilon}_T\}$. The covariance matrix $\mathbf{H}(h)$ is defined by

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

where $G_k(h, r)$, $0 < h \leq r$ is³

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

Closed form expressions for $\mathbf{H}(h)$, for $k = 1, \dots, 5$ can easily be obtained and are provided in Hamilton (2001).⁴⁵ This particular choice of covariance matrix perhaps seems somewhat restrictive. This is, however, not true. Hamilton (2001, p. 550-551) shows that

²Here \mathbf{g} is a $k \times 1$ vector of parameters and \odot denotes element-by-element multiplication i.e. $\mathbf{g} \odot \mathbf{x}_{t.}$ is the Hadamard product. $\boldsymbol{\beta}$ is a $(k, 1)$ vector of coefficients.

³ $G_k(h, r)$ can then 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 $k = 2, 3, \dots$

⁴The correlation between $m(\mathbf{z})$ and $m(\mathbf{w})$ is given by the volume of the intersection of a k dimensional unit spheroid centered at \mathbf{z} and a k dimensional unit spheroid centered at \mathbf{w} relative to the volume of a k dimensional unit spheroid. Hence, the correlation between $m(\mathbf{z})$ and $m(\mathbf{w})$ is zero if the Euclidean distance between \mathbf{z} and \mathbf{w} is ≥ 2 .

⁵The reader interested in a critical review on the choice of an appropriate covariance function is referred to Dahl and González-Rivera (2003)

if the "true" conditional mean function (data generating process) can be represented as Taylor series expansions or Fourier sine series expansions then it can be estimated consistently based on choice of $\mathbf{H}(h)$ given by (5). These results suggest that the class of nonlinear functions that can be estimated consistently using (5) is quite flexible.

Since it is not possible to directly observe $m(\mathbf{z})$ - for any choice of \mathbf{z} - the functional form of $\mu(\mathbf{x}_t)$ cannot be observed. Hence, inference about the unknown parameters of the model summarized by $(\boldsymbol{\beta}, \lambda, \mathbf{g}, \sigma)$ must be based on observing the realizations of y_t and \mathbf{x}_t only. For that purpose rewrite model (1) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (7)$$

where \mathbf{y} is a $(T, 1)$ vector with t th element equal to y_t , \mathbf{X} a (T, k) matrix with t th row equal to \mathbf{x}_t , and $\boldsymbol{\varepsilon}$ is a $(T, 1)$ random vector with t th element equal to $\lambda m(\mathbf{g} \odot \mathbf{x}_t) + \tilde{\varepsilon}_t$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}_T, \lambda^2 \mathbf{H} + \sigma^2 \mathbf{I}_T)$. To avoid identification issues and to simplify the analysis, the shape parameters \mathbf{g} in the random field model will be treated as fixed and focus will be on deriving the limiting properties of the estimates of $\boldsymbol{\beta}$ (the parameters of the linear part of the model), and $(\lambda_1, \sigma_1) = (\lambda^2, \sigma^2)$ (the parameters of the nonlinear part of the model). The following assumptions on the parameters entering the nonlinear part of (1) are imposed:

Assumption 1 The parameter vector $\mathbf{g} = (g_1, g_2, \dots, g_k)'$ in the random field model (1) consists of predetermined constants. In particular, $g_i = \frac{1}{2\sqrt{k}s_i^2}$, for $i = 1, \dots, k$, where $s_i^2 = \frac{1}{T} \sum_{t=1}^T (x_{ti} - \bar{x}_i)^2$, and \bar{x}_i is the sample mean of the i th explanatory variable.

Assumption 2 Let $\lambda_1 \in \Gamma_0$ and let $\boldsymbol{\theta} = (\lambda_1, \sigma_1)' \in \Theta \subseteq \mathbb{R}_+^2$, where Θ is a compact parameter space. In particular, there exist sufficiently small but positive real numbers $\underline{\boldsymbol{\theta}} = (\underline{\lambda}, \underline{\sigma})'$, and sufficiently large positive real numbers $\bar{\boldsymbol{\theta}} = (\bar{\lambda}, \bar{\sigma})'$, such that $\lambda_1 \in [\underline{\lambda}, \bar{\lambda}]$ and $\sigma_1 \in [\underline{\sigma}, \bar{\sigma}]$.

Assumption 1 is adapted from the Bayesian analysis in Section 5 in Hamilton (2001), and corresponds to fixing the parameters at the mean of a "prior" log-normal distribution. Assumption 2 is almost similar to Assumption 5 in McDonald and Newey (1988) where $\boldsymbol{\theta}$

is referred to as the distributional parameter vector. Assumption 2 is not very restrictive in the present context and is very common in large sample theory. It is important to notice that Assumption 2 exclude the case where the actual data is generated from a linear model, since $\lambda_1 = 0$ cannot be an element of Θ . This restriction turns out to simplify many proofs considerably. In addition, when $\lambda_1 = 0$, the limiting behavior of the estimated parameters has already been established by White (1980b) and Hamilton (2001). In the next section, we will show that the parameter space of β , denoted B , is also compact implying that the product space $\Theta \times B$ is compact.

2.2 The data generating process

We make the following definitions and assumptions regarding the explanatory variables and the actual data generating process.

Definition 1 The deterministic sequence $\{\mathbf{x}_t\}$, with $\mathbf{x}_t \in A$ and A a closed rectangular subset of \mathbb{R}^k , is said to be *dense* for A uniformly on the compact space $A \times \Gamma_0 \subset \mathbb{R}^k \times \mathbb{R}$ if there exists a continuous $f : A \rightarrow \mathbb{R}$ such that $f(\mathbf{x}) > 0$ for all \mathbf{x} and such that for any $\epsilon > 0$, and any continuous $\phi : A \times A \times \Gamma_0 \rightarrow \mathbb{R}$ there exists an N such that for all $T \geq N$,

$$\sup_{A \times \Gamma_0} \left| \frac{1}{T} \sum_{s=1}^T \phi(\mathbf{x}_t, \mathbf{x}_s; \lambda_1) - \int_A \phi(\mathbf{x}_t, \mathbf{x}; \lambda_1) f(\mathbf{x}) d\mathbf{x} \right| < \epsilon$$

Definition 2 Let A , Γ_0 and $\phi(\cdot)$ be given as in Assumption 2 and Definition 1 and let $l : A \times \Gamma_0 \rightarrow \mathbb{R}$ be an arbitrary continuous function. We say that $l(\cdot)$ is representable with respect to $\phi(\cdot)$ if there exists a continuous function $f : A \rightarrow \mathbb{R}$ such that

$$l(\mathbf{x}_t; \lambda_1) = \int_A \phi(\mathbf{x}_t, \mathbf{x}; \lambda_1) f(\mathbf{x}) d\mathbf{x}$$

Assumption 3 Let \mathbf{x}_t be dense according to Definition 1 and let y_t be generated according to $y_t = \psi(\mathbf{x}_t) + e_t$, for $t = 1, 2, \dots, T$, where $\psi : \mathbb{R}^k \rightarrow \mathbb{R}$ is given as $\psi(\mathbf{x}_t) = \mathbf{x}_t \beta + l(\mathbf{x}_t; \lambda_1)$ and $l(\mathbf{x}_t; \lambda_1)$ denotes a representable function given by Definition 2. Finally, let e_t be an i.i.d. Gaussian distributed error term with zero mean and variance σ_e^2 .

Note that $l(\mathbf{x}_t; \lambda_1)$ can be any smooth nonlinear function of \mathbf{x}_t . By defining the sample version of $l(\cdot)$ as

$$l_T(\mathbf{x}_t; \lambda_1) = \frac{1}{T} \sum_{s=1}^T \phi(\mathbf{x}_t, \mathbf{x}_s; \lambda_1),$$

it follows directly from Assumption 3 that $l_T(\mathbf{x}_t; \lambda_1) \xrightarrow{p} l(\mathbf{x}_t; \lambda_1)$ uniformly on $A \times \Gamma_0$ for $\forall t$, hereby providing a necessary link between the approximating random field model and the true data generating process. Hamilton (2001) discusses only pointwise convergence of $l_T(\cdot)$ to $l(\cdot)$, in \mathbf{x}_t for $\forall t$. By Assumption 3 we require uniform convergence in \mathbf{x}_t which essential implies that $l(\cdot)$ can be replaced by $l_T(\cdot)$ asymptotically, simplifying many of the proofs. It can be argued that denseness is in fact a weaker condition than i.i.d. and that the assumption will generally be satisfied for most exogenous/endogenous explanatory variables applied in econometrics/economics. Assumption 3 implies that all probability measures can be conditioned on \mathbf{X} , that is, we can take \mathbf{X} as if it was a matrix of constants. However, we believe that the results would apply directly to the case where \mathbf{x}_t are i.i.d. random variables (see also Hamilton's (2001) discussion). In addition we make the following standard assumptions regarding the existence of the second order moments.

Assumption 4 Let $\Psi(\mathbf{X}) = (\psi(\mathbf{x}_1), \dots, \psi(\mathbf{x}_T))'$ where $\psi(\mathbf{x}_t)$ is defined as in Assumption 3. Assume: *i.* $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}'\mathbf{X}$ converges to a finite nonsingular matrix. *ii.* $\lim_{T \rightarrow \infty} \frac{1}{T} \Psi(\mathbf{X})'\Psi(\mathbf{X})$ converges to a finite scalar uniformly in β . *iii.* $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}'\Psi(\mathbf{X})$ converges to a finite $(k,1)$ vector uniformly in β .

The conditions of Assumption 4 are essentially the requirements needed to ensure convergence and asymptotic normality of the non-locally misspecified OLS estimator, see, e.g., White (1980b).

2.3 The likelihood function

From (7) it is easily seen that the random field model can be viewed as a GLS model with non-spherical covariance function $\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T$ and joint log likelihood function

$$\begin{aligned} \ell(\lambda_1, \sigma_1, \boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T| \\ &\quad - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned} \quad (8)$$

For three reasons the parameters of the linear and nonlinear part will be treated differently analytically. First, because the main interest is the large sample behavior of the estimates of $\boldsymbol{\theta}$. Secondly, because the choice of the linear (or perhaps a known nonlinear) part can be done quite flexibly, as it will not affect the asymptotic representability of $l_T(\cdot)$. Thirdly, because it appears that the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be done faster and with good numerical precision separately in a two-stage procedure. However, before turning to estimation, an analytically more convenient representation of the likelihood function is needed and for that purpose the following lemmas and notation become useful (all proofs are provided in the mathematical appendix):

Lemma 1 Let $\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T$ be the covariance matrix of the random field model given by (1). Then there exists an orthogonal matrix \mathbf{P} , such that $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}_T$ and $\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T = \mathbf{P}'\mathbf{V}\mathbf{P}$, where $\mathbf{V} = \text{diag}(\lambda_1 h_1 + \sigma_1, \dots, \lambda_1 h_T + \sigma_1)$, and $0 \leq h_1 \leq \dots \leq h_T$ are the eigenvalues of \mathbf{H} . The column vectors of \mathbf{P}' are the eigenvectors of \mathbf{H} , corresponding to h_t , for $t = 1, \dots, T$.

Lemma 2 Let h_t be defined as in Lemma 1. For any sample \mathbf{X} of sample size T , $\sum_{t=1}^T h_t = T$.

Definition 3 Let \mathbf{P} and $\boldsymbol{\Psi}(\mathbf{X})$ be defined as in Lemma 1 and Assumption 4 respectively. Define *i.* $\mathbf{v} = (v_1, \dots, v_T)' = \mathbf{y} - \mathbf{X}'\boldsymbol{\beta}$, *ii.* $\mathbf{w} = (w_1, \dots, w_T)' = \mathbf{P}\mathbf{v}$, *iii.* $\mathbf{c} = (c_1, \dots, c_T)' = \boldsymbol{\Psi}(\mathbf{X}) - \mathbf{X}\boldsymbol{\beta}$. *iv.* $\mathbf{b} = (b_1, \dots, b_T)' = \mathbf{P}\mathbf{c}$.

Lemma 2 provides an upper bound on h_t , which is equal to the sample size. It furthermore states, that the sample and population average of h_t always will be identical and equal to unity. Lemma 1 and Definition 3 make it possible to write the log likelihood as a sample average. To see this note that Definition 3 implies that $\mathbf{v}'\mathbf{v} = \mathbf{w}'\mathbf{w}$ and by using Lemma

1 the last term of (8) can be rewritten as

$$\begin{aligned}
\frac{1}{2} \mathbf{v}' (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{v} &= \frac{1}{2} \text{tr} \left(\mathbf{v}' (\mathbf{P}' \mathbf{V} \mathbf{P})^{-1} \mathbf{v} \right) \\
&= \frac{1}{2} \text{tr} (\mathbf{P} \mathbf{v} \mathbf{v}' \mathbf{P}' \mathbf{V}^{-1}) \\
&= \frac{1}{2} \sum_{t=1}^T w_t^2 (\lambda_1 h_t + \sigma_1)^{-1}.
\end{aligned} \tag{9}$$

Similarly, the second term of (8) can be written as

$$\begin{aligned}
-\frac{1}{2} \ln |\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T| &= -\frac{1}{2} \ln |\mathbf{P}' \mathbf{V} \mathbf{P}| \\
&= -\frac{1}{2} \ln |\mathbf{V}| \\
&= -\frac{1}{2} \sum_{t=1}^T \ln (\lambda_1 h_t + \sigma_1).
\end{aligned} \tag{10}$$

Consequently, by ignoring the constant term and after dividing by T , (8) can be represented as

$$Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{2T} \sum_{t=1}^T q_t(\boldsymbol{\theta}, \boldsymbol{\beta}), \tag{11}$$

where $q_t(\boldsymbol{\theta}, \boldsymbol{\beta}) = \log(\lambda_1 h_t + \sigma_1) + w_t^2 (\lambda_1 h_t + \sigma_1)$. Notice that w_t , for each $t = 1, 2, \dots, T$, depends on the entire sample, i.e., $\{y_t, \mathbf{x}_t\}_{t=1}^T$. However, due to normality of \mathbf{v}_t (Assumption 3) and since $\text{var}(\mathbf{w}) = \mathbf{P}' \text{var}(\mathbf{v}) \mathbf{P} = \sigma_e^2 \mathbf{I}_T$, $q_t(\boldsymbol{\theta}, \boldsymbol{\beta})$ for $t = 1, 2, \dots, T$ will be independent but not identical distributed random variables. The representation of the log likelihood function given by (11) shows that it is a sum of double array independent random variables, discussed by, for example, White (1994) and Davidson (1994). We have not been able to find any Law of Large Number theorem directly applicable to (11), mainly because the eigenvalues as well as the moments of w_t generally are unbounded with unknown memory properties. As a consequence we will develop some new tools to establish the asymptotic properties in the following section and we suspect that this theory may have applications beyond the present one.

3 Asymptotics

The asymptotic theory of maximum likelihood estimators associated with random field models are particularly rich in the literature on spatial statistical analysis. A general treatment of the limiting behavior of the maximum likelihood estimator that applies to correctly specified random fields is provided by Sweeting (1980). However, as pointed out by Ying (1991, 1993) and Mardia and Marshall (1984), the elementary conditions/assumption of Sweeting's results are typically very hard to verify/justify. However, by applying Sweeting's results and assuming that sample points are taken on a grid with fixed spacing (such that as the sample size increases, the sampling domain also increases) Mardia and Marshall (1984) are able to establish consistency and asymptotic normality of the estimated parameters based on the correctly specified spherical random field regression model similar to Hamilton's model for $k = 3$. Ying (1991,1993) establishes the asymptotic properties of maximum likelihood estimator when the explanatory variables are sampled on a dense and compact grid using an Ornstein-Uhlenbeck type covariance function. Ying's result are very specific to the choice of covariance function since his results depends critically on being able to write the determinant and the inverse of the covariance matrix in an explicit form. Van der Vaart (1996) shows asymptotic efficiency of the maximum likelihood estimator using Ying's (1991,1993) setup. Loh and Lam (2000), and ABT and Welch (1998) partially generalize the findings by Ying and Van der Vaarts by including a broader class of covariance functions than the Ornstein-Uhlenbeck covariance function. The class of covariance functions they consider all satisfy the so-called "product correlation rule", which as in Ying's work makes it feasible to calculate simple analytical expression for the inverse and the determinant of these matrices. In the context of misspecification analysis, Currin et. al. (1991), Sacks, Schiller, and Welch (1989), and Sacks, Welch, Mitchell, and Wynn (1989) use the random field model to study deviations of the data generating processes from a purely deterministic data generating process in a computer experiment design. They assume that the deterministic function is equal to the first moment of the random field model implying that deviations from the deterministic data generating process would be equivalent to functional misspecification of the variance. Unfortunately, the main interest in these paper is not on

the asymptotic properties of the estimated parameters and their results are not directly applicable here.

It is clear that the results from the existing literature cannot be employed directly in the present context. Many of the results cited above, either on model misspecification or in terms of the class of covariance function used, need to be generalized in order to study the asymptotic properties of the estimated parameters in Hamilton's random field regression model. Furthermore, we will not be willing to restrict the explanatory variables to be sampled on a grid, and will consider data on the variable of interest to be generated from a wide range of smooth nonlinear functional forms. We will generalize some of the convergence results in Hamilton (2001) and show that these results are basically design free in terms of the choice of covariance functions. Many of the asymptotic results will be based on a new theorem we provide which links uniform convergence of a sequence of functions to uniform convergence of the associated sequence of the differentiated function (see, Theorem 5). Under a set of very weak (standard) regularity condition this result enables us to establish asymptotic consistency and normality of the estimated parameters in a very broad class of misspecified random field regression models.

3.1 Consistency

Contrary to Hamilton (2001), we suggest estimating the parameters of the linear part of the model by OLS in a first stage, initially ignoring the possible nonlinear part of the model. This approach is well known from estimation theory of the general GLS model. In the present context the two-stage procedure can furthermore be motivated by the following lemmas.

Lemma 3 Let the Assumptions 1 – 4 hold. Then the parameter space $\Theta \times B \subseteq \mathbb{R}^{k+2}$ is compact.

Lemma 4 Let the assumptions of Lemma 3 hold. Then the OLS estimator $\hat{\beta}$ is consistent with respect to a fixed population magnitude $\beta^* \in B \subseteq \mathbb{R}^k$ given as

$$\beta^* = \left(\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}' \mathbf{X} \right)^{-1} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}' \Psi(\mathbf{X}), \quad (12)$$

where B is a compact parameter space.

Accordingly, the first step OLS estimator will asymptotically provide a consistent estimate of β^* under rather weak conditions. Note that for the consistency result of Lemma 4 to hold, the parameter space B needs to be compact. However, under the assumptions of Lemma 3 this is true by Lemma A1 in McDonald and Newey (1988). To prove consistency of θ we will use the following very general theorem suggested by Wooldridge (1994), and reported below for convenience, where β is treated as a "nuisance" parameter:

Theorem 1 Let Θ and B be sets in finite-dimensional real spaces and let $\{Q_T(\theta, \beta)\}_T$ be the sequence of objective functions defined by (11). Then $\hat{\theta} = \arg \max Q_T(\theta, \beta^*)$ satisfies $\hat{\theta} \xrightarrow{p} \theta^*$ if the following conditions hold: *i.* Θ and B are compact sets. *ii.* $\hat{\beta} \xrightarrow{p} \beta^* \in \text{int}(B)$. *iii.* $Q_T(\theta, \beta)$ is a continuous measurable function for $\forall T$. *iv.* $\exists Q^*(\theta, \beta)$ such that $Q_T(\theta, \beta) \xrightarrow{p} Q^*(\theta, \beta)$ as $T \rightarrow \infty$ uniformly on $\Theta \times B$. *v.* There exists a unique maximizer $\theta^* \in \text{int}(\Theta)$ of $Q^*(\theta, \beta^*)$.

Note that conditions *i.* and *ii.* already have been established by Lemma 3 and 4, and condition *iii.* is satisfied trivially. The proof of the existence of a function $Q^*(\theta, \beta)$ such that $Q_T(\theta, \beta) \xrightarrow{p} Q^*(\theta, \beta)$ as $T \rightarrow \infty$ uniformly on $\Theta \times B$ can be completed in the following two steps. First, existence of a non-stochastic sequence of functions $Q_T^*(\theta, \beta)$ is established, where $Q_T^*(\theta, \beta)$ depends upon the sample \mathbf{X} , such that, $Q_T(\theta, \beta) \xrightarrow{p} Q_T^*(\theta, \beta)$, as $T \rightarrow \infty$ uniformly in \mathbf{X} on $\Theta \times B$. Secondly, a unique non-stochastic function $Q^*(\theta, \beta)$ is shown to exist, that does not depend upon the sample \mathbf{X} , where $Q_T^*(\theta, \beta) \xrightarrow{p} Q^*(\theta, \beta)$ uniformly in \mathbf{X} on $\Theta \times B$ for $T \rightarrow \infty$. To complete the second step a generalization of Theorem 4.4 p. 549 in Hamilton (2001) is needed, regarding the convergence of the continuous covariance function

$$p_t(\mathbf{z}, \mathbf{w}; \theta) = \text{E}((\xi_t(\mathbf{z}) - \mu(\mathbf{z}))(\xi_t(\mathbf{w}) - \mu(\mathbf{w}))), \quad (13)$$

where $\xi_t(\mathbf{x}_t) = \text{E}(\mu(\mathbf{x}_t) | y_t, \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \dots)$. Theorem 4.4 in Hamilton is obviously essential for his proof of consistency of the random field estimator of the conditional mean function, however, pointwise convergence of the covariance function $p_t(\cdot)$ as in Hamilton (2001) turns out not to be sufficient for establishing the existence of $Q^*(\theta, \beta)$

uniformly on $\Theta \times B$. Using the conditioning rules of multivariate normal random variables, $p_t(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta})$ given by (13) can be expressed as a functional recursion which enables us to establish its limit as T grows to infinity.

Theorem 2 Under Assumptions 2 and 3, let $p_t : A \times A \rightarrow \mathbb{R}$ for $t = 0, 1, 2, \dots, T$, be defined as

$$p_t(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) = p_{t-1}(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) - \frac{p_{t-1}(\mathbf{z}, \mathbf{x}_t; \boldsymbol{\theta})p_{t-1}(\mathbf{z}, \mathbf{x}_t; \boldsymbol{\theta})}{p_{t-1}(\mathbf{x}_t, \mathbf{x}_t; \boldsymbol{\theta}) + \sigma_1}, \quad (14)$$

where $p_0(\mathbf{z}, \mathbf{w}; \lambda_1)$ is a semi-positive continuous (covariance) function, i.e., $p_0(\mathbf{z}, \mathbf{w}; \lambda_1) = \lambda_1 \mathbb{E}(m(\mathbf{z})m(\mathbf{w}))$. Then, for all $(\mathbf{z}, \mathbf{w}) \in A^2$

$$\lim_{T \rightarrow \infty} \sup_{\Theta \times A^2} p_T(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) \rightarrow 0. \quad (15)$$

Theorem 2 establishes that $p_t(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta})$ converges uniformly to zero on $A^2 \times \Theta$ as $T \rightarrow \infty$ under a fairly weak set of conditions. This result is very important since it enable us to establish consistency of the estimated conditional mean function on $A^2 \times \Theta$.

Theorem 3 Let $p_0 : A \times A \rightarrow \mathbb{R}^2$ and $\tilde{\phi} : A \rightarrow \mathbb{R}$ be arbitrary continuous functions with $p_0(\cdot, \cdot; \lambda_1)$ semi-positive. For a given sample of T observations, construct

$$l_T(\mathbf{x}; \lambda_1) = \frac{1}{T} \sum_{t=1}^T p_0(\mathbf{x}, \mathbf{x}_t; \lambda_1) \tilde{\phi}(\mathbf{x}_t). \quad (16)$$

Next, consider a sequence of samples of size $T = 1, 2, \dots$, where the T th sample $\{y_t, \mathbf{x}_t\}_{t=1}^T$ is generated from (by Assumption 3)

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + l_T(\mathbf{x}; \lambda_1) + e_t, \quad (17)$$

$$\mathbb{E}(e_t e_s) = \begin{cases} \sigma_e^2 & t = s \\ 0 & \text{otherwise} \end{cases}, \quad (18)$$

for $t = 1, 2, \dots, T$. Let $\xi_T(\mathbf{z})$ and $p_T(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta})$ be generated from the recursions

$$\xi_t(\mathbf{z}) = \xi_{t-1}(\mathbf{z}) - \frac{p_{t-1}(\mathbf{z}, \mathbf{x}_t; \boldsymbol{\theta})(y_t - \xi_{t-1}(\mathbf{x}_t))}{p_{t-1}(\mathbf{x}_t, \mathbf{x}_t; \boldsymbol{\theta}) + \sigma_1}, \quad (19)$$

and (14) respectively, starting with $p_0(\mathbf{z}, \mathbf{w}; \lambda_1)$ for $\sigma_1 > 0$. If $\{\mathbf{x}_t\}$ is dense according to Definition 1, then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum \mathbb{E}[\xi_t(\mathbf{x}_t) - \mathbf{x}_t \boldsymbol{\beta} - l_T(\mathbf{x}_t; \lambda_1)]^2 = 0, \quad (20)$$

uniformly on $\Theta \times A$.

This theorem generalizes Theorem 4.7 in Hamilton (2001) as it guarantees uniform convergence of the estimated conditional mean function. By varying $\tilde{\phi}(\cdot)$, Theorem 3 describes the general class of nonlinear models for which it is possible to obtain a consistent estimator of $\psi(\mathbf{x}_t)$ using the random field regression model in (1) with covariance function $p_0(\cdot, \cdot; \lambda_1)$ used to start the iterations (14) and (19). The result given by equation (20) will be crucial in verification of condition *iv.* in Theorem 1. The next lemma and the following Theorem 4 defines the non-stochastic sequence $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ as the expectation of $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ and gives the conditions for the uniform convergence of $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$.

Lemma 5 Let Assumptions 1 - 4 hold and consider $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ as defined in (11). Then,

$$\mathbb{E}(Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})) = Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}), \quad (21)$$

for

$$Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{2T} \sum_{t=1}^T \log(\lambda_1 h_t + \sigma_1) - \frac{1}{2T} \sum_{t=1}^T \frac{\sigma_e^2 + b_t^2}{\lambda_1 h_t + \sigma_1}, \quad (22)$$

where h_t and b_t are defined in Lemma 1 and by Definition 3 respectively.

Theorem 4 Let Assumptions 1 - 4 hold and let $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ and $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ be given by (11) and (22) respectively. Then

$$\lim_{T \rightarrow \infty} \sup_{\Theta \times B} |Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) - Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})| \xrightarrow{p} 0. \quad (23)$$

Theorem 4 completes the first step in verifying condition *iv.* of Theorem 1. To complete the second step, the following two propositions (derived from Theorem 3) are essential.

Proposition 1 Let Assumptions 1-4 hold and let b_t and h_t be defined as in Lemma 5. Then,

$$\lim_{T \rightarrow \infty} \sup_{\Theta \times B} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \rightarrow 0, \quad (24)$$

$$\lim_{T \rightarrow \infty} \sup_{\Theta \times B} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2} \rightarrow 0. \quad (25)$$

Proposition 2 Define $R_T \equiv \frac{1}{T} \sum_{t=1}^T \ln(\lambda_1 h_t + \sigma_1)$, $U_T \equiv \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{\lambda_1 h_t + \sigma_1}$, $R_T^{i,j} = \frac{\partial^{i+j} R_T}{(\partial \lambda_1)^i (\partial \sigma_1)^j}$, and $U_T^{i,j} = \frac{\partial^{i+j} U_T}{(\partial \lambda_1)^i (\partial \sigma_1)^j}$. Furthermore, denote $R^{i,j} = \limsup_{T \rightarrow \infty} R_T^{i,j}$ and $U^{i,j} = \limsup_{T \rightarrow \infty} U_T^{i,j}$. Given the assumptions of Proposition 1, $R^{2,0}$, $R^{1,1}$, $R^{2,1}$, $R^{1,2}$, $U^{0,1}$, $U^{0,2}$ and $U^{1,1}$ all equal zero.

Using the notation introduced in Proposition 2, $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ can be expressed as

$$Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\frac{1}{2} \left(R_T + \sigma_e^2 R_T^{0,1} + U_T \right). \quad (26)$$

To determine the existence of the limit of $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ we therefore need to determine the existence of the limit of R_T , $R_T^{0,1}$ and U_T . Using Propositions 1 and 2 we can establish the following lemmas.

Lemma 6 Let $R_T, R_T^{i,j}, U_T, U_T^{i,j}$ be defined as in Proposition 2 and consider the collections of sequences $F^{I,J}$, defined as

$$F^{I,J} = \left\{ \{R_T\}_T, \{R_T^{i,j}\}_T, \{U_T\}_T, \{U_T^{i,j}\}_T; i = 0, 1, \dots, I, j = 0, 1, \dots, J \right\}. \quad (27)$$

Given the assumptions of Proposition 1, each element of $F^{4,4}$ is composed of a sequence of uniformly bounded functions on $\Theta \times B$.

Lemma 7 Let $F^{I,J}$ be defined as in Lemma 6. Given the assumptions of Proposition 1, each element of $F^{3,3}$ is composed of a sequence of equicontinuous functions on $\Theta \times B$.

Theorem 7.16 in Rudin (1976), states that uniform convergence of a sequence of functions implies the convergence of the sequence of the integrated functions. But the convergence of a sequence of the differentiated functions is not guaranteed. However, under some mild equicontinuity conditions, this result can be shown to hold. We have the following result.

Theorem 5 Let $\{f_n(x, y)\}_n$ and $\{D_x f_n(x, y)\}_n$ be two equicontinuous sequences of functions on a compact set $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^2$. Then, if $\{f_n(x, y)\}_n$ converges pointwisely, both $\{f_n(x, y)\}_n$ and $\{D_x f_n(x, y)\}_n$ converges uniformly. In particular, if $\lim_{n \rightarrow \infty} f_n(x, y) = f_0(x, y)$, then $\lim_{n \rightarrow \infty} D_x f_n(x, y) = D_x f_0(x, y)$ uniformly on $\mathcal{X} \times \mathcal{Y}$.

The main challenge in showing the convergence of $R_T, R_T^{0,1}$ and U_T is how to handle the eigenvalues h_t . As already mentioned very little is known about the limiting properties of h_t . Furthermore, since the eigenvalues will change not only with t , but also with T all the summands in the elements of $F^{I,J}$ will change with the sample size. This prevents the use of the standard convergence results on sums of infinite sequences. Based on Lemmas 6 and 7, Theorem 5, and Theorem 7.16 in Rudin (1976) the following theorem can be established.

Theorem 6 Let the assumptions of Proposition 2 hold. If just a single element in each of the two collections of sequences given as

$$\left\{ \{R_T\}_T, \left\{ R_T^{i,j} \right\}_T ; i = 0, 1, \dots, 4, j = 0, 1, \dots, 4 \right\}, \quad (28)$$

and

$$\left\{ \{U_T\}_T, \left\{ U_T^{i,j} \right\}_T ; i = 0, 1, \dots, 4, j = 0, 1, \dots, 4 \right\}, \quad (29)$$

converges uniformly on $\Theta \times B$, then each element of $F^{3,3}$ converges uniformly on $\Theta \times B$.

All the terms of $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ belongs to $F^{3,3}$ and since more than one of the elements converge by Proposition 2 there must exist a limiting function $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$, by Theorem 6 such that $\lim_{T \rightarrow \infty} Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}) = Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ uniformly on $\Theta \times B$. Consequently, we can conclude that condition *iv.* in Theorem 1 is satisfied under the Assumptions 1 - 4.

3.1.1 Identifiable uniqueness of $\boldsymbol{\theta}$

Finally, we need to justify the existence and uniqueness of the maximizer $\boldsymbol{\theta}^*$ of $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$, which is last condition in Theorem 1. In the standard likelihood framework under stationarity and ergodicity, the information inequality, see, e.g., Lemma 2.2 in Newey and McFadden (1994) usually guarantees this identification condition, under most distributional assumptions. However, in case of misspecification, the identification condition does not seem trivial. Typically identifiable uniqueness is simply assumed under the claim that it is a weak condition. In the random field regression model it is possible to be more explicit.

Theorem 7 Let the assumptions of Proposition 1 hold. If $\sigma_e^2 > \frac{1}{2}\sigma_1$, then $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*) = \lim_{T \rightarrow \infty} Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ is concave uniformly on $\Theta \times B$.

Theorem 7 completes the proof of the validity of the conditions of Theorem 1 implying that under Assumptions 1 – 4, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\beta}}$ will be consistent with respect to $\boldsymbol{\theta}^*$ and $\boldsymbol{\beta}^*$. It is not possible to check the condition of Theorem 7 prior to estimation. However, under Assumptions 1 – 4 a consistent estimator of σ_e^2 can be obtained according to

$$\widehat{\sigma}_e^2 = \frac{\frac{1}{T}\widehat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\mu}}}{\widehat{\sigma}_1\widehat{R}_T^{0,2}}, \quad (30)$$

where $\widehat{\boldsymbol{\mu}} = \mathbf{y} - \left(\mathbf{X}\widehat{\boldsymbol{\beta}}_T + \widehat{\lambda}_1\mathbf{H}'\left(\widehat{\lambda}_1\mathbf{H} + \widehat{\sigma}_1\mathbf{I}_T\right)^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_T\right)\right)$ and $\widehat{R}_T^{0,2} = R_T^{0,2}\left(\widehat{\boldsymbol{\theta}}\right)$, as shown by Qin, Dahl and González-Rivera (2004). Hence post estimation we recommend checking the condition $\widehat{\sigma}_e^2 > \frac{1}{2}\widehat{\sigma}_1$ in order to get an indication as to whether $\boldsymbol{\theta}$ is identified in the sample or not.

3.2 The asymptotic distribution of $\widehat{\boldsymbol{\theta}}$

The limiting distribution of $\widehat{\boldsymbol{\beta}}$ under Assumptions 1 - 4 has already been provided by White (1980b), and the main focus of this section will be to establish the asymptotic distribution of the second stage estimator of the vector $\boldsymbol{\theta}$. Following Proposition 2, we define $R_T = \frac{1}{T}\sum_{t=1}^T R_t$, $U_T = \frac{1}{T}\sum_{t=1}^T U_t$, etc. Note that $w_t \sim \text{IN}(b_t, \sigma_e^2)$, for $t = 1, \dots, T$. Accordingly, $w_t = b_t + \sigma_e z_t$, where z_t is *i.i.d.* $\text{N}(0, 1)$ for $\forall t, T$. Then (22) can be written as

$$\begin{aligned} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) &= -\frac{1}{2T}\sum_{t=1}^T \log(\lambda_1 h_t + \sigma_1) - \frac{1}{2T}\sum_{t=1}^T \frac{(b_t + \sigma_e z_t)^2}{\lambda_1 h_t + \sigma_1} \\ &= -\frac{1}{T}\sum_{t=1}^T \left(\frac{1}{2}R_t + \frac{1}{2}U_t + \sigma_e b_t R_t^{0,1} z_t + \frac{1}{2}\sigma_e^2 R_t^{0,1} z_t^2 \right), \end{aligned} \quad (31)$$

with associated gradient vector

$$\mathbf{D}_{\boldsymbol{\theta}}Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) = \begin{pmatrix} -\frac{1}{T}\sum_{t=1}^T \left[\frac{1}{2}R_t^{1,0} + \frac{1}{2}U_t^{1,0} + \sigma_e b_t R_t^{1,1} z_t + \frac{1}{2}\sigma_e^2 R_t^{1,1} z_t^2 \right] \\ -\frac{1}{T}\sum_{t=1}^T \left[\frac{1}{2}R_t^{0,1} + \frac{1}{2}U_t^{0,1} + \sigma_e b_t R_t^{0,2} z_t + \frac{1}{2}\sigma_e^2 R_t^{0,2} z_t^2 \right] \end{pmatrix}. \quad (32)$$

In relation to the first stage estimation, we similarly define

$$m_T(\boldsymbol{\beta}) = -\frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{x}_t \boldsymbol{\beta})^2, \quad (33)$$

as the objective function of the OLS estimation with gradient vector

$$D_{\boldsymbol{\beta}} m_T(\boldsymbol{\beta}) = \frac{2}{T} \sum_{t=1}^T \mathbf{x}_t (y_t - \mathbf{x}_t \boldsymbol{\beta}). \quad (34)$$

Assume that $\mathbf{x}_{\cdot i}$ denotes the i th column of \mathbf{X} and define $\mathbf{a}_{\cdot i} = \mathbf{P} \mathbf{x}_{\cdot i}$, for $\forall i = 1, \dots, k$ where \mathbf{P} is as defined in Lemma 1. Notice that $\sum_{t=1}^T a_{ti} a_{tj} = \mathbf{a}_{\cdot i}' \mathbf{a}_{\cdot j} = \mathbf{x}_{\cdot i}' \mathbf{P}' \mathbf{P} \mathbf{x}_{\cdot j} = \mathbf{x}_{\cdot i}' \mathbf{x}_{\cdot j} = \sum_{t=1}^T x_{ti} x_{tj}$ for $\forall i, j = 1, \dots, k$. Using this notation we can rewrite (34) as

$$D_{\boldsymbol{\beta}} m_T(\boldsymbol{\beta}) = \frac{2}{T} (\mathbf{a}'_{\cdot 1} \mathbf{w}, \dots, \mathbf{a}'_{\cdot k} \mathbf{w})' \quad (35)$$

$$= \frac{2}{T} \left(\sum_{t=1}^T a_{t1} (b_t + \sigma_e z_t), \dots, \sum_{t=1}^T a_{tk} (b_t + \sigma_e z_t) \right)'. \quad (36)$$

Let $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\beta}')$. The joint gradient vector can then be written as

$$\mathbf{g}_T(\boldsymbol{\zeta}) = (D_{\boldsymbol{\theta}} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})', D_{\boldsymbol{\beta}} m_T(\boldsymbol{\beta})')' \quad (37)$$

$$= \frac{1}{T} \sum_{t=1}^T \mathbf{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta}), \quad (38)$$

where

$$\mathbf{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta}) = \begin{pmatrix} -\frac{1}{2} R_t^{1,0} - \frac{1}{2} U_t^{1,0} - \sigma_e b_t R_t^{1,1} z_t - \frac{1}{2} \sigma_e^2 R_t^{1,1} z_t^2 \\ -\frac{1}{2} R_t^{0,1} - \frac{1}{2} U_t^{0,1} - \sigma_e b_t R_t^{0,2} z_t - \frac{1}{2} \sigma_e^2 R_t^{0,2} z_t^2 \\ a_{t1} b_t + \sigma_e a_{t1} z_t \\ \vdots \\ a_{tk} b_t + \sigma_e a_{tk} z_t \end{pmatrix}. \quad (39)$$

Furthermore, notice that

$$\mathbb{E}(\mathbf{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta})) = \begin{pmatrix} -\frac{1}{2} R_t^{1,0} - \frac{1}{2} U_t^{1,0} - \frac{1}{2} \sigma_e^2 R_t^{1,1} \\ -\frac{1}{2} R_t^{0,1} - \frac{1}{2} U_t^{0,1} - \frac{1}{2} \sigma_e^2 R_t^{0,2} \\ a_{t1} b_t \\ \vdots \\ a_{tk} b_t \end{pmatrix},$$

hence

$$\mathbf{g}_T(\boldsymbol{\zeta}) = \frac{1}{T} \sum_{t=1}^T (\bar{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta}) + \mathbf{E}(\mathbf{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta}))), \quad (40)$$

where

$$\bar{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sigma_e \begin{pmatrix} -R_t^{1,1} [b_t z_t + \frac{1}{2} \sigma_e (z_t^2 - 1)] \\ -R_t^{0,2} [b_t z_t + \frac{1}{2} \sigma_e (z_t^2 - 1)] \\ a_{t1} z_t \\ \vdots \\ a_{tk} z_t \end{pmatrix}.$$

Note that $\bar{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})$ is an independent $(k+2)$ - dimensional random vector with $\mathbf{E}(\bar{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})) = \mathbf{0}$ and $\text{cov}(\bar{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})) = \bar{\boldsymbol{\Sigma}}_t$.⁶ Using Hoadley's (1971) Theorem A.6 (p. 1990), the asymptotic distribution of $\mathbf{g}_T(\boldsymbol{\zeta})$ can now be established.

Theorem 8 Let $\mathbf{g}_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ be given by equation (40). Under Assumptions 1 - 4, $\sqrt{T} \mathbf{g}_T(\boldsymbol{\theta}^*, \boldsymbol{\beta}^*) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}^*)$ as $T \rightarrow \infty$, where $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\boldsymbol{\theta}^*, \boldsymbol{\beta}^*) = [\boldsymbol{\Sigma}_{11}^* \quad \boldsymbol{\Sigma}_{12}^* : \boldsymbol{\Sigma}_{12}^{*'} \quad \boldsymbol{\Sigma}_{22}^*]$ and

$$\begin{aligned} \boldsymbol{\Sigma}_{11}^* &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{12} \sigma_e^4 R^{0,4} \end{pmatrix}, \\ \boldsymbol{\Sigma}_{22}^* &= \sigma_e^2 \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}' \mathbf{X}, \\ \boldsymbol{\Sigma}_{12}^* &= \mathbf{0}. \end{aligned}$$

Theorem 9 Define $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\beta}')'$ and let $Q_T(\boldsymbol{\zeta})$ and $m_T(\boldsymbol{\beta})$ be given by (31) and (33) respectively. Under Assumptions 1 - 4 the following conditions are satisfied: *i.* $\hat{\boldsymbol{\zeta}}_T \xrightarrow{p} \boldsymbol{\zeta}^*$. *ii.* $Q_T(\boldsymbol{\zeta})$ and $m_T(\boldsymbol{\beta})$ are twice continuously differentiable. *iii.* $\sqrt{T} \mathbf{g}_T(\boldsymbol{\zeta}^*) = (\sqrt{T} \mathbf{D}_{\boldsymbol{\theta}} Q_T(\boldsymbol{\zeta}^*)', \sqrt{T} \mathbf{D}_{\boldsymbol{\beta}} m_T(\boldsymbol{\beta}^*)')'$ converges to a normal random variable $\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}^*)$ in distribution. *iv.* $\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 Q_T(\boldsymbol{\zeta})$, and $\mathbf{D}_{\boldsymbol{\beta}\boldsymbol{\beta}}^2 m_T(\boldsymbol{\beta})$ converges to nonsingular matrices for any $\boldsymbol{\zeta}$ in a neighborhood of $\boldsymbol{\zeta}^*$. Under the additional assumption, that $\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\beta}}^2 Q_T(\boldsymbol{\zeta})$ converges to a nonsingular matrix for any $\boldsymbol{\zeta}$ in a neighborhood of $\boldsymbol{\zeta}^*$ conditions *i.* - *iv.* imply that

$$\sqrt{T} (\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}^*) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{M}^*), \quad (41)$$

⁶For an explicit expression for $\bar{\boldsymbol{\Sigma}}_t$, see the proof of Theorem 8.

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

$$\begin{aligned} \mathbf{G}_T(\boldsymbol{\zeta}) &= \mathbf{D}_{\boldsymbol{\zeta}} \mathbf{g}_T(\boldsymbol{\zeta}) \\ &= \begin{pmatrix} \mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}_T(\boldsymbol{\zeta}) & \mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\beta}}^2 \mathbf{Q}_T(\boldsymbol{\zeta}) \\ 0 & \mathbf{D}_{\boldsymbol{\beta}\boldsymbol{\beta}}^2 \mathbf{m}_T(\boldsymbol{\beta}) \end{pmatrix}. \end{aligned}$$

In particular, as the primary interest is in $\boldsymbol{\theta}$, notice that,

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

where \mathbf{M}_{11} is the upper left corner of \mathbf{M} , with

$$\begin{aligned} \mathbf{M}_{11}^* &= (\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*))^{-1} \boldsymbol{\Sigma}_{11}^* (\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*))^{-1} + \\ &\quad T \sigma_e^2 (\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*))^{-1} \mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\beta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*) (\lim_{T \rightarrow \infty} \mathbf{X}' \mathbf{X})^{-1} (\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\beta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*))' (\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}^*(\boldsymbol{\zeta}^*))^{-1}. \end{aligned} \quad (42)$$

A consistent estimator of the covariance matrix \mathbf{M}_{11}^* can be obtained by using $\widehat{\mathbf{M}}_{11} = \mathbf{M}_{11}(\widehat{\boldsymbol{\zeta}}_T, \widehat{\sigma}_e^2)$ where $\widehat{\sigma}_e^2$ is given by equation (30). It can be expected that $\widehat{\mathbf{M}}_{11}$ will give an over confident estimate of the precision of the estimated parameters in small samples. As it is common, we therefore recommend to based the variance estimator on (42) using the sample counterparts $\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\theta}}^2 \mathbf{Q}_T(\widehat{\boldsymbol{\zeta}})$, $\mathbf{D}_{\boldsymbol{\theta}\boldsymbol{\beta}}^2 \mathbf{Q}_T(\widehat{\boldsymbol{\zeta}})$, $\mathbf{D}_{\boldsymbol{\beta}\boldsymbol{\beta}}^2 \mathbf{Q}_T(\widehat{\boldsymbol{\zeta}})$ and the sample variance of $\mathbf{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta})$ instead of $\frac{1}{T} \boldsymbol{\Sigma}^*$. Finally, we recommend substituting $\boldsymbol{\psi}(\mathbf{X})$ with $\widehat{\boldsymbol{\mu}}$ as defined in relation to equation (30).

Hamilton (2001) suggests a Bayesian method to construct confidence bands in association with the estimated conditional mean function. Utilizing the asymptotic results, in particular Theorem 9, we are able to construct confidence bands that, besides from being easier to compute, have a moderately better coverage of the true conditional mean function and do not rely on specifying "arbitrary" priors for the unknown parameters of the model. Level α confidence bands can be constructed by the following simple procedure: First, fix the parameter vector $\boldsymbol{\beta}$ at $\widehat{\boldsymbol{\beta}}_T$, and sample $\boldsymbol{\theta}^j$ from $\mathbf{N}\left(\widehat{\boldsymbol{\theta}}_T, \frac{1}{T} \widehat{\mathbf{M}}_{11}(\widehat{\boldsymbol{\zeta}}_T, \widehat{\sigma}_e^2)\right)$, where $\widehat{\mathbf{M}}_{11}$ is a consistent estimator of \mathbf{M}_{11}^* given by (42). Secondly, for each $\boldsymbol{\theta}^j$, generate the predictor $\widehat{\xi}_s^j(\mathbf{x}_s^*)$ from equation (3.4) on page 543 in Hamilton (2001). Thirdly, sort the values $(\widehat{\xi}_s^1(\mathbf{x}_s^*), \dots, \widehat{\xi}_s^N(\mathbf{x}_s^*))$ in an ascending order as $(\widehat{\xi}_s^{j_1}(\mathbf{x}_s^*), \dots, \widehat{\xi}_s^{j_N}(\mathbf{x}_s^*))$.

Table 1: Alternative nonlinear data generating processes, using $e_t \sim N(0, 1)$.

Index	True DGP
Model 1	$y_t = 2x_{1t}1_{\{x_{1t}>0\}} + 1.5x_{2t} + e_t$
Model 2	$y_t = \frac{2x_{1t}-1}{x_{1t}+3.5} - \exp(0.5x_{2t}) + e_t$
Model 3	$y_t = x_{1t}^2 + e^{0.5x_{1t}-1}x_{2t} + e_t$
Model 4	$y_t = 3 \sin(x_{1t} + x_{2t}) + e_t$

The confidence band is then given as $(\widehat{\xi}_s^{j_{\alpha/2N}}(\mathbf{x}_s^*), \widehat{\xi}_s^{j_{(1-\alpha/2)N}}(\mathbf{x}_s^*))$ for $s = 1, 2, \dots, S$, where \mathbf{x}_s^* is the vector of regressors that can take values not realized in the sample.

Another obvious and important application of the asymptotic normality of $\widehat{\zeta}_T$ is that it facilitates tests of the presence of nonlinearity in the actual data generating process through λ_1 , see, e.g., Hamilton (2001) and Dahl and González-Rivera (2003a).

4 Simulation Experiments

The analysis of actual data with the random field regression model is still in a preliminary state of development. In empirical applications the approach have been successfully employed to detect nonlinearities and modelling/forecasting US unemployment rates (Hamilton, 2001), oil prices (Hamilton, 2003), industrial production, Dahl and González-Rivera (2003a) and Dahl and Hylleberg (2004) and US GDP growth rates, Dahl and González-Rivera (2003b). Furthermore, Dahl (2002) provides an extensive comparison of the random field model to non-parametric approaches. He 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.⁷ However, based on simulations Dahl (2002) shows that the random field approach has superior predictive accuracy compared to the spline smoother, when the data is generated from popular econometric models such as LSTAR/ESTAR

⁷Using the results of Kimeldorf and Wahba (1971) and Wahba (1978, 1990) we show how this result generalizes to $\mathbf{x}_t \in \mathbb{R}^k$.

and various bilinear specifications.

In this section, we focus on analyzing the finite sample properties of the estimated parameters (by simulations) and comparing to the theoretical large sample results derived in the previous section. We will use four alternative data generating processes (DGPs) summarized in Table 1. Model 1 and 2 are additive models and Model 3 and 4 are nonadditive models. Model 1 is a threshold model, similar to the model used in Example 1 in Hamilton (2001). Although the model is non-differentiable, Hamilton shows that it is well approximated by the random field regression model. Model 2 and 3 have complicated nonlinear structures. Model 4 is a trigonometric model or first order Fourier expansion. Lemma 4.10, p. 551 in Hamilton (2001) shows the adequacy of using a random field regression model to this type of model. The explanatory variables \mathbf{x}_t are sampled independently from a uniform distribution, i.e., from $U(-3, 3)$. The coefficients in the four models are selected such that their nonlinearity can be reflected on the sampled region. As to the choice of the uniform distribution this turns out not to be important as long as the sample of the explanatory variables is sufficiently dense.

First, a small sample comparison of the proposed two-stage estimation routine and Hamilton's maximum likelihood estimation algorithm (MLE) is made. In Table 2, the mean squared errors (MSE) associated with 100 out-of-sample prediction points are reported, using the two estimation algorithms under each of the four alternative data generating processes. From Table 2 we see, as expected, that the two-stage estimation algorithm performs almost identical to the MLE, even in smaller samples. This result might favor the two-stage estimation approach, since it is somewhat faster than Hamilton's algorithm. In what follows only the two-stage estimation approach will be used. Next, we compare the simulated distribution of the estimated parameters with the asymptotic distribution described by Theorem 9. Again data is generated according to the four models described in Table 1. We let the sample size grow gradually from $T = 200$ to $T = 1000$. For each model and each T , we generate 1000 samples and estimate the nonlinear parameters based on each sample. The resulting simulated densities of $\hat{\lambda}_1$ and $\hat{\sigma}_1$ are then plotted against the asymptotic density, see Figure 1.

Figure 1 confirms that for larger samples the asymptotic density provides a fairly accurate approximation of the true sample density. It also becomes clear that the as-

Figure 1: Asymptotic and simulated densities associated with the estimated parameters $\hat{\lambda}_1$ and $\hat{\sigma}_1$ entering the nonlinear part of the random field regression model. Black solid line: Asymptotic densities using $\widehat{A\text{var}}$ (see below for definition). Solid line: Simulated densities when $T = 200$. Dotted line: Simulated densities when $T = 500$. Dashed lines: Simulated densities when $T = 1000$. The left diagrams shows the densities of $\hat{\lambda}_1$ and the right shows those of $\hat{\sigma}_1$ for Model 1 (first row) to Model 4 (fourth row), respectively.

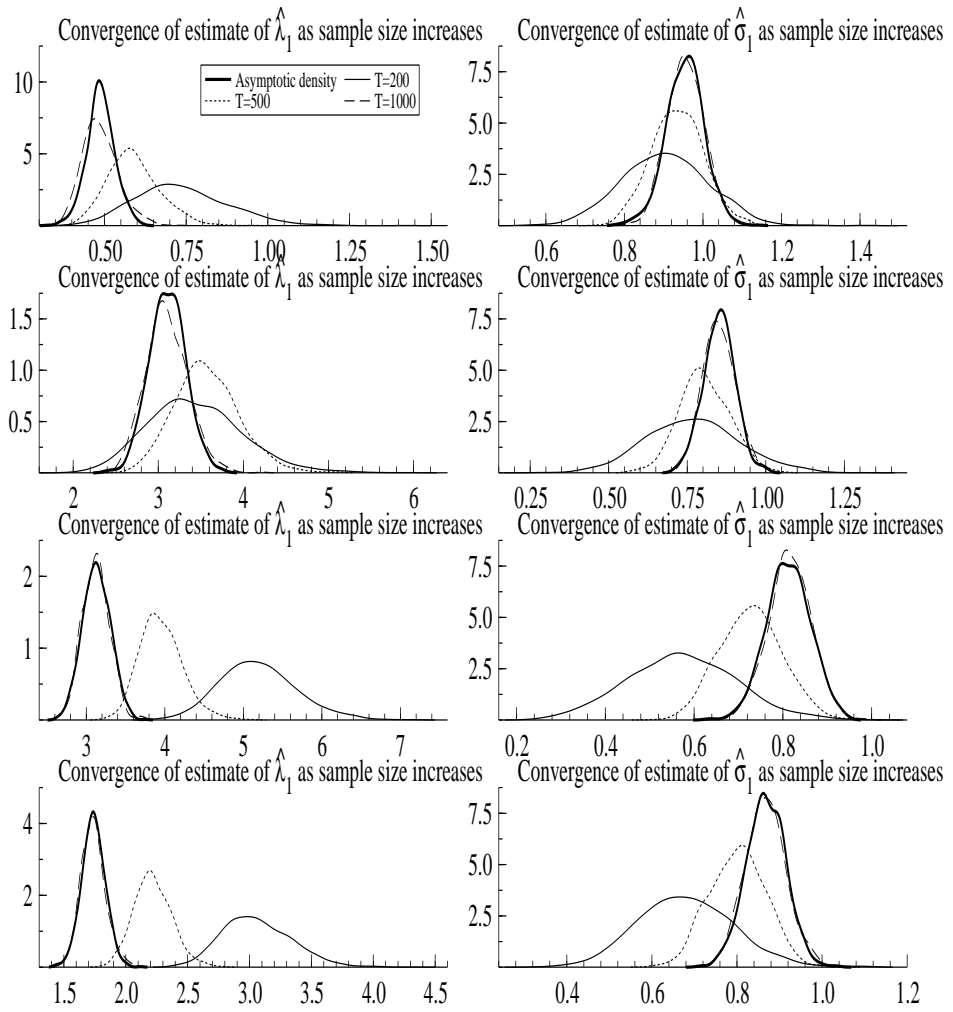


Table 2: Mean squared errors (MSE) associated with 100 out-of-sample prediction errors. MSE_2 and MSE_H are based the two-stage estimation procedure and Hamilton’s maximum likelihood estimator respectively. Standard errors are abbreviated s.e. The number of Monte Carlo replications equals 1000.

	MSE_2 (s.e.)	MSE_H (s.e.)
T = 200		
Model 1	1.075 (0.059)	1.071 (0.077)
Model 2	1.431 (0.111)	1.505 (0.253)
Model 3	1.268 (0.108)	1.224 (0.189)
Model 4	1.229 (0.091)	1.226 (0.135)
T = 500		
Model 1	1.055 (0.044)	1.058 (0.045)
Model 2	1.301 (0.094)	1.433 (0.139)
Model 3	1.069 (0.085)	1.079 (0.137)
Model 4	1.116 (0.073)	1.111 (0.085)

ymptotic standard errors may not be a good estimator of the standard errors of the estimated parameters in finite samples. In order to evaluate various alternative estimates of the variance/standard errors, we need to define a "true" standard error as a benchmark. Given a data generating process from Table 1, we do this by generating the empirical standard error of the estimated parameter using a Monte Carlo simulation design similar to the one already described, and take this empirical standard error as the "true" finite sample standard error. In Table 3, we let the sample size be $T = 100$, and perform 1000 Monte Carlo replications. In the comparison of alternative estimates we include two candidates based on the asymptotic variance matrix given by (42), denoted $A \text{ var} = A \text{ var} \left(\hat{\boldsymbol{\xi}}, \sigma_e^2 \right)$ and $\widehat{A \text{ var}} = \widehat{A \text{ var}} \left(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\mu}}, \hat{\sigma}_e^2 \right)$ respectively, where the latter is a consistent estimator of the asymptotic variance as discussed in the previous section. In addition, it seems natural to include the variance estimator based on the sample version of the Hessian matrix, as in Hamilton (2001). Finally, we also provide a simple

Table 3: Alternative estimates of the standard errors (s.e.) of the estimated parameters for $T = 200$ using 1000 Monte Carlo replications. "True" represents the Monte Carlo generated true finite sample s.e., $\widehat{A \text{ var}}$ denotes the consistent estimator of the asymptotic s.e., $A \text{ var}$ is the asymptotic s.e., *Bootstrap* is the bootstrapped s.e. and *Hessian* the s.e. derived from Hessian of the likelihood function. After each model label, a consistent estimator of σ_e^2 is provided in the parenthesis.

Index	"True"	$\widehat{A \text{ var}}$	$A \text{ var}$	<i>Bootstrap</i>	<i>Hessian</i>
Model 1 ($\hat{\sigma}_e^2 = 1.220$)					
s.e. of $\hat{\lambda}_1$ (= 0.847)	0.147	0.158	0.022	0.149	0.310
s.e. of $\hat{\sigma}_1$ (= 1.063)	0.111	0.126	0.157	0.119	0.138
Model 2 ($\hat{\sigma}_e^2 = 1.297$)					
s.e. of $\hat{\lambda}_1$ (= 3.455)	0.521	0.559	0.050	0.448	0.935
s.e. of $\hat{\sigma}_1$ (= 0.893)	0.144	0.145	0.185	0.129	0.185
Model 3 ($\hat{\sigma}_e^2 = 1.175$)					
s.e. of $\hat{\lambda}_1$ (= 6.026)	0.476	0.600	0.038	0.550	1.100
s.e. of $\hat{\sigma}_1$ (= 0.617)	0.121	0.142	0.147	0.105	0.151
Model 4 ($\hat{\sigma}_e^2 = 1.159$)					
s.e. of $\hat{\lambda}_1$ (= 3.232)	0.295	0.380	0.015	0.353	0.720
s.e. of $\hat{\sigma}_1$ (= 0.786)	0.111	0.129	0.141	0.103	0.142

parametric bootstrap estimator of the variance, which is used extensively in the spline smoother literature, see, e.g., Wahba (1990). In the bootstrap estimation, we generate samples of the dependent variable from the fitted random field model, as if the fitted random field model was the actual data generating process. This procedure is valid since the random field model gives an asymptotically consistent estimate of the mean function response. Based on the bootstrap samples we then calculate a sequence of estimated parameters and the empirical standard error of this sample will be referred to as the bootstrap standard errors. The results are summarized in Table 3.

The estimated standard errors based on the asymptotic variance matrices seems to

underestimate the actual standard errors of $\hat{\lambda}_1$ (as expected) but are too conservative regarding the precision of $\hat{\sigma}_1$, which perhaps is somewhat surprising. Hessian based estimators are in general overestimating the standard errors, particularly with respect to $\hat{\lambda}_1$ (by about 50 per cent). Overall, the consistent estimate of the asymptotic variance $\widehat{A\text{var}}$ and the bootstrap estimator seem to be the preferred estimators although the bootstrap estimator tends to slightly underestimate the actual standard error of $\hat{\lambda}_1$.

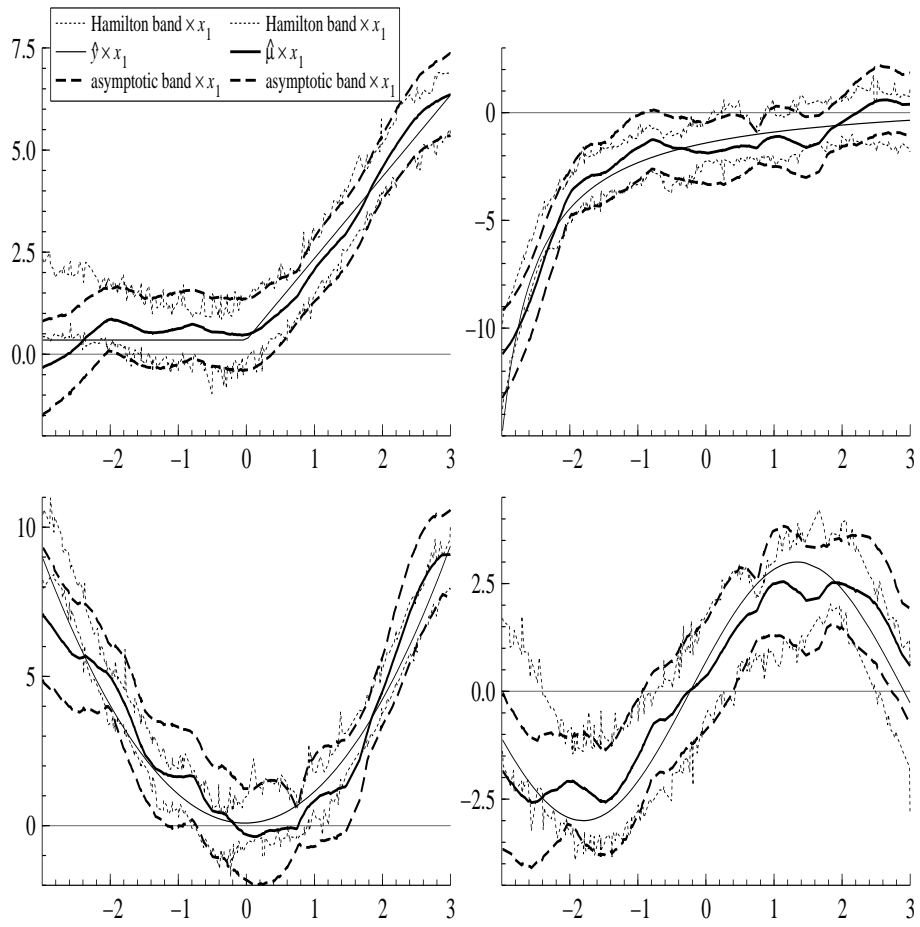
Finally, we illustrate how the asymptotic results of Theorem 9 can be used to construct prediction confidence bands that have very good coverage properties in terms of "containing" the true conditional mean function of the variable of interest. As a comparison, the Bayesian confidence bands proposed by Hamilton (2001) are also plotted. $\widehat{A\text{var}}$ described in the last simulation study is used as the estimator of the variance. In Figure 2, a 90% confidence band is plotted for each of the four models. We use 200 sample points for estimation purpose and use 5000 Monte Carlo replications to simulate the confidence bands. In order to contain the plots in two dimensions, x_{t2} is fixed at the sample average $\bar{x}_{.2}$.

From Figure 2 it seems that the confidence bands based on the asymptotic variance estimate in general are very close to the Bayesian confidence bands, but with perhaps a slightly better coverage of the actual mean function, particularly at the endpoints. These results are very encouraging since they indicate that by using the asymptotic properties of the estimated parameters of the random field regression model we should be able to uncover the shape of the unknown conditional mean function without having to make any choices with respect to bandwidth/kernel as in a non-parametric approach or with respect to priors as in the Bayesian approach.

5 Conclusion

In this paper, the large sample behavior of the maximum likelihood estimates of the unknown parameters in a non-locally misspecified random field regression model is characterized. We show how to write the log likelihood function in the form of a sum of a double array of independently but not identically distributed random variables and derive conditions under which the log likelihood function and its derivatives converges

Figure 2: Confidence bands when x_2 is fixed at the sample mean \bar{x}_2 of the second variable for all four models in Table 1. The diagrams from left to right and from top to bottom are for Model 1 through 4, respectively. Black solid line: the Monte Carlo sample mean of $\bar{\mu}(x_1, \bar{x}_2|\mathbf{y})$ as a function of x_1 . Black dashed line: 90% confidence band constructed from the asymptotic distribution of the parameters. Solid line: Conditional mean of the true DGP. Dotted line: 90% Bayesian confidence band as in Hamilton (2001).



uniformly to limits that are fixed and independent of the sample. Based on these uniform convergence results, we demonstrate how asymptotic consistency and normality of the likelihood based estimators can be established. Simulation studies shows that in samples of small to moderate size the estimator of the asymptotic variance as well as a bootstrapped variance estimator both are reasonable accurate estimators of the estimated nonlinear parameters in the random field model. Finally, we find that confidence bands constructed using the asymptotic distribution of the estimated parameters has a good coverage of actual conditional mean function of the variable of interest relative to existing approaches. In summary, our results indicate that classical statistical inference techniques in generally works very well for random field regression models in finite samples and that these models successfully can fit and uncover many types of nonlinear structures in data. The results are very encouraging and render the parametric random field model as a good alternative specification relative to its nonparametric counterparts, particularly in small samples.

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.
- Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.
- Currin, C., T. Mitchell, M. Morris, and D. Ylvisaker (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association* 86(416), 953–963.
- 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. González-Rivera (2003a). 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 G. González-Rivera (2003b). Identifying nonlinear components by random fields in the us gnp growth. implications for the shape of the business cycle. *Studies in Nonlinear Dynamics and Econometrics* 7, Article 2.
- Dahl, C. M. and S. Hylleberg (2004). Specifying econometric models by flexible regression models and relative forecast performance. *International Journal of Forecasting* 20, 201–217.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press Inc.
- Diebold, F., L. Ohanian, and J. Berkowitz (1998). Dynamic equilibrium economies: A framework for comparing models and data. *Review of Economic Studies* 65, 433–451.
- Domowitz, I. and H. White (1982). Misspecified models with dependent observations. *Journal of Econometrics* 20, 35–58.

- Gallant, A. and H. White (1988). *A Unified Theory of Estimation and Inference in Nonlinear Models*. Basil Blackwell, Oxford.
- Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica* 52(3), 681–700.
- Hall, A. and A. Inoue (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114, 361–394.
- Hamilton, J. D. (2001). A parametric approach to flexible nonlinear inference. *Econometrica* 69(3), 537–573.
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics* 113, 363–398.
- Hansen, L. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52, 557–590.
- Hansen, L. and T. Sargent (1993). Seasonality and approximation errors in rational expectations models. *Journal of Econometrics* 55, 21–55.
- Kitamura, Y. (1997). Comparing misspecified dynamic econometric models using non-parametric likelihood. *Discussion paper, Department of Economics, University of Minnesota*.
- 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.
- Maasoumi, E. and P. Phillips (1982). On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics* 19, 183–201.
- 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.
- McDonald, J. B. and W. K. Newey (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory* 4, 428–457.
- Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV,

Chapter 36. Amsterdam, North Holland.

- Qin, Y., C. M. Dahl, and G. González-Rivera (2004). The additive random field regression model. *Unpublished Manuscript, Purdue University*.
- Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.
- Sacks, J., S. Schiller, and W. Welch (1989). Designs for computer experiments. *Technometrics* 31(1), 41–47.
- Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). Design and analysis of computer experiments. *Statistical Science* 4(4), 409–423.
- Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics* 8, 1375–1381.
- van der Vaart, A. (1996). Maximum likelihood estimation under a spatial sampling scheme. *The Annals of Statistics* 24, 2049–2057.
- Wahba, G. (1990). *Spline Models for Observational Data*. Capital City Press, Montpelier, Vermont.
- Watson, M. (1993). Measures of fit for calibrated models. *Journal of Political Economy* 101, 1011–1041.
- White, H. (1980a). Nonlinear regression on cross-section data. *Econometrica* 48(3), 721–746.
- White, H. (1980b). Using least squares to approximate unknown regression functions. *International Economic Review* 21(1), 149–170.
- White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76(374), 419–433.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50(1), 1–25.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.
- 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.

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

Proof of Lemma 1 See, e.g., Magnus and Neudecker (1999) Theorem 13, p. 16.

Proof of Lemma 2 Notice from the definition of \mathbf{H} , that $\text{tr}(\mathbf{H}) = T$. By Lemma 1 we can write $\mathbf{H} = \mathbf{P}'\mathbf{V}_H\mathbf{P}$, where $\mathbf{V}_H = \text{diag}(h_1, \dots, h_T)$ and $\mathbf{P}\mathbf{P}' = \mathbf{I}$. Then it follows easily that $T = \text{tr}(\mathbf{P}'\mathbf{V}_H\mathbf{P}) = \text{tr}(\mathbf{V}_H\mathbf{P}\mathbf{P}') = \sum_{t=1}^T h_t$.

Proof of Lemma 3 See, e.g., White (1980) and Amemiya (1985).

Proof of Lemma 4 See, e.g., Lemma A1 in McDonald and Newey (1988).

Proof of Theorem 1 See, Wooldridge (1994).

Proof of Theorem 2 The proof follows the proof of Theorem 4.4 in Hamilton (2001), however, since Hamilton is concerned only with pointwise convergence, two modifications to his proof are needed. First, observe the following three properties of $p_t : i$. $\{p_t(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1)\}_t$ is a sequence of continuous functions on the compact set $A \times A$. *ii.* $\{p_t(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1)\}_t$ converges pointwise to a continuous function $p_\infty(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1)$ on $A \times A$. *iii.* $p_{t+1}(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1) \leq p_t(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1), \forall \mathbf{z} \in A, t = 0, 1, 2, \dots, T$. Given the conditions *i. – iii.*, $\lim_{T \rightarrow \infty} p_T(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1)$ exists uniformly on $A \times \Theta$, from Theorem 7.13 in Rudin (1976). Secondly, from the specification of $G_k(h, r)$ in Lemma 2.1, Hamilton (2001), it is obvious that $p_0(\mathbf{z}, \mathbf{w}; \lambda_1)$ is uniform continuous in \mathbf{z} and \mathbf{w} on $A \times A$. Applying these two modifications in Hamilton's proof and using the Cauchy-Schwarz inequality we obtain

$$p_t(\mathbf{z}, \mathbf{w}; \lambda_1, \sigma_1) \leq \sqrt{p_t(\mathbf{z}, \mathbf{z}; \lambda_1, \sigma_1)^2 p_t(\mathbf{w}, \mathbf{w}; \lambda_1, \sigma_1)^2},$$

for $\forall t$ which establishes the proof of uniform convergence in \mathbf{z} and \mathbf{w} . Convergence uniformly in λ_1 and σ_1 follows then by applying Assumption 2 (compactness of the parameter space).

Proof of Theorem 3 Define $\mathbf{L}_T(\lambda_1) = (l_T(\mathbf{x}_1.; \lambda_1), l_T(\mathbf{x}_2.; \lambda_1), \dots, l_T(\mathbf{x}_T.; \lambda_1))'$, $\boldsymbol{\xi}_T = (\xi_T(\mathbf{x}_1.), \xi_T(\mathbf{x}_2.), \dots, \xi_T(\mathbf{x}_T.))'$, and $\tilde{\boldsymbol{\phi}}_T = (\tilde{\phi}(\mathbf{x}_1.), \tilde{\phi}(\mathbf{x}_2.), \dots, \tilde{\phi}(\mathbf{x}_T.))'$, such that we can

write

$$\begin{aligned}\mathbf{L}_T(\lambda_1) &= \frac{1}{T} \mathbf{P}_0(\lambda_1) \tilde{\boldsymbol{\phi}}_T, \\ \boldsymbol{\xi}_T &= \mathbf{X}\boldsymbol{\beta} + \mathbf{P}_0(\lambda_1) (\mathbf{P}_0(\lambda_1) + \sigma_1 \mathbf{I}_T)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).\end{aligned}$$

By using

$$\mathbf{P}_T(\lambda_1) = \left[\mathbf{P}_0(\lambda_1) - \mathbf{P}_0(\lambda_1) (\mathbf{P}_0(\lambda_1) + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{P}_0(\lambda_1) \right],$$

we obtain

$$\begin{aligned}\boldsymbol{\xi}_T - \mathbf{X}\boldsymbol{\beta} - \mathbf{L}_T(\lambda_1) &= \mathbf{P}_0(\lambda_1) (\mathbf{P}_0(\lambda_1) + \sigma_1 \mathbf{I}_T)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \mathbf{L}_T(\lambda_1) \\ &= -\frac{1}{T} \mathbf{P}_T(\lambda_1) \boldsymbol{\phi}_T + \mathbf{P}_0(\lambda_1) (\mathbf{P}_0(\lambda_1) + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{e}.\end{aligned}$$

Hence, we can apply the result of Theorem 4.7 in Hamilton (2001) and combined with the result of our Theorem 2 establish that the convergence in (20) is indeed uniform in λ_1, σ_1 and \mathbf{X} on $\Theta \times B$ and A .

Proof of Lemma 5 First, notice given the denseness condition in \mathbf{X} the first term of (11) is a fixed constant and we can focus entirely on the second term. Let \mathbf{v} be given by Definition 3 and notice that

$$\mathbb{E} \left(\frac{1}{2T} \mathbf{v}' (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{v} \right) = \frac{1}{2T} \sum_{t=1}^T \mathbb{E} (w_t^2) (\lambda_1 h_t + \sigma_1)^{-1},$$

where w_t for $\forall t$ equals $w_t^2 = \left(\sum_{i=1}^T P_{ti} (y_i - \mathbf{x}_i \boldsymbol{\beta}) \right)^2 = \left(\sum_{i=1}^T P_{ti} v_i \right)^2$. Consequently,

$$\begin{aligned}\mathbb{E} (w_t^2) &= \mathbb{E} \left(\sum_{i=1}^T P_{ti} v_i \right)^2 \\ &= \text{var} \left(\sum_{i=1}^T P_{ti} v_i \right) + \left(\mathbb{E} \left(\sum_{i=1}^T P_{ti} v_i \right) \right)^2 \\ &= \sum_{i=1}^T P_{ti}^2 \text{var} (v_i) + \left(\sum_{i=1}^T P_{ti} (\psi(\mathbf{x}_i) - \mathbf{x}_i \boldsymbol{\beta}) \right)^2 \\ &= \sigma_e^2 + \left(\sum_{i=1}^T P_{ti} c_i \right)^2 = \sigma_e^2 + b_t^2.\end{aligned}$$

We can then write $\mathbb{E} \left(\frac{1}{2T} \mathbf{v}' (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{v} \right) = \frac{1}{2T} \sum_{t=1}^T \frac{\sigma_e^2 + b_t^2}{\lambda_1 h_t + \sigma_1}$, which completes the proof.

Proof of Theorem 4 First, define $f_T(\mathbf{w}, \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) \equiv Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) - Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ such that $f_T(\mathbf{w}, \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{1}{2T} \sum_{t=1}^T \frac{w_t^2 - \sigma_e^2 - b_t^2}{\lambda_1 h_t + \sigma_1}$. Secondly, notice that $\mathbb{E}[f_T(\mathbf{w}, \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\beta})] = 0$, and since $w_t \sim \text{IN}(b_t, \sigma_e^2)$ it follows immediately that,

$$\mathbb{E}[f_T(\mathbf{w}, \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\beta})^2] = \frac{1}{T^2} \sum_{t=1}^T \frac{2\sigma_e^4 + 4\sigma_e^2 b_t^2}{(\lambda_1 h_t + \sigma_1)^2},$$

By Chebyshev's inequality, $f_T(\mathbf{w}, \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{p} 0$ uniformly as $T \rightarrow \infty$, if

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T \left(\frac{2\sigma_e^4}{(\lambda_1 h_t + \sigma_1)^2} + \frac{4\sigma_e^2 b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \right) = 0, \quad (43)$$

uniformly in $\boldsymbol{\theta}, \boldsymbol{\beta}$ and \mathbf{X} . As $h_t > 0$, for $\forall t$,

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T \frac{2\sigma_e^4}{(\lambda_1 h_t + \sigma_1)^2} < \lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T \frac{2\sigma_e^4}{\sigma_1^2} = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{2\sigma_e^4}{\sigma_1^2} = 0,$$

by Assumption 2 uniformly in $\boldsymbol{\theta}, \boldsymbol{\beta}$ and \mathbf{X} . Finally, $\lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} = 0$ uniformly in $\boldsymbol{\theta}, \boldsymbol{\beta}$ and \mathbf{X} by Proposition 1, which completes the proof.

Proof of Proposition 1 Let $\psi(\mathbf{x}_t)$ be defined as in Assumption 3 and consider $\Psi(\mathbf{X}) = (\psi(\mathbf{x}_1), \dots, \psi(\mathbf{x}_T))'$. Using Equation (3.5) in Hamilton (2001) and Assumptions 3 and 4, we can write

$$\begin{aligned} \mathbf{S} &= \Psi(\mathbf{X}) - \left(\mathbf{X}\boldsymbol{\beta} + \lambda_1 \mathbf{H} (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) \\ &= \left(\mathbf{I}_T - \lambda_1 \mathbf{H} (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} \right) (\Psi(\mathbf{X}) - \mathbf{X}\boldsymbol{\beta}) - \lambda_1 \mathbf{H} (\lambda_1 \mathbf{H} + \sigma_1 \mathbf{I}_T)^{-1} \mathbf{e} \\ &= \mathbf{P}' \mathbf{V}_2 \mathbf{P} (\Psi(\mathbf{X}) - \mathbf{X}\boldsymbol{\beta}) - \mathbf{P}' \mathbf{V}_3 \mathbf{P} \mathbf{e}, \end{aligned}$$

where $\mathbf{V}_2 = \text{diag} \left(\frac{\sigma_1}{\lambda_1 h_1 + \sigma_1}, \dots, \frac{\sigma_1}{\lambda_1 h_T + \sigma_1} \right)$, $\mathbf{V}_3 = \text{diag} \left(\frac{\lambda_1 h_1}{\lambda_1 h_1 + \sigma_1}, \dots, \frac{\lambda_1 h_T}{\lambda_1 h_T + \sigma_1} \right)$, $\mathbf{e} = \mathbf{y} - \Psi(\mathbf{X})$, and $\mathbf{P}\mathbf{P}' = \mathbf{I}$. Then

$$\begin{aligned} \frac{1}{T} \mathbb{E}(\mathbf{S}'\mathbf{S}) &= \frac{1}{T} \mathbf{b}' \mathbf{V}_2^2 \mathbf{b} + \frac{1}{T} \mathbb{E}(\mathbf{e}' \mathbf{P}' \mathbf{V}_3^2 \mathbf{P} \mathbf{e}) \\ &= \frac{1}{T} \text{tr}(\mathbf{b}\mathbf{b}' \mathbf{V}_2^2) + \frac{1}{T} \text{tr}(\mathbf{P} \mathbb{E}(\mathbf{e}\mathbf{e}') \mathbf{P}' \mathbf{V}_3^2) \\ &= \frac{1}{T} \text{tr}(\mathbf{b}\mathbf{b}' \mathbf{V}_2^2) + \frac{1}{T} \text{tr}(\sigma_e^2 \mathbf{V}_3^2) \\ &= \frac{1}{T} \sum_{t=1}^T \frac{\sigma_1^2 b_t^2}{(\lambda_1 h_t + \sigma_1)^2} + \frac{1}{T} \sum_{t=1}^T \frac{\sigma_e^2 \lambda_1^2 h_t^2}{(\lambda_1 h_t + \sigma_1)^2}. \end{aligned}$$

From Theorem 3, we know that $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}(\mathbf{S}'\mathbf{S}) = 0$, uniformly in $\boldsymbol{\theta}, \boldsymbol{\beta}$ and \mathbf{X} and given Assumptions 1 – 4 this condition is satisfied if and only if (24) and (25) are true.

Proof of Proposition 2 First, notice that $R_T^{2,0} = -\frac{1}{T} \sum_{t=1}^T \left(\frac{h_t^2}{(\sigma_1 + \lambda_1 h_t)^2} \right)$ and from Proposition 1, $R^{2,0} = 0$. Next, notice that

$$\begin{aligned} R_T^{1,1} &= \frac{1}{T} \sum_{t=1}^T \frac{h_t}{(\lambda_1 h_t + \sigma_1)^2} \\ &\leq \frac{1}{T\sigma_1} \sum_{t=1}^T \frac{h_t}{(\lambda_1 h_t + \sigma_1)} \\ &\leq \frac{1}{\sigma_1} \sqrt{\frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2}} = \frac{1}{\sigma_1} \sqrt{R_T^{2,0}}, \end{aligned}$$

and consequently $R^{1,1} \leq \frac{1}{\sigma_1} \sqrt{\lim_{T \rightarrow \infty} R_T^{2,0}} = 0$. From $R_T^{2,1} = \frac{2}{T} \sum_{t=1}^T \left(\frac{h_t^2}{(\sigma_1 + \lambda_1 h_t)^3} \right) \leq \frac{2}{\sigma_1} R_T^{2,0}$, and $R_T^{1,2} = \frac{2}{T} \sum_{t=1}^T \left(\frac{h_t}{(\sigma_1 + \lambda_1 h_t)^3} \right) \leq \frac{2}{\sigma_1^2} \sqrt{R_T^{2,0}}$, we have $R^{2,1} = R^{1,2} = 0$. From Proposition 1, $U^{0,1} = \lim_{T \rightarrow \infty} -\frac{1}{T} \sum_{t=1}^T \left(\frac{b_t^2}{(\sigma_1 + \lambda_1 h_t)^2} \right) = 0$. Furthermore, since $0 < \frac{h_t}{\sigma_1 + \lambda_1 h_t} \leq \frac{1}{\lambda_1}$ we can write $U_T^{1,1} = \frac{1}{T} \sum_{t=1}^T \frac{h_t b_t^2}{(\lambda_1 h_t + \sigma_1)^3} \leq \frac{1}{\lambda_1} U_T^{0,1}$, and $U_T^{0,2} = \frac{2}{T} \sum_{t=1}^T \left(\frac{b_t^2}{(\sigma_1 + \lambda_1 h_t)^3} \right) \leq \frac{1}{\sigma_1} U_T^{0,1}$. It follows that $U^{1,1} = U^{0,2} = 0$, which completes the proof.

Proof of Lemma 6 Since it is trivial (but very space-consuming) to show that each of the 34 elements in $F^{4.4}$ is bounded on $\Theta \times B$ the proof is left out, but can be obtained from the authors upon request.

Proof of Lemma 7 Notice, that $F^{4.4}$ contains the derivatives of all the elements in $F^{3.3}$, and by Lemma 6, all elements in $F^{4.4}$ are uniformly bounded. The result of Lemma 7 then follows immediately.

Proof of Theorem 5 The objective is to show that the sequence $\{D_x f_n(x, y)\}_n$ converges to $D_x f_0(x, y)$. If so, it follows from equicontinuity of $\{D_x f_n(x, y)\}_n$ that the convergence is uniform on $\mathcal{X} \times \mathcal{Y}$. Now, suppose that $\{D_x f_n(x, y)\}_n$ does not converge to $D_x f_0(x, y)$. Then $\exists(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, and $\exists \varepsilon > 0$, such that for any $N_1 \in \mathbb{N}$, and $N' = N'(N_1) > N_1$ it holds that

$$|D_x f_{N'}(x^0, y^0) - D_x f_0(x^0, y^0)| > \varepsilon. \quad (44)$$

For $\forall n \in \mathbb{N} \cup \{0\}$, there exists a sufficiently small number $\Delta > 0$, such that,

$$f_n(x + \Delta, y) - f_n(x, y) = D_x f_n(\tilde{x}, y)\Delta, \quad (45)$$

where \tilde{x} is a value between $x + \Delta$ and x . By the equicontinuity of $\{D_x f_n(x, y)\}_n$, Δ can be made arbitrarily small, such that

$$|D_x f_n(x, y) - D_x f_n(\tilde{x}, y)| \leq \frac{\varepsilon}{4}, \quad (46)$$

is satisfied, for any $n \in \mathbb{N} \cup \{0\}$ and any $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Hence, by combining (45) and (46)

$$\left| D_x f_n(x, y) - \frac{f_n(x + \Delta, y) - f_n(x, y)}{\Delta} \right| \leq \frac{\varepsilon}{4}, \quad (47)$$

for $\forall n \in \mathbb{N} \cup \{0\}$ and $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$. Notice, that $\Delta = \Delta(\varepsilon)$, i.e., Δ depends on ε , but not N' . Therefore,

$$\left| D_x f_{N'}(x, y) - \frac{f_{N'}(x + \Delta, y) - f_{N'}(x, y)}{\Delta} \right| \leq \frac{\varepsilon}{4}, \quad (48)$$

$$\left| D_x f_0(x, y) - \frac{f_0(x + \Delta, y) - f_0(x, y)}{\Delta} \right| \leq \frac{\varepsilon}{4}, \quad (49)$$

and consequently

$$\begin{aligned} |D_x f_{N'}(x, y) - D_x f_0(x, y)| &\leq \left| \frac{f_{N'}(x + \Delta, y) - f_{N'}(x, y)}{\Delta} - \frac{f_0(x + \Delta, y) - f_0(x, y)}{\Delta} \right| \\ &\quad + \frac{\varepsilon}{2} \\ &= \left| \frac{f_{N'}(x + \Delta, y) - f_0(x + \Delta, y)}{\Delta} - \frac{f_{N'}(x, y) - f_0(x, y)}{\Delta} \right| \\ &\quad + \frac{\varepsilon}{2} \\ &\leq \left| \frac{f_{N'}(x + \Delta, y) - f_0(x + \Delta, y)}{\Delta} \right| + \left| \frac{f_{N'}(x, y) - f_0(x, y)}{\Delta} \right| \\ &\quad + \frac{\varepsilon}{2}. \end{aligned} \quad (50)$$

uniformly on $\mathcal{X} \times \mathcal{Y}$. By equicontinuity and convergence of $\{f_n(x, y)\}_n$, there exists a sufficiently large number $N_2 \in \mathbb{N}$, such that $\forall n \geq N_2$

$$|f_n(x, y) - f_0(x, y)| < \frac{\varepsilon}{4}\Delta,$$

or equivalently

$$\left| \frac{f_n(x, y) - f_0(x, y)}{\Delta} \right| < \frac{\varepsilon}{4},$$

uniformly on $X \times Y$. By choosing $N_1 \geq N_2$, and $(x, y) = (x_0, y_0)$, the inequality (50) reduces to

$$\begin{aligned} |\mathbf{D}_x f_{N'}(x^0, y^0) - \mathbf{D}_x f_0(x^0, y^0)| &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} \\ &= \varepsilon, \end{aligned}$$

and hence contradicts the statement in (44), which completes the proof.

Proof of Theorem 6 By combining Lemmas 6 and 7 and Theorem 7.16 p. 151 in Rudin (1976) we have that if $\{R_T^{I..J}\}_T$ and $\{U_T^{I..J}\}_T$ converges uniformly on $\Theta \times B$, then $\{R_T^{i..j}\}_T$ and $\{U_T^{i..j}\}_T$ converges uniformly on $\Theta \times B$ for all $i < 4$, and $j < 4$. Furthermore, by combining Lemmas 6 and 7 and Theorem 5 we have that if $\{R_T\}_T$ and $\{U_T\}_T$ converges uniformly on $\Theta \times B$, then $\{R_T^{i..j}\}_T$ and $\{U_T^{i..j}\}_T$ converges uniformly on $\Theta \times B$ for all $i, j = 1, 2, \dots, 4$, hereby completing the proof.

Proof of Theorem 7 Consider the sequence of functions $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$, given by equation (22) of which the limit is $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$. The Hessian matrix $\mathcal{H}_T(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ associated with $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ is given as

$$\mathcal{H}_T(\boldsymbol{\theta}, \boldsymbol{\beta}^*) = \begin{pmatrix} \mathbf{D}_{\lambda_1 \lambda_1}^2 Q_T^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) & \mathbf{D}_{\lambda_1 \sigma_1}^2 Q_T^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) \\ \mathbf{D}_{\lambda_1 \sigma_1}^2 Q_T^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) & \mathbf{D}_{\sigma_1 \sigma_1}^2 Q_T^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) \end{pmatrix},$$

and from Theorem 6

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathcal{H}(\boldsymbol{\theta}, \boldsymbol{\beta}^*) &= \begin{pmatrix} \mathbf{D}_{\lambda_1 \lambda_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) & \mathbf{D}_{\lambda_1 \sigma_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) \\ \mathbf{D}_{\lambda_1 \sigma_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) & \mathbf{D}_{\sigma_1 \sigma_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) \end{pmatrix} \\ &= \mathcal{H}(\boldsymbol{\theta}, \boldsymbol{\beta}^*), \end{aligned}$$

uniformly on $\Theta \times B$. Using Proposition 2 we can write

$$\begin{aligned} \mathbf{D}_{\lambda_1 \lambda_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) &= \frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^3} \left[\frac{1}{2} (\lambda_1 h_t + \sigma_1) - \sigma_e^2 - b_t^2 \right] \\ &= -\frac{1}{2} (R_T^{2..0} + \sigma_e^2 R_T^{2..1} + U_T^{2..0}) \end{aligned}$$

$$\begin{aligned}
D_{\sigma_1 \sigma_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) &= \frac{1}{T} \sum_{t=1}^T \frac{1}{(\lambda_1 h_t + \sigma_1)^3} \left[\frac{1}{2} (\lambda_1 h_t + \sigma_1) - \sigma_e^2 - b_t^2 \right] \\
&= -\frac{1}{2} (R_T^{0,2} + \sigma_e^2 R_T^{0,3} + U_T^{0,2})
\end{aligned}$$

$$\begin{aligned}
D_{\lambda_1 \sigma_1}^2 Q^*(\lambda_1, \sigma_1, \boldsymbol{\beta}^*) &= \frac{1}{T} \sum_{t=1}^T \frac{h_t}{(\lambda_1 h_t + \sigma_1)^3} \left[\frac{1}{2} (\lambda_1 h_t + \sigma_1) - \sigma_e^2 - b_t^2 \right] \\
&= -\frac{1}{2} (R_T^{1,1} + \sigma_e^2 R_T^{1,2} + U_T^{1,1})
\end{aligned}$$

and

$$\lim_{T \rightarrow \infty} \mathcal{H}(\boldsymbol{\theta}, \boldsymbol{\beta}^*) = \begin{pmatrix} \lim_{T \rightarrow \infty} -\frac{1}{2} U_T^{2,0} & 0 \\ 0 & \lim_{T \rightarrow \infty} -\frac{1}{2} (R_T^{0,2} + \sigma_e^2 R_T^{0,3}) \end{pmatrix} \quad (51)$$

where the existence of $\lim_{T \rightarrow \infty} -\frac{1}{2} U_T^{2,0}$ and $\lim_{T \rightarrow \infty} -\frac{1}{2} (R_T^{0,2} + \sigma_e^2 R_T^{0,3})$ is guaranteed by Assumption 2 and Theorem 6. For the matrix in (51) to be negative definite, the condition $\lim_{T \rightarrow \infty} -\frac{1}{2} (R_T^{0,2} + \sigma_e^2 R_T^{0,3}) < 0$ must hold (since $\lim_{T \rightarrow \infty} -\frac{1}{2} U_T^{2,0} < 0$, as $U_T^{2,0} > 0$ uniformly in T). This restriction can be written as

$$\begin{aligned}
\lim_{T \rightarrow \infty} (R_T^{0,2} + \sigma_e^2 R_T^{0,3}) &> 0 \quad (52) \\
\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left(\left(\frac{2\sigma_e^2}{(\sigma_1 + \lambda_1 h_t)^3} \right) - \left(\frac{1}{(\sigma_1 + \lambda_1 h_t)^2} \right) \right) &> 0 \\
\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{1}{(\sigma_1 + \lambda_1 h_t)^3} (2\sigma_e^2 - \lambda_1 h_t - \sigma_1) &> 0
\end{aligned}$$

Notice that

$$\begin{aligned}
\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{\lambda_1 h_t}{(\sigma_1 + \lambda_1 h_t)^3} &= \lambda_1 \lim_{T \rightarrow \infty} R_T^{1,2} \\
&= 0
\end{aligned}$$

Consequently, the restriction in (52) becomes

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{1}{(\sigma_1 + \lambda_1 h_t)^3} (2\sigma_e^2 - \sigma_1) > 0$$

Hence, a sufficient condition for negative definiteness of $\mathcal{H}(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ is that $\sigma_e^2 > \frac{1}{2} \sigma_1$.

Proposition A.1 Define $\overline{\Sigma}_{11t} = \text{cov}(D_{\theta}Q_t(\theta, \beta))$, and $\Sigma_{11} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \overline{\Sigma}_{11t}$.

Given Assumptions 1 - 4,

$$\Sigma_{11} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{12} \sigma_e^4 R^{0,4} \end{pmatrix}.$$

Using arguments similar to the ones used in Theorem 6, it is clear that $R^{0,4}$ exists.

Proof of Proposition A.1 Recall from the properties of a standard normal random variable that $E z_t^4 = 3$ and $\text{var}(z_t^2) = 2$. From straightforward calculations,

$$\overline{\Sigma}_{11t} = \begin{pmatrix} \sigma_e^2 b_t^2 (R_t^{1,1})^2 + \frac{1}{2} \sigma_e^4 (R_t^{1,1})^2 & \sigma_e^2 b_t^2 R_t^{1,1} R_t^{0,2} + \frac{1}{2} \sigma_e^4 R_t^{1,1} R_t^{0,2} \\ \sigma_e^2 b_t^2 R_t^{1,1} R_t^{0,2} + \frac{1}{2} \sigma_e^4 R_t^{1,1} R_t^{0,2} & \sigma_e^2 b_t^2 (R_t^{0,2})^2 + \frac{1}{2} \sigma_e^4 (R_t^{0,2})^2 \end{pmatrix}$$

We have the following limiting results regarding the components of Σ_{11} : First,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t^2 (R_t^{1,1})^2 &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2 b_t^2}{(\lambda_1 h_t + \sigma_1)^4} \\ &\leq C_1 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \\ &= 0 \end{aligned}$$

by Proposition 1, where C_1 is the uniform upper bound of $\frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2}$. Secondly,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (R_t^{1,1})^2 &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^4} \\ &\leq C_2 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2} \\ &= 0 \end{aligned}$$

by Proposition 1, where C_2 is the uniform upper bound of $\frac{1}{(\lambda_1 h_t + \sigma_1)^2}$. Thirdly,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t^2 R_t^{1,1} R_t^{0,2} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t b_t^2}{(\lambda_1 h_t + \sigma_1)^4} \\ &\leq C_3 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \\ &= 0 \end{aligned}$$

by Proposition 1, where C_3 is the uniform upper bound of $\frac{h_t}{(\lambda_1 h_t + \sigma_1)^2}$. Fourthly,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T R_t^{1,1} R_t^{0,2} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t}{(\lambda_1 h_t + \sigma_1)^4} \\ &\leq C_4 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t}{\lambda_1 h_t + \sigma_1} \\ &\leq C_4 \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2}} \\ &= 0 \end{aligned}$$

by Proposition 1, where C_4 is the uniform upper bound of $\frac{1}{(\lambda_1 h_t + \sigma_1)^3}$, and the second inequality is by Cauchy-Schwarz. Finally,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t^2 (R_t^{0,2})^2 &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^4} \\ &\leq C_2 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \\ &= 0. \end{aligned}$$

Consequently,

$$\Sigma_{11} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{12} \sigma_e^4 R^{0,4} \end{pmatrix}$$

where $R^{0,4}$ is given by Proposition 2, which completes the proof.

Proposition A.2 Define $\overline{\Sigma}_{12t} = \text{cov}(D_{\theta} Q_t(\boldsymbol{\theta}, \boldsymbol{\beta}), D_{\boldsymbol{\beta}} m_t(\boldsymbol{\beta}))$, and $\Sigma_{12} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \overline{\Sigma}_{12t}$. Given Assumptions 1 - 4, $\Sigma_{12} = \Sigma'_{21} = \mathbf{O}$.

Proof of Proposition A.2 Since $\text{cov}(z_{t,T}, z_{t,T}^2) = 0$, it follows that

$$\Sigma_{12t} = \begin{pmatrix} \sigma_e^2 b_t a_{t1} R_t^{1,1} & \sigma_e^2 b_t a_{t1} R_t^{0,2} \\ \sigma_e^2 b_t a_{t2} R_t^{1,1} & \sigma_e^2 b_t a_{t2} R_t^{0,2} \\ \vdots & \vdots \\ \sigma_e^2 b_t a_{tk} R_t^{1,1} & \sigma_e^2 b_t a_{tk} R_t^{0,2} \end{pmatrix}.$$

For all $j = 1, 2, \dots, k$,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t a_{tj} R_t^{1,1} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t a_{tj} h_t}{(\lambda_1 h_t + \sigma_1)^2} \\ &\leq \left(\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} \right)^{\frac{1}{2}} \left(\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{h_t^2 a_{tj}^2}{(\lambda_1 h_t + \sigma_1)^2} \right)^{\frac{1}{2}} \\ &= 0, \end{aligned}$$

by Proposition 1, uniform boundedness of $\frac{h_{t,T}^2}{(\lambda_1 h_t + \sigma_1)^2}$, the fact that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T a_{tj}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_{tj}^2$ converges to a constant by Assumption 4 and the Cauchy-Schwarz inequality. Using similar techniques it can be shown that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t a_{tj} R_t^{2,0} = 0$ for all $j = 1, 2, \dots, k$, which completes the proof.

Proposition A.3 Define $\Sigma_{22t} = \text{cov}(D_{\beta} m_t(\beta))$, and $\Sigma_{22} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \overline{\Sigma}_{22t}$. Given Assumptions 1 - 4, $\Sigma_{22} = \sigma_e^2 \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}' \mathbf{X}$.

Proof of Proposition A.3 The result follows since $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{a}'_i \mathbf{a}_{.j} = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{x}'_i \mathbf{x}_{.j}$ converges to constants for all $i, j = 1, 2, \dots, k$.

Proof of Theorem 8 Given Assumptions 1 - 4 it has already been established that $\overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})$ is an independent $(k+2)$ -dimensional random vector with $E(\overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})) = \mathbf{0}$ and $\text{cov}(\overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta})) = \overline{\Sigma}_t$. Furthermore, from Propositions A.1 - A.3 we have established that $\Sigma = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \overline{\Sigma}_t$, where Σ is a positive definite matrix. To complete the proof of Theorem 8 according to Hoadley's (1971) Theorem A.6, it needs only to be shown that

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T E(\boldsymbol{\gamma}' \overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta}))^4 = 0 \quad (53)$$

for all $\boldsymbol{\gamma} \in \mathbb{R}^{k+2}$. Condition (53) is satisfied since $E(\boldsymbol{\gamma}' \overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta}))^4$ will depend on the moments of z_t which are all finite (recall that z_t is a standard normal random variable). In particular, by using Propositions 1 and 2 it can be shown that $\frac{1}{T} \sum_{t=1}^T E(\boldsymbol{\gamma}' \overline{\mathbf{g}}_t(\boldsymbol{\theta}, \boldsymbol{\beta}))^4$ is uniformly bounded (and convergent) for any fixed $\boldsymbol{\gamma}$, which completes the proof. From (40) and consistency of the estimated parameters the result of Theorem 8 follows immediately.

Proof of Theorem 9 The proof of the theorem roughly follows Newey and McFadden (1994). First notice that condition *i.* was shown in Section 3 while condition *iii.* follows from Theorem 8. Condition *ii.* is trivially satisfied whereas the first part of condition *iv.* has been verified in the proof of Theorem 7. The second part of condition *iv.* holds under Assumptions 1-4 since

$$D_{\beta\beta}^2 m_T(\beta) = -\frac{1}{T} \mathbf{X}' \mathbf{X}.$$

Finally notice that from expanding $g_T(\widehat{\zeta}_T)$ about ζ^* gives,

$$g_T(\widehat{\zeta}_T) = g_T(\zeta^*) + \mathbf{G}_T(\bar{\zeta})(\widehat{\zeta}_T - \zeta^*),$$

where $\bar{\zeta}$ is the mean value of $\widehat{\zeta}_T$ and ζ^* . Multiplying both sides by \sqrt{T} and rearranging yields

$$\sqrt{T}(\widehat{\zeta}_T - \zeta^*) = -\mathbf{G}_T^{-1}(\bar{\zeta}) \left[\sqrt{T} g_T(\zeta^*) \right].$$

Consistency wrt. ζ^* implies that $\bar{\zeta} \xrightarrow{p} \zeta^*$. Hence by conditions *i.-iv.* and the Slutsky theorem,

$$\sqrt{T}(\widehat{\zeta}_T - \zeta^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_T^{*-1}(\zeta^*) \boldsymbol{\Sigma}^* \mathbf{G}_T^{*-1}(\zeta^*)').$$

where

$$\mathbf{G}_T^{-1}(\zeta) = \begin{pmatrix} (D_{\theta\theta}^2 Q_T(\zeta))^{-1} & -(D_{\theta\theta}^2 Q_T(\zeta))^{-1} D_{\theta\beta}^2 Q_T(\zeta) (D_{\beta\beta}^2 m_T(\beta))^{-1} \\ 0 & (D_{\beta\beta}^2 m_T(\beta))^{-1} \end{pmatrix}.$$

and $\boldsymbol{\Sigma}^*$ is defined in Theorem 8.